Trade, Globalization and Poverty

Edited by
Elias Dinopoulos, Pravin Krishna,
Arvind Panagariya, and Kar-yiu Wong

Routledge
Taylor & Francis Group

Trade, Globalization and Poverty

Trade, Globalization and Poverty contains papers in honor of Jagdish Bhagwati, the pre-eminent scholar, teacher and public policy intellectual in the area of international trade in the world today.

This collection brings together a stellar line of contributors from across the world, all of whom discuss the themes and arguments raised by Bhagwati's latest work. It is essential reading for students and academics involved with international trade and development economics.

Elias Dinopoulos is Professor of Economics at the University of Florida.

Pravin Krishna is Chung Ju Yung Distinguished Professor of International Economics at Johns Hopkins University, Washington DC and Research Associate at the National Bureau of Economic Research.

Arvind Panagariya is Professor of Economics and Jagdish Bhagwati Professor of Indian Political Economy at Columbia University, New York.

Kar-yiu Wong is Professor of Economics at the University of Washington.

Routledge Studies in International Business and the World Economy

1 **States and Firms**
Multinational enterprises in institutional competition
Razeen Sally

2 **Multinational Restructuring, Internationalization and Small Economies**
The Swedish case
Thomas Andersson, Torbjörn Fredriksson and Roger Svensson

3 **Foreign Direct Investment and Governments**
Catalysts for economic restructuring
Edited by John H. Dunning and Rajneesh Narula

4 **Multinational Investment and Economic Structure**
Globalization and competitiveness
Rajneesh Narula

5 **Entrepreneurship in a Global Context**
Edited by Sue Birley and Ian Macmillan

6 **The Global Structure of Financial Markets**
An overview
Edited by Dilip K. Ghosh and Edgar Ortiz

7 **Alliance Capitalism and Global Business**
John H. Dunning

8 **Multinational Enterprises from the Netherlands**
Edited by Roger van Hoesel and Rajneesh Narula

9 **Competition, Growth Strategies and the Globalization of Services**
Real estate advisory services in Japan, Europe and the United States
Terrence LaPier

10 **European Integration and Foreign Direct Investment in the EU**
The case of the Korean consumer electronics industry
Sang Hyup Shin

11 **New Multinational Enterprises from Korea and Taiwan**
Beyond export-led growth
Roger van Hoesel

12 Competitive Industrial Development in the Age of Information
The role of co-operation in the technology sector
Edited by Richard J. Braudo and Jeffrey G. MacIntosh

13 The Global Restructuring of the Steel Industry
Innovations, institutions and industrial change
Anthony P. D'Costa

14 Privatisation and Liberalisation in European Telecommunications
Comparing Britain, the Netherlands and France
Willem Hulsink

15 Multinational Corporations
Emergence and evolution
Paz Estrella Tolentino

16 Foreign Direct Investment in Emerging Economies
Corporate strategy and investment behaviour in the Caribbean
Lou Anne A. Barclay

17 European Integration and Global Corporate Strategies
Edited by François Chesnais, Grazia Ietto-Gillies and Roberto Simonetti

18 The Globalisation of Corporate R & D
Implications for innovation systems in host countries
Prasada Reddy

19 Globalization of Services
Some implications for theory and practice
Edited by Yair Aharoni and Lilach Nachum

20 A Century of Foreign Investment in the Third World
Michael J. Twomey

21 Global Capitalism at Bay
John H. Dunning

22 Foreign Direct Investment
Research issues
Edited by Bijit Bora

23 Ford and the Global Strategies of Multinationals
The North American auto industry
Isabel Studer Noguez

24 The World Trade Organization Millennium Round
Freer trade in the next century
Klaus Deutsch and Bernhard Speyer

25 Consultancy and Innovation
The business service revolution in Europe
Edited by Peter Wood

26 Knowledge Economies
Clusters, learning and co-operative advantage
Philip Cooke

27 The Role of Resources in Global Competition
John Fahy

28 Globalization, Employment and the Workplace
Diverse impacts
Edited by Yaw A. Debrah and Ian G. Smith

29 Transnational Corporations
Fragmentation amidst integration
Grazia Ietto-Gillies

30 Growth Theory and Growth Policy
Edited by Harald Hagemann and Stephan Seiter

31 International Business and the Eclectic Paradigm
Developing the OLI framework
Edited by John Cantwell and Rajneesh Narula

32 Regulating Global Trade and the Environment
Paul Street

33 Cultural Industries and the Production of Culture
Edited by Dominic Power and Allen J. Scott

34 Governing Interests
Business associations facing internationalization
Edited by Wolfgang Streeck, Jürgen Grote, Volker Schneider and Jelle Visser

35 Infrastructure Development in the Pacific Region
Edited by Akira Kohsaka

36 Big Business and Economic Development
Conglomerates and economic groups in developing countries and transition economies under globalisation
Edited by Alex E. Fernández Jilberto and Barbara Hogenboom

37 International Business Geography
Case studies of corporate firms
Edited by Piet Pellenbarg and Egbert Wever

38 The World Bank and Global Managerialism
Jonathan Murphy

39 Contemporary Corporate Strategy
Global perspectives
Edited by John Saee

40 Trade, Globalization and Poverty
Edited by Elias Dinopoulos, Pravin Krishna, Arvind Panagariya, and Kar-yiu Wong

Trade, Globalization and Poverty

Edited by
Elias Dinopoulos, Pravin Krishna,
Arvind Panagariya, and Kar-yiu Wong

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2008
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

This edition published in the Taylor & Francis e-Library, 2007.

Routledge is an imprint of the Taylor & Francis Group,
an informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available
from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN 0-203-93292-7 Master e-book ISBN

ISBN10: 0–415–77322–9 (hbk)
ISBN10: 0–203–93292–7 (ebk)

ISBN13: 978–0–415–77322–5 (hbk)
ISBN13: 978–0–203–93292–6 (ebk)

Contents

Figures

Tables

Contributors

Aisbett, Emma, Crawford School and Economics Program RSSS, Australian National University.

Beaulieu, Eugene, Department of Economics, University of Calgary.

Choi, Nakgyoon, Korea Institute for International Economic Policy.

Deardorff, Alan V., Department of Economics, University of Michigan.

Dehejia, Vivek H., Department of Economics, Carleton University.

Dinopoulos, Elias, Department of Economics, University of Florida.

Fujiwara, Kenji, Department of Economics, Kwansei Gakuin University.

Gungoraydinoglu, Ali, Department of Economics, University of Florida.

Harrison, Ann, Department of Agricultural and Resource Economics, University of California, Berkeley.

Krebs, Tom, Department of Economics, University of Mannheim.

Krishna, Pravin, School of Advanced International Studies and Department of Economics, Johns Hopkins University.

Lee, Hong Shik, Korea Institute for International Economic Policy.

Magee, Stephen P., Department of Finance, University of Texas, Austin.

Maloney, William, Office of the Chief Economist, Latin America and Caribbean Region, World Bank.

Markusen, James R., Department of Economics, University of Colorado, Boulder.

Mayer, Wolfgang, Department of Economics, University of Cincinnati.

Mourmouras, Alex, IMF Institute, International Monetary Fund.

Panagariya Arvind, Department of Economics, Columbia University.

Shimomura, Koji, RIEB, Kobe University.

Srinivasan, T.N., Department of Economics,Yale University.

Strand, Bridget, Department of Economics, University of Colorado, Boulder.

Syropoulos, Constantinos, Department of Economics, Drexel University.

Wilson, John Douglas, Department of Economics, Michigan State University.

Wong, Kar-yiu, Department of Economics, University of Washington.

Yoo, Kwang-Yeol, OECD, Paris and Korea Ministry of Finance and Economy.

Zakhilwal, Hazrat-Omar, Afghanistan Institute for Economic Development and Social Policy.

Zwane, Alix Peterson, Department of Agricultural and Resource Economics, University of California, Berkeley.

Acknowledgments

We would like to express our deep appreciation to Dean's Office of the School of Business Administration at the University of Florida, CIBER at Columbia University, and the following University of Florida centers for providing financial support for the January 2005 conference in honor of Jagdish Bhagwati's seventieth birthday: CIBER, Public Policy Research Center, Public Utility Research Center, Transnational and Global Studies Center, Center for European Studies, and the Center for Latin American Studies. We are also indebted to Taiba Batool of Routledge and Robert J. Bell of the University of Washington for their assistance in preparing the materials for this volume.

Introduction

Elias Dinopoulos, Pravin Krishna, Arvind Panagariya, and Kar-yiu Wong

This volume celebrates the seventieth birthday of Jagdish Bhagwati, the pre-eminent scholar, teacher, and public policy intellectual in the area of international trade in the world today. Jagdish was born on July 26, 1934 and did his schooling in Bombay. He went off to do his first degree at St Johns College at Cambridge, England during 1954–1956, where he first met his teacher, the great international economist Harry Johnson. It was during his second year at Cambridge that Jagdish wrote his famous article on immiserizing growth (Bhagwati, 1958). He graduated with a first in the Economics Tripos.

Jagdish spent the year 1956–1957 at MIT, where he met his future wife, Padma Desai. He then returned to England for postgraduate studies and spent the years 1957–1959 at Nuffield College, Oxford University. He moved back to India in 1961, joining first the faculty at the Indian Statistical Institute (ISI) and then, in 1963, that at the Delhi School of Economics. It was during these years that Jagdish wrote (jointly with V. K. Ramaswami) his path-breaking article (Bhagwati and Ramaswami, 1963) on the theory of distortions that laid to rest many arguments for protection. In a nutshell, the authors demonstrated that once a distortion was corrected at source, free trade still remained unambiguously beneficial.

In 1966–1967, Jagdish visited Columbia University and went on to join the faculty at MIT in 1968, where he later became the Ford International Professor of Economics. During his MIT years, Jagdish turned the work that he began at the ISI into a full-blown *Generalized Theory of Distortions* (Bhagwati, 1971), with virtually all scholars of international trade at the time deriving inspiration from it. In 1980, he commenced his work modeling political economy aspects of international trade policy (Bhagwati, 1982) and moved to Columbia University, where he is currently a Professor and continues to write prolifically on all aspects of globalization. His book *In Defense of Globalization* (Oxford University Press, 2004) offers a definite rebuttal of virtually all the important arguments made by globalization skeptics.

The accomplishments of Jagdish Bhagwati are too numerous to count here. We only note that he has deeply influenced the world of trade theory as well as the practice of trade policy and is widely recognized today as the world's foremost defender of free trade. For his scientific contributions, he is the leading candidate

for a Nobel Prize in the area of international trade, with Nobel Laureate Paul Samuelson calling the period spanning 1960s through 1980s the "Age of Bhagwati *et al*." in the history of international trade literature. As for Jagdish's influence on policy, Lawrence Summers, the former Treasury Secretary and President of Harvard University, has said

> There are literally billions of consumers who do not know his [Jagdish's] name whose real incomes have been higher because of the goods they have been able to purchase at a lower price because of the influence this man has had on the world trading system.

In a similar vein, the citation conferring the Distinguished Fellowship of the American Economic Association states

> Jagdish Bhagwati's intellectual arc has taken him from profound theoretical analyses of international trade to deep insights into the political economy of globalization. No economist now living has displayed so potent a combination of academic analysis and practical wisdom.

Jagdish has always occupied a special place in the lives and hearts of his students (not only those who were his classroom students but also those, like Arvind Panagariya, who learned international trade from his writing and found themselves generously adopted as his students) and vice versa. And since he also has had so many students, there have been numerous celebrations of his milestone birthdays. The present volume emerged from one of the three academic celebrations of his seventieth birthday, this one a conference at the University of Florida.

The chapters in this volume are organized into three parts corresponding to the themes of the festschrift conference. Part I consists of chapters on globalization, wages, and poverty, issues that were the subject of one of Jagdish's earliest contributions to the field of economics (*The Economics of Underdeveloped Countries* (1966), McGraw-Hill, 1966) and which continue to engage his research interests today, for example, in *Trade and Wages* (with Marvin Kosters, AEI Press, 1994) and *In Defense of Globalization*.

Part 1 begins with an evaluative survey (Chapter 1) by T. N. Srinivasan in which he highlights some of the analytical and measurement issues in exploring the trade-poverty nexus and the mechanisms that, in theory, link globalization, and poverty, as well as factors other than trade that might reinforce or mitigate the operation of these mechanisms. Srinivasan's discussion of the issues relating to the meaning and measurement of poverty also enable him to provide a critical assessment of the burgeoning empirical literature on globalization and poverty. He is particularly critical of international comparisons of poverty using a common definition such as $1 per day converted at the purchasing power parity (PPP) at the exchange rate. He rightly notes that such a definition requires an internationally accepted basket of goods that anyone above the poverty line should be able to buy. But a "poverty bundle common to all regions within a geographically and

culturally diverse country such as India, let alone for all countries of the world, cannot be meaningfully defined." He goes on to argue, "Deaton [Princeton economist Angus Deaton] is absolutely right in arguing that, because such an internationally accepted bundle does not exist, it does not make sense to simply convert $1/day to local currency values using PPP exchange rates with commodities weighted by their shares in the consumption of the poor. The reason is that doing so makes poverty lines move around with changes in PPP exchange rates arising from world market price changes that have no relevance to the poor."

In Chapter 2, Emma Aisbett, Ann Harrison, and Alix Peterson Zwane survey the evidence on the linkages between globalization and poverty, drawing on the collected works of Jagdish Bhagwati and the results of an National Bureau of Economic Research (NBER) study directed by Ann Harrison, *Globalization and Poverty* (2006). Their study argues for a more subtle understanding of the links between globalization and poverty than is offered by the traditional Heckscher–Ohlin framework that argues that unskilled workers in developing countries are more likely to gain from trade reform. Rather, they conclude, that there needs to be programs to promote human capital development and infrastructure for unskilled workers to have a likely share in the gains from globalization.

Chapter 3, by Eugene Beaulieu, Vivek Dehejia, and Hazrat Omar-Zakhilwal, considers an important hypothesis concerning trade and income distribution first articulated in Bhagwati and Dehejia (1994). The hypothesis is that trade liberalization has made many industries "footloose," that is, small shifts in cost can cause comparative advantage to transfer constantly from one country to another, which may increase labor turnover, which may prompt mobile workers to accumulate fewer skills, which may cause a reduction or stagnation in their real wages. If the effects of such mobility strike less-educated workers harder, the wage differential between educated and less-educated labor will rise. Using Canadian data, the authors argue that the observed relationship between trade and the relative wage of educated to less-educated workers have a closer fit to the Bhagwait-Dehejia hypothesis than the traditional Stolper-Samuelson predictions.

In Chapter 4, Tom Krebs, Pravin Krishna, and William Maloney argue for the importance of studying individual labor income risk and emphasize the need to go beyond aggregate wage statistics in studying the impact of international trade on the labor market. This study examines the effect that trade policy changes have on individuals of differing human capital levels (proxied by education), finding that the effects of trade policy changes on income risk are not monotonically linked with education. Income risk changes due to policy changes may simply be higher at higher education levels.

Part 2 of this volume consists of chapters studying multinational firms, the mechanisms and consequences of international technology transfer, and the political role of multinational firms in the international trade system, topics that are variously related to and inspired by Bhagwati's own work on the questions of factor mobility (Bhagwati and Brecher, 1980; Bhagwati *et al*., 2004), technological change, and political economy (Bhagwati, 1982; *The World Trading System at Risk* (Princeton University Press, 1991)).

Chapter 5, by Elias Dinopoulos, Ali Gungoraydinoglu, and Constantinos Syropoulos, addresses some questions raised by the signing of the General Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPs Agreement), which calls for all World Trade Organization (WTO) members to adopt a set of global minimum standards on intellectual property rights protection. The authors build a model to analyze the effects of strengthening intellectual property protection on growth and poverty. The model generates product-cycle trade; endogenous, long-run, scale-invariant Schumpeterian (R&D-based) growth; and an endogenous wage gap between Northern and Southern workers. The authors find that an increase in intellectual property protection worsens the wage-income inequality between North and South, increases the rate of international technology transfer, and has an ambiguous effect on long-run growth.

In Chapter 6, Kenji Fujiwara, Koji Shimomura, and Kar-yiu Wong analyze the process of technological evolution in an international economy. Motivated by the observation that many developing countries achieve technological catch-up and sometimes even surpass developed countries, they abandon the traditional assumptions in the literature where developed countries exclusively innovate and developing countries exclusively imitate technologies available elsewhere. Endogenizing the decision taken by firms to innovate or imitate advanced technologies allows them to derive conditions under which technological catch-up or surpass takes place.

In Chapter 7, James Markusen and Bridget Strand use recent empirical findings on multinational production, particularly on foreign manufacturing affiliates of US multinationals, to undermine a popular argument against globalization. According to the authors, opponents of globalization adopt a view not uncommon even in traditional economic thinking about multinationals, that international factor-price differences are the primary determinant of foreign direct investment (FDI). According to this view, firms will want to move activities that are intense in unskilled labor from countries where that labor is scarce to countries where that labor is abundant. And in doing so, antiglobalists claim, they undermine existing indigenous markets, destroy competition, and increase "exploitation" of the local labor force. In more recent "horizontal" or "market-seeking" view of FDI and multinationals, it is thought that multinationals create firm-specific (intangible) assets through investments in R&D and learning by doing that can be applied to a variety of locations without any additional costs due to their "public good" nature.

This view predicts that multinational activities will arise between countries at a similar stage in development. Markuson and Strand report that, in fact, US multinationals' investments in manufacturing affiliates and, indeed, investments from all parent countries flow primarily to high-income countries, that US outward FDI is primarily "horizontal" or "market-seeking," and that US manufacturing affiliates primarily seek skilled labor. This suggests that international factor-price differences play a subordinate role in FDI, and that globalization may not have the destructive effects its opponents claim.

In Chapter 8, Alan Deardorff analyzes the political economy of rules-setting in the global trading system and describes the role played by corporations in it.

He argues that, unlike in a competitive system, where the profit-maximizing efforts of firms contribute to social surplus, corporate influence on rules may have negative effects. Deardorff cites several examples, including the expansion of the WTO's ambit to cover intellectual property and services, as well the use of rules of origin in bilateral trade agreements, to make his case.

Part 3 of this volume consists of chapters focused on international trade policies and international institutions, areas in which Bhagwati made many contributions and had great influence, for example, *The Economics of Preferential Trade Agreements* (with Arvind Panagariya, AEI Press, 1996); Bhagwati, 1993; *Free Trade Today* (Princeton University Press, 2002; and *The World Trading System at Risk* (Princeton University Press, 2002)).

In Chapter 9, Arvind Panagariya provides a welfare-theoretic analysis of a preferential trade agreement between two small countries. In a literature that has generally been characterized by inconclusive results, this analysis of small unions provides a number of unambiguous theoretical results, including that a union member is necessarily hurt by its own preferential liberalization and that the higher its external tariffs and the larger its imports from the partner, the more it loses from its liberalization.

In Chapter 10, Wolfgang Mayer and Alex Mourmouras analyze the viability of conditionally assisted economic reform programs such as those supported by international financial institutions like the International Monetary Fund (IMF), which must cope with informational asymmetries and special interest politics. This chapter examines what conditions must be satisfied to make conditional assistance programs viable, that is, to assure that the assistance-receiving government not only takes the assistance but also implements reforms, without compromising the country's political and the international financial institution's financial stability.

In Chapter 11, Stephen P. Magee, Kwang-Yeol Yoo, Nakgyoon Choi, and Hong Shik Lee examine several versions of the hypothesis that the "United States is a large country in world trade," a hypothesis, which if validated would provide a theoretical basis for the use of trade barriers to improve US terms of trade and welfare. Using a variety of cross-sectional and time-series tests, however, they find scant empirical evidence supporting the large-country hypothesis, especially in recent years.

In Chapter 12, John Wilson reassesses Bhagwati's (1972, 1973) proposal for "taxing the brain drain" so that developing countries could receive revenue from taxes levied on emigrants residing in developed countries. Using recent theoretical developments on international taxation, Wilson finds that the proposal withstands some important arguments that have been made against it and remains remarkably valid more than thirty years after it was first put forth.

Bibliography

Jagdish Bhagwati (1958) "Immiserizing Growth: A Geometrical Note," *Review of Economic Studies*, 25(3): 201–205.

—— (1970) "Lobbying and Welfare," *Journal of Public Economic*s, 14: 355–363.

Jagdish Bhagwati (1971) "The Generalized Theory of Lobbying and Welfare," in Bhagwati, Jones, Mundell and Vanek (eds) *Trade, Balance of Payments and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*, Amsterdam: North Holland.

—— (1972) "The United States in the Nixon Era: The End of Innocence," *Daedalus*, 25–47.

—— (1982) "Directly Unproductive, Profit-Seeking (DUP) Activities," *Journal of Political Economy*, 90: 988–1002.

—— (1982) "Shifting Comparative Advantage, Protectionist Demands, and Policy Response," in Bhagwati (ed.) *Import Competition and Response*, National Bureau of Economic Research Conference Report, University of Chicago Press.

—— (1993) "Regionalism and Multilateralism: An Overview," in deMelo Panagariya (ed.) in *New Dimensions in Regional Integration*, Cambridge University Press.

Jagdish Bhagwati and Richard Brecher (1980) "National Welfare in an Open Economy in the Presence of Foreign Owned Factors of Production," *Journal of International Economics*, 10:103–115.

Jagdish Bhagwati and Vivek Dehejia (1994) "Free Trade and Wages of the Unskilled—Is Marx Striking Again? in Bhagwati, J. and Kosters, M. H. (eds) *Trade and wages: Leveling wages down*? Washington, DC: AEI Press.

Jagdish Bhagwati and William Dellafar (1973) "The Brain Drain and Income Taxation," *World Development*, 1: 94–100.

Jagdish Bhagwati and V. K. Ramaswami (1963) "Domestic Distortions, Tariffs and the Theory of Optimum Subsidy," *Journal of Political Economy*, 71(1): 44–50.

Jagdish Bhagwati, Arvind Panagariya and T. N. Srinivasan (2004) "The Muddles Over Outsourcing," *Journal of Economic Perspectives*, 18(4): 93–114.

Part 1

Poverty and wages

1 Globalization and poverty*

T. N. Srinivasan

Introduction

In his latest best seller, *In Defense of Globalization*, Jagdish Bhagwati (2004) devotes a chapter entitled, "Poverty: Enhanced or Diminished" to the effects of globalization on the poor. He points out that most policy makers in the twentieth century were acutely aware of the poverty in their midst and passionately wanted to eliminate it. In his view, the central question is not about lack of awareness of or commitment to eradicate poverty, but whether globalization as it is commonly understood is an integral part of any poverty eradication strategy as proponents of globalization insist.

Jagdish's colleague at Columbia University, Joseph Stiglitz, in his own earlier best seller, *Globalization and Its Discontents* (2002), categorically asserts

> Globalization today is not working for many of the world's poor. It is not working for much of the environment. It is not working for the stability of the global economy. The transition from communism to a market economy has been so badly managed that, with the exception of China, Vietnam, and a few Eastern European countries, poverty has soared as incomes have plummeted.
>
> (p. 214)

Jeffrey Sachs, another of Jagdish's colleagues at Columbia, directed the United Nations Millennium Project. The report of the project (referred to as the Sachs Report by the media), submitted to the UN Secretary-General on January 17, 2005 (UN Millennium Project, 2005a), recommends (Recommendation 8, p. xvi) that

> High-income countries should open their markets to developing country exports through the Doha trade round and help Least Developed Countries raise export competitiveness through investments in critical trade-related infrastructure, including electricity, roads and ports. The Doha Development Agenda should be fulfilled and the Doha Round completed no later than 2006.

It recognizes that

> Trade is among the most politically charged of international issues for development. Though hugely important, it is far from a magic bullet for achieving development. The slogan "trade, not aid" is misguided, particularly in the poorest countries. Trade reforms are complementary to other parts of development policy, such as infrastructure investments and social programs to develop a healthy and well educated workforce.
>
> (p. 46)

It further suggests that

> To establish an overarching framework for progress, we recommend that global political leaders first agree to a conveniently distant long-term target (for example, 2025) for the total removal of barriers to merchandise trade, a substantial and across-the-board liberalization of trade in services, and the universal enforcement of the principles of reciprocity and nondiscrimination.
>
> (p. 46)

It is clear that, its caveats notwithstanding, the Sachs report, and even more forcefully the other report of the Millennium project, *Trade, Development and WTO: An Action Agenda for the Doha Round* (UN Millennium Project, 2005b), by a task force coordinated by Patrick Messerlin and Ernesto Zedillo, both endorse Jagdish's analysis. Stiglitz's assertion has to be dismissed out of hand—the very exceptions he mentions, which include China and Vietnam (and I would add India and other East Asian countries—whose success was based on globalization according to Stiglitz himself on the same page) *are* the countries in which an overwhelming majority of the world's poor live. The Sachs report implicitly rejects Stiglitz's assertion in recognizing the importance of trade, though it does not resist striking down a straw man in saying it is far from a magic bullet. Related to Jagdish's central question is: are there examples of countries which pursued an inward-oriented (outward-oriented) development strategy and yet succeeded (failed) in significantly reducing poverty? Jagdish's chapter addresses this question as well. He points out that, first, countries such as China and India were able to reduce poverty significantly only after shifting to an outward-oriented development strategy from an earlier inward-oriented one. Second, he argues, drawing on his justly celebrated work on domestic distortions, that the potential causes of the failure, where it occurred, of an outward-oriented development strategy in reducing poverty are almost always the consequences of domestic distortions of various kinds.The literature on globalization and poverty is vast. Winters *et al*. (2004) survey the evidence. Brookings Institution's Trade Forum, held during May 13–14, 2004, was devoted to "Globalization, Poverty and Inequality: What Do We Know and Where We are Going." The papers presented at this forum, including, in particular, that of Goldberg and Pavcnik (2004), also survey some of the evidence. Part II of the Annual Report of the WTO (2003) was

devoted to trade and development, with a subsection on trade liberalization and poverty alleviation. Of course, given its mandate, the United Nations Conference on Trade and Development (UNCTAD) has published several reports on trade and development, including a recent publication *Trade Liberalization and Poverty in India* (www.unctadindia.org) launched at the Eleventh Session of UNCTAD at Sao Paolo, Brazil during June 13–18, 2004.

Jagdish's own contributions to poverty and development go way back to his *The Economics of Underdeveloped Countries* (McGraw-Hill, 1966), perhaps the only one of his many books that has photographs illustrating development projects.

Its opening chapter is entitled, "Poverty and Income Distribution." In this remarkable short book he discusses many aspects of trade and development that continue to be relevant. For example, Jagdish decries tariff escalation and tariff peaks in rich countries against developing country exports. The two issues are on the Doha Round Agenda. The book also shows that even Jagdish, prescient as he has always been, did not quite see *then* the problems with central planning in poor countries and with developing countries not having to reciprocate industrial country offers in bargaining over tariff reductions in the GATT! I cannot hope to do justice in this chapter to his many contributions which we are celebrating, or to add much to what has already been said by others in the vast literature on trade and poverty.

What I propose to do instead is to highlight some of the analytical and measurement issues in exploring the trade-poverty nexus. The issues relating to the meaning and measurement of poverty are important to recognize and allow for in assessing the burgeoning empirical literature on globalization and poverty. Of course, any empirical econometric analysis, to be credible, has to be founded on sound theory. This being the case, the mechanisms that, in theory (or theories), are behind the globalization–poverty liberalization links have to be well-understood, as well as factors other than trade that might reinforce or blunt the operation of these mechanisms. Unfortunately, the cross-country regressions that are proliferating like amoeba show very little understanding either of the data and measurement issues that plague poverty estimates or of the need for a sound analytical foundation before putting data from diverse countries, sources and time periods into a common regression.

Poverty: concepts and measurement

Conceptual issues[1]

Any study of poverty has to begin with what Sen (1981) calls an exercise of "identification," a method that distinguishes the poor from the nonpoor in a population that includes the poor. The most widely used method of identification is to define a norm or set of norms and classify anyone who does not meet the norm or norms as poor. Both the poverty norm and the population will be distinguished by time and spatial dimensions. Given a norm for identifying whether each member of a specified population at a specified time and place is

poor, there is the further issue of deriving a poverty measure that aggregates the member-specific poverty status over the specified population, over space and over time.

Sen (1981, p. 21) sees the measurement of poverty as "an exercise of description assessing the predicament of people in terms of the prevailing standards of necessities. It is primarily a factual rather than an ethical exercise, and the facts relate to what is regarded as deprivation." By basing measurement on social norms, that is, "prevailing standards of necessities," this view would seem to eschew *welfarism*; that is, the idea that an individual's welfare, as judged by that individual, should be the basis from which any assessment of the state of affairs in a population, (e.g. the extent of poverty in that population) is to be made. However, in his subsequent discussion of the income based method of identification, Sen (1981, p. 27) argues that "if the poverty level income can be derived from behavior norms of society, a person with a higher income who is choosing to fast on a bed of nails can, with some legitimacy, be declared to be nonpoor." If indeed such a person can be presumed to fast because he *prefers* it to consuming the bundle constituting the prevailing standards of necessities, then classifying him as nonpoor is *welfarist*.

Whether or not one takes a welfarist view, one could reasonably argue that the identification exercise should preferably encompass an individual's lifetime: an individual who meets the prevailing poverty norm by borrowing in the present should nevertheless be deemed poor if servicing the debt will push him into poverty in the future. A welfarist approach involves deep ethical and philosophical problems regarding what constitutes individual welfare, and whether a minimum degree of comparability of the chosen welfare measure across individuals can be postulated so that a meaningful aggregation over individual members of a population is possible. Sen himself (1973, 1976, 1981, 1984, 1985, and 1987), among others, has written extensively on these matters. However, these are essentially theoretical issues. There is not much an empirical study can contribute in this regard. However, an empirical study can explore the implications of alternative poverty norms, whether or not they have any welfarist connotations.

Measurement issues[2]

In the poverty literature, it is customary to define a threshold level of per person real income or consumption as the poverty line, with anyone having real income or consumption below the poverty line deemed as poor. This does not necessarily imply that dimensions of poverty other than income or consumption are not as important or more important, particularly deprivations in health, education, and democratic rights. However, there are difficult measurement issues associated with all of these dimensions.

The best and most easily understood starting point for deriving a consumption-based poverty line is a "poverty consumption" bundle of goods and services for a representative (in size and age-gender composition) household. It is common to assume that part of the bundle would be provided free or at subsidized prices by

the state. Valuing the private component of the bundle at appropriate prices yields the poverty line. This valuation of the private component is in effect what an expert group did for India's Planning Commission in 1962, in defining poverty lines for rural and urban households in India. There is unavoidable arbitrariness in determining which goods and services (and in what amounts) are to be included in a poverty bundle. Nevertheless, given a poverty bundle for a representative household, appropriate adjustments for differences of any other household in its size and age-gender composition could be made to arrive at a household-specific poverty line. A household would be deemed to be poor if it does not have the resources, measured in terms of either income or total consumption expenditure, to buy the private component of its poverty bundle at the prices it faces. In such a definition, a household that can afford to but does not buy its poverty bundle is not deemed poor.

Clearly, if an annual survey collects data on the resources that each household commands and the prices it faces, it is a straightforward matter to estimate the number of the poor. As long as the constituents and the nonprivate component of the poverty bundle remain the same over time and space and surveys in different regions and time periods continue to collect household-specific data on resources and prices, estimating the poor in each region and time period continues to be relatively uncomplicated.

The price data collected in different household expenditure surveys differ in their coverage, completeness, and representativeness. In some, only total expenditure data are collected. In others, total expenditures on and quantities purchased of each commodity are collected so that unit values could be computed. In still others, prices paid are also collected along with quantities purchased. But I know of no survey that collects data on prices actually paid by household in each of their transactions involving purchasing included in their consumption. The Indian National Sample Survey (NSS) does collect price data (although it is not transaction specific).

Deaton and Tarozzi (2005) use more than 7 million pieces of price (more precisely, unit value) information from two rounds (1987–1988 and 1993–1994) of NSS expenditure surveys. In his analysis of the unit value data from the NSS for the state of Maharashtra in 1983, Deaton (1997) found that they matched independently collected market price data. Survey data confirm that households living in the same region and canvassed at the same time reported paying different prices for the same commodity even after allowance was made for possible differences in quality and other factors. Indeed, such differences raise serious questions about the common assumption in the analysis of household surveys that households purchase homogeneous commodities in competitive markets in which the "law of one price" holds. Apart from such interhousehold variation in prices, there are spatial and intertemporal variations as well. Also, there could be (and often are) interhousehold, interregional, and intertemporal variations in access to the state-provided component of the poverty bundle. I am unaware of any poverty count that allows for these variations.

Clearly, it is impractical to update poverty lines through revaluation of a *given* poverty bundle at prices that are specific to each household, region, and period of

time. As such, a common practice is to use some price index to adjust some poverty line (not necessarily one derived from valuing a poverty bundle) at base year (or base region) prices to arrive at a poverty line for a different year (or region). The fact that poor (rural) households face different prices compared with nonpoor (urban) households could be taken into account in such an approach by using poor-specific rural and urban prices indexes to update poverty lines (or alternatively to deflate consumption expenditures). For example, in India, because a large proportion of rural poor are believed to be landless agricultural laborers, the consumer price index for agricultural laborers has been used for updating the rural poverty line in official estimates of poverty. A simple average of the consumer price index for industrial workers and that for urban nonmanual workers is used to update the poverty line. However, commodity weights used in constructing these indexes are outdated, and the price quotations used are not representative of the relevant transactions. Using commodity weights and unit values based on the household surveys, Deaton and Tarozzi (2005, p. 405) recomputed the poverty estimates. They find that, in contrast to the diverging trends in the official estimates, "between 1987–1988 and 1993–1994, there was no great difference in the rate of decline of urban and rural poverty." Clearly, the choice of price index matters. But whatever index is used, as Deaton points out, the basic, standard textbook index number problem remains and cannot be wished away.

A poverty bundle common to all regions within a geographically and culturally diverse country such as India, let alone for all countries of the world, cannot be meaningfully defined. If such a bundle could be defined, then the national poverty line at any point in time would be the value of that bundle at the prices in local currency that households face in that nation at that point in time.

There is no need for any exchange rate in such a calculation. Deaton is absolutely right in arguing that, because such an internationally accepted bundle does not exist, it does not make sense to simply convert $1/day to local currency values using purchasing power parity (PPP) exchange rates with commodities weighted by their shares in the consumption of the poor. The reason is that doing so makes poverty lines move around with changes in PPP exchange rates arising from world market price changes that have no relevance to the poor. For example, the poverty line for one country would be shifted by a change in the world price of a commodity that is not consumed by the poor in that country but consumed by the poor in some other country, because such a price change affects the PPP exchange rate. In any case, global poverty counts are based on neither a common global poverty bundle nor conversions to local currency values using PPP exchange rates with commodity weights more relevant to the poor.

An international poverty line for base year 1985 was chosen ($1/day at 1985 PPP dollars) as being representative of poverty lines in use in low-income countries. In making this choice, poverty lines in local currency (consumption expenditures per person per day) in use in 1985 were presumably converted to US dollar terms using the then-available PPP exchange rate for each currency. Because these were apparently found to cluster around one dollar, $1/day in constant 1985 PPP dollars was seen as representative of the poverty lines then in use.

It should be obvious that even assuming that local currency poverty lines in 1985 represented the value of a national poverty bundle, it cannot be claimed that the $1/day at 1985 PPP dollars poverty line is representative of national poverty lines even in that base year. Moreover, revisions of the PPP rate to reflect, on one hand, better and more accurate information, but on the other hand, the change in base year play havoc with the poverty counts. Deaton's critique is more than enough to persuade any serious analyst that these poverty counts are virtually meaningless (Deaton, 2000).

Data issues[3]

Most poverty estimates are based on household survey data. On the other hand, some of the data that are used in the analysis of determinants of poverty at the national or regional level (rather than at the level of a household) are from national accounts statistics or other sources. In particular, growth estimates are almost always based on real GDP data from national accounts. Very often, the mutual consistency of diverse data sources cannot be assured. Also, data from different sources could differ in their biases and measurement errors. I will illustrate some of these data issues by drawing on the debate on the discrepancy data from national accounts and national sample survey. Before I do so, let me highlight some problems of international and intertemporal comparability and those of drawing distributional inferences from household survey data from sample surveys when the survey is not necessarily designed for that purpose.

Growth rates of real income are often compared within countries over different periods and across countries. As is well-known, the relative price structure (across commodities and sectors) would shift over time as development proceeds and as different sectors experience different rates of technical progress within and across countries. Under these circumstances, the growth rate of conventional real income estimates over a given period would depend on whether the initial or the terminal year is used as the base and on the method of deflation (e.g. single or double deflation). Also, if coverage improves over time or if downward biases in income measurement go down, growth in income as measured would overstate the "true" growth. Equally, understatement of growth could also arise. This is not to say that these problems cannot be addressed. For example, through chain linking the problem of using a single base year could be alleviated. And as the base year is changed to a more recent one, the opportunity is often used also to expand coverage and use more recent information and an ad hoc revision is made in the past data for the difference in coverage as compared to the old base year.

In international comparisons, in addition, an exchange rate conversion is involved. Thus, if some (official, black market, PPP or whatever) exchange rate is used to compare the levels or income at a point in time between two countries and growth rates for each for a period preceding that time are computed using a constant domestic relative price structure, anomalous results can arise. A prime example of this is provided by World Bank (1999a, Tables 1.1 and 1.4). According to these tables, in 1997 India and China have GNP per capita of US$370 and US$860,

respectively (based on exchange rates of the Bank's "Atlas" method) and the average annual rate of growth of GNP per capita during 1965–1997 was 2.3% in India and 6.8% in China. Taken together, these levels and rates of growth would imply that China's per capita GNP in 1965 was about 90% of India's! No knowledgeable observer of the two countries would subscribe to this ranking. Ahmed (1994) illustrates a similar problem by comparing the actual PPP exchange rate with converted GDPs of a year with a forecast obtained by applying domestic GDP growth to the PPP-converted GDP of some previous years. The differences seem substantial for many countries (see also Heston, 1994). The easy accessibility of PPP-based macroeconomic data put together by Robert Summers and Alan Heston, ostensibly for more than 100 countries and for more than four decades since 1950, spawned the "cross-country growth-regression" industry. The practitioners of this technique often ignore the fact that most of the data are mere extrapolations.

For many small open developing economies with severely distorted foreign trade sectors, the opportunity costs of traded commodities are their world market prices, although domestic resource allocation decisions would be driven by distorted domestic prices. If distortions are severe, and domestic transformation of imported inputs into tradable outputs is inefficient, it is possible that value added at world prices (computed by valuing tradable inputs and outputs at world prices) in some activities could be negative, and indeed were found to be negative (Soligo and Stern, 1965 for Pakistan, Bhagwati, 1968 for India) in some developing countries. Of course, value added at domestic distorted prices would always be positive. From a welfare perspective, as the manuals on project appraisal (e.g. Little and Mirrless, 1974) argued, only value added at world prices is the appropriate measure. Even if activities with negative value added at world prices are unimportant from the perspective of the economy as a whole, still the levels and growth rates of national income (i.e. value added) at world prices would differ from those at domestic prices (Bhagwati and Hansen, 1973).

Estimates of the extent of poverty (income- or consumption based) and inequality, are mostly based on data for distribution of income or consumption. Many countries also collect income and consumption data at regular intervals through household surveys. Most often survey-based estimates of consumption or income and those derived from national accounts data differ substantially. Deaton eloquently describes the problems in estimating income in developing countries through a survey as follows:

> The concept of income is itself extraordinarily complex, and most people in developing countries have little reason to distinguish between business and personal cash transactions . . . Even in developed countries the measurement of self-employment income is notoriously inaccurate. The problems are not entirely solved even by the detailed questioning of more sophisticated surveys, in which the surveyor, not the respondent, calculates income. And the national accounts data for household saving are not themselves reliable enough to provide a good cross-check that will show what sort of surveys do best or how they should be redesigned to do better.
>
> (Deaton, 1989, p. 63)

Apart from the problem of measuring income and expenditures reliably through surveys, there are others related to survey design. For example, in the Indian National Sample Survey, the focus varied from one annual round to the next. Although, consumption expenditure data are canvassed, even in a round in which it is not the focus, from a subset of households, the sampling and nonsampling errors of estimated average consumption will be higher as compared to founds in which the primary focus is consumption expenditure, both for the reason of a lower sample size and because it not being the primary focus of the survey. The data collected on consumption expenditure are likely to be of poorer quality as well. Also, these surveys are often designed only for estimating a national or a regional average, accurately. This means that using the empirical distribution across households in the sample, as an estimate of the distribution among the entire population of households, would be subject to large sampling errors. The same problem arises with respect to inequality measures. For example, a sample size that is adequate for estimating the mean consumption with reasonable precision could be inadequate to estimate mean incomes of individual decile groups with the same precision. Also the definition of a household in most surveys, that is, a household consists of all individuals who usually eat out of the same kitchen in the reference period, is not appropriate for analyzing the distribution of income and expenditure. The more appropriate unit is a family or family like group whose members pool their incomes and expenditures. Although the survey household is most often a family, this is not always the case.

At least four possible size distributions can be generated from survey data, namely, distribution of households by total or per head of household income, of persons by total or per head of household income. Whether one reports per head or per household distribution is extremely important since the two would differ substantially and inferences could be very sensitive to the distribution reported. Of course, whether the same weight should be assigned to all persons in a household regardless of their sex and age is another issue, though poverty incidence or intensity is unlikely to be sensitive to the use of adult (male or female) equivalence scales. Often data on inequality measures such as the Gini coefficient are published without explicitly specifying to which of the four distributions the coefficient refers. Even by his own very minimal standards, namely, that data be *national* in coverage, concepts of income and recipient unit be the same over time, and the data base be a national census or household survey, Fields (1994) finds that half of the 70 data sets for developing economies he examined had to be rejected. Of course, if the problems highlighted by Deaton (1989) about income measurement in surveys are taken as seriously as they should be, and if stricter and perhaps more appropriate standards of reliability and clarity are applied, even more data sets will be rejected. Further, the surveys that are not rejected might be unrepresentative in that they are more likely to be from richer and advanced countries of the world.

Deninger and Squire (DS) (1996) have updated Fields (1994). By drawing together more than 2600 observations on inequality measures, such as Gini coefficients and quintile shares, from a wide range of studies for 112 developed and developing countries and for the period 1947–1994, they have extracted what

they deem a "high-quality" set of 700 observations and labeled them as acceptable for users. The DS data set has been widely used, along with the Summers–Heston data by the (mindless) cross-country regression industry. Atkinson and Brandolini (1999), who are very appreciative of the work of DS, subject the DS "acceptable" data for a subset of OECD countries to a careful scrutiny as to their sources, methods of compilation, definitions of income and the unit (e.g. household, individual, etc.) of measurement and other relevant aspects. Among their many conclusions, the following is worth noting: "we are not convinced that at present it is possible to use secondary data sets *safely without some knowledge of* the underlying sources, and we strongly caution against mechanical use of such *data sets*" (Atkinson and Brandolini, 1999, p. 35, emphasis added). Few practitioners of cross-country regression go beyond mechanical use. This does not mean that it is not possible to do so: the message of Atkinson and Brandolini is simply that one has to satisfy oneself that the inequality data are reasonably comparable in concept and coverage across countries and over time before embarking on an explanatory or casual analysis which make use of them.

Estimates of private consumption expenditure from household surveys often differ from those from national accounts. The ratio of per capita consumption expenditure from India's National Sample Survey (NSS) and National Accounts (NAS) fell from 75% in 1974 to 50% in 1998! In the 1950s and 1960s, the two estimates were much closer to each other, rarely differing by more than 5% (Mukherjee and Chatterjee, 1974; Srinivasan *et al*., 1974). As noted in the two studies, there are many legitimate reasons for the difference. Still if the difference increases over time, there is a strong presumption that one or both estimation methods have deteriorated in reliability. However, if the increasing discrepancy is solely or largely due to increasing nonresponse of and underreporting of their consumption by nonpoor households, it would not affect poverty analysis. On the other hand, if it is largely due to underreporting by all groups, it would. Let me illustrate with estimates of trends in poverty at the national level.

The trend in estimated poverty (head-count ratio of the proportion of poor in the Indian population) in NSS data differs significantly from that in poverty estimated from proportionally adjusted so that its mean is the same as the mean household consumption expenditure in NAS. The former shows no trend in poverty during the postreform 1990s, while the latter shows a declining trend. Clearly, the policy inference about reforms would be different depending on which data set is used.

Globalization, growth, and poverty[4]

Jagdish argues that globalization influences poverty through its influence on growth. He points out that some types of growth will help the poor more than others, depending on the presence of other factors including policy distortions that reinforce or attenuate the effect of growth on the poor. It is also clear from his analysis of the linkage between trade and growth, that the effect of globalization on growth could vary across countries and over time for similar reasons. Above all, since trade and other policies, as well as their outcomes in terms of growth and poverty are endogenous, without a well-specified analytical and

econometric framework, it is hard to draw valid inferences. Much of the cross-country regression literature that purports to show or deny the beneficial effects of globalization suffer from this disability.

Several plausible links in the globalization–growth–poverty reduction chain can be postulated in theory, yet the reality is far more complicated and many links could be absent in some countries at some points in time. Agenor (2004), for example, finds that globalization may have a U-shaped effect on poverty: while extensive integration reduces poverty, small amounts of globalization may hurt the poor.[5] Even in theory, not all links need be unidirectional—thus, it is possible that in some links, globalization influences growth positively but the character of growth increases poverty. This being the case, protesters find it easy enough to blame the process of globalization for any observed or imagined deterioration in the condition of the poor rather than look for the missing links or for other factors that could have muted or outweighed the beneficial effects of globalization. Enthusiasts find it equally easy to argue that observed outcomes deviate from globalizations' predicted contributions to poverty reduction only because globalization has not gone far enough. It would seem that both sides focus selectively on some aspects of globalization while ignoring that other processes besides globalization also influence the observed outcomes.[6]

There are several mechanisms through which globalization could be expected to contribute to poverty reduction (Winters *et al*., 2004). In theory, greater international integration should play an important part in reducing poverty around the world. In practice, it has had mixed effects due to domestic policy distortions, continued industrial country protectionism, and limited labor market integration across countries. In practice, many of the ways in which globalization is perceived as harmful to the poor are not intrinsic aspects of global integration. They reflect, rather, domestic policy failures such as segmented and distorted internal markets as well as industrial country protectionism and limited labor market integration across countries. It is useful to discuss separately

1 globalization and growth
2 growth and poverty reduction
3 globalization and poverty reduction
4 globalization for the poor.

While there are many mechanisms in theory for expecting greater integration to increase growth and reduce poverty, the theory is not without caveats, and the empirical evidence is not conclusive. The effect of globalization on inequality, in particular, is ambiguous.

Globalization and growth

Theory

There are essentially three sources of economic growth: growth in inputs of production; improvements in the efficiency of allocation of inputs across economic

activities; and innovation that creates new products, new uses for existing products, or brings about more efficient use of inputs. The combination of changes in these three dimensions that brings about higher long-run growth (as opposed to short-run transition effects) depends on the economy's characteristics.

Whether or not a change in rate of accumulation of a factor of production or the efficiency of factor allocation, for example, has long-run or only transitional effects on growth depends in part on the technology of production. An exogenous change in the rate of investment or opening the economy to foreign trade has only a transitional effect on growth in a Solow type two-factor (capital and labor) constant–returns–to–scale growth model if the marginal product of capital declines to zero as capital increases indefinitely relative to labor. On the other hand, if the technology is such that the marginal product of capital is bounded away from zero, transitional as well as steady state growth effects could arise from an exogenous change in investment or foreign trade policy.[7]

Being open to trade and investment contributes to each of the three sources of growth. Domestic resources are allocated more efficiently when the economy can specialize in those activities in which it has comparative advantage. By being open to capital, labor, and other resource flows, an economy is able to augment relatively scarce domestic resources and use part of its abundant resources elsewhere where they earn a higher return. Clearly, efficiency of resource use in each nation and across the world is enhanced by the freedom of movement of resources. Finally, the fruits of innovation anywhere in the world become available everywhere in such an open world. Empirical studies suggest total factor productivity (TFP) in poor countries, which do not have domestic research and development (R&D) capacities, is higher when their trade with industrialized countries who account for the bulk of R&D in the world is greater.[8]

Theory also suggests that globalization and growth have a self-reinforcing relationship, in that higher growth spurs a larger volume of trade flows. While the decision to alter policies to further integration is a policy change that harnesses trade as an "engine of growth," trade also serves as the "handmaiden" of growth once policies support freer interchange of goods and services.[9]

Domestic institutions, however, can offset the contributions of liberalization to growth by limiting labor flexibility, segmenting internal markets, and failing to provide the social infrastructure for education. The traditional argument about static factor price effects and gains from trade, for example, assumes that resources move smoothly and costlessly from import competing to exporting activities. Obviously, if resources cannot or do not move, exporting industries would not expand, while import competing industries surely contract because of increased competition from imports after trade liberalization, thus creating unemployment. This somewhat extreme but elementary argument has been raised by Stiglitz (2002, p. 59) against trade liberalization when he says that: "It is easy to destroy jobs, and this is often the immediate impact of trade liberalization, as the inefficient industries [those created under the protectionist walls] close down under pressure from international competition." Since he assumes that no new, more efficient jobs would be created, he concludes that "moving resources from

low-productivity uses [in inefficient industries] to zero productivity [to unemployment] does not enrich any country." True, but neither does keeping factors in less productive uses forever. There may be a rationale for credibly committing to phase in trade liberalization over a period while at the same time removing impediments to labor mobility, but certainly not postponing liberalization indefinitely, as is sometimes argued.[10]

Globalization's effect on short-term growth also depends upon the exact forms of globalization and preexisting market distortions. Removing barriers and controls on financial capital flows, for example, may improve resource allocation and give more people access to better-functioning credit markets in the long run. In the short-term, on the other hand, it can lead to crisis and lower growth in countries with fragile domestic financial sectors. Capital controls have been advocated by many in the wake of the Asian financial crisis, and some find that controls were useful in helping some countries recover faster than those that had freer capital movements.[11]

Empirical evidence

Theoretical formulations of a trade-growth linkage are foundations of the globalization-growth links. Empirical demonstrations of the linkage go back to the careful and nuanced cross-country studies in the late 1960s and 1970s sponsored by OECD (Little *et al*., 1970) and NBER (Bhagwati, 1978; Krueger, 1978). More recent studies based on simple cross-country regressions (e.g. Sachs and Warner, 1995) asserting the same linkage have been controversial, though the questions are more about the magnitude than the sign of the linkage between trade and growth.[12] Wacziarg and Welch (2003) find that the Sachs and Warners' cross-sectional results are somewhat weaker when the sample period is extended through the 1990s. Their estimates, however, show that openness has positive effects on growth and investment rates within countries.

World Bank (2002, table 1) data show that both China and India enjoyed historically unprecedented average annual rates of growth of GDP of around 10% and 6%, respectively, as the two countries engaged in opening their economies to foreign trade and investment over 1980–2000. The effect is not entirely attributable to "globalization," as both countries also engaged in domestic economic reforms allowing a greater role for markets and the private sector in the economy, but integration no doubt played a large role.

Growth and poverty reduction

Aggregate growth is undoubtedly an instrument for poverty reduction, and it is associated with improvements in the minimum standard of living over some time horizon. Besley and Burgess's (2003) estimates for the elasticity of poverty with respect to income per capita vary widely across country samples, but all are negative, implying that growth reduces poverty.

This association between growth and poverty reduction, however, could take a long time to be seen. Inequality—which is different conceptually and empirically

though as salient as poverty in the eyes of critics—does not necessarily drop with growth. Although the poor may be becoming better off over time, the rich could gain even more. There are many possible mechanisms through which aggregate growth could affect, positively or negatively, poverty at national or subnational levels on the one hand, and on the other, how levels and trends in poverty could influence growth, again in either direction. The basic mechanisms behind growth and poverty reduction do not fully overlap, and some policies meant to encourage growth will have little or negative effect on the poor in the short-run. These lags can affect the political feasibility of reforms.

There is some evidence that regional disparities widened in China and India as these nations liberalized their foreign trade and introduced other reforms. To a certain extent this is natural: those regions (and individuals) which are better placed initially to take advantage of the opportunities opened up by reforms or, for that matter, by any other factor, such as, for example, the information technology revolution, are likely to grow faster (and richer). For example, India's phenomenal success in software is still confined to a few cities in the south and west. The real issue is not one of increasing regional disparities but of whether the socioeconomic system would enable the initially disadvantaged regions and individuals to catch up. If it does not, the social and political consequences could be serious and could lead to secessionist threats.

Globalization and poverty reduction

Some individuals or groups in a society may be poor and remain poor because they are disadvantaged in the social and political processes of their society. For example, social institutions such as the caste system in India or forms of racism elsewhere have denied equal access to sociopolitical processes to lower castes and racial minorities. While recognizing the importance of addressing this, we have nothing to add. The economic mechanisms for alleviating poverty could be divided into two broad categories: increasing the resources held by the poor through redistribution and affecting the economic environment that perpetuates poverty. We focus primarily on several ways in which globalization affects the latter category, as these economic changes have far deeper impacts on reducing poverty over the long run.[13] While critics focus on globalization's sometimes negative effects on redistribution, they overlook its potential to reduce economic constraints that continue to limit the prospects of the poor.[14] Market integration and increased migration limit the extent of poverty-alleviating redistribution in several ways, but their effect on removing the market distortions that perpetuate poverty over the long run far outweighs these consequences.

Needless to say, in the developing world, market distortions are ubiquitous, and their impact on the extent and depth of poverty are often serious. Whether an individual (or a household) has adequate resources to purchase the poverty bundle at the relevant prices at a point in time depends, of course, on what he/she (or his/her household) can earn from his/her assets (land, financial and physical capital) and most importantly from his/her (allowing for skills and educational

attainments) labor. The functioning of asset and labor markets, as well as product markets for goods and services bought or sold, obviously influence the earnings from assets and their purchasing power.[15] In this analysis we echo Tinbergen's emphasis that the central issue in alleviating poverty is not to simply offset it with transfers, but to tackle the more difficult issues of removing obstacles the poor face in trading, saving, and investing in their assets for a higher standard of living.[16]

The critics of globalization point to some forms of exploitation of workers in developing countries, such as the frequent use of child labor, the damage that a rapid exposure to global agricultural markets can cause for developing country farmers, and other aspects of the increasingly integrated global economy. However, these are not intrinsic or permanent consequences of globalization. They often reflect, rather, domestic institutional policy failures as well as continued industrial country protectionism and the reality that the most powerful forces for alleviation of poverty take time to work.

Globalization and redistribution

The critics of globalization generally point to a decline in state spending on price supports and services for the poor as countries become more internationally integrated. Clearly, in a globalized economy subject to the discipline of international capital markets, fiscal deficits are not sustainable, and taxation of mobile factors will induce their flight. Again, if the growth-accelerating effects of globalization are strong, the revenue expansion (with a buoyant fiscal system) from growth would enable financing of the needed services for the poor and avoidance of taxing mobile factors.

Increased migration flows are also often thought to reduce domestic demand for redistribution, as the original population may not want to finance services for new immigrants to their country. The rise of nationalist parties as well as anti-immigrant violence in several industrial countries reinforces this perception. Razin (2002) shows that immigration generally moves the political voting equilibrium toward less redistribution, unless the new migrants join forces with existing low-income voters in their destination country.

The ramifications of less redistribution for poverty, however, are small relative to potential gains in other areas. Government-led redistribution is one of the more direct, but generally less effective means of reducing poverty in the long run. It has limited effectiveness as a means of affecting the resources that poor households command. It will reduce poverty in the short run if individuals are poor because the assets they own are too meager. Unless redistribution of income is sustained indefinitely, its poverty reduction effect will be temporary if the conventional belief that the marginal propensity to consume of the poor is close to unity is correct. On the other hand, if credit markets are absent so that investment is constrained by resources owned and marginal returns to investment diminish, the rich would have a lower marginal return to investment than the poor if assets are unequally distributed. A redistribution of resources to the poor from the rich would raise the average rate of return to investment and hence the rate of growth of the economy.[17]

Subsidy policies may not even achieve the short-term goals of providing more resources to the poor, as subsidies are not often targeted at the poor or there is hijacking of subsidies intended for the poor by the nonpoor. The cost of transferring a dollar to the poor through subsidy schemes (particularly poorly targeted ones) also often exceeds a dollar by a substantial margin. The Indian public distribution system (PDS), through which fixed amounts per person of foodgrains and a few other essential commodities are sold at subsidized prices, for example, has a negligible impact on rural poverty and the central government alone spends four rupees to distribute one.[18]

Lessening market distortions such as those listed in the next subsections, in contrast, has a dynamic effect in that it not only increases the value of present resources but encourages greater investment and future accumulation.

Globalization and the productivity of the poor

The chief asset of the poor is their labor. In poor countries such as India, more than half (56%) of the work force was self-employed, and only 7% were regular employees in 1999–2000—and a larger proportion of the self-employed were poor than in the population as a whole. This being the case, raising productivity to create a sustained increase in real returns to labor in wage and self-employment would contribute significantly to poverty alleviation. Domestic public policy has a large role to play: increasing the human capital endowments of the poor, perhaps by providing incentives for investment in human capital or through public expenditure on, and improving the access of the poor to public education and health care programs, raises the productivity of their labor.

Globalization also contributes to this goal in several ways. First, the growth associated with globalization will generally create an outward shift in the demand for wage labor and for goods and services produced by the self-employed. Second, returns to the abundant factor, which in most poor countries is unskilled labor, would rise with trade liberalization.[19] While multinational companies naturally take advantage of the less developed country's abundance of unskilled workers to pay less than they would pay similar workers in their home countries, these wages are often higher than the wages paid by domestic companies in the host countries. Third, more integrated labor markets are important to ensure that workers receive the best return for their work.[20] This last aspect of globalization has not yet been realized—in fact, national integration is still limited by poor infrastructure, explicit restriction on movement (as in China), and linguistic differences across regions. Unlike commodities, the cost of whose movement within and between countries is primarily determined by costs of transportation and insurance, the cost of mobility of labor involves social and legal barriers as well as economic ones.

Globalization and returns from accumulation

Globalization can also benefit the poor by creating strong competitive pressures for improved financial intermediation. More efficient financial intermediation

would have large and long-term benefits for the poor by facilitating their investment in both physical and human capital. Although this is again partly a matter for domestic policymakers, international capital market integration may provide an added incentive to move financial sector reforms faster. Banks facing international competition in their traditional markets may be faster to move into microlending or services for small depositors. Similarly, the competition for investment under globalization encourages governments to focus more closely on providing better opportunities for investment in human capital.

The share of savings used to finance direct investment in physical assets depends in large part on the functioning of the financial system and access to it, which together influence the cost of financial intermediation. The costs faced by the poor are high, and a large share of savings and investment by households in developing countries is currently in the form of physical assets which they finance on their own without involving financial intermediaries. This share could be as high as 80%, as in Ghana, or around 50%, as in India. These assets include mostly those related to their production activities and also dual use (i.e. production and consumption) assets.

Even though the poor do not save enough to invest in financial markets (particularly in equity markets), they do invest their meager financial savings in the form of deposits in commercial banks, purchase of life insurance policies, and also lending in informal credit markets. Clearly, the returns they realize on such investments depend on the functioning of the financial sector, including the banking system.

The more important effect of improved financial intermediation—via domestic policy changes or spurred by international competition—would be in providing greater opportunities for financing education. Just as labor is the major asset owned by the poor, it is their investment in accumulation of human capital that is likely to be the major component of their investment. Although their poverty limits their saving and investment in any form, it is particularly limiting when it comes to human capital accumulation. Indeed, a major reason that the incidence of child labor is very high in many poor countries of South Asia and sub-Saharan Africa is the poverty of the parents of working children. Such parents cannot afford to forego the income from the work of a child (directly from paid work or indirectly in terms of unpaid contribution to the household's farm or non-farm enterprise). The out-of-pocket costs to the parents for sending their children to school are often substantial.[21]

Lack of investment in education has three serious consequences. First, the earning prospects of uneducated (or less educated) children in their adult working life would be less compared to their competitors in labor markets. Second, unless the labor market conditions improve in their adult life as compared to those that prevailed in their childhood, they are likely to end up as poor as their parents were and, as such, unlikely to educate their own children. The prospect of perpetuation of poverty across generations in such circumstances cannot be ruled out. Third, since some minimal education is often needed for an individual to participate effectively in political and social processes that make decisions affecting his

social and economic prospects, these individuals may be in effect unable to exercise their right to participate.

Apart from its beneficial effect on investment in human capital through improved financial intermediation, globalization has a direct effect on the demand for educated labor in poor countries through outsourcing of some service activities by industrialized countries. Already, China and India have benefited from such outsourcing. Unfortunately, there is a protectionist backlash in the United States against outsourcing. Hopefully, policymakers will not cave in to such pressure.

Globalization and product market efficiency

Product market efficiency, affected by domestic policies as well as international integration, determines the "terms of trade" the poor face in attempting to exchange their production for consumption and investment. Needless to say, the extent of integration of national markets and also the competitiveness of exports in world markets depends in large part on whether or not transport and communications infrastructure exists, and functions efficiently, to minimize costs of transportation and of acquiring market intelligence. Insufficient integration would mean the existence of price differentials across markets that cannot be arbitraged away. Also given the uncertainties, not only about harvests, but also about the prices that would rule at harvest time or at any time thereafter when the harvested output would be sold, it matters whether national markets for forward transactions exist and how costly it is to store commodities for later sale. Globalization could have a particularly significant impact on poverty by affecting the prices farmers in developing countries receive for their products and pay for inputs. There are currently two obstacles: lack of competition among intermediaries who aggregate primary products for international trading and continued industrial country subsidies and protection for their farmers.[22]

Economic institutions

Globalization can be helpful in mitigating or even overcoming the institutional and market failures that affect the poor adversely.

Governance quality

Tackling corruption is a major challenge of governance in developing countries, as processes of adoption, enforcement, and effectiveness of policy interventions are often distorted by endemic corruption. Inefficient and corrupt bureaucracies raise transactions costs in asset markets important for the poor. Land and tenancy markets, for example, have higher transactions costs than other asset markets because of the difficulty in establishing a claim of ownership in land when records are poorly maintained and officials who have the authority to certify ownership are corrupt. It is also a phenomenon that has been with us for ages. The

Arthasastra, a Sanskrit treatise on statecraft dated to fourth century BC India, lists more than fifty ways in which officials could be corrupt.[23]

Any extra impetus that globalization provides toward more transparent institutions will contribute to poverty reduction. The prospect of losing in the race to attract international capital flows, for example, can act as an important impetus to curb corruption. Wei (2000a) finds that corruption's effects on international investment are as if governments were imposing a tax—investors are significantly less likely to invest in more corrupt countries. The direction of causation between corruption and globalization, however, is likely to go in the opposite direction for trade: corrupt bureaucrats' bribes act as tariffs limiting imports and exports.[24]

Insurance and credit

Limited access to domestic credit and insurance markets exposes the poor to substantial uncertainty and potentially to consumption volatility. Domestic financial markets in many developing countries are simply not efficient enough to find it worthwhile to extend services to the poor because it is costly relative to returns to provide small loans to many poor in contrast to providing large loans to few rich. As mentioned previously, globalization can create an impetus to improve financial market and offer such credit and insurance opportunities to the poor.

In addition to being dependent on agriculture, the rural poor also have to cope with uncertainties, some of which relate to the environment for production and consumption (e.g. weather, disease vectors) and others which are idiosyncratic (e.g. health and mortality shocks to humans and livestock). Further, the agricultural production process is one in which inputs have to be committed in advance of the realization of an uncertain harvest, while the process of consumption is more certain and evenly paced over time. In an economy closed to international agricultural commodity markets, shocks to domestic output and demand have to be absorbed through price changes. To the extent such shocks are not highly correlated across economies, world output and consumption would be more stable than their domestic counterparts. Thus, integration with world markets in effect would provide insurance against price effects of domestic shocks, though not necessarily against their income effects.

It is clear that, even if production and consumption processes were to be free of any risk and uncertainty, still the lack of synchronization between the two would require means for smoothing consumption over time. It is also clear that achieving such smoothing would be less expensive, compared to each individual holding inventories of inputs and consumption goods, if access to smoothly and efficiently functioning credit markets is available. The need for credit is enhanced also if purchased inputs (e.g. fertilizers and pesticides, energy and fuels, hired labor etc.) account for a large share of production costs, as in the case of the cultivation of the so-called Green Revolution varieties of crops. With well-functioning insurance markets, insurable risks would be addressed. However, uninsurable (or more precisely, insurable only at a high cost) risks are also significant in rural areas of poor countries.

For well-known and well-understood reasons of moral hazard, absence of collateralizable assets, poorly functioning legal system for enforcement of contracts and the seizing and sale of whatever collateral that has been pledged, formal credit and insurance markets in poor countries are either virtually absent or costly, if not altogether out of reach, of the poor. On the other hand, informal arrangements substitute in part for transactions in formal markets (Townsend, 1994; Udry, 1993). However, the cost of informal transactions is not necessarily low and, in any case, informal arrangements are nowhere near adequate to substitute fully for the incomplete and imperfect functioning of credit and insurance markets.

Globalization for the poor

The globalization discussed in the last section looks substantially different than the partially integrated goods, capital, and factor markets we have today. The most notable departures, and those with the largest negative effects on the poor, are the continued industrial country protection of agricultural and other markets and the lack of integration in labor markets. The industrial countries' practice of protecting vulnerable markets may be thought of as politically difficult to change, but that does not justify acceptance. The answer may be to devote more attention to strengthening international trade policy-making bodies.

Labor market integration has been similarly underemphasized in the current wave of globalization. As if these two changes were not substantial enough, it is also important to emphasize that globalization alone will not bring about the changes in the economic and institutional environment that are most effective for reducing poverty. It creates additional pressure for such reforms, but these are ultimately domestic decisions.

Notes

* Samuel C. Park, Jr Professor of Economics, Yale University, New Haven, Connecticut USA. This paper draws extensively from Srinivasan (1993), Srinivasan (2001a), Srinivasan (2001b), and Srinivasan and Wallack (2004).

1 This section draws from Srinivasan (1993).
2 This section draws from Srinivasan (2001a).
3 This section draws from Srinivasan (2001b).
4 This section draws from Srinivasan and Wallack (2004).
5 The paper presents a variety of theoretical reasons for this finding, though the composite index of "globalization"—a weighted average of trade and financial openness—is difficult to interpret, however, in the context of his cross-country study.
6 Deardorff (2003) goes so far as to suggest that globalization's critics are motivated by a different understanding of how the world works than the globalization supporters.
7 Srinivasan (1995).
8 For example, Coe *et al*., 1998.
9 The phrases "engine of growth" and "handmaiden of growth" are associated with Dennis Robertson (1940) and Irwin Kravis (1970), respectively.
10 Tinbergen (1962) suggests, for example, that subsidies be given to retrain workers and facilitate the transfer of capital from declining industries to new, more dynamic sectors. (p. 42).

11 Mussa (2000). See Kaplan and Rodrik (2002) on the second point.
12 Rodriguez and Rodrik (2000) are the most recent critics, though Warner (2003) refutes the Rodriguez-Rodrik critique. He essentially argues that their finding that the effect of trade openness on growth is not statistically robust is due to the fact that the forms of protectionism vary across countries so that any single indicator of openness will not describe the effective level of integration for all countries.
13 Besley and Burgess (2003) discuss the poverty reduction benefits of a similar set of policies, though they do not address the link with globalization.
14 Critics often associate globalization with a reduction in the state's ability to redistribute income. We discuss this point further in the next subsection.
15 Clearly, if there are no distortions in all these markets and all individuals and households face the same prices, the extent of poverty would be determined by the distribution of assets and labor in the economy.
16 See, for example, Tinbergen (1962), Section 1.4 or Tinbergen (1975), Section 10.2.
17 For surveys of relevant analytical and policy issues, see Aghion *et al*. (1999) and Bénabou (1996).
18 Rao and Radhakrishna (1997) analyzed India's PDS from the national and international perspective. They found that in 1986–1987, PDS and other consumer subsidy programs accounted for only about 3% of the per capita expenditure of the poor, and their impact on poverty and nutritional status of the poor was minimal.
19 There are, empirically, important exceptions to this general theoretical expectation: Harrison and Hanson (1999) present evidence that trade openness in several Latin American and Asian countries has been associated with an increased return to skilled labor relative to unskilled labor.
20 Wage labor market policies are obviously not a solution for all of poverty. Only a small part of the labor force in many developing countries (less than 20% in India and South Asia, for example) is in formal wage and salary employment. An overwhelming majority of the labor force is in self-employment (often in subsistence farming, in handicraft activities and household-based production for local markets). For them it is not so much the functioning of labor markets but that of product and credit markets that is more relevant.
21 National and international attempts to eradicate child labor through restrictions on imports or consumer boycotts of goods produced by children are not likely to succeed unless the basic cause of child labor, namely, the poverty of parents, is addressed.
22 See McMillan *et al*. (2002) on how domestic market imperfections limited farmers' benefits from liberalization of the cashew sector in Mozambique.
23 On the other hand, China, a country in which corruption is thriving, has attracted large flows of investment, particularly from overseas Chinese who apparently are better at operating in a corrupt system! Corrupt practices by transnational enterprises, often with the connivance of the governments in countries of their origin, have received attention in the literature.
24 Wei (2000b).

Bibliography

Agènor, Pierre-Richard (2004) Does Globalization Hurt the Poor? *International Economics & Economic Policy*, 1(1): 21–51.

Aghion, Philippe, Eve Caroli, and Cecilia Garcia-Penalosa (1999) "Inequality and Economic Growth: The Perspective of the New Growth Theories," *Journal of Economic Literature*, XXXVII (December): 1615–1660.

Ahmed, Sultan (1994) "Improving Inter-spatial and Inter-temporal Comparability of National Accounts," *Journal of Development Economics*, 44(1): 53–75.

Atkinson, A. B. and A. Brandolini (1999) *Promise and Pitfalls in the Use of "Secondary" Data Sets: Income Inequality in OECD Countries*, Oxford: Nuffield College, mimeo.

Bénabou, Roland (1996) "Inequality and Growth," *NBER Macroeconomics Annual*, 11, MIT Press: 11–74.

Besley, Timothy, and Robin Burgess (2003) "Halving Global Poverty," *Journal of Economic Perspectives*, 17(3): 3–22.

Bhagwati, Jagdish (1966) *The Economics of Underdeveloped Countries*, New York: McGraw-Hill.

—— (1968) "The Theory and Practice of Commercial Policy," Frank Graham Memorial Lecture, Princeton University, 1967, International Finance Section, Princeton.

—— (1978) *Foreign Trade Regimes and Economic Development: Anatomy of Exchange Control Regimes*, Cambridge: Ballinger.

—— (2004) *In Defense of Globalization*, New York: Oxford University Press.

Bhagwati, Jagdish and Bent Hansen (1973) "Should Growth Rates be Evaluated at International Prices?" in J. Bhagwati and R. S. Eckaus (eds), *Development and Planning*, Cambridge, MA: MIT Press, Chapter 3: 53–68.

Coe, David T., Elhanan Helpman, and Alexander W. Hoffmaister (1998) "North-South R&D Spillovers," *Economic Journal*, 107 (January): 134–149.

Deardorff, Alan (2003) "What Might Globalization's Critics Believe," Occasional Paper 20, Bureau of Economic Studies, Macalester College, St Paul, MN.

Deaton, Angus (1989) "Saving in Developing Countries: Theory and Review, Proceedings of the World Bank Annual Conference on Development Economics," Supplement to the *World Bank Economic Review*: 61–96.

—— (2000) "Counting the World's Poor: Problems and Possible Solutions. 2000," Program in Development Studies of Economics, Princeton University, Princeton, NJ, Processed.

Deaton, A. and A. Tarozzi (2005) "Prices and Poverty in India," in A. Deaton and V. Kozel (eds), *The Great Indian Poverty Debate*, New Delhi, Macmillan Ltd: 381–411.

Deninger, K. and L. Squire (1996) "New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10: 565–591.

Fields, Gary (1994) "Data for Measuring Poverty and Inequality Changes in the Developing Countries," *Journal of Development Economics*, 44(1): 87–102.

Goldberg, Pinelopi and Nina Pavcnik (2004) "Trade Inequality and Poverty: What Do We Know?" in C. Graham and S. Collins (eds), *Brookings Trade Forum*: *2004*, Washington, DC: Brookings Institution: 223–270.

Harrison, Anne, and Gordon Hanson (1999) "Who Gains From Trade Reform? Some Remaining Puzzles," *Journal of Development Economics*, 59: 129–144.

Heston, Alan (1994) "A Brief Review of Some Problems in Using National Accounts Data in Level of Output Comparisons and Growth Studies," *Journal of Development Economics*, 44(1): 87–102.

Kaplan, Ethan and Dani Rodrik (2002) "Did Malaysian Capital Controls Work?" in S. Edwards and J. Frankel (eds), *Preventing Currency Crises in Emerging Markets*, Chicago, University of Chicago Press: 393–440.

Kravis, Irving (1970) "Trade as a Handmaiden of Growth: Similarities between the Nineteenth and Twentieth Centuries," *Economic Journal*, 80 (December): 850–872.

Krueger, Anne (1978) *Foreign Trade Regimes and Economic Development: Liberalization Attempts and Consequences*, Cambridge: Ballinger.

Little, I. M. D. and J. A. Mirrless (1974) *Project Appraisal and Planning for Developing Countries*, London: Heinemann.

Little, I. M. D., T. Scitovsky, and M. F. G. Scott (1970) *Industry and Trade in Some Developing Countries*, Oxford: Oxford University Press.

McMillan, Margaret, Rodrik, Dani, and Karen Horn Welch (2003) "When Economic Reform Goes Wrong: Cashews in Mozambique," in S. Collins and D. Rodrik (eds), *Brookings Trade Forum*: *2003*, Washington, DC, Brookings Institution: 97–165.

Mukherjee, M. and G. S. Chatterjee (1974) "On the Validity of NSS Estimates of Consumption Expenditure," in T. N. Srinivasan and P. K. Bardhan, (eds) *Poverty and Income Distribution in India*, Calcutta: Indian Statistical Institute: 139–147.

Mussa, Michael (2000) "Factors Driving Global Economic Integration," Paper presented in Jackson Hole, Wyoming at a symposium sponsored by the Federal Reserve Bank of Kansas City on *Global Opportunities and Challenges*, August 25, 2000.

Rao, K. S. and R. Radhakrishna (1997) *India's Public Distribution System in a National and International Perspective*, Washington, DC: World Bank.

Razin, Assaf (2002) "Tax Burden and International Migration: Political-Economic Theory and Evidence," *Journal of Public Economics*, 85(2): 167–190.

Robertson, D. H. (1940) "The Future of International Trade," in *Essays in Monetary Theory*, Stockholm: 209–226.

Rodriguez, Fransisco and Dani Rodrik (2000) "Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence," *NBER Macroeconomics Annual*,15.

Sachs, Jeffrey and Andrew Warner (1995) "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity*, 1: 1–118.

Sen, Amartya (1973) *On Economic Inequality*, London: Oxford University Press.

—— (1976) "Poverty: An Ordinal Approach to Measurement," *Econometrica*, 44: 219–231.

—— (1981) *Poverty and Famines*, Oxford: Clarendon Press.

—— (1984) *Resources, Values and Development*, Parts IV and V, Delhi, Oxford University Press.

—— (1985) "A Sociological Approach to the Measurement of Poverty: A Reply [Poor, Relatively Speaking]," *Oxford Economic Papers*, 37(4): 669–676.

—— (1987) *The Standard of Living*, Cambridge: Cambridge University Press.

Soligo, R. and J. Stern (1965) "Tariff Protection, Import Substitution, and Investment Efficiency," *Pakistan Development Review*, 5(2) (summer): 249–270.

Srinivasan, T. N. (1993) "Rural Poverty: Conceptual, Measurement, and Policy Issues," in M. G. Quibria, (ed.) *Rural Poverty in Asia: Priority Issues and Poverty Options*, Hong Kong: Oxford University Press: 103–123.

—— (2001a) "Comments on 'Counting the World's Poor," *The World Bank Research Observer*, 16(2): 157–168.

—— (2001b) "Growth and Poverty Alleviation: Lessons from Development Experience," Paper prepared for IMF Institute Seminar on *Growth, Poverty and Inequality: Theory, Empirics and Policy*, August 2, 2002, published in French with the title *Croissance et allégement de la pauvreté: les leçons tirées de l'expérience du développement* in *Revue d'économic du développemet*, January–February: 115–168.

Srinivasan, T. N., P. N. Radhakrishnan and A. Vaidyanathan (1974) "Data on Distribution of Consumption Expenditure in India: An Evaluation," in T. N. Srinivasan and P. K. Bardhan (eds) *Poverty and Income Distribution in India*, Calcutta: Indian Statistical Institute: 148–166.

Srinivasan, T. N. and Jessica Seddon Wallack (2004) "Globalization, Growth and the Poor," *De Economist*, 152(2): 251–272.

Stiglitz, Joseph E. (2002) *Globalization and Its Discontents*, New York: W. W. Norton.

Tinbergen, Jan (1962) *Shaping the World Economy*, New York: The Twentieth Century Fund.
—— (1975) *Income Distribution*, New York: Elsevier.
Townsend, Robert (1994) "Risk and Insurance in Village India," *Econometrica*, 62(3): 539–591.
Udry, Chris (1993) "Case Studies: Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy," Chapter 5 in Karla Hoff, Avishay Braverman, and Joseph E. Stiglitz (eds) *The Economics of Rural Organization: Theory, Practice and Policy*, Washington, DC: Oxford University Press for the World Bank.
Wacziarg, Romain and Karen Welch (2003) "Trade Liberalization and Growth: New Evidence," Working Paper 10152, National Bureau of Economic Research, Cambridge, MA.
Warner, Andrew (2003) "Once More Into the Breach: Economic Growth and Global Integration," Working Paper No.34, December 2003, Center for Global Development.
Wei, Shang-Jin (2000a) "How Taxing is Corruption on International Investors?" *Review of Economics and Statistics*, February, 82(1): 1–11.
—— (2000b) *Natural Openness and Good Government*, Harvard University. Mimeo.
UNDP (1996) *Human Development Report*, New York: United Nations Development Program.
—— (1997) *Human Development Report*, New York: United Nations Development Program.
UN Millennium Project (2005a) *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals*, Report to the UN Secretary-General, New York: Millennium Project.
—— (2005b) *Trade for Development: Achieveing the Millenium Development Goals*, Report of the Task Force on Trade, UN Millennium Project, Lead authors Ernesto Zedillo, Patrick Messerlin and Julia Nielson, London: Earthscan.
Winters, L. Alan, Neil McCulloch, and Andrew McKay (2004) "Trade Liberalization and Poverty: The Evidence So Far," *Journal of Economic Literature*, XLII(1): 72–115.
World Bank (1999) *World Development Indicators*, Washington, DC: World Bank.
—— (2002) *World Development Indicators*, Washington, DC: World Bank.
WTO (2003) *World Trade Report*, Geneva: World Trade Organization.

2 Globalization and poverty

What is the evidence?

Emma Aisbett, Ann Harrison, and Alix Peterson Zwane

Introduction

This chapter reviews the evidence on the linkages between globalization and poverty, drawing on the collected works of Jagdish Bhagwati and the results of an National Bureau of Economic Research (NBER) project directed by Ann Harrison, *Globalization and Poverty*. We focus on two measures of globalization: trade integration (measured using tariffs or trade flows), and international capital flows. Many economists have used the Heckscher–Ohlin framework in international trade to argue that the unskilled or the poor in countries with a comparative advantage in unskilled labor are most likely to gain from trade reform. Our first conclusion is that such a simple interpretation of general equilibrium trade models is likely to be misleading. Second, the evidence discussed suggests that the poor are more likely to share in the gains from globalization when there are complementary policies in place. Such complementary policies include programs to promote human capital development, infrastructure development, credit and technical assistance to farmers, and macroeconomic stability. Third, we find that trade and foreign investment reforms have produced benefits for the poor, particularly those in exporting sectors or sectors which receive foreign investment. Fourth, financial crises are very costly to the poor. Finally, the collected evidence suggests that globalization produces both winners and losers among the poor. The fact that some poor individuals are made worse off by trade or financial integration suggests the need for carefully targeted safety nets. We emphasize the heterogeneity of results across different countries and settings, but also present cross-country evidence, which suggests that the path from globalization to poverty reduction via the growth effects of trade reforms is likely to be important.

> Today, thanks to television, we have what I call the paradox of inversion of the philosopher David Hume's concentric circles of reducing loyalty and empathy… What the Internet and CNN have done is to take Hume's outermost circle and turn it into the innermost. No longer can we snore while the other half of humanity suffers plague and pestilence and the continuing misery of extreme poverty.
>
> (Jagdish Bhagwati, *In Defense of Globalization*, 2004, p. 18)

The impact of globalization on the poor has been one of the most intensely fought areas of the globalization debate, and with good reason. Sustainable poverty reduction is one of the most important challenges facing the world today, and globalization is a powerful force affecting the well-being of the world's people. How globalization can be used to promote growth and reduce poverty in developing countries has been an ongoing theme in Professor Jagdish Bhagwati's work. This chapter was written in honor of Professor Bhagwati, who is without a doubt the most influential academic in this field.

One of Professor Bhagwati's many outstanding contributions as an academic has been his commitment and ability to communicate his extraordinary understanding to a broad audience. He has written extensively on the topic of globalization and poverty; including in his most recent book, *In Defense of Globalization*[1] Professor Bhagwati has been particularly effective at reminding his audience of important insights which are at risk of being lost due to either misinformation or misunderstanding of the scholarly debate by the general public.

This study provides an economic perspective on how globalization affects poverty in developing countries. Our objective is twofold. First, we seek to highlight Professor Bhagwati's writings on the linkages between globalization and poverty reduction. Second, we survey recent empirical evidence on those linkages, drawing primarily from a forthcoming book, *Globalization and Poverty*, commissioned by the NBER. The fifteen studies and fifteen discussions that are part of the NBER project, which was directed by Ann Harrison, ask the following questions: how has global economic integration affected the poor in developing countries? Do trade reforms that eliminate or reduce import protection lead to rising or falling poverty? Has increasing financial integration led to more or less poverty? How have the poor fared during currency crises? Do agricultural support programs in rich countries hurt the poor in developing countries, as some critics argue? Or do such policies in fact provide assistance by reducing the cost of food imports? Finally, does food aid help or hurt the poor? Although the concept of "globalization" is quite broad, we focus on two important aspects: (1) international trade in goods and (2) capital flows—including foreign investment, portfolio flows, and aid. Of course, this definition is not all-encompassing: economic aspects of globalization have also affected information flows, migration, and trade in services. However, we focus primarily on trade and capital flows, as these have been the focus of Professor Bhagwati's voluminous writings and are at the center of intense policy debates.

Several recent surveys seek to identify the relationship between globalization and poverty (see for example, Goldberg and Pavcnik, 2004, and Ravallion, 2004, Winters *et al*., 2004). However, the authors of these surveys acknowledge that they can only review the *indirect* evidence regarding the linkages between globalization and poverty. There have been almost no studies which test for the direct linkages between the two. Winters *et al*. (2004) write in their insightful and comprehensive *Journal of Economic Literature* (JEL) survey that "there are no direct studies of the poverty effects of trade and trade liberalization" (JEL, p. 73). Goldberg and Pavcnik's (2004) excellent review points out that "while the

literature on trade and inequality is voluminous, there is virtually no work to date on the relationship between trade liberalization and poverty." The few studies which do examine the links between globalization and poverty, including several cited in the Winters *et al*. (2004) survey and Ravallion (2004), typically use computable general equilibrium models to disentangle the complex linkages between trade reform and poverty. However, while such research provides an important contribution to our understanding of the channels through which globalization or future reforms could affect poverty, it is extremely important to be able to look at actual *ex post* evidence of the impact of trade and investment reforms on the poor.

There are several reasons why the links between globalization and poverty have not been adequately explored in the past. One reason is that academic researchers who address questions of poverty and globalization have typically chosen not to achieve mastery of both subdisciplines. Other reasons for the limited evidence are the methodological problems associated with linking trade to poverty outcomes. Simply producing comparable measures of poverty over time within a single country is considered an accomplishment (see Deaton, 2001). On the trade side, measuring and properly identifying the effects of trade policy on growth has spawned an enormous and acrimonious debate. Thus, it is not surprising that attempting to directly relate measures of globalization and poverty poses a significant challenge. Yet there is a pressing need for some answers. If globalization is accompanied by increasing inequality, but *both* the incomes of the rich and poor are rising, this is a very different picture than if globalization has led to absolute income gains for some income groups but real income losses for others.

What are the mechanisms through which globalization could affect poverty? One important possible mechanism is through globalization's impact on growth. As we discuss later in this chapter, growth is typically good for the poor: if globalization increases a country's growth rate, then that growth is likely to reduce poverty. Apart from its impact via aggregate growth, trade reform could directly affect the welfare of the poor by changing the relative prices they face as consumers and producers. If liberalization leads to falling prices for goods purchased by poor consumers, this could reduce poverty. If globalization raises the prices of goods produced by the poor—such as agricultural goods or textiles and apparel—then poverty is also likely to decline. In addition, international trade could affect poverty through its impact on the incomes and employment opportunities of poor wage earners.

What does the evidence show on the linkages between globalization and poverty? Many economists have used the Heckscher–Ohlin framework in international trade to argue that the unskilled or the poor in countries with a comparative advantage in unskilled labor are most likely to gain from trade reform. The first lesson is that the poor do not always gain from trade. Why not? One reason is that labor is not nearly as mobile as simple trade models assume; for comparative advantage to increase the incomes of the unskilled, they need to be able to move out of contracting sectors and into expanding ones. Another reason is that developing countries have historically protected their unskilled-intensive sectors

(although this is less true in the case of agriculture), which implies that in the manufacturing sector trade reforms frequently result in less protection for unskilled workers relative to skilled labor. A third reason is that even sectors which are relatively unskilled-intensive in a global context may require workers with more skills than the poor in developing countries typically possess.

A second lesson that emerges from a review of country case studies is that the poor are more likely to share in the gains from globalization when there are complementary policies in place. Studies prepared for *Globalization and Poverty* on India (Topalova, 2006) and Colombia (Goldberg and Pavcnik, 2006) suggest that globalization is more likely to benefit the poor if trade reforms are implemented in conjunction with labor market deregulation. In Zambia, poor farmers are only expected to benefit from greater access to export markets if they also have access to credit, technical know-how, and other complementary inputs (Balat and Porto, 2006). The studies also point to the importance of social safety nets. In Mexico, if poor corn farmers did not receive income support from the government, their real incomes would have been halved during the 1990s (Ashraf *et al*., 2006). In Ethiopia, if food aid had not been well targeted, globalization would have had little impact on the poor (Levinsohn and McMillan, 2006).

Third, the evidence suggests that trade and foreign investment reforms in a number of countries have contributed towards reducing poverty. In Mexico, the poor in the most globalized regions have weathered macroeconomic crises better than their more isolated neighbors (Hanson, 2006). In India, opening up to foreign investment was associated with a decline in poverty. The study on Zambia suggests that poor consumers gain from falling prices for the goods they buy, while poor producers in exporting sectors benefit from trade reform through higher prices for their goods. In Colombia, increasing export activity was associated with an increase in compliance with labor legislation and a fall in poverty. In Poland, unskilled workers—who are the most likely to be poor—have gained from Poland's accession to the European Union.

Fourth, both the cross-country and individual case studies suggest that financial crises are very costly to the poor. For the NBER project, Prasad *et al*. (2006) study financial deregulation across countries and find that lower income countries who embark on financial globalization are likely to experience higher consumption and output volatility. Their work reinforces the need for complementary policies, such as the creation of reliable institutions and macroeconomic stabilization policies (including the use of flexible exchange rate regimes). While financial crises resulting from unrestricted capital flows are associated with a higher likelihood of poverty, foreign direct investment inflows are associated with a reduction in poverty. The poverty-reducing effects of Foreign Direct Investment (FDI) are clearly documented in the studies on India and Mexico.

The final lesson from the breadth of cross-country studies and individual country experiences is that globalization produces both winners and losers among the poor. It should not be surprising that the results defy easy generalization. Even within a single region, two sets of farmers may be affected in opposite ways. In Mexico, while small corn farmers saw their incomes fall by half in the 1990s,

large corn farmers gained. Across different countries, poor wage earners in exporting sectors or in sectors with incoming foreign investment gained from trade and investment reforms; conversely, poverty rates increased in previously protected sectors that were exposed to import competition. Within the same country or even the same region, a trade reform may lead to income losses for rural agricultural producers and income gains for rural or urban consumers of those same goods.

Part 2 of this chapter summarizes the results from the cross-country studies, while Part 3 describes the results of the country case studies, which analyze the impact of globalization on employment opportunities and labor income of the poor, as well as on consumption and production opportunities for the poor. The studies that address the impact of capital flows on the poor are summarized in Part 4. Since the evidence suggests that globalization creates winners as well as losers among the poor, this chapter moves in Part 5 to a discussion of why globalization's critics seem all too aware of the costs of globalization and generally fail to see the benefits. We argue that this is due to the use of different methodologies in estimating poverty and inequality, the concerns of globalization's critics about the short term costs versus the longer term gains from trade reform, their rejection of a perfectly competitive framework, and different interpretations regarding the evidence. Another reason is the lack of knowledge on the possible linkages between globalization and poverty reduction, a missing link which Professor Jagdish Bhagwati's books and the forthcoming NBER volume seek to address.

Aggregate linkages and cross-country evidence

Professor Bhagwati has argued in many of his articles, books, and lectures that trade generally enhances growth, and growth reduces poverty. He reminds his readers that for trade to permanently affect growth rates it must act through at least one of two growth fundamentals: accumulation and innovation. He then points out that trade can have a positive impact on both of these fundamentals through a variety of channels, including specialization, scale economies, increased competition, incentives for macroeconomic stability, and increased marginal efficiency of imported capital. Thus, through its impact on growth, trade is on average good for the poor. However, Professor Bhagwati strikes two notes of caution. First, growth was (rightfully) never meant to be an end in itself. Rather, growth is the most broadly effective poverty reduction strategy that has been found. Second, growth can bypass, or even immiserize the poor: not all growth is created equal.[2] Bhagwati also suggests that government policies can affect income distribution, which has important implications for poverty. In *In Defense of Globalization* he contrasts the outward orientation of the East Asian countries with the inward orientation of the Indian government's policies up until the last decade or so. The former resulted in strong, labor intensive growth; the latter with weak, capital intensive growth. Not surprisingly, the former policy was far more effective at reducing poverty. Some of India's policies since independence

have, however, been great successes for the poor. In particular, Bhagwati notes the policies that successfully enabled small farmers to participate in the green revolution. These policies prevented what could have been seriously immiserizing impacts of a major source of growth in the economy. Thus, Bhagwati reminds us that a belief in the superiority of markets over bureaucrats as a means of allocating resources is not equal to advocating *laissez-faire*, hands-off, passive strategies for poverty reduction.

What does the evidence on the relationship between openness and poverty indicate? The most direct approach to answering this question would be to examine the aggregate relationship between different poverty measures and globalization. In Tables 2.1 through 2.4, we present evidence on the linkages between openness, GDP growth, and different measures of poverty. We begin by revisiting the evidence on the linkages between trade and growth; these results are presented in Tables 2.1 and 2.2.

We use two different measures of openness to trade: (1) the ratio of trade (X + M) to GDP in nominal terms and (2) average tariffs, defined as import revenues divided by imports. We find that an increase in openness—using these two measures—is associated with an increase in aggregate income or an increase in aggregate income growth. To address concerns regarding endogeneity, we measure openness either as the three year lag of trade shares or tariffs or the contemporaneous value for openness instrumented using lagged values. We also explore the robustness of the results to including other controls, such as country fixed effects or policy variables likely to be correlated with trade policies. Additional extensions, using growth of GDP per capita as the dependent variable instead of income per capita, yielded similar results. The evidence presented here is consistent with Professor Bhagwati's assessment of the enormous literature on cross-country trade and growth regressions. Bhagwati (2004) notes that not all of the literature shows a positive relationship between trade and growth, but that the evidence "by and large, is consonant with the views of the free trade proponents." (*In Defense of Globalization*, p. 64)

In the course of writing this chapter and completing the NBER study, we were surprised to learn that there has been almost no research on the association between globalization and measures of poverty based on household survey data (for the problems associated with using national income data—an approach adopted by Dollar and Kraay (2002, 2004) to measure poverty—see Deaton, 2003). One likely reason is that there are very few data points available over time and across countries. In columns (5) and (10) of Tables 2.1 and 2.2, we redo the basic specifications, but restrict the sample to the observations for the country-years where there exists poverty data based on the household surveys. Once we restrict the sample to the observations with information on poverty, the link between openness to trade and GDP per capita in levels or growth rates weakens significantly. Other policies continue to matter in the restricted sample, including inflation—which is negatively associated with growth—and currency crises, which also negatively affect incomes per capita. The weakness of the association between openness and growth in this small sample suggests that efforts to find

Table 2.1 Incomes and trade shares in a cross-section of countries

	Dependent variable: Ln income per capita ($1993 PPP)									
	Reduced form					*Instrumental variables*				
	(1) OLS	*(2) OLS*	*(3) OLS*	*(4) OLS*	*(5) OLS*	*(6) IV*	*(7) IV*	*(8) IV*	*(9) IV*	*(10) IV*
3-year lag trade share	0.907 [0.036]***	0.514 [0.037]***	0.214 [0.038]***	0.203 [0.035]***	0.081 [0.074]					
Trade share						0.978 [0.037]***	0.857 [0.057]***	0.426 [0.067]***	0.402 [0.064]***	0.248 [0.167]
Ln (1 + CPI)			−0.027 [0.013]**	−0.090 [0.018]***	−0.020 [0.019]			−0.035 [0.014]**	−0.094 [0.018]***	−0.034 [0.018]*
Government expenditure/ nominal GDP			0.159 [0.174]	0.264 [0.142]*	0.625 [0.261]**			0.230 [0.180]	0.298 [0.142]**	0.840 [0.219]***
Currency crisis			−0.039 [0.011]***	−0.039 [0.013]***	0.013 [0.022]			−0.070 [0.012]***	−0.060 [0.013]***	−0.000 [0.018]
GDI/real GDP			0.298 [0.126]**					0.116 [0.126]		
Fraction of population literate			−0.208 [0.146]					−0.171 [0.146]		
Constant	7.239 [0.028]***					7.173 [0.030]***				
Country fixed effects?	NO	YES	YES	YES	YES	NO	YES	YES	YES	YES
Time fixed effects	NO	NO	YES	YES	YES	NO	NO	YES	YES	YES
Observations	3,294	3,294	1,996	2,657	308	3,288	3,288	1,996	2,657	308
R^2	0.17	0.93	0.96	0.95	0.99					
Restricted sample?	NO	NO	NO	NO	YES	NO	NO	NO	NO	YES

Notes
Restricted sample is country-year observations for which poverty (head count) data is available; All regressions exclude OECD high-income countries; Huber robust standard errors in square brackets. Significantly different from zero at 90% (*) 95% (**) 99% (***) confidence; In IV regressions, trade share instrumented using three-year lagged value.

Table 2.2 Income per capita and average import tariffs in a cross-section of countries

	Dependent variable: Ln income per capita ($1993 PPP)									
	Reduced form					*Instrumental variables*				
	(1) OLS	*(2) OLS*	*(3) OLS*	*(4) OLS*	*(5) OLS*	*(6) IV*	*(7) IV*	*(8) IV*	*(9) IV*	*(10) IV*
3-year lag average I tariff	−3.586 [0.377]***	−0.721 [0.142]***	−0.298 [0.117]**	−0.137 [0.119]	−0.250 [0.281]					
Average import tariff						−4.830 [0.441]***	−4.830 [0.441]***	−0.635 [0.328]*	−0.338 [0.379]	−1.831 [1.563]
Ln (1 + CPI)			−0.047 [0.014]***	−0.095 [0.016]***	−0.050 [0.017]***			−0.055 [0.013]***	−0.106 [0.013]***	−0.054 [0.015]***
Goverenment expenditure/ nominal GDP			0.360 [0.224]	0.379 [0.181]**	0.418 [0.290]			0.410 [0.245]*	0.395 [0.161]**	0.676 [0.349]*
Currency crisis			−0.026 [0.013]*	−0.018 [0.014]	0.040 [0.029]			−0.027 [0.014]**	−0.018 [0.016]	0.003 [0.034]
GDI/real GDP			0.345 [0.157]**					0.572 [0.171]***		
Fraction of population literate			0.398 [0.211]*					0.482 [0.216]**		
Constant	8.368 [0.045]***					8.108 [0.194]***				
Country fixed effects?	NO	YES	YES	YES	YES	NO	YES	YES	YES	YES
Time fixed effects	NO	NO	YES	YES	YES	NO	NO	YES	YES	YES
Observations	1,617	1,617	1,261	1,485	212	1,415	1,415	1,125	1,306	189
R^2	0.07	0.96	0.97	0.96	0.99					
Restricted sample?	NO	NO	NO	NO	YES	NO	NO	NO	NO	YES

Notes
Restricted sample is country-year observations for which poverty (head count) data is available; All regressions exclude OECD high-income countries; Huber robust standard errors in square brackets. Significantly different from zero at 90% (*) 95% (**) 99% (***) confidence; In IV regressions, average import tariff instrumented using three-year lagged value.

any direct relationship between openness and poverty reduction are likely to be plagued by limited data availability. Nevertheless, we present those results below.

We examine the relationship between measures of openness, GDP growth, and poverty in Tables 2.3 and 2.4.

Measures of poverty are derived from household sample surveys made available by the World Bank. We use two different measures of poverty: the percentage of households living on less than $1 a day in purchasing power parity (PPP) terms, and the level of income earned by the poorest decile. The evidence in Tables 2.3 and 2.4 suggest that growth is indeed good for the poor. We use several different measures of income: contemporaneous income, income lagged three periods, and contemporaneous income instrumented using annual average levels of precipitation and temperature. Across all specifications, aggregate income or aggregate income growth (not shown here) is associated with a reduction in the percentage of the population that is poor.

The strong association between aggregate growth and poverty reduction is consistent with Professor Bhagwati's interpretation of the evidence. With regard to empirical evidence on the relationship between growth and poverty reduction, Bhagwati quotes from the work of Xavier Sala-i-Martin (2002). Sala-i-Martin concluded strongly that poverty in Asia fell because Asian countries grew, while poverty in Africa increased dramatically because African countries did not grow. According to Bhagwati (*In Defense of Globalization*, 2004, p. 65) Sala-i-Martin's findings are "as strong a corroboration as I can find of my 1960s conjecture that growth must be reckoned to be the principle force in alleviating poverty."

Although the results presented in Tables 2.1 through 2.4 suggest a strong link from trade integration to aggregate incomes and from income growth to poverty reduction, the evidence on direct linkages between trade shares or tariffs and poverty outcomes is quite weak, and disappears if we control for country fixed effects. Nevertheless, the association always goes in the same direction: greater openness, measured as either an increase in trade shares or a reduction in tariffs, is associated with a reduction in poverty. All the results that are statistically significant suggest that greater openness is associated with reduction in the percentage of the population living on less than one PPP dollar or two PPP dollars a day. However, the results are not robust to instrumental variable (IV) estimation and controlling for country fixed effects. Similar results were found when using different poverty measures—such as the percentage of the poor living on less than 2 PPP dollars per day, or the incomes of the poorest quintile or decile.

To summarize, there is certainly no evidence in the aggregate data that trade reforms are bad for the poor. This is true if one uses trade shares—not an ideal measure of trade policy since trade shares measure outcomes, not policies—or tariffs, which are a more appropriate measure of trade policy. In a comparable exercise using country-level poverty headcounts and trade shares, Ravallion (forthcoming) reaches a similar conclusion; he argues that there is no robust relationship between poverty and globalization in the aggregate data.[3]

However, the cross-country results presented in this volume and in earlier studies should be considered as a first step in this research.[4] Owing to limited

Table 2.3 Head count poverty ($1 per day) and trade shares in a cross-section of countries

	Dependent variable: Ln fraction of households living on less than $1 per day ($1993 PPP)									
	Reduced form						*Instrumental variables*			
	(1) OLS	*(2) OLS*	*(3) OLS*	*(4) OLS*	*(5) OLS*	*(6) OLS*	*(7) IV*	*(8) IV*	*(9) IV*	*(10) IV*
3-year lag trade share	−1.921 [0.385]***	−1.772 [0.502]***	−1.517 [0.484]***	0.418 [0.931]	0.685 [1.209]	0.579 [1.086]				
Trade share							−2.164 [0.760]***	−1.767 [0.524]***	−3.609 [2.240]	2.261 [3.576]
Ln (1 + CPI)		−0.348 [0.396]	−0.879 [0.396]**		0.291 [0.382]	−0.022 [0.367]	0.824 [0.557]	−0.845 [0.349]**	−0.014 [0.391]	−0.145 [0.366]
Government expenditure/ nominal GDP		−2.945 [3.757]	−8.275 [4.172]**		5.299 [4.503]	−3.058 [4.222]	−5.266 [5.629]	−7.589 [3.918]*	−1.935 [5.929]	−1.270 [4.237]
Currency crisis		−0.153 [0.402]	0.118 [0.446]		−0.143 [0.455]	−0.281 [0.444]	−0.609 [0.624]	0.150 [0.423]	−0.187 [0.375]	−0.386 [0.373]
GDI/real GDP		−0.117 [2.486]			0.696 [5.203]		1.171 [3.606]		4.806 [3.997]	
Fraction of population literate		1.997 [0.912]**			1.673 [6.827]		11.779 [4.606]**		2.397 [7.459]	
3-year lag ln income per capita		−2.225 [0.252]***			−5.154 [1.646]***		−6.405 [2.070]***		−0.380 [3.926]	
Constant	−1.874 [0.252]***									
Country fixed effects?	NO	NO	NO	YES	YES	YES	NO	NO	YES	YES
Time fixed effects	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES
Observations	349	284	325	349	284	325	229	325	229	325
R^2	0.06	0.38	0.14	0.65	0.67	0.67				

Notes
All regressions exclude OECD high-income countries; Huber robust standard errors in square brackets. Significantly different from zero at 90% (*) 95% (**) 99% (***) confidence; In IV regressions, trade share instrumented using 3-year lagged value and income instrumented using precipitation and temperature.

Table 2.4 Head count poverty ($1 per day) and average import tariffs in a cross-section of countries

	Dependent variable: Ln fraction of households living on less than $1 per day ($1993 PPP)									
	Reduced form						*Instrumental variables*			
	(1) OLS	*(2) OLS*	*(3) OLS*	*(4) OLS*	*(5) OLS*	*(6) OLS*	*(7) IV*	*(8) IV*	*(9) IV*	*(10) IV*
3-year lag average I tariff	7.543 [1.229]***	0.418 [1.618]	5.606 [1.490]***	0.811 [2.167]	0.741 [4.736]	1.018 [4.305]				
Average import tariff							−1.038 [2.266]	6.158 [1.801]***	1.549 [22.870]	8.242 [22.706]
Ln (1 + CPI)		0.339 [0.305]	−0.021 [0.302]		0.429 [0.378]	0.169 [0.329]	0.687 [0.434]	0.062 [0.290]	0.039 [0.431]	0.018 [0.196]
Government expenditure/ nominal GDP		−6.406 [3.812]*	−7.992 [3.763]**		2.795 [5.966]	1.124 [5.750]	−2.080 [3.407]	−9.186 [3.775]**	0.344 [6.494]	−0.402 [4.544]
Currency crisis		−0.411 [0.563]	−0.299 [0.610]		−0.967 [0.601]	−1.083 [0.645]*	−0.302 [0.621]	−0.129 [0.553]	−0.698 [0.604]	−0.730 [0.579]
GDI/real GDP		−1.229 [1.673]			2.870 [4.822]		−1.906 [1.948]		10.863 [9.068]	
Fraction of population literate		0.908 [0.888]			3.717 [7.905]		2.279 [2.113]		−7.041 [14.028]	
3-year lag ln income per capita		−1.896 [0.263]***			−3.662 [1.666]**		−2.551 [1.044]**		0.843 [6.771]	
Constant	−3.966 [0.225]***									
Country fixed effects?	NO	NO	NO	YES	YES	YES	NO	NO	YES	YES
Time fixed effects	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES
Observations	223	202	217	223	202	217	152	194	152	194
R^2	0.07	0.36	0.19	0.70	0.73 0.75					

Notes

All regressions exclude OECD high-income countries; Huber robust standard errors in Square brackets. Significantly different from zero at 90% (*) 95% (**) 99% (***) confidence; In IV regressions, average import tariff instrumented using 3-year lagged value and income instrumented using precipitation and temperature.

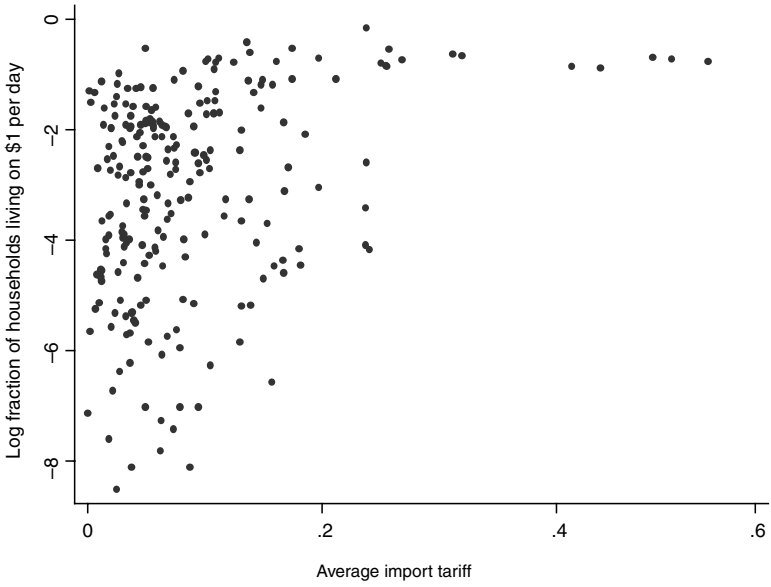

Figure 2.1 Correlation between fraction of households living on $1 per day and average import tariffs.

data availability as well as the concerns expressed by Deaton (2003), it should not be surprising that a number of the results using aggregate data are somewhat fragile. The cross-country evidence presented in Figure 2.1, for example, shows that there is a positive relationship between globalization and poverty reduction, but this association disappears in Figure 2.2 if we control for country fixed effects.

Second, it is difficult to find appropriate instruments for trade policy at the country level, or to adequately control for other changes that are occurring at the same time. Even the inclusion of additional controls is likely to be problematic, since other variables—such as the quality of institutions—are likely to be collinear with measures of trade policy. (Some researchers actually define institutional quality or rule of law using trade policy as an input.)

Third, even if cross-country studies point to a positive relationship between globalization and overall growth, such growth may lead to unequal gains across different levels of income. If the growth effects on average are small and there are large distributional consequences, trade-induced growth could be accompanied by a decline in incomes for the poor. Finally, even if cross-country studies overcome this problem by directly testing for the relationship between poverty and

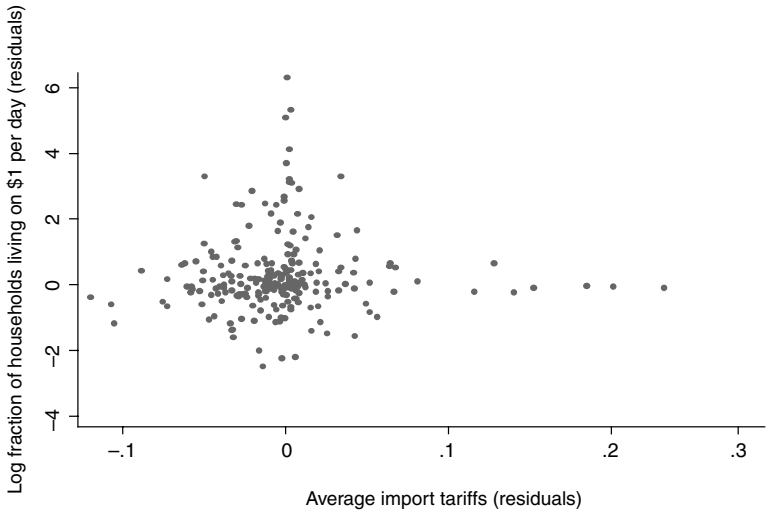

Figure 2.2 Correlation between fraction of households living on less than $1 per day and average import tariffs controlling for country fixed effects.

trade reform, there may be significant underlying heterogeneity across different segments of the population. (see also Ravallion, 2004). Aggregate poverty could move in one direction or remain unchanged while poverty increases in some parts of a country and declines in others.

Trade, growth, and poverty reduction: single country evidence

Professor Bhagwati is generally skeptical of cross-country studies of the relationship between trade and growth. A series of matching case studies is his preferred method of accounting for both the diversity of experience of different countries, and the important idiosyncrasies of each case. In Chapter 5 of *In Defense of Globalization*, he supports his stated preference for evidence gathered from in-depth case studies over cross-country regressions by referring to several of them. The most substantial of these studies were those by the Organisation for Economic Co-operation and Development (OECD) and NBER in the 1960s and 1970s. Bhagwati himself co-directed the NBER study. He reports that both these studies found overwhelming evidence in favor of outward orientation in trade and FDI, and rejected inward looking, autarkic, or import substitution trade strategies for developing countries. These studies were instrumental in overturning the previous prevailing wisdom in favor of import substitution or 'infant industry' protection policies.

Looking at the most recent experience, Bhagwati suggests that the best way to understand the link between trade and poverty reduction is to examine the last quarter century in China and India. Thus he says

> Both [China and India] shifted to outward orientation roughly two decades ago, and this contributed to their higher growth in the 1980s and 1990s. China adopted aggressively outward-oriented economic policies in 1978. India also began opening its insular economy in a limited fashion in the 1980s and more systematically and boldly in the 1990s . . . real income (gross domestic product) grew at an annual rate of 10 percent in China and 6 percent in India during the two decades ending in 2000 . . . poverty declined from an estimated 28 percent in 1978 to 9 percent in 1998 in China. Official Indian estimates report that poverty fell from 51 percent in 1977–78 to 26 percent in 1999–2000.
>
> (*In Defense of Globalization*, 2004, pp. 64–65)

Consistent with Professor Bhagwati's emphasis on country case studies, this section reviews the evidence from ten country case studies included in *Globalization and Poverty*. These case studies take as their point of interest the distributional effects of globalization. In other words, they emphasize how changes in trade policy or factor flows could have very different effects across different segments of the population. This is an important question not only for the design of social safety nets, but also because even if globalization raises aggregate incomes, it may not raise the incomes of all of the population.

Impact of globalization on employment and labor incomes of the poor

Apart from its impact on poverty via growth, trade policy can directly affect the poor through its impact on wages. The standard story is the following: the poor are assumed to be owners of (generally unskilled) labor, but not of capital. Thus trade will benefit the poor if it increases the relative returns to labor: real wages. This is the Stolper-Samuelson theorem: when a developing country increases its trade with a richer, relatively more capital abundant country, the less skilled in the developing country should gain relative to the more skilled. In other words, we would expect trade reforms in developing countries to be inherently pro-poor, since these countries are more likely to have a comparative advantage in producing goods that use unskilled labor.

As Don Davis and Prachi Mishra (2006) point out in their contribution to *Globalization and Poverty*, however, this popular story—which suggests that opening up to trade should increase the incomes of the poor in low income countries—is based on a very narrow interpretation of the standard Heckscher–Ohlin (HO) model. Davis and Mishra show that in a world of many factors and many goods, a poor country might no longer have a comparative advantage in producing unskilled-intensive goods. This idea is easy to understand

in the context of three countries—for example, the United States, Mexico, and China. Although Mexico might have a comparative advantage in producing goods that used unskilled labor *vis-à-vis* the United States, its comparative advantage changes if we allow for the possibility of trade with China.

Many of the contributors to the NBER project do not use the HO model as their framework, but instead refer to the specific sector (SS) model, which may be more appropriate in the short run. In the SS framework, workers or machines may be "attached" to a specific sector or industry, and consequently any reduction in protection to sector X will lead to a fall in the incomes of workers who are unable to relocate elsewhere. The mechanism is the following: a fall in protection is assumed to put downward pressure on the price of the previously protected good, which in turn shifts labor demand downwards. It is important to remember, however, that the reverse is also true: any increase in export activity in sector Y would then be beneficial to workers attached to that sector. The specific sector model suggests that workers may gain from globalization depending on which sectors (import-competing or exporting) they are attached to; this is very different from the HO framework, which suggests that winners and losers from globalization can be identified by their skill levels, regardless of where they work. If the HO assumption of perfect labor mobility across sectors is violated, which the evidence on India and Poland suggests, then the SS model may be the more appropriate framework—particularly in the short run. Milanovic and Squire (2006), in their contribution to that project, also analyze the impact of globalization on inequality in the context of an SS framework.

Four country cases in the NBER study examine the relationship between trade reform and labor market outcomes: the studies on Colombia, India, Mexico, and Poland. Goldberg and Pavcnik (2006) investigate the impact of a large reduction in average tariffs in Colombia between 1984 and 1998 on a variety of urban labor market outcomes: the probability of becoming unemployed, minimum wage compliance, informal sector employment, and the incidence of poverty. Analyzing the relationship between globalization and these different labor market outcomes is useful since poverty is highly correlated with unemployment, informal sector employment, and noncompliance with the minimum wage.

The Colombian experience suggests that individuals in sectors with increasing import competition are likely to become poorer, while those in sectors where exports are growing are less likely to be poor. Increasing import competition increases the likelihood of unemployment and informality, and is associated with higher incidence of poverty. Export growth is associated with the opposite: falling informal sector employment, rising minimum wage compliance, and falling poverty. These results suggest that workers cannot easily relocate away from contracting towards expanding sectors in the context of trade reforms, contradicting the assumption of perfect labor mobility in the HO framework.

The Colombian trade reforms suggest the importance of complementary reforms for minimizing the adverse effects on the poor. Trade reforms are only associated with negative labor market outcomes in the absence of labor market reforms; when trade reform is accompanied by labor market reforms, the adverse

impact of tariff reductions disappears. This is exactly the conclusion reached by Topalova (2006) in her study relating the impact of trade reform in India to poverty.

Topalova's study for the NBER project on globalization and trade reform in India is a particularly important one. One-third of the world's poor live in India. In the 1990s, India embarked on a remarkable trade reform, reversing decades of protectionist policies, which had led to average tariffs in excess of 90%. Using household data that spans the period before and after the reform period, Topalova relates changes in tariffs to changes in the incidence of poverty. In particular, she uses the interaction between the share of a district's population employed by various industries on the eve of the economic reforms and the reduction in trade barriers in these industries as a measure of a district's exposure to foreign trade. Because industrial composition is predetermined and trade liberalization was sudden and externally imposed, she argues that it is appropriate to causally interpret the correlation between the changes in the levels of poverty and trade exposure.

Topalova finds that trade liberalization benefited less those individuals living in poverty in the rural districts where industries more exposed to trade reforms were concentrated. The effect is significant and large in magnitude. A district experiencing the mean level of tariff changes saw a 2% increase in poverty, accounting for a setback of about 15% of India's progress in poverty reduction over the 1990s.

As Topalova points out, she does not study the level effect of liberalization on poverty in India, but rather the relative impact on areas more or less exposed to liberalization. Trade reform was probably associated with the overall decline in poverty in India observed during this same period, but this is an aggregate result which the cross-country studies described earlier are designed to address. However, the evidence on poverty linkages suggests that the rural poor gained less, compared to either other income groups or the urban poor. Topalova's study also discusses why restrictions on labor mobility in rural areas have impeded adjustment, driving home the point that rural India was more consistent with the SS framework in the short run.

While the studies on Colombia and India suggest that the gains from trade reforms were less likely to benefit the poor, the evidence for Mexico and Poland suggests the opposite. Hanson (2006), in his country study on the Mexican experience, explores the different outcomes for individuals born in states with high exposure to globalization versus individuals born in states with low exposure to globalization between 1990 and 2000. He finds that the income of individuals in high-exposure states increased relative to the income of individuals in low-exposure states. While labor incomes in the 1990s deteriorated in both regions, caused in part by Mexico's peso crisis in 1995, the deterioration was much less severe in states with high exposure to globalization.

While poverty was falling dramatically in India during this period, between 1990 and 2000 poverty in Mexico increased. In the states with low exposure to globalization, poverty increased from 32% to 40%; in the states with high

exposure, poverty increased only slightly, from 21% to 22%. If we take the difference in the increase in poverty within each region over the 1990s, we find that poverty increased by 8% in low exposure states and by only 1% in high-exposure states. The "difference-in-difference" estimator is the differential in these two changes (i.e. 8–1 equals 7 percentage points) and is the basis for Hanson's conclusions that the incidence of wage poverty in low exposure states increased relative to that in high-exposure states by approximately 7%. During Mexico's globalization decade, poverty increased less in the more globalized states.

How can we reconcile the findings on Mexico and India? As pointed out by Hanson, the peso crisis in Mexico in 1995 is one major reason for the aggregate increase in poverty, in contrast to India which experienced no major adverse macroeconomic shock during this period. In addition, Hanson defines high globalization states to include those with a high proportion of *macquiladoras*—production activities designated for exports—and foreign direct investment. Topalova also finds, consistent with Hanson, that activity associated with exports and FDI is positively correlated with poverty reduction. Consequently, both studies consistently show that export activity and FDI is correlated with beneficial outcomes for the poor.

Goh and Javorcik (2006) examine the relationship between tariff changes and wages of workers in Poland. Controlling for a variety of firm and worker characteristics, the authors exploit the significant trade reforms that occurred in Poland during the 1990s, when the country moved from a closed to a very open economy, particularly *vis-à-vis* the European Union. One advantage of choosing Poland is the fact that the changes in its tariffs can be treated as exogenous, as they were stipulated by the Association Agreement between the European Community and Poland signed in 1991. This agreement also predetermined the schedule of tariff reductions, which took place during 1994–2001.

Goh and Javorcik demonstrate that labor mobility is fairly restricted in Poland, placing their analysis also in the context of a specific sector framework. Their results suggest that workers in sectors that experienced the largest tariff declines experienced the highest increases in wages, after controlling for worker characteristics such as education and experience, as well as sector-specific and time-specific effects. These results are remarkable. They posit that the reason why tariff declines led to wage increases is that firms were forced to increase productivity, and that those productivity increases were shared with the workers in the form of higher wages. They also present evidence showing—consistent with Topalova and previous productivity studies—that tariff reductions were indeed accompanied by significant increases in total factor productivity.

These micro-level results showing a positive relationship between tariff reductions and productivity increases are consistent with the more aggregate evidence on the positive relationship between openness to trade and aggregate growth. Their results are also consistent with the other country studies that show that increasing export activity is correlated with wage increases. In a diversity of country settings—Poland, Colombia, India, and Mexico—this volume documents that exporting activities are associated with increasing incomes for the unskilled and the poor.

Impact of globalization on poverty via prices of production and consumption goods

In many developing countries, wage income is not the primary source of income for the rural poor. In their contribution to the NBER study, Balat and Porto (2006) calculate that in Zambia wages accounted for only 6% of income for the rural poor in 1998. In Zambia, where 72% of the population was living below the poverty line in 1998, most of the rural poor either consumed their agricultural output, sold their crops, or derived income from other sources. Consequently, globalization could affect poverty by affecting the prices of goods consumed by the poor (the consumption channel) and goods produced by the poor (the production channel).

In many cases, the urban poor are net consumers of agricultural products and the rural poor are net producers of those same products; in this case, an increase in agricultural prices caused, for example, by a removal of export taxes could lead to an increase in urban poverty but a decline in rural poverty. As an illustration, China's accession to the WTO, which is associated with liberalization of the agricultural sector, is expected to contribute to an increase in rural poverty but a decline in urban poverty over the next several years (Ravallion, 2004).

These linkages are explored to various degrees in the studies on Ethiopia, Mexico, and Zambia. In Mexico, Ashraf *et al*. (2006) explore the impact of liberalizing Mexico's corn market on the incomes of the poor rural farmers. The evidence suggests that during the 1990s, imports of both white and yellow corn increased, and prices of Mexican corn fell. The income from corn production among poor farmers also fell, both as a share of total income and in absolute terms. The 50% decline in income from corn production would have translated into an equivalent decline in real income if poor farmer incomes had not been supplemented with remittances and transfers through government programs such as Progressa.

In their study of Ethiopian rural grain producers, Levinsohn and McMillan (2006) explore the impact of food aid on both consumption and production of the rural poor. This is an important contribution because some critics have argued that food aid further exacerbates poverty by depressing incomes of rural producers. While Levinsohn and McMillan confirm that a more optimal arrangement would be to buy food from local producers and distribute it to poor consumers, they also show that the net impact of food aid on the poor in Ethiopia has been positive. This is because the poor in Ethiopia are primarily net consumers, rather than net producers of food, and consequently food aid has alleviated poverty. As pointed out by Pande in her excellent discussion of this paper, these results are contingent on food aid actually reaching the poor. Levinsohn and McMillan show that this is often the case.

For Zambia, Balat and Porto (2006) calculate the impact of liberalizing the market for maize, which was heavily subsidized to both consumers and producers. They find that the resulting price increase led to consumption losses, which were offset by domestic market liberalization. They also measure the potential increase in income due to switching from production for home consumption to

production and wage activities associated with production of cash crops. Balat and Porto estimate that rural Zambians would gain substantially from expanding into the production of cash crops, particularly in the production of cotton, tobacco, and maize.

However, Balat and Porto also caution that such gains can only be achieved if other complementary policies are in place. These would include extension services, infrastructure, irrigation, access to credit and finance, education, and health services. Balat and Porto also point to the fact that Zambia needs to have access to international agricultural markets in order to realize potential gains.

Another paper in the NBER volume explicitly addresses the issue of industrial country distortions by measuring the impact of OECD support policies for domestic agriculture on incomes in developing countries. Ashraf *et al*. calculate a country-specific measure of OECD support to measure whether industrial country policies directly affect income and poverty in developing countries. The vast majority of least developed countries have historically been net importers of food, particularly cereals, which are among the most heavily subsidized crops. As net food importers, they may be hurt by higher commodity prices and could possibly gain from rich country subsidies (see also Panagariya 2002, 2004; Valdes and McCalla, 1999). Even within exporting countries, the poorest members of society may be net purchasers of food. Ashraf *et al*. find that for countries with food export shares greater than 48%, OECD subsidies reduce income per capita. Again, the picture is decidedly mixed, with net consuming countries gaining from subsidized imports and net producing countries losing as a result of the same subsidies.

Capital flows and the poor

Another avenue through which globalization could affect the welfare of the poor is via the liberalization of international capital markets. In the 1980s and 1990s, developing countries became increasingly open to international capital flows, measured either using policy instruments such as capital controls or *ex post* capital flows. In theory, openness to capital flows could alleviate poverty through several channels. If greater financial integration contributes to higher growth by expanding access to capital, expanding access to new technology, stimulating domestic financial sector development, reducing the cost of capital, and alleviating domestic credit constraints, then such growth should reduce poverty. Access to international capital markets should also allow countries to smooth consumption shocks, reducing output or consumption volatility. However, as the evidence will show, access to capital flows can also exacerbate volatility.

In his discussion of the impact of capital flows on poverty reduction in chapter 12 of *In Defense of Globalization*, Professor Bhagwati is careful to distinguish between FDI and shorter-term portfolio investment. He argues strongly for the advantages of foreign direct investment to developing countries, and particularly to low-skilled and thus poor workers. Despite this, he does not favor current proposals for a multilateral investment treaty. Such a treaty, he argues, would have

little positive impact on investment flows reaching developing countries. Its primary impact, therefore, would be to increase the share of the benefits for foreign investment that accrue to the investor, rather than the host country.

Professor Bhagwati's position on capital controls (relating primarily to shorter-term investments) is more controversial, and one in which he was a pioneering voice in the 1990s. In 1998, Professor Bhagwati began arguing that full and rapid capital account liberalization was not in developing countries' best interests.[5] Indeed, he placed much of the blame for the Asian financial crisis on the International Monetary Fund's (IMF) efforts to make these countries liberalize their capital accounts too quickly. The emerging evidence on this issue provides support for Bhagwati's keen insights. In this section, we summarize the results of two studies on capital flows and poverty outcomes that will appear in the NBER study *Globalization and Poverty*.

The first study, by Prasad *et al*. (2006) begins by examining the relationship between financial integration and growth. Reviewing over a dozen studies and examining the data themselves, they find that there is no clear relationship between the two. This suggests that the impact of financial integration on poverty outcomes—via possible growth effects—is likely to be small. They suggest that since there are no clear linkages between financial integration and growth in the aggregate cross-country evidence, that the direct linkages between financial and poverty are also likely to be difficult to find.

They also explore another link whether financial integration has smoothed or exacerbated output and consumption volatility. They point out that greater macro-economic volatility probably increases both absolute and relative measures of poverty, particularly when there are financial crises. Since the poor are likely to be hurt in periods of consumption volatility, such income smoothing could be beneficial to the poor. However, Prasad *et al*. find evidence that suggests the opposite: financial globalization in developing countries is associated with higher consumption volatility, not lower volatility. More specifically, the data suggests that more financially integrated developing countries have experienced an increase in consumption volatility, relative to both industrial countries and to other developing countries. They posit the existence of a threshold effect: beyond a certain level of financial integration (50% of GDP), financial integration significantly reduces volatility. However, most developing countries are well below this threshold.

Much of the increases in consumption volatility identified by Prasad *et al*. for less financially integrated countries occurred in the context of currency crises. In recent years, a number of countries have experienced a massive and largely unanticipated collapse of their exchange rate. One study in the NBER volume, by Duncan Thomas and Elizabeth Frankenberg (2006), examines the impact of such a crisis on the poor. Using longitudinal household survey data from the Indonesia Family Life Survey (IFLS), Thomas and Frankenberg examine the immediate and medium term effects of the East Asian crisis on multiple dimensions of well being. In IFLS, the same households were interviewed a few months before the onset of the crisis, a year later and again two years after that, which provides

unique opportunities for measuring the magnitude and distribution of the effects of the crisis on the population.

Thomas and Frankenberg demonstrate that in the first year of the crisis, poverty rose by between 50% and 100%, real wages declined by around 40% and household per capita consumption fell by around 15%. However, focusing exclusively on changes in real resources is complicated by the fact that measurement of prices in an environment of extremely volatile prices is not straightforward. Moreover, it misses important dimensions of response by households. These include changes in leisure (labor supply), changes in living arrangements (household size and thus per capita household resources), changes in assets and changes in investments in human capital. These responses are not only quantitatively important but also highlight the resilience of families and households in the face of large unanticipated shocks as they draw on a wide array of mechanisms to respond to the changes in opportunities they face.

While the volatility of bank borrowing and portfolio flows may be costly to the poor, many of the authors in *Globalization and Poverty* emphasize the benefits from another type of inflow: FDI. Prasad and his co-authors emphasize that the composition of capital flows can have a significant impact on a country's vulnerability to financial crises. They also document that FDI flows are significantly less volatile than other types of flows. Studies on Mexico (Hanson), India (Topalova), Poland (Goh and Javorik), and Colombia (Goldberg and Pavcnik) all demonstrate that incoming foreign investment is associated with a significant reduction in poverty.

Criticism: continuing in spite of the evidence?

The previous sections of this chapter have considered the relationship between globalization and poverty from a number of different perspectives. In the current section we review the broad trends in poverty and relate the trends to the vigorous public debate on globalization and poverty.

Trends in poverty

Table 2.5 provides a comparison of the most widely cited current estimates of the world poverty headcount and incidence. It further shows that different authors have produced very different estimates for the level and trend in poverty headcounts.

The sources of these differences are methodological, with key issues including whether national accounts data or household survey data is used to calculate average income; how PPP is calculated; and whether world population or only developing country population is used in the denominator for calculation of poverty incidence. Other important issues that lead to differences in reported world poverty figures and trends include the choice of base year, poverty line, and time span. These issues are discussed in some detail in Aisbett's (2006) chapter in the NBER study.

Table 2.5 Comparison of recent world poverty estimates

1998 Headcount (billion)	*1998 Incidence (%)*	*Average change in headcount 1987–1998 (million p.a.)*	*Average change in incidence 1987–1998 (% p.a.)*	*Poverty line ($/day)*	*Source*
1.20	24.0	+1.4	−0.40	1.08	Chen and Ravallion (2000), Table 2
2.80	56.0	+22.9	−0.46	2.15	Chen and Ravallion (2000), Table 3
1.10	22.1	−6.8	−0.55	1.08	Chen and Ravallion (2004), Table 3
2.71	54.8	+21.8	−0.48	2.15	Chen and Ravallion (2004), Table 4
0.35	6.7	−3.3	−0.19	1.08	Sala-i-Martin (2002a), Table 3a
0.97	18.6	−20.0	−0.77	2.15	Sala-i-Martin (2002a), Table 3a
0.46	9.2	−30.8	−0.90	1.08	Bhalla (2003), Table 1, PWTv6
0.37	7.4	−22.6	−0.67	1.08	Bhalla (2003), Table 1, WBPPP

Notes
Average Change—total change in the headcount over the period 1987–1998, divided by 11 years.
WBPPP—World Bank Purchasing Power Parity conversion using base year 1993. Uses Elteto, Koves and Szulc method.
PWTv6—Penn World Tables Purchasing Power Parity conversion using base year 1996. Uses Geary-Khamis method.
Chen and Ravallion (2000, 2004) are the "World Bank" figures referred to below. 1998 values for Chen and Ravallion (2004) are the linear combination of their reported numbers for 1996 and 1999. For a discussion of the differences in method between Chen and Ravallion's 2000 and 2004 estimates, see Chen and Ravallion (2004).
Total Headcount and Average Change for Bhalla (2003) were calculated from his reported incidence figures, using the same population size as Chen and Ravallion (2000).
Sala-i-Martin's incidence is based on the total world population, rather than the population of developing countries as used by the other authors.

Table 2.5 also shows that despite the differences in headcount trends and incidence levels, all authors agree that the incidence of poverty was falling in the world over the period 1987–1998. The difference between the unambiguous progress when poverty is measured as incidence, and the highly ambiguous "progress" when poverty is measured as headcount is of some importance to the globalization debate. As pointed out by Aisbett, there is a diversity of opinions on whether headcount or incidence is the appropriate measure of poverty. She further argues that this diversity of opinion, combined with the sometimes contradictory trends that headcount and incidence suggest, is one of the reasons that there is so much disagreement about whether world poverty has been increasing during the period of globalization.

Globalization's critics

In light of these trends, which suggest falling poverty, why does there continue to be so much criticism of globalization? This is the central question of Aisbett (2006), and she argues that there are several parts to the answer. The first part is that people have a natural tendency to weight the information that they receive according to their prior beliefs and values. Thus, evidence which is objectively "mixed" is quite likely to be interpreted by one type of person as very positive, and by another as very negative. The mere fact that the evidence on globalization's impact on the poor is not unequivocally positive will lead people with negative priors to believe it is negative.

The second part of Aisbett's answer is to examine what types of beliefs and values lead people to a more negative interpretation of the evidence on globalization and poverty. The values which she identifies include concern over inequality, independent of poverty. In particular, globalization's critics feel differently about the polarization of the income distribution and inequality in the gains that different groups receive from globalization.

As first pointed out by Kanbur (2001), critics of globalization also tend to focus on shorter term impacts, while globalization's proponents are more concerned about the longer term. Critics of globalization also focus on the losses experienced by subgroups of the poor, even when at the country level poverty has declined. Aisbett suggests a number of explanations for this value preference, including recent evidence from behavioral experiments. She notes that the results of these experiments suggest that people concerned for subgroups may simply be displaying a very common human characteristic. After conducting experiments based on hypothetical allocation decisions (unrelated to globalization), Baron (2003, p. 1) finds that

> People are reluctant to harm some people in order to help others, even when the harm is less than the forgone help (the harm resulting from not acting). The present studies use hypothetical scenarios to argue that these judgments go against what the subjects themselves would take to be the best overall outcome.

Of more relevance to Bhagwati's work, however, is what Aisbett refers to as "beliefs about the process of globalization." She argues that many people believe that the current form of globalization is based on processes which distill both political and market power upward and away from the poor. In particular they believe that corporate and commercial lobbies have disproportionate access to the international organizations such as the WTO and IMF, and that rich countries exploit their power within these international organizations. This belief about the processes through which globalization occurs is partly what predisposes them to interpret the available evidence negatively.

The belief that globalization favors the rich and powerful is fuelled by a number of key examples of successful lobbying on the part of corporate interests. They include the Trade Related Aspects of Intellectual Property Rights (TRIPs) agreement, proposals for a multilateral agreement on investment, and the IMF push for capital account convertibility. What makes these examples interesting is that many economists and proponents of globalization actually agree with the critics on both their causes and consequences for the poor. Few, however, have recognized the threat that they pose to globalization. Just as the excesses of early capitalism may have contributed to the communist movement, so too could the excesses of globalization lead to a backlash and return to protectionism.

The prime exception, of course, has been Jagdish Bhagwati. He has both identified and responded to the threat that these negative manifestations of globalization represent; and his response has been appropriately targeted towards a nonacademic audience. His efforts include many books on this issue, including *In Defense of Globalization*, as well as letters and articles in *Foreign Affairs* and *The Financial Times*.[6]

Conclusion

In the last two decades, the percentage of the world's population living on less than $1 dollar a day has been cut in half, falling from 40% to 21%. Many countries, including China, have made tremendous strides in reducing not only the percentage of the population living in poverty but also the absolute number of individuals living on less than $1 a day. While the income gap between the richest and poorest countries has increased, population-weighted measures of inequality show a significant decline. At the same time, developing countries increased their trade shares and slashed their tariffs. To what extent can we claim that increasing globalization is responsible for the fall in the incidence of poverty?

The first theme that emerges from the forthcoming book, *Globalization and Poverty*, is that the relationship between changes in globalization and changes in poverty is a complex one. In many cases, the outcome depends not just on trade reform or financial globalization but on the interaction of those policies with the rest of the environment. It is misleading to examine the impact of trade reform on poverty without taking into account the complementarity between trade or financial globalization and other changes in the environment. Financial globalization

is more likely to have a benign impact on growth and poverty reduction if it is accompanied or preceded by the development of good institutions and governance, as well as macroeconomic stability (including the use of flexible exchange rates). In India and Colombia, trade reforms have been associated with an increase in poverty only in regions with inflexible labor laws. Consequently, reaching any conclusions without taking into account the role of labor market legislation—and its contribution to inhibiting labor mobility in those countries—would be highly misleading.

The importance of complementary policies in ensuring that globalization benefits all segments of a population has long been a theme in the writings of Jagdish Bhagwati. In his discussion of growth and poverty, he points out that the relationship between growth and poverty reduction is generally a positive one, but that governments play an important role in affecting the strength of that association.[7] He suggests that key government policies that can make growth more pro-poor include ensuring access to credit and political voice.[8] Professor Bhagwati also uses the success of the East Asian Tigers to illustrate the importance of complementary policies. He notes that these countries were particularly successful at achieving growth and poverty reduction through their outward-oriented trade policies because of their complementary emphasis on high rates of investment, high literacy rates, expansion of higher education, and the use of export income to import capital that embodied advanced technology.

A second lesson that emerges from our review of the evidence is that globalization leads to clearly identifiable winners among the poor. Across several different continents, export expansion has been accompanied by a reduction in poverty. The evidence also points to the beneficial effects of FDI. While the macroeconomic evidence suggests that FDI is a less volatile source of capital than other types of inflows, the microeconomic evidence for India, Mexico, Poland, and Colombia indicates that higher inflows of foreign investment are associated with a reduction in poverty.

Third, it is also possible to identify some losers from globalization among the poor. Poor workers in import-competing sectors—who cannot relocate possibly due to the existence of inflexible labor laws—are likely to be hurt by globalization. Financial crises also affect the poor disproportionately, as indicated by the cross-country evidence and the erosion of real wages following currency crises in Indonesia and Mexico. In Mexico, poor corn farmers have been negatively affected by increasing import competition. However, transfer programs which redistribute income have been successful in preventing the erosion of their real income.

Professor Bhagwati pointed out many years ago that increased trade can, in theory, reduce incomes. Indeed, this very paradox was coined "immiserizing growth" by Bhagwati in his seminal 1958 paper on the topic. This work was particularly relevant to developing countries that feared that their increases in exports were in fact causing immiserizing growth by leading to a large fall in their

export prices.[9] In his most recent book Bhagwati's suggestion is, once again, that there is a role for government intervention. As he says

> So when you depress your export prices by selling more because you are a major supplier, restrain yourself; push in other directions. A suitable policy can always nip the immiserizing growth paradox in the bud.
>
> (*In Defense of Globalization*, p. 55)

The fourth lesson that emerges from a review of cross-country evidence and country case studies is that simple interpretations of general equilibrium trade models such as the Heckscher–Ohlin framework are likely to be incorrect. Many economists predicted that developing countries with a comparative advantage in unskilled labor would benefit from globalization through increased demand for their unskilled-intensive goods, which in turn would reduce inequality and poverty. The theoretical discussions as well as the empirical evidence presented in this chapter suggest that this interpretation of trade theory is too simple and frequently not consistent with reality. Cross-country evidence (see Easterly (forthcoming) and Milanovic and Squire (forthcoming)) suggests that globalization has been accompanied by increasing inequality within developing countries. However, the micro studies are more mixed, with evidence from Colombia consistent with rising inequality accompanying trade reforms, the evidence on India suggesting no relationship between trade reform and inequality, and the evidence on Poland suggesting that trade reforms have contributed to falling inequality.

The heterogeneity in outcomes suggests that careful targeting is necessary to address the poor who are likely to be hurt by globalization. This includes the poor in countries hit by financial crises, as well as the smallest farmers who cannot compete with the more efficient larger farmers or with expanding import competition. Clearly, the concerns of globalization's critics have been heard, but much remains to be done.

Notes

1 See also his recent article co-authored with T. N. Srinivasan that appeared in the *American Economic Review* (2002), entitled "Trade and Poverty in Poor Countries."
2 See *In Defense of Globalization*, Chapter 5.
3 Possibly the only exception to these general conclusions is Agenor (2002b), who finds that that poverty increased in countries more open to trade. However, his sample is limited to a sample size of 30 observations. In a similar paper using a somewhat larger sample, Agenor (2002a) finds no significant relationship between trade shares and a headcount measure of poverty.
4 See, for example, Dollar and Kraay (2002, 2004), and Ravallion (2004).
5 Bhagwati (1998a).
6 See, for example, IDG pp. 82–83, 165, 182–185, 199–207; Bhagwati (1998a); Bhagwati (1998b).
7 See *In Defense of Globalization*, pp. 54–60.
8 With regard to the provision of credit, however, Bhagwati suggests that markets should be used wherever possible (*In Defense of Globalization*, 2004, p. 58). Markets, he says, are more egalitarian allocation mechanisms than corrupt officials.

9 The possibility that a large negative terms of trade shock could lead to losses from trade has recently been revived by Samuelson in an article that appeared in the *Journal of Economic Perspectives*.

Bibliography

Aisbett, Emma (2006) "Why are the Critics so Convinced that Globalization is Bad for the Poor?," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.

Ashraf, Nava, Margaret McMillan and Alix Peterson Zwane (2006) "My Policies or Yours: Have OECD Agricultural Policies Affected Incomes in Developing Countries?," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.

Balat, Jorge, and Guido Porto (2006) "Globalization and Complementary Policies. Poverty Impacts in Rural Zambia," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.

Baron, J. (2003) "Blind justice: fairness to groups and the do-no-harm principle," Mimeo, University of Pennsylvania. Available at www.psych.upenn.edu/~baron/eq.pdf, last accessed January 5, 2005.

Bhagwati, Jagdish (1998a) "The Capital Myth: The Difference between Trade in Widgets and Dollars," *Foreign Affairs*, 77(3): 7–13.

Bhagwati, Jagdish (1998b) "Letter to the Financial Times on Multilateral Agreement on Investment (MAI)," Available at www.columbia.edu/~jb38/index_papers.html, last accessed January 10, 2005.

Bhagwati, Jagdish (2004) *In Defense of Globalization*, New York: Oxford University Press.

Bhagwati, J. and Srinivasan, T. N. (2002) "Trade and Poverty in the Poor Countries," *AEA Papers and Proceedings*, 92(2): 180–183.

Bhalla, S. (2002) *Imagine There's No Country: Poverty, Inequality, and Growth in the Era of Globalization*, Washington, DC: Institute for International Economics.

Bhalla, S. (2003) "Crying Wolf on Poverty: Or How the Millennium Development Goal Has Already Been Reached," Institute for International Economics Working Paper No. 0403. Available at www.iie.com/publications/papers/bhalla0403.pdf, last accessed January 8, 2005.

Chen, S. and Ravallion, M. (2000) "How Did the World's Poorest Fare in the 1990s?" World Bank Development Research Group Working Paper w2409, Available at http://econ.worldbank.org/view.php?type=5&id=1164, last accessed January 8, 2005.

Davis, Don and Prachi Mishra (2006) "Stolper-Samuelson is Dead and Other Crimes of Both Theory and Data," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.

Deaton, A. (2001) "Counting the World's Poor: Problems and Possible Solutions," *World Bank Research Observer*, 16(2): 125–147.

Deaton, Angus (2003) "Measuring Poverty in a Growing World (or measuring growth in a poor world)," NBER Working Paper No. 9822. Available at http://papers.nber.org/papers/w9822, last accessed January 8, 2005.

Dollar, David (2001) "Globalization, Inequality and Poverty Since 1980," Background paper, World Bank, Washington, DC. Available at www.worldbank.org/research/ global

Dollar David and Aart Kraay (2001) "Trade, Growth and Poverty," World Bank Development Research Group Working Paper 2615. Available at http://econ.worldbank.org/files/24986_wps2615.pdf, last accessed January 8, 2005.

Dollar, D. and Kraay, A. (2002) "Spreading the Wealth," *Foreign Affairs*, 81(1): 120–133.
Easterly, William (2006) "Globalization, Prosperity, and Poverty," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.
Goh, Chor-Ching, and Beata Smarzynska Javorcik (2006) "Trade Protection and Industry Wage Structure in Poland," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.
Goldberg, Penny and Nina Pavcnik (2004) "Trade, Inequality, and Poverty: What Do We Know? Evidence from Recent Trade Liberalization Episodes in Developing Countries," NBER Working Paper No. 10593. in Susan Collins and Carol Graham (eds) *Brookings Trade Forum*, Washington, DC: Brookings Institution Press.
Goldberg, Penny and Nina Pavcnik (2006) "The Effects of the Colombia Trade Liberalization on Urban Poverty," in Ann Harrison (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.
Hanson, Gordon (2006) "Globalization, Labor Income, and Poverty in Mexico," in Ann Harrison (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.
Harrison, Ann (2006) *Globalization and Poverty*, University of Chicago Press for NBER.
Kanbur, Ravi (2001) "Economic Policy, Distribution and Poverty: The Nature of the Disagreements," *World Development*, 29(6): 1083–1094.
Krueger, A. (1983) *Trade and employment in developing countries, 3: Synthesis and conclusions*, Chicago and London: University of Chicago Press.
Levinsohn, James and Margaret McMillan (2006) "Does Food Aid Harm the Poor: Household Evidence from Ethiopia," in Ann Harrison, (ed.) *Globalization and Poverty*, Forthcoming, University of Chicago Press for NBER.
Milanovic, Branko (2002) "True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone," *The Economic Journal*, 112 (January): 51–92.
Milanovic, Branko and Lyn Squire (2006) "Does Tariff Liberalization Increase Inequality? Some Empirical Evidence," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.
Milanovic, Branko (2005) *Worlds Apart: Measuring Global and International Inequality*, Princeton University Press.
Panagariya, A. (2002) "Trade and food security: Conceptualizing the Linkages," College Park, MD: University of Maryland, mimeo.
Panagariya, A. (2004) *Comments on Subsidies and Trade Barriers*, Copenhagen: Consensus Opponent Note.
Prasad, Eswar S., Kenneth Rogoff, Shang-Jin Wei, and M. Ayhan Kose (2006) "Effects of Financial Globalization on Developing Countries: Some Empirical Evidence," in Ann Harrison (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.
Sala-i-Martin, X. (2002a) "The World Distribution of Income (estimated from individual country distributions)," NBER Working Paper No. 8933. Available at http://papers.nber.org/papers/w8933, last accessed January 8, 2005.
Sala-i-Martin, Xavier (2002b) "The Disturbing 'Rise' of Global Income Inequality," NBER Working Paper No. w8904. Available at http://papers.nber.org/papers/W8094, last accessed January 8, 2005.
Thomas, Duncan and Elizabeth Frankenberg (2006) "Financial Crises and Poverty: The Case of Indonesia," in Ann Harrison (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.

Topalova, Petia (2006) "Trade Liberalization, Poverty and Inequality: Evidence from Indian Districts," in Ann Harrison, (ed.) *Globalization and Poverty*, University of Chicago Press for NBER.

Valdes, A. and A. F. McCalla (1999) *Issues, interests and options of developing countries*, World Bank: Washington, DC.

Winters, Alan L., Neil McCulloch, and Andrew McKay (2004) "Trade Liberalization and Poverty: the Evidence So Far," *Journal of Economic Literature*, XLII (March): 72–115.

3 International trade, labor turnover, and the wage premium

Testing the Bhagwati–Dehejia hypothesis for Canada

Eugene Beaulieu, Vivek H. Dehejia, and Hazrat-Omar Zakhilwal

Introduction

In this chapter we examine the impact of international trade on the absolute and relative wages of educated and less-educated workers in Canada over 1993–1996. We show that after correcting for the relative supply effect of educated to less-educated workers the wage differential would have been on an upward trend. Moreover, after controlling for other relevant factors influencing real wages, trade had a statistically significantly positive impact on the wages of both educated and less-educated workers. However, the impact on the educated workers was four times stronger, roughly the same as the impact of technology on relative wages. We show that the observed relationship between trade and the relative wage of educated to less-educated workers does not fit the Stolper–Samuelson theoretical explanation. The observed results are more in line with the Bhagwati–Dehejia hypothesis, which posits a link from trade to wages through volatility, labor turnover, and jobless spells.

During the 1980s and 1990s the wage differential between skilled and less-skilled workers widened in almost all developed countries. Countries with relatively inflexible labour markets experienced an increase in unemployment of less-skilled workers over the same period. Two widely suggested causes of the relative wage change phenomenon are trade liberalization and skill-biased technological change. The former is usually assumed to work through the Stopler–Samuelson Theorem (SST) that implies that the reduction in trade barriers will increase the real return to the factor that is relatively abundant in the country, and decrease the real return to the relatively scarce factor. Since the North is abundant in skilled labor and the South in unskilled labor, the SST seems to provide a *prima facie* plausible explanation for the observed trends.

A consensus emerged from the empirical literature that both skilled-biased technological change and international trade were affecting relative wages in the same direction. However, although international trade was found to play a role empirically, the empirical evidence was not consistent with the SST. There are two important problems that the SST has in explaining wage inequality. First, although there is convincing empirical evidence that increased trade volumes are

associated with increased wage inequality, the SST connects relative output prices to real factor prices and is silent on the relationship between trade volumes and wages. There is no evidence that changes in goods prices increased wage inequality. Second, a number of studies have shown that international trade increased wage inequality in both skill-abundant and skill-scarce countries. Therefore, there was a need for an alternative explanation that was trade-dependent but did not rely on SST. This is what Bhagwati and Dehejia (1994) did in their widely cited article, "Freer Trade and Wages of the Unskilled—Is Marx Striking Again?" The proposed explanation that they put forward has come to be known in the empirical trade literature as the "Bhagwati–Dehejia hypothesis" (BDH).[1]

The BDH is based on the hypothesis that trade liberalization has made many industries "footloose" (i.e. small shifts in costs can cause comparative advantage to shift suddenly from one country to another), hence making comparative advantage "kaleidoscopic" (i.e. one country may have comparative advantage in X and another in Y one day, and next day it may suddenly be reversed). This in turn leads to increased labor turnover. The added turnover means that mobile workers could be accumulating fewer skills causing a reduction or stagnation in the real wages of the affected workers. However, it is assumed that the less-educated workers will be affected by more than the educated workers, because the former embody sector-specific human capital acquired on the job, whereas the latter embody education-based human capital, which is less sector specific. These factors as a whole provide a trade-depended explanation for the observed wage differential between educated and less-educated labor.

The purpose of this chapter is to examine the empirical relationship between trade and wages in Canada within a well-grounded theoretical framework, with a view to casting light on the empirical relevance of the BDH. We present a simple theoretical model of the BDH and derive some reduced form estimating equations from it. The theoretical model predicts a causal relationship between trade volumes and relative wages. The model also provides a mechanism through which trade volumes affect relative wages. The model predicts that high trade volatility increases labour turnover and that increased turnover will increase the wage premium of skilled over unskilled. Panel labor force data are the appropriate data for examining turnover and wages. We use the Survey of Labor and Income Dynamics (SLID) data from the period 1993–1996 to study the impact of trade liberalization on absolute and relative wages of educated and less-educated workers in Canada. The educated workers consist of people who received a university degree, certificate or diploma, ranging from below Bachelor's to PhD as defined by the SLID survey.

We find that after controlling for some of the most likely factors influencing real wages, international trade had a significantly positive impact on the wages of both educated and less-educated workers. However, the impact on educated workers seems to have been some four times stronger, roughly the same as the impact of technology on relative wages. We also find that the observed relationship between trade and the relative wage of educated to less-educated workers does not fit the SST explanation. Rather, the theoretical explanation provided by BDH is more in line with the results observed in this chapter.

Very little research has been conducted examining the issue of trade and wages for Canada, which seems surprising considering the importance of trade to the Canadian economy. The only paper that explicitly investigates the trade effect alongside the technological change effect on the relative wages of nonproduction (skilled) to production (unskilled) workers is that of Baldwin and Rafiquzzaman (1998). They investigate whether technological change and trade could have been responsible for increased wage inequality. Their conclusion is that both are at work. Their study, however, is limited in the sense that it does not take other factors—particularly the labor supply effect into account.

Unfortunately, like most of the studies done for the United States, Baldwin and Rafiquzzaman investigate the causal links between the volume of trade and relative wages without grounding their empirical findings in a theoretical framework. It is also fair to say that their study was geared more toward investigating the impact of technology rather than that of trade. This paper, nonetheless, investigates the channels through which trade is hypothesized to have contributed to an increase in the wage differential. The by-product obtained from investigating the channels such as the impact of trade liberalization on labor turnover and of labor turnover on skill accumulation will certainly add to important empirical findings about the Canadian labor market. Moreover, we are not aware of any papers published in peer reviewed journal articles that examine trade and wage differentials in Canada using microeconomic data.

Freeman and Needels (1993) and Murphy *et al*. (1998), whose main focus is on the relative wage effect of the relative labor supply of educated workers in Canada and in the United States, argue that neither trade nor technology is the culprit in influencing relative wages in either country. They maintain that both over time and between countries the variation of rate of growth in relative wages is due to variation in the relative supply of more educated workers alone. For example, the more conspicuous rise in the educated workers' relative wages in the United States and a less-evident rise (or, no rise at all) in Canada has more to do with a relatively higher growth in the relative supply of educated workers in Canada over the period under investigation than anything else.

While the adverse effect on the educated workers' relative wages of their relative supply may not be disputed, the finding of a nonincrease or even a decrease in the wage premium of the educated cannot be counted as an evidence against the positive impact of trade liberalization on the wage differential because it could just be that the supply side might have overwhelmed the demand side. We find that trade has a significantly positive impact on the wages of both educated and less-educated workers; however, as just noted, the impact on the educated workers is four times stronger, roughly the same as the impact of technology on relative wages. Moreover, since the results in this chapter show that the widening of educated/less-educated wage differential does not come even partially at the expense of less-educated workers, the result does not fit the SST trade explanation, but is, at a minimum, consistent with the BDH.

The remainder of this chapter is structured as following: First we present the model through which we intend to fit the trade explanation of the educated/less-educated

wage differential into a theoretical framework. Then we examine the nature of the education premium in Canada and examine the relationship between trade and the wages and wage differential of the educated and the less-educated. We also investigate the impact of trade volatility on labor turnover and the impact of jobless spells on educated/less-educated skill accumulation.

The model: kaleidoscopic comparative advantage and labor turnover

The BDH alternative trade explanation is based on the hypothesis that trade liberalization has made many industries "footloose," hence, making comparative advantage "kaleidoscopic." "Footloose" is a situation in which small shifts in costs can cause comparative advantage to shift suddenly from one country to another, while "kaleidoscopic" refers to a situation in which one country may have comparative advantage in X and another in Y one day, and next day it may suddenly be reversed.

This hypothesis above in turn leads to four main consequences

1 increased *ceteris paribus* labor turnover;
2 the added turnover means that the mobile less-educated labor could be accumulating less skills due to the "rolling-stone-gathers-no-moss" effect, causing a reduction or stagnation in real wages of less-educated workers (the educated workers are assumed to be shielded from the "rolling-stone-gathers-no-moss-effect" for reasons explained below);
3 longer jobless spells for the unskilled as against the more skilled, reinforcing the flatter earnings profile for the former group;
4 these factors as a whole provide a trade-dependent explanation for the observed wage differential between skilled (or educated) and less-skilled (or less-educated) labor.

Bhagwati and Dehejia, nevertheless, admit that they "doubt that this alternative explanation can carry the weight that the technical-change (and technological) explanation probably does, but it could well be a contributory factor of some, perhaps also growing importance."

The hypothesis that trade liberalization has made comparative advantage "kaleidoscopic" is based on the observed evidence that the world economy is now increasingly integrated and that the convergence of technology among the Organization for Economic Cooperation and Development (OECD) countries and the spread of global multinational corporations have brought many modern industries to most countries. Therefore, according Bhagwati and Dehejia (1994), more industries, are now "footloose" than before. The evidence of increasing globalization is documented by many researchers such as Baumol *et al*. (1989), Frankel (1994), and Dunning *et al*. (1990).

One of the most significant characteristics of the globalization of production is the extent to which it aids the mobility of assets, notably money capital and innovatory capacity, and of intermediate products, notably technology and management skills, across national boundaries. This mobility immediately

offers the owners of these assets and products a wider option in their location of use—hence described as "footloose."

Dehejia (1996) presents a formal model and presents numerical simulation results, which exemplify the BDH that builds on the outline of the model presented by Bhagwati and Dehejia (1994). We are not concerned in this chapter with theoretical model-building as such. Rather, we deploy a basic model, reflecting conventional wisdom as well as the BDH, which is amenable to testing.

The basic model we use to motivate our analysis is based on a constant elasticity of substitution (CES) production function of the type used by Bound and Johnson (1992) in which the output of each of the j industries (Y_j) depends on physical capital intensity (k_j) and a constant elasticity of substitution aggregator of the i education groups (L_i) where ϕ_{ij} is the share of workers belonging to education group i hired by industry j.

$$Y_j = F(k, L, t, \tau, T) = k_j \left[\sum_i \phi_{ij} \left(A_i(t, \tau, T) L_{ij} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{3.1}$$

$A_i(t, T, \tau)$ captures both a technological change, which is a function of time t, and learning by doing which in turn is a function of potential experience defined as age minus years of schooling minus six and is denoted by T and on-the-job-tenure τ. They all augment the services of education group i workers. The elasticity of intralabor substitution, σ, is assumed to be equal across industries.

Adding the process of learning by doing, however, distinguishes our model from the one used by Bound and Johnson. Moreover, we know of no other model that enters so explicitly the process of learning by doing the way we have (below). In a competitive market in each industry the real wage rate of each type of labor is equal to its marginal product so that

$$\begin{aligned} w_{ij}(t, \tau, T) &= \frac{\partial Y_j}{\partial L_{ij}} \\ &= k_j \phi_{ij} \left[A_i(t, \tau, T) \right]^{\frac{\sigma-1}{\sigma}} \left[\sum_i \phi_{ij} (A_i(t, \tau, T) L_{ij})^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}} L_{ij}^{\frac{-1}{\sigma}} \end{aligned} \tag{3.2}$$

Dividing and multiplying (3.2) by $Y_j^{\frac{1}{\sigma}}$, the wage equation can be simplified.

$$w_{ij}(t, \tau, T) = \left[\frac{Y_j}{L_{ij}} \right]^{\frac{1}{\sigma}} \phi_{ij} \left[k_j A_i(t, \tau, T) \right]^{\frac{\sigma-1}{\sigma}} \tag{3.3}$$

Let us define

$$A_i(t, \tau, T) = N_i(t) e^{\delta_i \left[\sum_l^M \tau_l + E(T - \sum_l \tau_l) \right]} \tag{3.4}$$

N is a technology argument that is a function of time t, and δ_i captures learning by doing by worker i which can be convex, concave, or linear in potential

experience T and of on-the-job-tenure τ depending on whether the partial derivatives of δ_i with respect to T and τ are smaller, greater, or equal to zero, respectively. $l = 1,2,\ldots, M$ is the number of jobs individual i has had over the period of potential experience T. This implies that E is an operator that can take different values depending on whether an individual is educated or less-educated. Bhagwati and Dehejia (1994) assume that E takes on the value of 1 if an individual is educated and 0 if less-educated. The assignment of the binary values is based on the assumption of "rolling-stone-gathers-no-moss" effect with educated labor being shielded from that effect. Taking the logarithms of (3.4) we can write:

$$\ln A_i(t, \tau, T) = \eta_i(t) + \delta_i\left(\sum_{l}^{M} \tau_l + E\left(T - \sum_{l} \tau_l\right)\right) \tag{3.5}$$

In (3.5) η_i captures worker type i skilled-biased technological change and δ_i captures the learning by doing of worker i.

Under free trade, the home and foreign countries' aggregate demand for the output of industry j produced at home (Q_j) relative to the same good produced in the foreign country is assumed to be

$$Q_j = \theta_j P_j^{-\varepsilon} \tag{3.6}$$

where P_j is the price of good j produced at home relative to that in the foreign country; θ_j is an exogenous demand shift parameter reflecting consumer taste and other factors and is assumed to be identical across countries; and ε is the absolute price elasticity of product demand for each industry. Substituting (3.4) and (3.6) into (3.3), then taking its logarithms, yields the following wage equation for workers of type i in any given industry j:

$$\ln w_{ijj} = c + \frac{1}{\sigma}\ln \theta_j - \frac{\epsilon}{\sigma}\ln P_j - \frac{1}{\sigma}\ln L_{ijj} + \ln k_j + \ln \phi_{ijj}$$

$$+ \frac{\sigma - 1}{\sigma}\left[\eta_i(t) + \delta_i\left(\sum_{l}^{M} \tau_l + E\left(T - \sum_{l} \tau_l\right)\right)\right] \tag{3.7}$$

Equation (3.7) is quite comprehensive in the sense that it relates the wage rate to most of the widely discussed worker-type-specific and firm-specific factors set forth to date. For example, the wage rate of worker type i in industry j is positively related to (1) the demand shift for good j (θ_j); (2) capital intensity in industry j (k_j); (3) a positive demand shift for worker type i in industry j (ϕ_{ij}); (4) worker-type-i skill-biased technological change (η_i); (5) learning by doing (δ_i). It is negatively related to (6) the increase in supply of worker of education type i who has industry j specific skills (L_{ij}); and (7) the loss of comparative advantage in production of good j at home represented by an increase in P_j.

It is important to note, however, that when σ goes to infinity—that is when labour types are perfect substitutes across industries—industry-specific demand

shocks such as changes in θ_j and P_j and labor-specific supply shocks such as changes in L_{ij} will have a negligibly small impact on the wages of workers within an industry. This is because the impact of such shocks will spread out across all types of workers across all industries.

In order to derive an equation for the relative wage rate of labor of one education type to some other education type from (3.3)—say the relative wage rate of labor of education type i to education type k—we can write:

$$\frac{w_{ij}}{w_{kj}}(t, \tau, T) = \left[\frac{L_{ji}}{L_{kj}}\right]^{\frac{-1}{\sigma}} \left[\frac{A_i(t, \tau, T)}{A_k(t, \tau, T)}\right]^{\frac{\sigma-1}{\sigma}} \tag{3.8}$$

Suppose that worker i is educated and worker k is less-educated. Let us also assume that E takes a value of 1 for worker i and 0 for worker k as suggested above. Then $A_i(t, \tau, T)$ and $A_k(t, \tau, T)$ are reduced to $\eta_i(t) + \delta_i T$ and $\eta_i(t) + \delta_i \sum_l^M \tau_l$ for educated and less-educated workers, respectively.

Let us further assume that labor is immobile across countries but freely mobile within a country across industries. This ensures that for a given potential experience and on-the-job-tenure the wage rate is the same across industries for workers of the same education level—the subscript j is therefore discarded hereafter. After substituting for $A_i(t, \tau, T)$ and $A_k(t, \tau, T)$ and taking logarithms equation (3.8) becomes

$$\ln \frac{w_i}{w_k}(t, \tau, T) = \ln \frac{\phi_i}{\phi_k} - \frac{1}{\sigma} \ln \left[\frac{L_i}{L_k}\right] + \frac{\sigma - 1}{\sigma} \times [\eta_i(t) - \eta_k(t)] + \frac{\sigma - 1}{\sigma}\left[\delta_i T - \delta_k \sum_l \tau_l\right] \tag{3.9}$$

Suppose worker i is educated (skilled) and worker k is less-educated (less-skilled) then there are four important explanations nested in the above equation for the widening of skilled to less-skilled workers' wage differential. The first and second terms on the right-hand side provide the demand and supply explanations, respectively. The demand shift in favor of educated workers increases the wage differential, whereas, the increase in the relative supply of educated workers decreases it.

The third term furnishes the skill-biased technological change explanation—that is if $\eta_i(t) > \eta_k(t)$ the wage differential will increase over time. The fourth and last term on the right-hand side is the Bhagwati–Dehejia "rolling-stone-gathers-no-moss-effect-with-skilled-worker-sheilded-from-that-effect" trade-dependent explanation as discussed above.

For an illustration let us hold constant the impact of all other variables and suppose that $\delta_i(\tau) = \delta_k(\tau) = \delta\tau$. Next suppose that during time period T both educated and less-educated workers experience labor turnovers and resulted

jobless spells for a subperiod $\check{T}$—so that $\check{T} = T - \sum_l \tau_l$. Since educated workers are assumed to be unaffected by the turnover and the resulting jobless spell, their log real wage rate due to learning by doing will have increased by δT the end of T. Whereas, that of the less-educated workers will have increased only by $\delta\sum_l \tau_l$. The log relative wage rate will, therefore, have widened by $(\sigma-1)\check{T}$.

A distinguishing feature of this model is that it does not rule out other explanations but rather adds the BDH for an additional role just as suggested by Bhagwati and Dehejia (1994). We can even further simplify (3.9) by following Murphy *et al*. (1998) in assuming that the shift in the demand for products is felt proportionately by all type of workers and that the ratio of one type of labor to another does not change across industries. Equation (3.9) can be rewritten as

$$\ln \frac{w_i}{w_k}(t, \tau, T) = C - \frac{1}{\sigma} \ln \left[\frac{L_i}{L_k}\right] + \frac{\sigma - 1}{\sigma} [\eta_i(t) - \eta_k(t)] + \frac{\sigma - 1}{\sigma}\left[\delta_i T - \delta_k \sum_l \tau_l\right] \tag{3.10}$$

where C is a constant. Numerous testable forms of (3.10), (3.9) and (3.7) are essentially what we intend to confront with the data for empirical investigation.

The empirical analysis

The previous section presented a theoretical model that predicts that trade volatility increases labor turnover and that increased labor turnover impacts wages. The theory predicts that skilled workers out-perform unskilled workers in a labor market characterized by increased turnover. We analyze a panel survey of Canadians from 1993 to 1996 and take a reduced form approach to examine the empirical veracity of the theoretical model. We examine whether trade volumes had an impact on relative wages in Canada over this period covering the initiation of the NAFTA trade agreement. Then we examine whether trade volatility had an impact on labor turnover in the Canadian labor market over this period and finally we examine whether jobless spells affected relative wages.

In the following analysis, educated workers are defined as those who received university degrees, certificates or diplomas, ranging from college graduates to PhD. Conversely, less-educated workers include those who did not receive university degrees or diplomas, or, certificates or diplomas from community colleges.

Table 3.1 presents summary statistics for some of the key variables in the sample. It reveals that real wages (in constant 1992 dollars deflated in the consumer price index) have declined for both educated and less-educated workers over the four-year period. More noteworthy is the fact that the relative wages of

Table 3.1 Descriptive statistics of educated to less-educated workers

Variables		*1993*	*1994*	*1995*	*1996*
Real wage rate	• educated	19.13	18.72	18.46	18.26
		(0.18)	(0.17)	(0.18)	(0.12)
	• less-educated	12.25	12.107	12.05	12.01
		(0.05)	(0.05)	(0.05)	(0.04)
	• educated to less-educated	1.561	1.546	1.532	1.521
Years of schooling	• educated	17.45	17.47	17.56	17.63
		(0.05)	(0.04)	(0.04)	(0.03)
	• less-educated	12.21	12.33	12.44	12.57
		(0.02)	(0.02)	(0.02)	(0.02)
	• educated to less-educated	1.429	1.416	1.412	1.402
Years of experience	• educated	13.31	12.70	12.76	12.49
		(0.23)	(0.23)	(0.23)	1
	• less-educated	12.59	12.30	12.32	(0.14)
		(0.09)	(0.10)	(0.23)	12.46
	• educated to less-educated	1.057	1.032	1.036	(0.06)
					1.002
Job tenure (years)	• educated	9.802	8.277	7.629	6.962
		(0.82)	(2.08)	(2.06)	(1.28)
	• less-educated	6.803	5.99	5.648	5.523
		(0.82)	(0.79)	(0.79)	(0.52)
	• educated to less-educated	1.441	1.380	1.350	1.260
Labor supply (thousands)	• educated	6,415	6,840	7,060	7,279
		(12.33)	(12.79)	(13.32)	(8.59)
	• less-educated	8,248	7,992	7,868	7,866
		(6.88)	(7.29)	(7.90)	(5.40)
	• educated to less-educated	0.777	0.855	0.897	0.925
Unemployment rate (percentage)	• educated	8.085	7.611	6.809	7.054
		(0.04)	(0.03)	(0.03)	(0.02)
	• less-educated	13.72	12.78	11.94	12.45
		(0.03)	(0.03)	(0.03)	(0.02)
	• educated to less-educated	0.589	0.595	0.570	0.567
Full-time/part-time ratio		3.956	3.956	3.676	3.279
Unionization rate (percentage)	• educated	46.00	45.00	43.00	40.00
	• less-educated	30.00	28.00	28.00	27.00
	• educated to less-educated	1.533	1.607	1.535	1.481
Number of observations	All	16,734	16,977	15,982	36,297
	• male	8,749	8,853	8,196	18,522
	• female	7,985	8,124	7,786	18,270
	• educated	2,262	2,391	2,327	5,686
	• less-educated	14,472	14,586	13,655	30,611

Source: Author's weighted calculation from the SLID.

Note
Standard errors in parenthesis.

educated to less-educated workers have fallen slightly between 1991 and 1994. This is in contrast to what has been observed in the United States. The *years of schooling* row of the table reveals that between 1993 and 1996 the average years of schooling completed by both educated and less-educated workers increased by some 1% and 3%, respectively. This is also reflected by the decreasing values for the *years of experience* row for both types of workers since potential experience is defined as age minus six minus years of schooling. This upward trend in education acquisition, perhaps, is a response to an increasing demand for more educated labor represented by a decreasing relative unemployment rate of the educated to less-educated workers.

The row on *job tenure* shows that job tenure decreased quite dramatically for both types of workers, but particularly more for educated workers. Job tenure is the duration of the current employment spell. This is consistent with the hypothesis of an increasing turnover due to the emergence of trade-liberalization-pushed kaleidoscopic comparative advantage that made industries footloose. However, other factors cannot be ruled out either.

The important aspect in the *labor supply* and *unemployment rate* rows to observe are the relative labor supply and relative unemployment rate of educated-to less-educated workers. Notice that the relative labor supply rose whereas the relative unemployment rate declined. In a simple demand and supply diagram the first pushes the relative wages of the educated workers down; the latter suggests that the sluggish demand for less-skilled workers has responded with a quantity adjustment (i.e. higher unemployment) rather than a price adjustment (i.e. lower wages). Both of these factors, therefore, imply that the relative real wage rate of educated workers would have been a lot higher in their absence.

The *full-time/part-time* row shows that the ratio declined slightly over the 1993–1996 period. The falling relative full-time/part-time jobs rate is consistent with increasing incidences of labor turnover, which in turn could be caused (in addition to other factors) by trade volatility due to trade liberalization. The *unionization rate* row shows that the unionization rates declined for both educated and less-educated workers. However, the unionization rate of educated workers declined by more than it did for less-educated workers (indicated by the falling ratio of educated to less-educated workers' unionization rates).

Trade and the education premium in Canada

Little research has been done investigating the changes in wages of educated and less-educated workers in Canada but in the literature there is a general consensus that there has been a positive shift in relative demand for the educated (or, skilled) workers (Gera *et al*., 1999). The disagreement, nevertheless, surfaces when it comes to explaining the factors behind the positive demand shift.

The two most familiar explanations for a rightward shift in the relative demand for skilled workers in Canada are trade liberalization and a skilled-biased technological change—the latter being more popular than the first one. To investigate

the relative contribution of trade and technologies to changes in the educated and less-educated workers' wages and wage differential we run the following multivariate regression:

$$W_{it} = \alpha_1 + \beta_1(\text{TRADE}_{jt}) + \beta_1(E^*\text{TRADE}_{jt}) + \pi_1(\text{TECH}_{jt}) + \pi_2(E^*\text{TECH}_{jt}) + \eta(\text{CAPITAL}_{jt}) + \theta X_{it} + \mu Z_i + \psi Y_t \quad (3.11)$$

where W_{it} is the real wage rate of worker i in time t; E is a dummy variable that takes a value of 1 if individual i is educated (i.e. ever received a university degree, certificate or diploma, ranging from below Bachelor's to PhD.) and 0 if less-educated (otherwise). Since less-educated workers are the reference group, the coefficients on E and any continuous variable interacted with E measure the differential effect of being an educated worker relative to less-educated worker.

TRADE_j is the variable representing trade intensity by 3-digit SIC level (the subscript j represents industries) and is equal to total trade by industry j (exports plus imports) divided by total output by industry j. $E^*\text{TRADE}_j$ is TRADE_j interacted with E. TECH_j is a technology variable for industry j. It is a dummy variable taking a value of 1 if an industry is technology-intensive and 0 otherwise. Technology intensity by industry is measured based on a technology use survey conducted by Statistics Canada. CAPITAL_{jt} is the physical capital intensity in industry j in time t. X is a vector of labor-market-specific characteristics such as changes in the labor supply and unemployment rates of educated and less-educated workers; Z is a vector of individual-specific characteristics such as potential experience, on-the-job-tenure, gender, if full time and if unionized; Y controls for a time trend or business cycle.

The results from estimating equation (3.11) are presented in Table 3.2.

The difference between columns 1, 2, 3, and 4 are that in the former two we estimate semi-log multivariate regressions, whereas, in the latter two we take the log of all variables except the dummy variables. In addition, in columns 1 and 3 the variable CAPITAL is not included, whereas, in the columns 2 and 4 it is. The data for CAPITAL for 1996 was not available, thus, the regressions that include CAPITAL are run on fewer observations. The coefficients on TRADE and E^*TRADE are positive in all four regressions implying that trade has had a positive impact on the real wages of both educated and less-educated workers. This perhaps is due to trade putting pressure on domestic industries to become more competitive and therefore more productive, enhancing the marginal productivity of labor. The significantly positive coefficients on E^*TRADE support the hypothesis that trade widens the educated/less-educated workers' wage differential.

According to the results in Table 3.2, a 1% increase in the volume of trade of goods to output ratio is associated with an increase in the educated/less-educated wage gap by about 2–3%. However, it is important to reiterate that the widening of the educated/less-educated workers wage differential does not come at the expense of the less-educated workers as both workers benefit from trade. However, in relative terms the gains to educated workers is larger than for less-educated workers.

Table 3.2 Results of regression (3.11): log of real wages is the dependent variable

	(1) *Semi-log ML estimates*	*(2)* *Semi-log ML estimates*	*(3)* *Log-linear estimates*	*(4)* *Log-linear estimates*
Intercept	2.055362	2.034451	1.7132	1.7743
	(98.294)	(73.354)	(37.9066)	(30.0616)
TRADE	0.009787	0.005437	0.0077	0.00366
	(5.438)	(2.074)	(5.2367)	(1.7234)
*E**TRADE	0.022638	0.020211	0.03413	0.03252
	(6.684)	(4.611)	(8.5179)	(6.2003)
TECH1	0.064665	0.07007	0.0498	0.05095
	(5.834)	(4.87)	(5.6332)	(4.4368)
*E**TECH1	0.149657	0.133147	0.1307	0.11500
	(5.949)	(4.075)	(4.5482)	(3.07224)
EXP	0.018865	0.019901	0.0595	0.07637
	(19.276)	(15.497)	(4.9261)	(5.0631)
EXP^2	−0.000324	−0.000339	0.00199	−0.00242
	(−17.888)	(−14.306)	(0.6396)	(−0.6112)
TENURE	0.002589	0.002592	0.0071	0.00862
	(33.345)	(26.07)	(0.9375)	(0.8698)
$TENURE^2$	−0.000004052	−0.00000401	0.01192	0.01460
	(−19.407)	(−15.287)	(10.0465)	(9.4727)
FT	0.104002	0.093069	0.1074	0.1153
	(10.269)	(6.864)	(8.1129)	(6.5842)
LS	0.000018374	0.000016149	0.0472	0.03692
	(4.68)	(3.074)	(8.9918)	(5.3657)
SEX	0.228312	0.231183	0.2558	0.2506
	(37.256)	(28.76)	(36.5369)	(27.1739)
UNEMP	−0.021394	−0.021087	−0.01855	−0.01881
	(−26.098)	(−20.025)	(−18.8315)	(−14.9259)
UNION	0.095322	0.093077	0.1200	0.10948
	(15.047)	(11.418)	(17.5823)	(12.3101)
Y1994	0.002841	0.006118	0.00637	0.00842
	(0.231)	(0.492)	(0.6748)	(0.8926)
Y1995	−0.009203	−0.004431	0.00245	0.0059
	(−0.731)	(−0.347)	(0.2509)	(0.6039)
Y1996	−0.01744	—	−0.0107	—
	(−1.63)		(−1.2993)	
CAPITAL	—	0.011643	—	0.013559
		(2.017)		(3.0335)
R^2	0.3669	0.3699	0.4409	0.4485
Durbin–Wat	2.103	2.1028	1.395	1.369
Number of observations	19,039	11,049	15,539	8,982

Note
t-ratios in parenthesis.

Moreover, as expected the coefficients on technology are quite significant for both educated and less-educated workers and that its impact on the educated is some three to four times higher than that of the less-educated. Nevertheless, the magnitude of the impact of technology on the relative wages of educated to less-educated workers is similar to that of trade. This result is in line with that found by Baldwin and Raiquzzaman (1998), but is in sharp contrast to that by Gera *et al*. (1999) who find that technology has a much more favorable effect on the relative wages than trade.

Gera *et al*. (1999) find strong evidence that advanced technologies are biased toward the use of skilled labor and thus conclude that skill-biased technological change is perhaps the most important factor in shifting the skilled labor relative demand curve to the right. Similarly, Baldwin and Rafiquzzaman (1998) find that both trade and technology are contributing factors toward the widening wage differential phenomenon. As they put it

> The past twenty years have seen a change in earnings inequality, both in the United States and Canada. The debate over the causes of increasing inequality has focused on whether it is changes in trade patterns or whether it is technological change that is at fault. This paper has demonstrated that both are at work.

The coefficients on all other variables, with the exception of labour supply, are in the expected direction in all of the regressions. Looking at Table 3.2, column 1, they could be interpreted as following: holding everything else constant, a year added to potential experience raises the real wage of all workers by about 1.8%; a month added to on-the-job tenure pushes the real wage up by 0.25%; a full-time job pays an hourly wage that is 10% higher than a comparable part-time job; on average men's wage is 22% higher than that of women; 1% increase in national unemployment rate suppresses real wages by 2.1%; unionized jobs pay 9.5% more than nonunionized; the coefficient on capital intensity, as expected, is positive. The results in columns 2, 3, and 4 are identical to those in column 1 in terms of signs and significance, although values for some variables are slightly different.

The puzzling part, however, is the positive coefficient on the labor supply (LS) variable: an increase in the labor supply pushes real wages up. When we separated the educated and less-educated it was found that the positive effect of labor supply on the real wages of educated workers is 6.5 times stronger than that of less-educated. This perhaps is a support to some sort of Lucas-type positive externality attached to the size of the skilled labor stock. However, the positive coefficient on the less-educated labor supply is puzzling.

The results from estimating (3.11) support the proposition that trade plays a significant role in the widening of educated/less-educated wage differential. However, since the widening does not come even partially at the expense of less-educated workers the result fails to fit the SST trade explanation. Instead, below we turn to the model we developed above to explain the relationship of trade and wages as evidenced by the result here.

Trade volatility and labor turnover

In order to investigate the effect of trade volatility on labor turnover we run the following regression:

$$\text{TURNOVER}_{it} = \alpha_1 + \beta_1(\text{CVTRADE}_{jt}) + \beta_1(E^*\text{CVTRADE}_{jt}) + \pi_1(\text{TECH}j) + \eta(\text{CAPITAL}_{jt}) + \theta X_{it} + \mu Z_i + \psi Y_t \quad (3.12)$$

where TURNOVER is a dummy variable taking a value of 1 if worker i experiences a labour turnover and 0 otherwise. CVTRADE is the coefficient of variation of the TRADE variable, representing trade volatility.

The results from estimating equation (3.12) are reported in Table 3.3.

Table 3.3 Results of regression (3.12) from logistic procedure of maximum likelihood estimation: the dependent variable is labor turnover

	(1)	*(2)*
Intercept	−0.9273	−0.6389
	(42.247)	(11.439)
E	0.4053	0.4099
	(24.551)	(13.168)
CVTRADE	0.0131	0.0153
	(32.340)	(25.416)
TECH1	0.2712	0.2793
	(28.666)	(16.242)
EXP	−0.0106	−0.0128
	(35.096)	(29.782)
TENURE	0.0323	0.0309
	(2255.806)	(1387.345)
FT	−0.2865	−0.363
	(20.046)	(17.482)
LS	0.000267	0.000254
	(102.608)	(50.449)
SEX	0.1564	0.1582
	(13.852)	(7.889)
UNEMP	−0.0393	−0.0418
	(42.011)	(27.716)
UNION	0.0641	0.0809
	(1.836)	(1.702)
Y1994	0.0861	0.0737
	(1.854)	(1.356)
Y1995	0.2044	0.1893
	(10.147)	(8.562)
Y1996	0.4143	—
	(56.079)	
CAPITAL	—	−0.067
		(5.590)
Number of observation	19,083	11,067

Note
Chi-square statistics in parenthesis.

As in Table 3.2, the second column of Table 3.3 includes the variable CAPITAL whereas the first one does not. In both columns the coefficients on CVTRADE (coefficient of variation of the TRADE variable) are significantly positive implying that the trade variation significantly intensifies the incidence of labor turnover. The result is in line with that reported by Heisz (1996) in which it was shown that over the period we suspect trade liberalization to have made industries "footloose" (perhaps, through kaleidoscopic comparative advantage) job turnover rates in almost all industries have increased. Similarly, Baldwin and Rafiquzzaman (1994) report a marked increase in labor turnover in all manufacturing industries—the highest turnover occurring in industries that are relatively more exposed to international competition. The coefficients on other variables can be interpreted as the following. Holding everything else constant the incidence of labor turnover is higher among educated workers, men and unionized workers; labor turnover rates increases with the intensity in technology; it falls with potential experience but increases with on-the-job tenure. An increase in labor supply also increases labor turnover but higher unemployment rate reduces it. Full-time workers experience lower incidence of labor turnover than part-time workers.

Jobless spells and the educated/less-educated relative wages

In the previous section it was shown that increased trade volatility intensified labor turnover (or jobless spell). In this section we examine whether the duration of jobless spells slows down the skill accumulation of less-educated workers by more than that of educated workers as suggested by the BDH.

To investigate the impact of increased labor turnover and jobless spell on skill accumulation and hence on growth profile of relative wage rates we estimate an equation based on (3.7). The results are presented in Table 3.4.

As in Tables 3.2 and 3.3, the second column of Table 3.4 includes the variable CAPITAL whereas the first one does not. The significantly positive differential impact of potential experience (the coefficients on E*EXP) on the wages of educated workers is what was hypothesized. However, the positive coefficients on E*Tenure are contrary to what was suggested a priori. This perhaps implies that either, unlike we suggested, it is the educated workers whose knowledge is more industry specific, or that the econometric estimators employed are not fitting the type of data being utilized.

One of the problems inherent in panel data is heterogeneity. Although in Table 3.4 we control for observed heterogeneity among workers, we do not do so for unobserved heterogeneity. Unobserved heterogeneity is a time-invariant latent individual effect correlated with the explanatory variables. If there is no unobserved heterogeneity present in the data then the result obtained from regression (3.7) would be unbiased and consistent, otherwise it will not be (Greene, 1997). However, with a panel data it is likely the case that the data is tormented by it.[2] For example, a higher wage rate associated with higher tenure may not be due

Table 3.4 Results of wage regressions based on (3.7): the dependent variable is the log real wage

	(1)	*(2)*
Intercept	1.9545	2.0105
	(73.7144)	(101.565)
EXP	0.0170	0.01581
	(13.632)	(16.559)
EXP^2	−0.00028	−0.00025
	(−8.740)	(−10.183)
*E**EXP	0.01224	0.01546
	(2.818)	(5.068)
*E**EXP^2	−0.00015	−0.00029
	(−1.207)	(−3.242)
TENURE	0.00258	0.00255
	(22.662)	(28.629)
$TENURE^2$	−4.23E-06	−4.15-E-06
	(−14.091)	(−17.353)
*E**TENURE	0.00305	0.0031
	(5.982)	(8.119)
*E**$TENURE^2$	−9.42E-06	−9.28E-06
	(−6.404)	(−8.248)
FT	0.10515	0.10392
	(6.785)	(9.071)
LS	4.32-E-05	4.76E-05
	(7.532)	(11.236)
SEX	0.24309	0.24939
	(27.677)	(37.487)
UNEMP	−0.01482	−0.01511
	(−12.375)	(−16.323)
UNION	0.12595	0.14391
	(14.389)	(21.549)
Y1994	0.00760	0.00473
	(0.836)	(0.5182)
Y1995	0.01020	−0.0077
	(1.081)	(0.742)
Y1996	—	−0.00775
		(−0.984)
CAPITAL	0.03156	—
	(8.365)	
R^2	0.4697	0.4592
Number of observations	9,575	16,804

Note
t-ratios in parenthesis.

entirely to the skill that a worker accumulates through on-the-job-learning-by-doing but it might be due to his/her latent individual-specific ability, and because of that ability the worker may have a longer tenure on the job in the first place. The same argument goes for education: the fact that an educated worker commands a higher

wage rate may not be due to his/her high education but rather to his/her individual specific ability which probably also has helped him/her to achieve higher education.

One way of correcting for unobservable heterogeneity is the Fixed Effect approach.[3] This eliminates the individual effects in the sample by transforming the data into deviations from individual means and, therefore, is dubbed the Within Group Estimator (WGE). Thus we run the following WGE Fixed Effect

Table 3.5 Results of wage regression based on (3.13): dependent variable is log real wage

	(1) *Fixed-effect* *WGE estimates*	*(2)* *Fixed-effect* *WGE estimates*
Intercept	0.793819 (27.877)	−0.511095 (−9.898)
E	0.777371 (10.257)	−0.673385 (−4.006)
Tenure	0.020224 (57.867)	0.013803 (22.789)
Tenure^2	−0.000008237 (−4.868)	0.000021733 (7.325)
*E**tenure	0.005807 (6.604)	0.011495 (6.78)
*E**tenure^2	−0.0000445 (−8.973)	−0.000002559 (−0.283)
Exp	0.090366 (43.916)	0.095702 (38.735)
Exp^2	−0.004949 (−40.195)	−0.003401 (−21.164)
*E**exp	0.148269 (21.874)	0.115023 (11.131)
*E**exp^2	−0.00649 (−13.872)	−0.004627 (−6.169)
Spell	—	−0.006165 (−8.604)
spell^2	—	0.000005397 (5.861)
*E**spell	—	0.003607 (1.761)
*E**spell^2 −0.00000431	—	(−2.223)
R^2	0.1752	0.1393
Durbin–Wat	1.1185	1.3266
Number of observations	85,980	28,305

Note
t-ratios in parenthesis.

regression model:

$$w_{it} = \alpha_1 + \rho_1(\text{tenure}_{it}) + \rho_2(E^*\text{tenure}_{it}) + \sigma_1(\text{tenure}_{it})^2 + \sigma_1(E^*\text{tenure}_{it})^2 + \psi_1(\exp_{it}) + \psi_2(E^*\exp_{it}) + \xi_1(\exp_{it})^2 + \xi_2(\exp_{it})^2 \quad (3.13)$$

where $w = W_i^t - W_{I^*}$; tenure = $\text{TENURE}_i^t - \text{TENURE}_{i^*}$; exp = $\text{EXP}_i^t - \text{EXP}_{I^*}$. W_i^t is the wage rate of an individual in group i in time t and W_{i^*} is the mean wage of an individual belonging to group *i*. TENURE_i^t, TENURE_{I^*}, EXP_i^t, and EXP_{I^*} have similar interpretation. The coefficients of (3.13) are of a different scale, however, their interpretations are similar to that of regression (3.7). The results of equation (3.13) are presented in Table 3.5, column 1.

The significantly positive coefficients on both ***e**tenure*** and ***e**exp*** make the result identical to that obtained reported in Table 3.4 and the issue that jobless spell has dissimilar effect on the educated and less-educated is therefore not resolved. However, the issue can be resolved by including "jobless spell" as an explicit variable. In Table 3.5 column 2 we report the WGE Fixed Effect regression result of (3.8) in which spell = $\text{SPELL}_i^t - \text{SPELL}_i$, whereas, SPELL_i^t is jobless spell of an individual in group *i* in time *t* and SPELL_I is the mean jobless spell of group *i*. The significantly negative coefficients on spell and the significantly positive coefficients on *E**spell clearly are testament to the hypothesis that a jobless spell slows down the skill accumulation of less-educated workers by more than that of educated workers.

Concluding remarks

In this chapter we studied the trend of educated and less-educated workers' absolute and relative wages over the period covered by the Survey of Labor and Income Dynamics (SLID) data and investigated if they are causally linked to international trade. We also provided a trade dependent theoretical explanation for the causal links between the two variables: trade and the educated/less-educated wages.

We showed the widening of educated/less-educated wage differential was an occurring phenomenon (albeit not as strong as that in the United States or some other developed countries) in at least the groups of workers that are more likely to be exposed to international competition brought about by trade. We also demonstrated that the differential would have been significantly higher in the absence of changes in the relative supply of educated workers and in the absence of quantity adjustment (increasing relative unemployment of less-educated workers) rather than price adjustment (wage changes).

We identified trade as a significant contributor to the rising education premium and showed that its impact on relative wages of educated to less-educated workers was just as a great as that of technology. Although the result implicated trade as a possible cause of the widening in educated/less-educated wage differential, it was found that trade was not necessarily harmful to less-educated workers.

The real wages of both types of workers respond positively to increased trade liberalization—it is just that the educated benefit by more than the less-educated. This finding adds some valuable information for social policies that are designed to counteract or alleviate the effect of trade liberalization on affected workers.

Moreover, we found that trade volatility (represented by the coefficient of variation of the trade variable) is a statistically significant determinant of labor turnover. We found that jobless spells (due to increased labor turnover) have affected educated workers more favorably than less-educated workers and that resulted in the widening of their relative compensation. These results are consistent with the alternative theoretical explanation for trade and wages developed by Bhagwati and Dehejia. In future research we plan to extend our empirical analysis to more recent years, and use an instrumental variables approach, which may allow us to make stronger claims of putative causal relationships along the lines of the BDH.

Notes

1 See, for example, Feenstra and Hanson (1996).
2 Hausman and Taylor (1981); Osberg (1986).
3 Greene (1997); Johnston and DiNardo (1998).

Bibliography

Acemoglu, D. (1998) "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence," NBER Working paper series 6658.

Baldwin, J. R. and M. Rafiquzzaman (1998) "The Effect of Technology and Trade on Wage Differentials between Non-Production and Production Workers in Canadian Manufacturing," Research paper series N. 98, Analytical Studies Branch, Statistics, Ottawa, Ontario, Canada.

Bartel, A. P. and G. J. Borjas (1981) "Wage Growth and Job Turnover: an Empirical Analysis," in Wherwin Rosen (ed.) *Studies in Labor Markets*, Chicago, NBER and University of Chicago Press: 65–90.

Berman, E., J. Bound and Z. Griliches (1994) "Changes in the Demand for Skilled Labour within U.S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturing," *Quarterly Journal of Economics*, 109: 367–397.

Bhagwati, J. and V. H. Dehejia (1994) "Free Trade and Wages of the Unskilled—Is Marx Striking Again?," in Bhagwati, J. and Kosters, M. H. (eds), *Trade and wages: Leveling wages down?*, Washington, DC: AEI Press: 36–75.

Bhagwati, J. and M. H. Kosters (eds) (1994) *Trade and wages: Leveling wages down?*, Washington, DC: AEI Press: 76–107.

Bound, J. and G. Johnson (1992) "Changes in the Structure of Wages during the 1980s: An Evaluation of Alternative Explanation," *American Economic Review*, 82: 371–392.

Borjas, G. J., R. B. Freeman and L. F. Katz (1992) "On the labour market effects of immigration and trade," in G. B. Borjas and R. B. Freeman (eds) *The Economic Effects of Immigration in Source and Receiving Countries*, Chicago: Chicago University Press.

Buckberg, E. and A. Thomas (1996) "Changes Wage dispersion in the 1980s: Resurrecting the Role of Trade through the Effects of Durable Employment," IMF Staff Papers, June: 336–353.

Chipman, J. S. (1969) "Factor Price Equalization and the Stolper–Samuelson Theorem," *International Economic Review*, 10: 399–406.

Davis, S. J. and J. Haltiwanger (1991) *Manufacturing plants: 1963–1986*, NBER working paper series 3722.

Deardroff, A. V. and D. S. Hakura (1994) Trade and Wages—What are the questions?, in Bhagwati, J. and M. H. Kosters (eds) *Trade and wages: Leveling wages down?*, Washington, DC: AEI Press: 76–107.

Dehejia, V. H. (1996) "International Trade and the Domestic Wage Structure: Alternative Explanations," *Konjunkturpolitik*, 42(2–3): 155–180.

Feenstra, R. C. and G. H. Hanson (1996) "Foreign Investment, Outsourcing, and Relative Wages," in: R. C. Feenstra, G. M. Grossman, and D. A. Irwin (eds) *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*, Cambridge, MA: MIT Press: 89–127.

Francois, J. F. and D. Nelson (1998) "Trade, Technology, and Wages: General Equilibrium Mechanics," *The Economic Journal*, 108: 1483–1499.

Freeman, R. and K. Needels (1991) "Skill differentials in Canada in an era of rising labor market inequality," NBER working paper series: 3827.

Freeman, R. B. (1995) "Are Your Wages Set in Beijing?" *Journal of Economic Perspectives*, 9: 15–32.

Gera S., W. Gu and Z. Lin (1999) "Technology and the demand for skills: An industry-level analysis," Industry, Ottawa, Ontario, Canada.

Gorecki, P. K. (1989) "Job Turnover in Canada's Manufacturing Sector," Research paper series N. 22, Analytical Studies Branch, Statistics, Ottawa, Ontario, Canada.

Gottschalk, P. and T. Smeeding (1997) "Cross National Comparisons of Earnings and Income Inequality," *Journal of Economic Literature*, 35: 633–687.

Gray T. and J. Johnson (1997) "Technology-Induced Wage Premia in Canadian Manufacturing Plants during the 1980," Research paper series N. 92, Analytical Studies Branch, Statistics, Ottawa, Ontario, Canada.

Greene, W. H. (1997) *Econometric Analysis*, Upper Saddle River: Prentice Hall.

Greenhalgh, C. and G. Mavrotas, 1996, Job training, new technology and labour turnover, *British Journal of Industrial Relations*, 34: 131–150.

Hausman, J. A. and W. E. Taylor (1981) "Panel Data and Unobservable Individual Effects," *Econometrica*, 49: 1377–1398.

Heisz, Andrew (1996) "Changes in Job Tenure and Job Stability in Canada," Research Paper Series N. 95, Analytical Studies Branch, Statistics, Ottawa, Ontario, Canada.

Johnston, J. and J. DiNardo (1997) *Econometric Methods*, 4th edn, New York: The McGraw-Hill Companies, Inc.

Katz, L. and K. M. Murphy (1992) "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107: 35–78.

Krueger, A. B. (1993) "How Computers Have Changed the Wage Structure: Evidence From Microdata," *The Quarterly Journal of Economics*, 108: 33–60.

Krugman, P. and R. Lawrence (1994) "Trade, Jobs and Wages," *Scientific American*, April: 44–49.

Kuhn, P. and A. Sweetman (1998) "Wage Loss Following Displacement: The Role Of Union Coverage," *Industrial and Labor Relations Review*, 51: 384–400.

Lang, K. (1998) "Journal of The Effect of Trade Liberalization on Wages and Employment: The Case Of New-Zealand," *Labour Economics*, 16: 792–814.

Lawrence, R. and M. Slaughter (1993) "International Trade And American Wages in the 1980s: Giant Sucking Sound or Small Hiccup?" Brookings Papers on Economic Activity: *Macroeconomics*, 2: 161–226.

Leamer, E. (1985) "International Trade Theory: The Evidence," in Grossman, G. M. and Rogoff, K. (ed.) *Handbook of International trade*, III: 1339–1394.

Leamer, E. (1993) "Wage Effects of a U.S.-Mexican Free Trade Agreement," in Garber, P. M. (ed.), *The Mexico-U.S. free trade agreement*, Cambridge: MIT Press: 57–127.

Lichtenberg, F. R. (1987) "The Comparative Advantage of Educated Workers in Implementing New Technology," *Review of Economics and Statistics*, 69: 1–11.

Lucas, R. E. 1993, Making a miracle, *Econometrica*, 61: 251–272.

McPherson, D. A. and J. B. Stewart (1990) "The Effect of International Competition on Union and Nonunion Wages," *Industrial and Labor Relations Review*, 43: 435–446.

Magee, S. P. (1980) "Three Simple Tests of the Stolper–Samuelson Theorem," in Peter Oppenheimer (ed.), *Issues in International Economics*, London: Oriental Press, 138–153.

Marcotte, D. E. (1995) "Declining Job Stability: What We Know And What It Means," *Journal of Policy Analysis and Management*, 14: 590–598.

Matyas, L. and P. Sevestre (1996) *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Dordrecht: Kluwer Academic Publishers.

Mincer, J. (1981) "Labor Mobility and Wages," in Sherwin Rosen (ed.) *Studies in labor Markets*, Chicago NBER and University of Chicago Press: 21–63.

—— (1986) "Wage changes in job changes," NBER working paper series 1907.

—— (1991) "Human Capital, Technology and the Wage Structure: What Do Time Series Show?" NBER working paper series No. 3581.

Murphy, K. and R. Topel (1997) "Unemployment and Nonemployment," *American Economic Review*, 87: 295–300.

Murphy, K. M. and F. Welch (1991) "The Role of International Trade in Wage Differentials," in Kosters, M. H. (ed.) *Workers and their Wages: Changing Patterns in the United States*, AEI Studies, no. 520, Washington, DC:AEI Press, 39–69.

Murphy, K., Riddell, W., and P. Romer (1998) "Wages, Skills and Technology in the United States and Canada," NBER working paper series: 6638.

Nye, W. W. (1996) "Firm-Specific Learning-by-Doing in Semiconductor Production: Some Evidence from the 1986 Trade Agreement," *Review of Industrial Organization*, 11: 383–394.

OECD (1997) Trade, Earnings and Employment: Assessing the Impact of Trade with Emerging Economics on OECD Labour Markets, *Employment Outlook*, OECD.

Osberg, L., R. L. Mazany, R. Apostle and D. Clairmont (1986) "Job Mobility, Wage Determination and Market Segmentation in the Presence of Sample Selection Bias," *Canadian Journal of Economics*, 19: 319–346.

Ragan, J. F. and S. P. Smith (1981) "In Turnover Rates on Male/Female Pay Differentials," *The Journal of Human Resources*, 16(3): 343–365.

Ramey, V. (1994) "Time-Series Evidence on the Sources of Wage Inequality," *American Economic Review*, 84: 11–16.

Ravenga, A. L. (1992) "Exporting Jobs? the Impact of Import Competition on Employment and Wage in U.S. Manufacturing," *Quarterly Journal of Economics*, 107: 255–284.

—— (1997) "Employment and Wage Effects of Trade Liberalization: The Case of Mexican Manufacturing," *Journal of Labor Economics*, 15: S20-S43.

Rowthorn, R. and R. Ramaswamy (1998) "Growth, Trade and De-Industrialisation," IMF working paper, Research Department, International Monetary Fund.

Sachs, J. D. and H. J. Shatz (1994) "Trade and Jobs in U.S. Manufacturing," *Brookings Papers on Economic Activity*, 34: 1–84.

Saeger, S. (1997) "Globalisation And Deindustrialisation: Myth and Reality in the OECD," *Weltwirtschaftliches Archiv*, 133: 579–608.

Slaughter, M. J. (1998) "International Trade and Labour-Market Outcomes: Results, Questions and Policy," *The Economic Journal*, 108: 1452–1462.

Stern, R. M. (1994) *The Stolper–Samuelson Theorem: A Golden Jubilee*, Ann Arbor: The University of Michigan Press.

Swinnerton, K. and H. Wial (1995) "Is Job Stability Declining in the Us Economy?" *Industrial and Labour Relations Review*, 48: 293–304.

Wood, A. (1994) *North-South Trade, Employment and Inequality: Changing Fotunes in a Skill-Driven World*, Oxford: Clarendon Press.

—— (1995) "How Trade Hurt Unskilled Workers," *Journal of Economic Perspectives*, Summer, 9(3): 57–80.

—— (1997) "Openness and Wage Inequality in Developing Countries: Latin American Challenge to East Asian Conventional Wisdom," *World Bank Economic Review*, 11: 33–57.

—— (1998) "Globalisation and the Rise in Labour Market Inequalities," *The Economic Journal*, 108: 1463–1482.

Young, A. (1991) "Learning by Doing and the Dynamic Effects of International Trade," *The Quarterly Journal of Economics*, May, CVI: 369–405.

4 Human capital, trade liberalization, and income risk*

Tom Krebs, Pravin Krishna, and William Maloney

Introduction

Using data from Mexico, this chapter studies empirically the link between trade policy and individual income risk and the extent to which this varies across workers of different human capital (education) levels. Longitudinal income data on workers are used to estimate time-varying individual income risk parameters in different manufacturing sectors in Mexico between 1987 and 1998, a period in which the Mexican economy experienced substantial changes in trade policy. In a second step, the variations in trade policy—across different sectors and over time—are used to estimate the linkage between trade policy and income risk for workers of varying education levels. Our findings are as follows. The level of openness of an economy is not found to be related to income risk for workers of any type. Furthermore, changes in trade policy (i.e. trade policy reforms) are not found to have any affect on the risk to income faced by workers with either low or high levels of human capital. However, workers with intermediate levels of human capital are found to experience a statistically and economically significant increase in income risk immediately following liberalization of trade. Our findings thus point to an interesting nonmonotonicity in the interaction between human capital, income risk, and trade policy changes.

In recent years, the impact of the increased "openness" of countries on factor markets has been actively debated in the theoretical and empirical literature in international trade.[1] This literature has focused primarily on how greater openness to trade might differentially impact the *level* of the returns earned by different factors of production. More specifically, guided by the logic of the well-known Stolper–Samuelson theorem, which predicts an increase in returns to abundant factors with trade, analysts have sought to examine how trade may have impacted workers of different levels of human capital (education)—looking, for example, to see whether wage inequality between skilled and unskilled workers has increased in countries with a relative abundance of skilled workers (developed countries) in the manner predicted by the theory.[2]

Recently, however, the literature has also begun discussing the important question of how openness may impact income *volatility* at the aggregate (see, for instance, Rodrik, 1997) as well as at the individual level (see Krebs *et al*., 2005). Various channels through which trade reform might affect individual income risk

have been articulated. For example, lowering trade barriers leads to an increase in foreign competition in the import-competing sectors and is likely to induce a reallocation of capital and labor across firms and sectors. In the short run, the resulting turbulence may raise individual labor income risk.[3] Rodrik (1997), going beyond the short-term reallocational effects of trade reform on income risk, has additionally argued that increased foreign competition following trade reform will increase the elasticity of the goods and the derived labor demand functions. If higher demand elasticity translates any given shock into larger variations in wages and employment, lower trade barriers may lead to increased individual income risk.[4] On the other hand, it has also been suggested that the world economy is likely to be less volatile than the economy of any single country, which leads to goods prices that are more stable worldwide than in any single autarkic economy. This opens up the possibility that greater openness may reduce the variance in individual incomes. Thus, overall, the relationship between openness and individual income risk is theoretically ambiguous, requiring empirical analysis for its resolution. In a recent paper, Krebs *et al*. (2005) have investigated this trade-income risk question empirically and have reported economically significant impacts of trade policy (changes) on the volatility of worker incomes.[5]

The possibility that trade policy may affect income risk in labor markets raises additional questions. Is the effect uniform across workers? Or, for instance, are workers with higher human capital better able to insulate themselves against trade policy changes? Evaluating empirically the interaction between human capital, trade policy, and income risk is important for a number of reasons. First, this helps us evaluate the merits of particular theoretical arguments that assume differential effects of openness on variables such as labor turnover for workers of different human capital levels. For instance, the well-known argument by Bhagwati and Dehejia (1994) regarding "kaleidoscopic comparative advantage" argues that the observed trend of increased inequality in wages between workers of high and low levels of education may be explained by the differences in the way that openness impacts labor turnover rates (a component of income risk) for these two groups. Second, from a public policy perspective, it is important to know how policy changes impact different segments of the population; public opposition to "globalization" is driven at least in part by concerns that openness affects most adversely the poorest workers. Finally, such an analysis will deepen our understanding of the trends in the supply of human capital in the labor market.

In this chapter, we study *empirically* the effects of trade policy on individual income risk for workers of different levels of human capital using the following approach. For each industry (sector), we use longitudinal data on individual earnings for different types of workers to estimate time-varying parameters of individual income risk using an approach similar to Carroll and Samwick (1997), Meghir and Pistaferri (2004), and Storesletten *et al*. (2004). More specifically, we focus on the variance of (unpredictable) changes of individual income as a measure of income risk, and carefully distinguish between transitory and persistent income shocks. The distinction between transitory and persistent income

shock is important since workers can effectively self-insure against transitory shocks through own savings, which implies that the effect of these types of shocks on workers' consumption and welfare are quite small (Aiyagari, 1994; Heaton and Lucas, 1996; Levine and Zame, 2002).

In contrast, highly persistent or permanent income shocks have a substantial effect on the present value of future earnings, and therefore lead to significant changes in consumption even if workers have own savings. Thus, from a welfare point of view, persistent income shocks matter the most, and we therefore focus on the relationship between trade policy and the persistent component of income risk. More specifically, after obtaining the estimates of the persistent component of income risk for each industry and each year, we use these estimates in conjunction with tariff data (as a proxy for trade policy) to study empirically the effect of trade policy on income risk.

Our previous discussion highlights the need for longitudinal information on incomes at a disaggregated level (individual or household)[6] in countries that have undergone discernable (and, ideally substantial) changes in their external regime. Unfortunately, countries that maintain detailed longitudinal records on individual incomes have rarely undertaken major trade reforms and countries that have undertaken extensive trade policy reforms have rarely collected data on individuals of requisite scope and quality. In this chapter, however, we focus on one country that satisfies both criteria, namely Mexico. As is well known, the Mexican economy experienced substantial changes in trade policy in the late 1980s and in the later half of the 1990s.[7] Moreover, as we discuss in detail later in this chapter, the Mexican government, since the mid-1980s, has conducted quarterly longitudinal income surveys that comprehensively surveyed workers in all manufacturing sectors of the economy—providing the unique data source that we use in our study.

Our empirical results for the Mexican case can be summarized as follows. The level of openness of an economy is not found to be related to income risk for workers of any type. Furthermore, changes in trade policy (i.e. trade policy reforms) are not found to have any affect on the risk to income faced by workers with either low or high levels of human capital. However, workers with intermediate levels of human capital are found to experience a statistically and economically significant increase in income risk immediately following liberalization of trade. Our findings thus point to an interesting nonmonotonicity in the interaction between human capital, income risk, and trade policy changes. Finally, the welfare costs of the increased income risk following trade policy reforms are substantial—amounting to reduction of between 1% and 2% of permanent income.[8]

The rest of the paper proceeds as follows. We first describe the estimation procedure and data that we use to estimate individual income risk. We then discuss the empirical methodology we use to estimate the relationship between income risk and trade policy. We also present a theoretical framework that will be used to translate changes in income risk into changes in welfare. Finally, we discuss our results and end with some concluding observations.

Income risk

The first stage of our analysis concerns the estimation of individual income risk. Our estimation strategy follows earlier approaches taken in the literature estimating US labor income risk (Carroll and Samwick, 1997; Meghir and Pistaferri, 2004; and Storesletten *et al*., 2004) with some important differences which we discuss in detail below. As in these papers, we define income risk as the variance of (unpredictable) changes in individual income, and carefully distinguish between transitory and persistent income shocks. This separation is essential from a welfare point of view since self-insurance through saving works well for transitory income shocks, but not for persistent ones (Aiyagari, 1994; Heaton and Lucas, 1996; and Levine and Zame, 2002). For this and other reasons (to be discussed in detail below), we eventually focus on persistent shocks and their relation to trade policy.

Data

In Mexico, the National Urban Employment Survey (ENEU) conducts extensive *quarterly* household interviews in the 16 major metropolitan areas and is available from the mid-1980s (we use data from 1987 to 1998 in our study). The sample is selected to be geographically and socioeconomically representative. The survey questionnaire is extensive in scope and covers all standard elements such as participation in the labor market, earnings, and so forth. The ENEU is structured so as to track a fifth of each sample across a five quarter period. To construct the panels, workers were matched by position in an identified household, level of education, age, and sex to ensure against generating spurious transitions. Taken together, we have 44 complete panels of 5 periods (i.e. quarters) each, spanning a total of 12 years (48 quarters). The number of individuals surveyed in any given calendar year is approximately 100,000. Table 4.1 presents a summary description of the workers surveyed by the ENEU. Data on sectoral (i.e. industry) trade barriers and other sectoral and macroeconomic variables were obtained from the World Bank.

Specification

Our survey data provide us with earnings (wage rate times number of hours worked) of individuals. As in previous empirical work, we assume that the log

Table 4.1 ENEU Worker Survey—Summary (1987–1998)

Variables	
Mean age	32
Mean years of education	8
Fraction high school and above	17
Fraction wage earners	65
Fraction self employed	25

of this labor income of individual i employed in industry j in period t, $log\ y_{ijt}$, is given by

$$\log y_{ijt} = \alpha_{jt} + \beta_t \cdot x_{ijt} + u_{ijt} \tag{4.1}$$

where α_{jt} and β_t denote time-varying coefficients, x_{ijt} is a vector of observable characteristics (such as age and education), and u_{it} is the stochastic component of earnings. The stochastic component u_{ijt} represents individual income changes that are *not* due to changes in the return to observable worker characteristics. For example, income changes that are caused by an increase in the skill (education) premium are not contained in u_{ijt} In this sense, u_{ijt} measures the unpredictable part of changes in individual income. It is to be noted that we allow the fixed effects α_{jt} to vary across sectors, but that the coefficient β_t is restricted to be equal across sectors. The latter assumption is made in order to ensure that the number of observations is large compared to the number of parameters to be estimated.

We assume that the stochastic term is the sum of two (unobserved) components, a permanent component ω_{ijt} and a transitory componentη_{ijt}.

$$u_{ijt} = \omega_{ijt} + \eta_{ijt} \tag{4.2}$$

Permanent shocks to income are fully persistent in the sense that the permanent component follows a random walk

$$\omega_{ij,t+1} = \omega_{ijt} + \epsilon_{ij,t+1} \tag{4.3}$$

where the innovation terms, $\{\epsilon_{ijt}\}$, are independently distributed over time and identically distributed across individuals. It is to be noted that we allow the parameters to depend on time t and industry j, but not on individual i. We further assume that $\epsilon_{ij,t+1} \sim \mathrm{N}(0, \sigma^2_{\epsilon j,t+1})$. Transitory shocks have no persistence, that is, the random variables η_{ijt} are independently distributed over time. Clearly, η_{ijt} captures both temporary income shocks and measurement error. We assume that they are normally distributed with zero mean and a variance that is independent of i, but may depend on time or industry: $\eta_{ijt} \sim \mathrm{N}(0, \sigma^2_{\eta jt})$.

Our specification for the labor income process is in accordance with the empirical work on US labor income risk. For most part, however, the previous literature has confined attention to the special case of time-independent variances (homoscedastic case). As we discuss later, the introduction of time-variation in the parameters $\sigma^2_{\epsilon jt}$ and $\sigma^2_{\eta jt}$ makes the estimation of these parameters more challenging. Finally, we should note that, in principle, both $\sigma^2_{\epsilon jt}$ and $\sigma^2_{\eta jt}$ represent measures of individual income risk. In this chapter, we will focus on $\sigma_{\epsilon jt}$ and its relationship to trade policy. This choice is motivated by the following two considerations. First, as mentioned before, transitory income shocks are unlikely to generate consumption volatility since self-insurance through own-saving is highly effective, and the welfare effects of these shocks are therefore small

(Aiyagari, 1994; Heaton and Lucas, 1996; and Levine and Zame 2002). Second, the term $\sigma^2_{\eta jt}$ will absorb the measurement error in income, and therefore overstate the degree of transitory income risk (and this is indeed reflected in our data, as we discuss in the fifth section).

Estimation

Consider the change in the residual of income of individual i between period t and $t+n$:

$$\begin{aligned}\Delta_n u_{ijt} &= u_{ij,t+n} - u_{ijt} \\ &= \epsilon_{ij,\,t+1} + \cdots + \epsilon_{ij,\,t+n} + \eta_{ij,\,t+n} - \eta_{ijt}\end{aligned} \tag{4.4}$$

Thus, we have the following expression for the variance of income changes:

$$\text{var}\left[\Delta_n u_{ijt}\right] = \sigma^2_{\epsilon j,\,t+1} + \cdots \sigma^2_{\epsilon j,\,t+n} + \sigma^2_{\eta\, jt} + \sigma^2_{j,\,t+n} \tag{4.5}$$

We use the moment restrictions (4.5) to estimate the parameters $\sigma^2_{\epsilon jt}$ and $\sigma^2_{\eta jt}$ using Generalized method of moments (GMM),[9] where the sample analogs to the moment conditions are formed by using the estimates of u_{ijt} obtained as residuals from regressions of labor income on observable characteristics as specified in (4.1)—an approach also used by Meghir and Pistaferri (2004), Storesletten *et al*. (2004) and Gourinchas and Parker (2002). Specifically, the estimator is obtained by minimizing:

$$\sum \left(\text{var}[\Delta_n u_{ijt}] - \left(\sigma^2_{\epsilon j,\,t+1} + \cdots \sigma^2_{\epsilon j,\,t+n} + \sigma^2_{\eta jt} + \sigma^2_{\eta j,\,t+n}\right)\right)^2 \tag{4.6}$$

The first-order conditions corresponding to the parameters $\sigma^2_{\epsilon j,t}$ and $\sigma^2_{\eta j,t}$ are given by

$$\begin{aligned}\forall t\ \frac{\partial \sum}{\partial \sigma^2_{\epsilon j,\,t}} &= 0 \\ \forall t\ \frac{\partial \sum}{\partial \sigma^2_{\eta j,\,t}} &= 0\end{aligned} \tag{4.7}$$

Note that in general there are many more moment conditions (4.5) than there are parameters to be estimated. More precisely, for each time period t and each industry j, there are two parameters ($\sigma^2_{\epsilon jt}$ and $\sigma^2_{\eta jt}$), but n moment conditions given by (4.5).

Note also that the objective function (4.6) is quadratic, which implies that the first-order conditions associated with the corresponding minimum-distance problem are linear in $\sigma^2_{\epsilon jt}$ and $\sigma^2_{\eta jt}$—a feature that facilitates the estimation

substantially. Specifically, the first order conditions can be organized into a linear equation system

$$\mathbf{A} \cdot \sigma = \mathbf{B} \tag{4.8}$$

where $\epsilon = (\sigma^2_{\epsilon,2} \cdots \sigma^2_{\epsilon,t} \cdots \sigma^2_{\epsilon,T}, \sigma^2_{\eta,2}, \sigma^2_{\epsilon,t}, \sigma^2_{\eta,T})'$ is a $2(T-1)$-dimensional vector of income parameters (T being the total number of time periods). Estimates of these income parameters can then easily be obtained through matrix inversion: $\sigma = \mathbf{A}^{-1}\mathbf{B}$.

Some intuition for the way in which our approach separates transitory from permanent income shocks can be obtained from the following simple example. Suppose that risk is time-invariant, $\sigma^2_{\epsilon jt} = \sigma^2_{\epsilon j}$ and $\sigma^2_{\eta jt} = \sigma^2_{\eta j}$, an assumption that has been made by most of the previous empirical literature on income risk. In this case, the moment restrictions (4.5) becomes the following:

$$\text{var}[\Delta_n u_{ijt}] = 2\sigma^2_{\eta j} + n\,\sigma^2_{\epsilon j} \tag{4.9}$$

Thus, the variance of observed n-period income changes is a linear function of n, where the slope coefficient is equal to $\sigma^2_{\epsilon j}$. The insight that the random walk component in income implies a linearly increasing income dispersion over time is the basis of the estimation method used by several authors. For example, Carroll and Samwick (1997) estimate σ^2_{ϵ} by performing Ordinary Least Squares (OLS) regressions of the left-hand side of (4.9) on n. While the preceding example, with time-invariant parameters, serves to illustrate the intuition underlying the estimation procedure, it should be clear that our exercise is more general in the sense that it allows for arbitrary time-variation in the income risk parameters.

Estimation using ENEU data

The preceding section provided a detailed description of a general econometric methodology that may be used to estimate time-variant income risk parameters given longitudinal data on individual incomes. We note here some additional issues that arise in applying this methodology to our data, with particular emphasis on the type of income risk accounted for by our estimation procedure.

In forming the sample analogs to the moment conditions (4.5), we can only use information on individuals who are present in a given manufacturing industry in both time periods t and $t+n$. In doing so, we pick up shocks to workers who retain their jobs but experience income changes due to changes in their wage rates or the number of hours worked. Moreover, we also account for changes in income experienced by displaced workers who are reemployed in the same industry. In contrast, displaced workers who are reallocated to a different manufacturing industry are not taken into account.[10] However, the exclusion of such workers should not be expected to cause too great an underestimation of our income risk parameters as the fraction of displaced manufacturing workers who make

transition from one manufacturing sector to another is very small. In our data, on average less than 10% of displaced workers undergo such a transition. This is consistent with observations from the United States that most job creation and destruction takes place within industries.

It is worth noting that there are very few labor force participants in our survey who do no work and receive zero wages in any given quarter, which is mainly a consequence of the lack of any government-provided unemployment insurance in Mexico and the very active informal labor market. More importantly, the proportion of workers who are unemployed for longer than the five quarter-periods is extremely small (implying that forming the moment conditions as we do above does not cause problems in the consistent estimation of persistent income shocks faced by workers).

Finally, we should mention that the variability in income experienced by workers in our data set derives from differential changes in the number of hours worked and also from both upward and downward changes in their real wages. Thus, despite the often cited downward rigidity of wages, our sample includes large numbers of workers whose real wages declined. Specifically, Mexico experienced very high inflation rates during our sample period with annual declines in *aggregate* real wage as high as 25% during this time (see, for instance, Hanson, 2003), implying that the wage rates of some individual workers declined by an even larger amount.

Trade reform and income risk

The procedure outlined in the previous section provides us with estimates of individual income risk, $\sigma^2_{\epsilon jt}$, for each industry (i.e. manufacturing sector) j and time period, (i.e. quarter, t). These time-varying, industry-specific estimates in conjunction with observations on trade policy, τ_{jt}, allow us to estimate the relationship between income risk, $\sigma^2_{\epsilon jt}$, and openness, τ_{jt}. Consider the following linear specification allowing for industry fixed-effects and aggregate time effects

$$\sigma^2_{\epsilon jt} = \alpha_0 + \alpha_{1j} + \alpha_{2t} + \alpha_\tau \tau_{jt} + \alpha_\delta \Delta\tau_{jt} + \nu_{jt} \tag{4.10}$$

where the coefficients α_{1j} capture the industry fixed-effects, the α_{2t}'s pick up aggregate trends, the coefficient α_τ measures the effect of openness on income risk and α_δ captures the effects of changes (in the preceding year, say) in trade policy, $\Delta\tau_{jt}$. The inclusion of industry dummies in the specification above allows us to control for any fixed industry-specific factors that may affect the level of riskiness of income in that industry. Moreover, the inclusion of time dummies controls for any changes in macroeconomic conditions that affect the level of income risk. While this ensures that our estimation results are not driven by changes in *macroeconomic* conditions (business cycle effects and/or long-run structural changes) unrelated to trade policy, it also means that identification of the relationship between $\sigma^2_{\epsilon jt}$ and τ_{jt} will have to be based on the differential rate of change in trade barriers across sectors over time (or the vector of observations

on tariffs in the panel corresponding to (4.10) will be perfectly collinear with the time-dummy vector). This, however, does not pose problems for our estimation since trade barriers in Mexico and their changes over time do in fact do exhibit substantial cross-sectional variation.[11]

Several econometric issues arise in the estimation of (4.10). One concern is that the left hand side variable, income risk, is estimated and not observed. This is not a substantial problem by itself as it is well known that while "measurement error" in the dependent variable does reduce precision, it does not bias our estimates. A concern arises, however, from the fact that the estimates of $\sigma^2_{\epsilon jt}$ have different standard errors across industries, that is, the specification we have described above suffers from a heteroscedasticity problem. Further, since the industries all belong to the same macroeconomic environment, there is a possibility of contemporaneous correlation in their σ's even after controlling for observable macroeconomic factors as in (4.10'), that is, $\text{Cov}(v_{jt}v_{j\hat{t}}) \neq 0$. Finally, serial correlation in income volatility within an industry is a possibility, that is, $\text{Cov}(v_{jt}v_{jt'}) \neq 0$. Given the possible presence of heteroscedasticity, spatial correlation and serial dependence, consistent estimates of the standard errors associated with the coefficient estimates in (4.10) above are obtained by using robust estimation techniques.

Income risk and welfare

The preceding discussion has outlined our approach to estimating the relationship between trade policy and income risk. We now turn to the analysis of the link between income risk and welfare, which is provided by a simple dynamic model with incomplete markets along the lines of Constantinides and Duffie (1996) and Krebs (2004). The model extends the basic insights of the large literature on the permanent income hypothesis to a general equilibrium setting with isoelastic preferences and incomplete markets. It remains tractable enough to permit closed-form solutions for equilibrium consumption and welfare that are simple and transparent. Clearly, our goal here is not to provide a complete assessment of the effects of income risk on welfare taking into account all possible channels, but rather to articulate a simple framework that allows us to obtain *indicative* estimates of welfare change through the *income risk* channel. The model structure and assumptions underlying our approach and the limitations of our methodology are discussed below in detail.

The model features long-lived households (workers) that make consumption/saving choices in the face of uninsurable income shocks. These income shocks are permanent, which implies that self-insurance is an ineffective means to smooth out income fluctuations. In other words, the effect of permanent income shocks on consumption is substantial. In accordance with Constantinides and Duffie (1996) and Krebs (2004), we consider an exchange economy and do not model the labor-leisure choice. In this section, we briefly discuss the basic assumptions of the model and state the main welfare results.

Model

Time is discrete and open-ended. Income of household i employed in industry j in period t is denoted by y_{ijt}. Income is random and defined by an initial level $\tilde{y}_{ij_0}$ and the law of motion

$$\tilde{y}_{ij,t+1} = (1 + \mu_{j,t+1})(1 + \theta_{ij,t+1})\tilde{y}_{it} \tag{4.11}$$

where $\mu_{j,\,t+1}$ is a mean growth-rate effect common across workers in the sector and $\theta_{ij,\,t+1}$ is an individual-specific shock to the growth rate of income. We assume that $\log(1+\theta_{ij,\,t+1})$ is normally distributed with time- and industry-dependent variance σ^2_{jt}.

Although the distribution of individual-specific shocks may change over time, the shocks are unpredictable in the sense that current and future shocks are uncorrelated. To ensure that workers are *ex ante* identical, we also assume that the distribution of shocks is identical across workers. Each household begins life with no initial financial wealth. Households have the opportunity to save, but not borrow, at the common risk-free rate r_t. Hence, the sequential budget constraint of worker i reads

$$a_{ij,t+1} = (1 + r_t)a_{ijt} + y_{ijt} - c_{ijt} \quad a_{ijt} \geq 0,\ a_{ij0} = 0 \tag{4.12}$$

Here c_{ijt} denotes consumption of household i in period t and a_{ijt} his asset holdings at the beginning of period t (excluding interest payment in this period). Note that by assuming the nonnegativity of a_{ijt}, we have automatically ruled out Ponzi schemes.

Households have identical preferences that allow for a time-additive expected utility representation:

$$U(\{c_{ijt}\}) = E\left[\sum_{t=0}^{\infty} \beta^t u(c_{ijt})\right] \tag{4.13}$$

Moreover, we assume that the one-period utility function, u, is given by $u(c) = \frac{c^{1-\gamma}}{(1-\gamma)}$, $\gamma \neq 1$, or $u\,(c) = \log c$, that is, preferences exhibit constant degree of relative risk aversion γ.

Welfare

We can derive an explicit formula for equilibrium welfare that depends on the preference parameters β and γ and the income parameters μ_{jt} and σ^2_{jt}, where σ^2_{jt} is the variance of the log-normally distributed income shocks η. We also show that the variance σ^2_{jt} of the income process (4.11) can be identified with the variance $\sigma^2_{\epsilon jt}$ of the permanent component of our empirical specification (4.1). This provides a tight link between the empirical results obtained above and the welfare

analysis conducted in this section. We now briefly outline and discuss the main welfare results.

For simplicity, assume that the income parameters are time-independent: $\mu_{jt} = \mu_j$ and $\sigma^2_{\epsilon jt} = \sigma^2_{\epsilon j}$. Suppose now that trade reform changes the tariff rate in a particular industry j from τ to $(1+a_\tau)\tau$ permanently. Suppose also that the change in the tariff rate leads to a corresponding permanent change in income risk from σ^2_ϵ to $(1+\Delta_\sigma)\sigma^2_\epsilon$. Clearly, this change in income risk induced by trade reform corresponds to the long-run effect that is associated with the level term, τ_{jt}, on the right hand side of our regression (4.10). We can find the welfare effect of the change in risk, Δ_σ, by calculating the compensating variation in lifetime consumption, Δ_c. That is, we can ask by how much we have to change consumption in each period and state of the world to compensate the household for the change in income risk. We can show that this compensating differential, expressed as percent of lifetime consumption, is given by

$$\Delta_c = \left(\frac{1-\beta(1+\mu)^{1-\gamma}\exp\left(.5\left((1-\gamma)^2-(1-\gamma)\right)(1+\Delta_\sigma)\sigma^2_\varepsilon\right)}{1-\beta(1+\mu)^{1-\gamma}\exp\left(.5\left(1-\gamma)^2-(1-\gamma)\right)\sigma^2_\varepsilon\right)} \right)^{\frac{1}{1-\gamma}} - 1$$

$$\text{if } \gamma \neq 1$$

$$\Delta_c = \exp\left(\frac{\beta}{(1-\beta)^2} \frac{\sigma^2_\varepsilon \Delta_\sigma}{2} \right) - 1 \quad \text{if } \gamma = 1 \tag{4.14}$$

Equation (4.14) shows how to translate long-run changes in labor income risk, Δ_σ, into equivalent changes in lifetime consumption, Δ_c. It provides the answer to the following question: how much lifetime consumption are risk averse workers willing to give up in return for not having to experience the increase in income risk that is caused by a change in trade policy. Note that (4.14) is the result of an *ex ante* welfare calculation under rational expectations. More specifically, (4.14) assumes that workers do not know who will lose and who will gain from trade reform, but they know to what extent trade reform creates winners and losers (the effect of trade reform on the income risk parameters is known *ex ante*).

The welfare expression (4.14) assumes that the change in σ^2_ε is permanent. However, we are also interested in the welfare effect of an increase in income risk from σ^2_ϵ to $(1+\Delta_\sigma)\sigma^2_\epsilon$ for n periods. In this case, the welfare effect is given by

$$\Delta_c = \left[\left(\frac{1-x}{1-x'} \right)\left(1-x'^{n+1}\right) + xx'^n \right]^{\frac{1}{\gamma-1}} - 1 \quad \text{if } \gamma \neq 1 \tag{4.15}$$

$$\Delta_c = \exp\left(\frac{\beta(1-\beta^n)}{2(1-\beta)^2} \sigma^2_\epsilon \Delta_\sigma \right) - 1 \quad \text{otherwise}$$

where we introduced the following notation:

$$x = \beta(1 + \mu)^{1-\gamma}\exp(.5((1-\gamma)^2 - (1-\gamma))\sigma_\epsilon^2)$$

$$x' = \beta(1 + \mu)^{1-\gamma}\exp(.5((1-\gamma)^2 - (1-\gamma))(1 + \Delta_\sigma)\sigma_\epsilon^2)$$

The welfare expressions (4.14) and (4.15) have some intuitive properties. First, the welfare effect of a change in income risk is a nonlinear and increasing function of the initial level of income risk. Put differently, if workers are already exposed to a large amount of income risk, then increasing income risk hurts a lot. Second, the welfare effects are increasing in the risk aversion parameter γ: the more risk averse the workers are, the stronger is the welfare effect of a change in income risk. Finally, the welfare effects are the same for all workers regardless of their wealth. This property is the result of the joint assumption of homothetic preferences and an income process defined as in (4.11).

The welfare expressions (4.14) and (4.15) form the basis for our quantitative welfare analysis of trade reform. In order to conduct such an analysis, we need information about the income parameters μ, σ_ϵ^2, and Δ_σ and the preferences parameters β and γ. Our empirical analysis provides estimates of the income parameters. For the preference parameters, we choose an annual discount factor of $\beta = 0.96$ and allow the degree of risk aversion γ to (separately) take values 1 and 2. These values for the preference parameters are in line with the values used in the macroeconomic literature (Cooley, 1995).

Results

In the first step of our analysis, we use data on individual income changes from workers in different manufacturing sectors in Mexico and the methodology outlined in the second section to estimate quarterly income risk parameters in each of these sectors during the time period 1987–1998. Tables 4.2 and 4.3 provide the average estimate of σ_ϵ^2 and σ_η^2 for each year (averaged across industries) and for each industry (averaged over time) respectively.

The mean value (across industries and over time) of the quarterly variance of the persistent shock, σ_ϵ^2, is estimated to be 0.0065 or 0.026 in annual variance. As expected, given the extent of measurement error in the income data, the estimated variances of transitory shocks are much larger in magnitude.

Human capital categories

We separate workers into three human capital (education) categories. In the first group, we have workers with zero to six years of education. The second group consists of those workers with more than six years of education but who have not graduated high school. Finally, the third group consists of workers who are all high school graduates. As indicated in Table 4.4, in our data set, the first group

Table 4.2 Estimates of persistent and transitory income shocks[a] annual averages (1987–1998)

Year	σ^2_ϵ	σ^2_η	*Sample size*
1987	0.011	0.096	19,136
	(0.003)	(0.002)	
1988	0.005	0.101	35,397
	(0.003)	(0.002)	
1989	0.004	0.103	28,203
	(0.002)	(0.001)	
1990	0.014	0.098	35,167
	(0.002)	(0.001)	
1991	0.001	0.103	37,344
	(0.002)	(0.001)	
1992	0.006	0.106	54,022
	(0.001)	(0.001)	
1993	0.007	0.112	78,741
	(0.001)	(0.001)	
1994	0.006	0.110	121,716
	(0.001)	(0.001)	
1995	0.014	0.118	164,212
	(0.001)	(0.001)	
1996	0.000	0.107	172,766
	(0.001)	(0.001)	
1997	0.006	0.104	172,870
	(0.001)	(0.001)	
1998	0.008	0.097	158,707
	(0.001)	(0.001)	

Note

a Figures shown are annual averages (across industries and quarters) of the point estimates of the persistent shock and the transitory shock. The figures in parentheses are the averages of the corresponding standard errors. Sample size denotes the numbers of workers surveyed in the respective year.

comprises approximately 36% of the workforce, the second group comprises 46% and workers in the highest human capital category the remaining 17% of the workforce.

Table 4.2 also presents estimates (averaging over industries and time) of the quarterly variance for the different human capital categories. An indicated there, on average, workers with intermediate levels of human capital face the highest level of income risk in our data. The quarterly variance of the persistent shocks to income that they face is 0.0075. In comparison, the quarterly variance is estimated to be 0.0045 for workers in the low human capital category and 0.006 for workers in the high human capital category.

A similar regression analysis (not reported here) was conducted for transitory income-shock parameters, σ^2_η, but we did not find any statistically significant relationship between transitory shocks to income and trade policy. One explanation

Table 4.3 Estimates of persistent and transitory income shocks[a] industry averages (1987–1998)

Industry	σ^2_ϵ	σ^2_η	*Industry*	σ^2_ϵ	σ^2_η
311	0.013	0.131	352	0.020	0.111
	(0.0004)	(0.0003)		(0.0025)	(0.0019)
313	0.012	0.088	353	0.002	0.081
	(0.0007)	(0.0005)		(0.0009)	(0.0007)
321	0.005	0.097	356	0.006	0.079
	(0.0006)	(0.0005)		(0.0016)	(0.0011)
322	0.012	0.124	369	0.011	0.113
	(0.0008)	(0.0006)		(0.0014)	(0.0011)
323	0.008	0.107	371	0.003	0.110
	(0.0022)	(0.0015)		(0.0031)	(0.0025)
324	0.004	0.088	381	0.006	0.125
	(0.0002)	(0.0001)		(0.0006)	(0.0004)
331	0.004	0.120	382	−0.002	0.098
	(0.0027)	(0.0020)		(0.0015)	(0.0011)
332	0.019	0.121	383	0.008	0.056
	(0.0017)	(0.0013)		(0.0002)	(0.0002)
341	0.004	0.102	384	0.004	0.073
	(0.0016)	(0.0012)		(0.0002)	(0.0001)
342	0.011	0.134	390	0.005	0.143
	(0.0016)	(0.0012)		(0.0062)	(0.0047)
351	0.012	0.107			
	(0.0029)	(0.0023)			

Note

a Figures shown are averages over time of the point estimates of the persistent shock and the transitory shock for the respective industries. The figures in parentheses are the averages of the corresponding standard errors.

Table 4.4 Educational categories—summary statistics (1987–1998)

Educational categories	*Percentage of labor force*	σ^2_ϵ
0–6 years of education	36	0.0045
6–12 years of education	46	0.0075
12 years of education and above	18	0.0060

of this negative finding is that our estimates of transitory income shocks are contaminated by measurement error in income.

As indicated earlier, to relate trade policy to idiosyncratic income risk, the specification we use is

$$\sigma^2_{\epsilon jt} = \alpha + \alpha_j + \alpha_t \tau_{jt} + \alpha_\delta \Delta\tau_{jt} + v_{jt} \tag{4.10}$$

In (4.10) we have included on the right-hand side the following variables: τ is the *ad valorem* sectoral tariff rate, $\Delta\tau$ is the change in the tariff over the preceding

Table 4.5 Trade policy and income risk—panel estimates[a]

Variables	*Low HC*	*Intermediate HC*	*High HC*
	σ_ϵ^2 *vs*	σ_ϵ^2 *vs*	σ_ϵ^2 *vs*
τ	0.077	0.127	0.180
	(0.097)	(0.104)	(0.181)
$\Delta\tau$	−0.041	−0.119	−0.081
	(0.125)	(0.03)	−(0.056)
Time effects	Included	Included	Included
Industry effects	Included	Included	Included
N	450	450	450
R^2	0.18	0.25	0.2

Note

a Figures in parentheses are robust standard error estimates obtained by allowing for heteroscedasticity, contemporaneous correlation of errors across industries and serial correlation within industries.

year, α_j is an industry fixed- effect, and α_t is a time dummy that captures general macroeconomic events in the economy. The effect of the tariff level on income risk is given by the coefficient α_τ and the effect of tariff changes on income risk is given by the coefficient α_δ.

Regression results are presented in Table 4.5.[12]

We note first that the estimate of α_τ is insignificant for all the human capital categories and we are therefore unable to reject that the *mean* effect of the tariff level on income risk is zero for all three groups. However, trade policy *changes* have statistically and economically significant short run effect on income risk for individuals with intermediate levels of human capital ($\hat{\alpha}_\delta = -0.219$, with an estimated standard error of 0.102). This estimate indicates that, on average, lowering the tariff rate by 5% would, for a year, raise σ_ϵ from a mean level of 0.0065 to 0.013—a substantial increase in the risk to income faced by individuals. It is worth emphasizing that the effect of trade policy changes on risk to income faced by very low income workers is highly insignificant. Individuals in the top human capital category see only marginally (statistically) significant effects.

Welfare analysis

Using the theoretical results derived in the fourth section and the empirical estimates obtained from the estimation of (4.10), we can estimate the welfare costs of trade policy changes due to any changes in income risk faced by workers. For the preference parameters we choose an annual discount factor of $\beta = 0.96$ and a degree of risk aversion of $\gamma = 1$ or $\gamma = 2$. As mentioned before, these values for the preference parameters are in line with the values used in the macroeconomic literature (Cooley, 1995). Consider a tariff reform which involves a

lowering of the tariff level by 5%. Given our empirical estimates, this would raise σ^2_ϵ in the short run (i.e. for one year following the reform) from a mean level of 0.065 to 0.013. The corresponding welfare cost of this change is calculated to be 0.98% of permanent consumption if the coefficient of risk aversion $\gamma = 1$ and is calculated to be 1.96% of lifetime consumption if the $\gamma = 2$ instead (always using an annual discount factor of $\beta = 0.96$).

Our findings can be summarized as follows. The level of openness of an economy is not found to be related to income risk for workers of any type. Furthermore, changes in trade policy (i.e. trade policy reforms) are not found to have any affect on the risk to income faced by workers with either low or high levels of human capital. However, workers with intermediate levels of human capital are found to experience a statistically and economically significant increase in income risk immediately following liberalization of trade. Our findings thus point to an interesting nonmonotonicity in the interaction between human capital, income risk, and trade policy changes. Finally, the welfare costs associated with the estimated increases in income risk, for workers with intermediate levels of human capital are substantial.

Conclusions

This chapter studies empirically the relationship between trade policy and *individual* income risk. The analysis proceeds in three steps. First, longitudinal data on are used to estimate individual income risk of manufacturing workers in various human capital categories. Second, the variation in income risk and trade barriers—both over time and across sectors—is used to arrive at estimates of the relationship between trade policy and individual income risk for these different workers. Finally, using the estimates of this relationship between trade policy and income risk, a simple dynamic general equilibrium model with incomplete markets is used to obtain estimates of the welfare costs of the effects of trade policy on income risk.

However, it is worth pointing out that our welfare analysis here focuses *exclusively* on the link between trade policy and individual income risk, and other possible channels through which trade policy may affect the economy are not studied here. More specifically, we would expect trade reform to have positive effects on the efficiency of resource allocation and economic growth, and such effects are important factors that ought to be taken into account when evaluating the total costs and benefits of trade reform. In addition, our welfare calculations are based on a simple theoretical model whose limitations include its neglect of the effect of income risk on labor supply and capital accumulation. Moreover, our calculations do not take into account that the welfare cost of an increase in income risk might be partially offset by a rise in transfer payments from the government or firms.[13] Finally, while our estimates of income shocks were obtained using observations on individuals over a limited time period, our welfare analysis assumes that shocks that are highly persistent through our sample period are

equally persistent beyond this period. Thus, the welfare results presented in this chapter have to be interpreted with caution keeping in mind our exclusive focus on the link between trade policy and income risk and the methodological limitations noted above.

Notes

* Paper prepared for presentation at the Festschrift conference organized in the honor of Jagdish Bhagwati, *Globalization: Prospects and Problems*, University of Florida, Gainesville, January 28–30, 2005. We wish to thank our discussant, J. David Richardson, for many valuable comments on an earlier draft of the paper.

1 For a general discussion of the debate, see for instance, Bhagwati (2001) and Rodrik (1997).

2 Early papers in this area include Lawrence and Slaughter (1993) and Borjas *et al*. (1992). See Richardson (1995) for a survey discussion.

3 See, for instance, the analysis of policy change by Fernandez and Rodrik (1991), in which *ex ante* identical workers experience *ex-post* different outcomes since some workers retain their jobs while others are forced to move to other firms.

4 While Rodrik (1997) appears to have in mind mostly aggregate volatility, it can be seen that with heterogeneous impact of such a change on firms (and thus individuals) in the economy, individual income volatility will be raised as well. See, for instance, the analysis of Melitz (2003) for an example of an aggregate policy shock affecting an entire sector leading to heterogeneous outcomes for individual firms within that sector.

5 In this chapter, we draw substantially on this earlier paper by us, specifically in presenting the main issues and in describing the methodological approach.

6 It should be clear that our need for longitudinal data follows from our desire to study how trade policy impacts the magnitude and frequency of individual income shocks (changes). This is a quite distinct task from that of measuring the impact of trade policy on the distribution of income levels.

7 In an early wave of trade reforms in the late 1980s, tariffs were cut from an average of about 40% to about 15%.

8 At this stage, it is worth pointing out that our welfare analysis focuses *exclusively* on the link between trade policy and individual income risk, and that other possible channels through which trade policy may affect the economy are not studied here. More specifically, we would expect trade reform to have positive effects on the efficiency of resource allocation and economic growth, and these effects are important factors that should be taken into account when evaluating the total costs and benefits of trade reform.

9 More specifically, we follow the bulk of the literature and use the equally weighted minimum distance (EWMD) estimator. Altonji and Segal (1996) suggests that the EWMD estimator (identity weighting matrix) is superior to the two-stage GMM estimator (optimal weighting matrix) once small-sample bias is taken into account.

10 This allows us to circumvent the extremely difficult problem of assigning industries (and thus trade policy) to individuals who transit to different industries. Including individuals who make transitions to the service (nontradables) sector by using the procedure of counting them as belonging to the manufacturing sector in which they are first observed does not result in any qualitative difference in our reported results.

11 For instance, in Mexico, tariffs varied between 80% and 20% prior to the trade reforms of 1987 and ranged between 20% and 10% by 1994—implying a variation in tariff changes across sectors that is quite substantial.

12 See Krebs *et al*. (2005) for a detailed qualitative discussion of why policy endogeneity and the possible self-selection of workers into sectors is only a minimal concern in this context.
13 Being that such transfers are provided by entities within the economy, they should perhaps nevertheless be counted as costs, even if the risk to workers is fully offset by these payments.

Bibliography

Aiyagari, R. (1994) "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109: 659–684.

Altonji, J., and L. Segal (1996) "Small-Sample Bias in GMM estimation of covariance structures," *Journal of Business and Economic Statistics*, 14: 353–366.

Bhagwati, J. (2001) *Free Trade Today*, Princeton, NJ: Princeton University Press.

Bhagwati, J., and Dehejia, V. (1994) "Freer Trade and Wages of the Unskilled: Is Marx Striking Again?" in Bhagwati and Kosters (eds) *Trade and Wages*, Washington, DC: American Enterprise Institute Press, 36–75.

Carroll, C. and Samwick, A. (1997) "The Nature of Precautionary Wealth," *Journal of Monetary Economics*, 40: 41–72.

Constantinides, G., and Duffie, D. (1996) "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104: 219–240.

Cooley, T. (1995) *Frontiers of Business Cycle Research*, Princeton, NJ: Princeton University Press.

Fernandez, R., and Rodrik, D. (1991) "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty," *American Economic Review*, 81(5): 1146–1155.

Gourinchas, P., and Parker, J. (2002) "Consumption over the Life-Cycle," *Econometrica*, 70: 47–89

Hanson, G. (2003) "What Has Happened To Wages in Mexico since NAFTA? Implications for Hemispheric Free Trade," *National Bureau of Economic Research*, working paper No 9563.

Heaton, J., and D. Lucas (1996) "Evaluation the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 104: 443–487.

Krebs, T. (2004) "Consumption-Based Asset Pricing Models with Incomplete Markets," *Journal of Mathematical Economics*, 40: 191–206.

Krebs, T., Krishna, P., and Maloney, W. (2005) "Trade Policy, Income Risk and Welfare," *National Bureau of Economic Research*, working paper No 11255.

Krusell, P. and Smith, A. (1998) "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106: 867–896.

Lawrence, R., and Slaughter, M. (1993) "International Trade and American Wages in the 1980s: Giant Sucking Sound or Small Hiccup?" in Martin Neil Baily and Clifford Winston (eds) *Brookings Papers on Economic Activity: Microeconomics 2*, 1993, 161–211.

Levine, D. and Zame, W. (2002) "Does Market Incompleteness Matter?" *Econometrica*, 71: 1695–1725.

Melitz, M. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71: 1695–1726.

Meghir, C. and L. Pistaferri (2004) "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72: 1–32.

Richardson, David, J. (1995) "Income Inequity and Trade: How to Think, What to Conclude," *Journal of Economic Perspectives*, 9 (Summer 1995): 33–55.

Rodrik, D. (1997) *Has Globalization Gone Too Far?*, Washington, DC: Institute for International Economics.

Storesletten, K., Telmer, C., and A. Yaron (2004) "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, 695–717.

Part 2

International technology transfer and multinational firms

5 Patent protection and global Schumpeterian growth*

Elias Dinopoulos, Ali Gungoraydinoglu, and Constantinos Syropoulos

Introduction

Arguably, patent protection of intellectual property rights has acquired the same importance in the "new" knowledge-based global economy as the tariff had in the "old" materials-based global economy. Since patent regimes regulate the creation and international transfer of new products and processes, changes in patent protection can have profound effects on global economic efficiency and income distribution between innovating advanced countries (the "North") and imitating developing countries (the "South").

The economic effects of global patent protection are elusive and difficult to explore for a number of reasons. First, changes in global patent protection generate income-transfer effects. The vast majority of new products are discovered by firms in the North and copied by firms in the South. Since patents discourage imitation of new technologies and thereby provide temporary monopoly power to inventors, higher patent protection results in short-run income transfers from Southern consumers to Northern firms.[1] Second, patent-regime changes affect the knowledge-spillover process within and between countries. For example, although it is possible for technologies with expired patents to generate information that enhances the discovery of new inventions, the threat of litigation against possible patent infringement may limit the usefulness of technical information on patented products aimed at the discovery of new products. In addition, changes in patent protection may in principle induce firms to switch to, or from, other mechanisms that provide intellectual property protection (such as trade secrecy). Third, changes in patent protection generate resource–reallocation effects. By changing the length of temporary monopoly power, patent protection affects the allocation of economic resources between manufacturing of new products and R&D investment. An increase in patent length, for instance, could reduce the amount of R&D investment and the rate of innovation by transferring resources from R&D investment to manufacturing of patented products. For these reasons, it is difficult to evaluate the effects of patent protection without considering formal dynamic economic models.

It is not surprising then that the signing of the General Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPs Agreement)

in the Uruguay Round—which calls for all World Trade Organization (WTO) members to adopt a set of global minimum (Northern) standards on intellectual property rights protection—has been met with skepticism by prominent economists.[2] What is surprising is the relative scarcity of formal economic analyses of the role of patent protection on global growth and poverty. Several new studies have built dynamic growth models to analyze the determinants of international technology diffusion (Eaton and Kortum, 1996, 1999), the effects of globalization on growth and global income distribution (Dinopoulos and Segerstrom, 2006; Grossman and Helpman, 1991; Segerstrom *et al.*, 1990; Sener, 2006), the effects of patent policy on welfare (Iwaisako and Futagami, 2003), and the welfare effects of differential intellectual property protection between countries (Grossman and Lai, 2004). However, these studies either did not explicitly incorporate the role of finite-length patents into the analysis, or they have abstracted from the scale-effects problem.[3]

This chapter develops a North–South model of Schumpeterian (R&D-based) growth and finite-length global patent protection. Schumpeterian growth is a particular type of growth that is generated by the endogenous introduction of new products and/or processes and is based on the process of creative destruction (Schumpeter, 1942). The model generates product-cycle trade, endogenous international technology transfer, and a North–South wage gap. Northern firms develop higher-quality final-consumption products by devoting labor into R&D activities. Each new product is protected by an international and perfectly enforceable patent of finite time duration $T > 0$. When a patent expires, the product becomes generic and all firms in the North know how to produce it. As a result, at each instant in time, a fraction of Northern industries produces generic products under conditions of pure competition, obtaining zero economic profits. Southern firms target Northern generic products to appropriate their technology and transfer the location of production to the South. In addition, Northern firms target industries without patent protection that produce generic products, either in the South or the North, to discover higher-quality products that would replace the production of generic ones. Within the context of the present model, the assumption of international patent enforcement is convenient but not necessary. If the market size of the North is at least as large as the market size of the South, and the North adopts a policy that prohibits the importation of Southern generic products in industries with active patents, then Southern firms will not have an incentive to imitate Northern products protected by patents. As a result, unilateral enforcement of Northern patents establishes an effective global patent independently of the Southern patent regime. For this reason, the present model is not appropriate for analyzing whether or not intellectual property protection must be part of global trade negotiations.

The model highlights the role of patents in governing the evolution of knowledge spillovers across industries protected by patents and those industries whose patents have expired. By introducing a parameter that assigns different weights to these two types of industries, the model generates endogenous long-run scale-invariant Schumpeterian growth. The long-run rate of innovation and

per capita growth is proportional to the exogenous rate of population growth (as in models developed by Jones, 1995; Kortum, 1997; and Segerstrom, 1998); however, per capita growth also depends on the length of global patents. The dependence of long-run growth on patent length, which is a novel result, is governed by the structure of knowledge spillovers (Proposition 1). If knowledge spillovers that affect the evolution of R&D difficulty are symmetric between products with active patents and products with expired patents, then changes in patent protection do not affect the rate of long-run growth. However, if knowledge spillovers from products with active patents differ from those of products without active patents, then changes in patent protection affect long-run growth.

In addition to the long-run growth effects, patent protection has a permanent impact on the rate of imitation and on global income distribution measured by the North–South wage gap (Proposition 3). An increase in the patent length raises permanently the rate of imitation. The effect of patent length on the North–South wage gap can be positive or negative depending on the initial measure of Northern industries under patent protection relative to the parameter measuring the degree of asymmetric knowledge spillovers. In Proposition 3 we show that, if the initial fraction of Northern industries with patents is sufficiently high (say, if more than one half of the high-tech industries enjoy patent protection), then an increase in the global patent length affects adversely the degree of global inequality by raising the relative wage of Northern workers. Such parameter values are also consistent with the notion that stricter intellectual property protection (measured by an increase in the length of global patents) reduces the long-run global rate of innovation and growth and worsens the global income distribution (Proposition 1).

The model also permits an analysis of the dynamic effects of globalization. Motivated by the entrance of China in the world trading system since 1978, and following Dinopoulos and Segerstrom (2006), we analyze the effects of one dimension of globalization, namely the geographic expansion of the world trading system. This dimension of globalization can be captured in the model by analyzing the effects of an increase in the size of the South, measured by the level of its population. Proposition 2 shows that a permanent expansion of the size of the South worsens the long-run wage-income distribution between North and South—by permanently raising the wage of Northern workers relative to the wage of Southern workers—and increases the rate of technology transfer from North to South without affecting the long-run rate of global innovation and growth. This result differs from the one obtained by Dinopoulos and Segerstrom (2006) where globalization just reduces the North–South wage gap.

The next section of the chapter describes the basic elements of the model. Then we establish the uniqueness of the steady-state equilibrium and derives explicit solutions for the endogenous variables. Next we use the equations derived previously to analyze the steady-state properties of the model and to describe the basic findings. Finally, we summarize the conclusions and provide several suggestions for future research.

The model

This chapter develops a North–South model of scale-invariant, endogenous, steady-state Schumpeterian (R&D-based) growth. Specifically, we generalize the North–South model of endogenous growth developed by Segerstrom *et al.* (1990) by adding an endogenous resource-using imitation process and by introducing positive global population growth. The model uses the deterministic R&D technology introduced in the seminal work of Romer (1990) in a context of a quality-ladders model of Schumpeterian growth. The use of a deterministic R&D process simplifies the analysis and differentiates the model from other quality-ladder growth models which view the discovery process as sequential and stochastic R&D races. We remove the scale-effects property by assuming that R&D becomes more difficult over time as in Jones' (1995) version of Romer's (1990) model. However, unlike Jones' model, we assume that the rate at which R&D difficulty increases over time depends only on the flow of patent creation. This assumption is consistent with the notion that patents not only exclude imitation, but also reduce the degree of knowledge spillovers that might be used in the discovery of other products.[4] This structural asymmetry across industries suffices to generate endogenous long-run Schumpeterian growth which depends on the patent length.

The model differs from the closed-economy model of Dinopoulos and Syropoulos (2006) in both the mechanism that generates endogenous growth and the questions analyzed. It complements a few recently developed growth models of North–South trade and growth by Dinopoulos and Segerstrom (2006) and Sener (2006) and the North–North models developed by Eaton and Kortum (1996, 1999), among others, by incorporating the role of finite-length patents into the analysis explicitly.

Consumer behavior

The global economy consists of two regions: the innovating North and the imitating South both of which engage in free trade. There is a fixed measure of identical dynastic households with infinitely lived members. The size of each household grows exponentially at the exogenous rate $g_L > 0$ as new household members are born continually. By normalizing the initial size of each household to unity, one can write the size of each household, measured by the number of its members, at time t as $e^{g_L t}$. Denote with $L_N(t) = \bar{L}_N e^{g_L t}$ the level of population in the North at time t, where $\bar{L}_N$ is a parameter capturing the number of Northern households at time zero. Similarly, denote with $L_S(t) = \bar{L}_S e^{g_L t}$ the corresponding level of Southern population at time t, where $\bar{L}_S$ is the initial Southern population. The world population level at time t is then given by $L(t) = \bar{L}e^{g_L t} = L_N(t) = L_S(t) = (\bar{L}_N + \bar{L}_S)e^{g_L t}$. Assuming that each household member is endowed with one unit of labor that is supplied inelastically to the market, $L(t)$ is also equal to the global supply of labor at time t.

There is a continuum of industries indexed by $\theta \in [0, 1]$ and producing final-consumption goods whose quality can be improved through innovative R&D. The knowledge-creation process and the international technology-transfer mechanism will be described later. Products of different quality levels in each industry θ are indexed by j, which is restricted to integer values and denotes the number of innovations in each industry. The quality level of a product in industry θ is given by $\lambda^{j(\theta,t)}$, where $\theta > 1$ is the quality increment generated by each innovation—which is identical across all industries—and $j(\theta, t)$ is the number of all innovations in industry θ at time t. Each identical dynastic household maximizes the following discounted lifetime utility:

$$U = \int_0^\infty e^{-(\rho - g_L)t} \ln u(t)\, dt \tag{5.1}$$

where $\rho > g_L$ is the subjective discount rate of a typical household member; and the instantaneous per capita utility function at time t is defined by

$$\ln u(t) = \int_0^1 \ln\left[\sum_j \lambda^j q(j, \theta, t)\right] d\theta \tag{5.2}$$

Equation (5.2) defines the standard quality-augmented Cobb-Douglas utility function across all industries $\theta \in [0,1]$. Variable $q(j, \theta, t)$ is the per capita quantity demanded of a product in industry θ at time t and quality level λ^j.

The consumer maximization problem is solved in three steps. First, the consumer allocates her budget across products within each industry by spending all her income on the product with the lowest quality-adjusted price $p(j, \theta, t)/\lambda^j$; we assume that, if two products within an industry have the same quality-adjusted prices, the consumer buys only the highest-quality product although she is formally indifferent. Second, the consumer allocates her budget on lowest quality-adjusted price products across all industries. The solution to this static maximization problem yields the per-capita demand function for the lowest quality-adjusted product in industry θ at time t

$$q(\theta, t) = \frac{c(t)}{p(\theta, t)} \tag{5.3}$$

where $c(t)$ is per capita consumption expenditure at time t. Third, the consumer maximizes (5.1) subject to the standard intertemporal wealth constraint

$$\dot{z}(t) = r(t)z(t) + w - c(t) - g_L z(t)$$

where $z(t)$ denotes the level of consumer assets, $r(t)$ is the market interest rate, and w is the consumer's income (wage). The solution to this dynamic maximization problem determines the optimal division between her consumption and savings. Using (5.3) in (5.2) and the resulting expression in (5.1)

generates an expression that depends on per capita consumption expenditure and product prices. The solution to this maximization problem yields the standard differential equation

$$\frac{\dot{c}(t)}{c(t)} = r(t) - \rho \tag{5.4}$$

Equation (5.4) implies that, at the steady-state equilibrium with constant per capita consumption expenditure, $c(t)$, the market interest rate equals the constant subjective discount rate (i.e. $r(t) = \rho$).

Product markets

By assumption, only Northern firms engage in innovative R&D and compete with each other in a Bertrand price-competition fashion. When a Northern firm discovers a new product in industry θ, it becomes the only firm in the world that knows how to produce the state-of-the-art quality product in this industry. Henceforth, we call this firm "Northern quality leader." We assume that a Northern quality leader obtains a perfectly enforceable patent of finite duration $T > 0$ that applies to its newly discovered product. When a patent expires, the product becomes "generic" only in the North—every firm in the North knows how to produce generic products. As a result, all generic products could be produced under competitive conditions in the North. Southern firms select generic products as targets for copying their technology and for transferring their production to the low-wage South. When a Southern firm copies a Northern generic product by expending resources to imitative R&D, it enjoys a manufacturing cost advantage that enables it to drive its Northern competitors out of the market. For notational purposes, we call these firms "Southern quality leaders." A Southern quality leader enjoys global monopoly power for a limited time because a Northern firm will eventually discover a new higher-quality product in industry θ that will replace its Southern competitor through limit pricing.

We assume that labor is the only factor production and that it is inelastically supplied. One unit of labor produces one unit of output independently of the geographic location of production and of the product quality level; therefore, marginal (and average) manufacturing costs are equal to the wage rate in each of the two regions. We focus on the case of product-cycle trade and international technology transfer in the presence of a North–South wage gap. Under such circumstances, the following condition must be satisfied at the steady-state equilibrium

$$w_S\lambda > w_N > \mathrm{w}_S \tag{5.5}$$

where w_S and w_N are the long-run equilibrium Southern and Northern wages respectively. Inequality (5.5) implies a positive North–South wage gap $w_N > w_S$, that is not very large relative to the size of each quality improvement $\lambda > 1$.

Inequality $w_S\lambda > w_N$ imposes an upper bound on the wage gap that allows a Northern firm with the state-of-the-art quality product to drive its Southern competitor out of the market.

A Northern quality leader producing the state-of-the-art quality product j in industry θ faces competition from a competitive Northern fringe producing generic product $j - 1$ if the product has not been copied by a Southern firm.[5] By charging price $p_N = \lambda w_N$, this firm can drive its Northern competitors out of the market and enjoy temporary global monopoly power for a period $T > 0$. We assume that, if product $j - 1$ has been successfully copied by a Southern firm, the Northern quality leader follows the following trigger pricing strategy: To drive the Southern quality leader out of the market it first charges a limit price $P'_N = \lambda w_S < \lambda w_N$, and then, when it learns that its Southern competitor is out of business, it charges a price $p_N = \lambda w_N$. Assuming that there are substantial costs associated with reentry in the South, this trigger strategy allows the Northern quality leader to charge a higher price except for an instance in time when the new product is discovered.[6]

The above reasoning implies that a Northern quality leader earns the flow of global profits $\pi_N = (p_N - w_N)(q_N L_N + q_S L_S)$, where $p_N = \lambda w_N$ is the price charged. Variables q_N and q_S are the per capita quantities demanded by Northern and Southern consumers described in (5.3). For notational purposes denote with $c = (c_N \bar{L}_N + c_S \bar{L}_S)/\bar{L}$ the per capita global consumption expenditure. Using (5.3), we can write the flow of profits π_N as

$$\pi_N = \frac{(\lambda - 1)}{\lambda} cL(t) \tag{5.6}$$

where $L(t) = \bar{L}e^{g_L t}$ is the level of world population at time t.

A Southern quality leader drives its Northern competitors out of the market by charging the limit price $p_S = w_N$. We assume that there are no reentry costs in generic products, since they are produced under conditions of pure competition in the North. The typical Southern quality leader earns a flow of global profits $\pi_S = (p_S - w_S)(q_N L_N + q_S L_S)$, where $p_S = w_N$. Using (5.3), we can write the flow of Southern profits as

$$\pi_S = \frac{(w_N - w_S)}{w_N} cL(t) = \left(1 - \frac{1}{\omega}\right) cL(t) \tag{5.7}$$

where $\omega = w_N/w_S > 1$ is the relative wage of Northern workers, c is per capita global consumption expenditure, and $L(t) = \bar{L}e^{g_L t}$ is the level of world population at time t.

The temporary flow of global monopoly profits provides strong incentives for Northern firms to engage in innovative R&D. Following the standard practice of the literature on North–South growth models, we will focus on the balanced-growth equilibrium properties of the model. The process of innovation mimics Romer's (1990) pioneering work on endogenous technological progress with

some additional features that generate endogenous long-run Schumpeterian growth. A Northern firm i, that employs l_i units of labor to engage in innovative R&D for a time interval dt, produces with certainty $dA_i = \alpha l_i\, dt/e^{\varphi x(t)}$ number of new designs, where α is an innovative-R&D productivity parameter; and $e^{\phi x(t)}$ captures the R&D difficulty in the long-run equilibrium, where $\varphi > 0$ is a parameter. We may think of dA_i as the instantaneous flow of patents (new product designs) that is directly proportional to the amount of labor devoted to innovative R&D, and inversely proportional to the R&D difficulty parameter. The economy-wide rate of patents is given by $dA = \alpha L_A dt/e^{\varphi x(t)}$, where $d_A = \Sigma_i dA_i$ is the aggregate flow of new products and $L_A = \Sigma_i \ell_i$ is the aggregate labor devoted to innovative R&D.

We assume that the steady-state evolution of $x(t)$ is given by $\dot{x}(t) = [v_p(t)]^{\beta}\dot{A}(t)$, where $0 < v_p(t) < 1$ is the measure of industries with active patents at time t (that is, the measure of industries with Northern quality leaders); $\dot{A}(t) = dA(t)/dt$ is the steady-state instantaneous flow of new products per industry and the economy as a whole since the measure of industries is of unit length by assumption; as it will become clear later, $\beta \in (-1,\ 1)$ is a parameter that determines the long-run correlation between the patent length and Schumpeterian growth. This specification of $x(t)$, especially the term $[v(t)]^{\beta}$, captures two competing features of patent-based intellectual property protection. On the one hand, patents facilitate the innovation process by making public the knowledge embedded in each patent application relative to other industries. This feature is captured by values of parameter β that are strictly negative—an increase in the measure $v_p(t)$ of industries protected by patents reduces the rate of increase in the R&D difficulty and makes the discovery of new products relatively easier. On the other hand, patents reduce the flow of knowledge spillovers in a number of ways. For example, firms may build patent portfolios to exclude other firms from discovering similar products; or, alternatively, they may threaten to resort to patent-infringement litigation, which hinders the research efforts of incumbents, and so forth. This feature is captured by strictly positive values of parameter β. In this case, an increase in the measure of industries protected by patents $v_p(t)$ accelerates the rate of increase in the R&D difficulty and makes the discovery of new products more difficult. Finally, by setting $\beta = 0$ one obtains the case of structural symmetry across all industries, independently of whether they are protected by patents or not.

In the steady-state equilibrium, the measure of industries protected by patents $v_p(t)$ is bounded and must, therefore, be constant over time. In addition, the requirement of a bounded (not exploding) per capita long-run growth rate requires that the flow of patents, $\dot{A}(t)$, to be constant over time as well. Ignoring initial level conditions, the steady-state value of R&D difficulty is given by

$$x(t) = \left[\left(v_p\right)^{\beta}\dot{A}\right]t \tag{5.8}$$

and the long-run innovation rate can be written as

$$\dot{A}(t) = \frac{\alpha L_A(t)}{e^{\varphi x(t)}} \tag{5.9}$$

Equations (5.8) and (5.9) have several desirable properties. First, differentiating (5.9) with respect to labor devoted to R&D yields $\alpha e^{-\varphi x(t)}$, the expression for the productivity of R&D researchers. This expression implies that the productivity of R&D labor (that is, the flow of patents per researcher) decreases over time, and that the innovative R&D labor requirement increases over time in the steady-state equilibrium. Second, (5.8) and (5.9) imply that parameter β relates the steady-state flow of patents $\dot{A}$ (and long-run Schumpeterian growth) to the patent length T. To see this rewrite equation (5.9) as $\dot{A}(t) = \alpha[L_A(t)/L(t)]L(t)e^{-\varphi x(t)}$, where the term in square brackets is the share of world labor devoted to innovative R&D. This share will be constant in the steady-state equilibrium, and therefore the term $L(t)e^{-\varphi x(t)} = \overline{L}e^{[tg_L - \varphi x(t)]}$ will also be constant over time. This means that $g_L = \varphi x(t)/t$ in the steady-state equilibrium.

Substituting $x(t)$ from (5.8) in this expression yields the following steady-state condition

$$\dot{A}(t) = \frac{g_L}{(\nu_p)^{\beta}\varphi} \tag{5.10}$$

According to (5.10), the steady-state rate of innovation is proportional to the rate of population growt g_L; inversely proportional to parameter φ, which is related to the R&D difficulty; and depends negatively on the measure of industries with Northern quality leaders raised to the power β. The last term provides the endogenous link between patent coverage and the rate of innovation $\dot{A}$. If $\beta = 0$, the steady-state rate of innovation is exogenous. The possibility of asymmetric knowledge spillovers is captured by $\beta \neq 0$ which links the steady-state rate of innovation to the market-determined measure of industries protected by patents and generates endogenous scale-invariant Schumpeterian growth. The rest of the chapter investigates the nature of this relationship and identifies the general-equilibrium forces that generate it.

All firms maximize (expected) discounted profits, in the presence of free entry into each innovative activity in the North. We also assume that Northern firms engage in R&D to discover higher-quality products by targeting only industries producing generic products. This assumption simplifies the analysis by eliminating the risk of default from products protected by patents, and can be justified by the existence of patent infringement law suits by firms with active patents against challengers. Since the technology of generic products is assumed to be public knowledge in the North, the possibility of patent-infringement law suits against challengers does not arise.

Consider now the profit-maximization decision of Northern firm i when choosing the amount of innovative-R&D labor. Let $V_A(\theta, t)$ denote the market value of a patent at time t in industry θ. This market value is given by

$$V_A(\theta, t) = \int_0^T \pi_N(t+s)e^{-r(t)s}\,ds \tag{5.11}$$

where $\pi_N(\theta, t)$ is the flow of global monopoly profits of a Northern quality leader described by (5.6). In the steady-state equilibrium, the market interest rate equals

the constant subjective discount rate; that is, $r(t) = \rho$. Substituting (5.6) into (5.11) and integrating appropriately yields the steady-state market value of a typical patent

$$V_A(t) = \left[\frac{1 - e^{-(\rho - g_L)T}}{\rho - g_L}\right]\frac{(\lambda - 1)}{\lambda}cL(t) \tag{5.12}$$

The structural symmetry across all industries implies that the value of a patent is identical across all industries, which explains why θ does not appear in V_A. At the steady-state equilibrium, the value of a typical patent grows at the rate of population growth, and is increasing in T. In the absence of patent protection ($T = 0$), the value of a patent vanishes; when the length of the patent becomes infinitely large ($T \to \infty$) the value of an invention equals the flow of steady-state global profits discounted by the effective discount rate, $\rho - g_L > 0$. In addition to the patent length, the value of a patent depends positively on the patent breath, measured by the quality increment λ and the rate of population growth g_L, and negatively on the subjective discount rate ρ. Equation (5.12) captures a standard property of patent protection: an increase in the patent length T increases the incentive to innovate by raising the market value of an invention.

A Northern firm i that devotes l_i units of labor to innovative R&D for a time interval dt, discovers with certainty $dA_i = \alpha \ell_i dt / e^{\varphi x(t)}$ new patentable product designs. This action generates a market value $V_A dA_i = [V_A \alpha \ell_i dt]/e^{\varphi x(t)}$, and total costs $(1 - \tau_A) w_N \ell_i dt$, where $\tau_A > 0$ is an *ad valorem* subsidy to R&D on innovation. Combining the benefits and costs of innovative R&D yields the following expression for the corresponding discounted net profits:

$$\left[V_A \alpha e^{-\varphi x(t)} - (1 - \tau_A) w_N\right] \ell_i dt$$

Free entry into innovative R&D results in nonpositive economic profits. The term inside the square brackets must be nonpositive, that is,

$$V_A \alpha e^{-\varphi x(t)} \leq (1 - \tau_A) w_N \tag{5.13}$$

This condition holds with equality for $\ell_i > 0$ (and $dA_i > 0$), and with inequality for $\ell_i = 0$ (and $dA_i > 0$). Romer (1990) derives (5.13) from the labor market equilibrium. The left-hand side of (5.13) is the value of the marginal product of labor of an R&D worker (V_A is the value of R&D "output" and $\alpha e^{-\varphi x(t)}$ coincides with the marginal product of labor in research); the right-hand side of (5.13) is the subsidy-adjusted wage rate (cost per worker). Consequently, the above equation states that each research firm hires workers up to the point where the value of the marginal product of labor is equal to the subsidized wage of labor.[7]

Substituting $x(t)$ from (5.8) and V_A from (5.12) into (5.13), and using the steady-state condition $L(t)e^{-\varphi x(t)} = \bar{L}e^{\left[t g_L - \varphi x(t)\right]} = \bar{L}$ generates the following

innovative-R&D condition:

$$\alpha \bar{L}\left[\frac{1-e^{-(\rho-g_L)T}}{\rho-g_L}\right]\frac{(\lambda-1)}{\lambda}c \leq (1-\tau_A)w_N \tag{5.14}$$

This condition defines a positive linear relationship between per capita consumption expenditure and the Northern wage rate. As per capita consumption increases, the innovation price (the reward to innovation) rises and so must the wage of Northern workers in order to restore the zero discounted profits condition. The market value of a patent plays a role very similar to the role terms of trade play in the static Ricardian trade model. Factors that raise the value of patents and innovation—such as the size of the global market $\bar{L}$, parameter α, the length of patents T, the rate of population growth g_L, and the size of innovations λ—raise the value of Northern wage income w_N relative to per capita global consumption c; thus, patent protection benefits Northern workers.

Imitation

The process of international technology transfer is endogenous and depends on the amount of Southern resources devoted to imitative R&D. At the steady-state equilibrium, a fraction of Northern industries manufactures state-of-the-art quality products whose patent protection has expired under conditions of perfect competition. Southern firms employ imitative-R&D labor to copy generic products and transfer their production to the low-wage South.

We model the process of endogenous imitation along the lines considered in Grossman and Helpman (1991, Chapter 11) and Dinopoulos and Segerstrom (1999). In particular, we assume that a Southern firm j that hires ℓ_j units of Southern labor to engage in imitative R&D for the time interval dt succeeds in copying $dM_j = \mu \ell_j dt/L(t)$ generic products manufactured in the North, where μ is an imitation productivity parameter. The level of global population $L(t) = \bar{L}e^{g_L t}$ appears in the denominator of the imitation production function in order to capture the notion that copying products becomes more difficult over time as the size of the global economy, measured by the level of population, increases.[8] Summing up over all firms engaged in imitative R&D yields the following aggregate rate of imitation

$$\dot{M}(t) = \frac{\mu L_M(t)}{L(t)} \tag{5.15}$$

where $\dot{M}(t) = dM/dt$ is the economy-wide rate of imitation, $dM = \Sigma_j dM_j$, and $L_M(t) = \Sigma_j \ell_j$. Equation (5.15) states that the aggregate rate of imitation is a function of the share of global resources devoted to imitative R&D.

The next step is to analyze the R&D choice of Southern imitators. Denote with $V_M(\theta, t)$ the expected discounted profits of a successful imitator j that manages to copy the state-of-the-art quality product in industry θ at time t. By employing ℓ_j

units of Southern labor in imitative R&D for the time interval dt, a firm can copy $dM_j = \mu \ell_j \, dt/L(t)$ Northern generic products and thereby create a market value of $V_M \mu \ell_j dt/L(t)$. The cost of this investment equals the subsidy-adjusted wage bill $(1 - \tau_M) w_S \ell_j dt$, where τ_M is an *ad valorem* subsidy to imitative R&D. As in the case of innovative R&D, we assume that there is free entry into the imitative R&D process that generates nonpositive net benefits $V_M[\mu \ell_j \, dt/L(t)] - (1 - \tau_M) w_S \ell_j \, dt \leq 0$ from imitative R&D, thus yielding the imitative R&D condition

$$V_M(\theta, t) \frac{\mu}{L(t)} \leq (1 - \tau_M) w_S \tag{5.16}$$

Equation (5.16) states that the value of the marginal product of labor devoted to imitative R&D cannot exceed the subsidy-adjusted wage in the South.

We assume that there is a global stock market capable of channeling global savings to Northern and Southern firms engaging in R&D. The stock market valuation of Southern monopoly profits yields another market-clearing condition which relates the flow of Southern monopoly profits π_S to the "price" of imitative R&D, V_M. The stock market valuation of Northern monopoly profits, V_A, is given by (5.12). Following the standard practice of Schumpeterian growth models, we derive the stock market valuation of Southern monopoly profits by setting up the no-arbitrage stock-market equilibrium condition.

Let ν_N be the measure (and set) of Northern industries producing generic products under perfect competition, and denote with ν_S the measure (set) of industries with Southern quality leaders. In other words, $1 = \nu_P + \nu_N + \nu_S$, where ν_P is the measure of industries with active (as opposed to inactive) patents. At each instant in time, a Southern quality leader earns the flow of profits $\pi_S dt$ during an interval of length dt. This firm faces a "creative destruction" risk of default that the next higher-quality product will be discovered by a Northern firm engaged in innovative R&D. We calculate this risk as follows: In an interval of length dt, $\dot{A}(t)dt$ new higher quality products are discovered in the North. With random selection (due to the structural symmetry across all industries), a Southern quality leader loses all its profits with probability $\dot{A}(t)dt/(\nu_S + \nu_N)$; this probability equals the instantaneous flow of innovations per industry without patent protection. If this event occurs, the Southern quality leader loses its monopoly power and suffers a loss equal to $(0 - V_M)$. If this event does not occur, the firm experiences a capital gain (loss) equal to $dV_M = \dot{V}_M \, dt$; therefore, the no-arbitrage condition, expressed in terms of rates of return, is

$$\frac{\pi_S dt}{V_M} + \left[1 - \frac{\dot{A} dt}{(\nu_S + \nu_N)}\right] \frac{\dot{V}_M}{V_M} \, dt + \frac{\dot{A} dt}{(\nu_S + \nu_N)} \frac{(0 - V_M)}{V_M} = r(t) \, dt \tag{5.17}$$

The left-hand side of equation 5.17 is the expected rate of return to a dollar invested in a stock issued by a Southern quality leader. The right-hand side is the return to a dollar invested in a completely diversified stock portfolio that yields

the riskless rate of return.[9] Taking limits as dt approaches zero and solving for the market value of a Southern quality leader generates

$$V_M = \frac{\pi_S}{r + \dfrac{\dot{A}}{v_S + v_N} - \dfrac{\dot{V}_M}{V_M}} \tag{5.18}$$

the standard stock-market valuation of temporary monopoly profits. Equation (5.18) discounts the flow of profits by the market interest rate plus the risk of default, which is related to the rate of innovation, minus the growth rate of the firm value. In the steady-state equilibrium, (5.18) takes the form

$$V_M = \frac{[1 - (1/\omega)]cL(t)}{\rho + \dfrac{\dot{A}}{(\nu_S + \nu_N)} - g_L} \tag{5.19}$$

where the steady-state expressions $r(t) = \rho$, $\dot{V}_M / V_M = g_L$, and $\pi_S = [1-(1/\omega)]cL\,(t)$ have been substituted into (5.19).[10] Substituting (5.19) into the zero-profit condition (5.16) gives the imitative-R&D condition

$$\mu[1 - (1/\omega)]c = (1 - \tau_M)w_S\left[\rho + \frac{\dot{A}}{(\nu_S + \nu_N)} - g_L\right] \tag{5.20}$$

Labor markets

We assume that there is perfect labor mobility across all industries and activities within each trading region. The wage rate is flexible and adjusts instantaneously to ensure that full-employment prevails and that the demand for labor equals its supply in each instant of time. These assumptions generate two full-employment labor conditions: one for the North and another for the South.

Consider the Northern labor market first. There are three components of the demand for Northern workers at the steady-state equilibrium: labor employed by firms engaged in innovative R&D, labor employed by Northern quality leaders to manufacture their products, and labor devoted to manufacturing of generic products produced in the North. Equations (5.8) and (5.9) imply that the demand for innovative-R&D labor is $L_A(t) = \dot{A}\, e^{t(\varphi \dot{A}\, \nu_P^{\beta})}/\alpha$. Consider now the demand for manufacturing labor. Each Northern quality leader manufactures $cL(t)/\lambda w_N$ units of output and there are ν_P industries with Northern quality leaders. Since, by assumption, each worker produces one unit of output, the aggregate demand for manufacturing labor employed by Northern quality leaders is $\nu_P cL(t)/\lambda w_N$. Observing that the competitive price in each Northern industry producing generic products is $p_N = w_N$ (instead of $p_N = \lambda w_N$), and that there are ν_N Northern industries producing generic products, the aggregate demand for labor in this

sector has to be $v_N cL(t)/w_N$. Setting the aggregate demand for labor in the North equal to its supply generates the Northern full-employment of labor condition

$$L_N(t) = \frac{\dot{A}\, e^{t(\varphi \dot{A} \nu^{\beta}_P)}}{\alpha} + \nu_P \frac{cL(t)}{\lambda w_N} + \nu_N \frac{cL(t)}{w_N} \tag{5.21}$$

Dividing the above equation by the level of population and using (5.10), yields the per capita version of *Northern full-employment condition*

$$\frac{\overline{L}_N}{\overline{L}} = \frac{g_L}{(\nu_P)^{\beta} \varphi \overline{L}} + \left[\frac{\nu_P}{\lambda} + \nu_N\right] \frac{c}{w_N} \tag{5.22}$$

Similar considerations apply to the Southern labor market, where labor is employed either by firms engaged in imitative R&D or by Southern quality leaders manufacturing state-of-the-art quality products. The demand for Southern labor in imitative R&D is given by equation (5.15) and equals $L_M(t) = \dot{M}L(t)/\mu$ Each of the v_S Southern quality leaders charges a price $p_S = w_N$ and produces $q_S = cL\,(t)/p_S$ units of output to serve the global market; therefore, the economy-wide demand for manufacturing labor is $v_S cL(t)/w_N$. Setting the aggregate demand for Southern labor equal to its supply, $L_S(t)$, yields the Southern full employment of labor condition

$$L_S(t) = \frac{\dot{M}}{L(t)\mu} + \nu_S \frac{cL(t)}{w_N} \tag{5.23}$$

Dividing both sides by the level of global population yields the per capita version of the *Southern full-employment condition*

$$\frac{\overline{L}_s}{\overline{L}} = \frac{\dot{M}}{\mu} + \nu_s \frac{c}{w_N} \tag{5.24}$$

The derivation of the Southern full-employment condition completes the description of the model.

The steady-state equilibrium

In this section, we establish the existence of the unique steady-state equilibrium and describe its properties, under the assumption that T is strictly positive. We focus on a balanced-growth equilibrium in which each variable grows at a constant rate over time. Several variables are constant in the steady-state equilibrium, including the market interest rate $r(t) = \rho$, per capita consumption expenditure c, all product prices, wage rates w_S and w_N, the rate of innovation (which equals the steady-state flow of patents) $\dot{A}$, the rate of imitation $\dot{M}$, and the measures of industries with Northern quality leaders v_P, with Southern quality leaders v_S, and with Northern firms producing generic products v_N. Although

per capita variables are constant over time, several variables (quantities produced, resources allocated to various activities, the flow of Southern and Northern profits, and the market value of quality leaders) grow at the constant rate of population growth. Lastly, the steady-state utility of each consumer grows at the same growth rate as total factor productivity generated by the process of creative destruction as new higher-quality products are discovered and produced in the North or South.

Following the standard methodology of Schumpeterian growth models, we let Southern labor serve as the *numeraire* (i.e. $w_S = 1$) so that $\omega = w_N > 1$ captures the relative wage of Northern workers as well as representing a measure of the North–South wage gap. In the steady-state equilibrium, the measure of industries with Northern quality leaders is related to the strictly positive patent length as follows:[11]

$$\nu_P = \int_0^T \dot{A}\, ds = \dot{A}T \tag{5.25}$$

Equation (5.25) states that, since patent protection is finite and the rate of patents is constant over time, the measure of industries with active patent protection is equal to the rate of innovation times the patent length $T > 0$. Equations (5.10) and (5.25) yield the steady state solution for the measure of Northern quality leaders

$$\nu_P = \left[\frac{Tg_L}{\phi}\right]^{\frac{1}{1+\beta}} \tag{5.26}$$

where the parameter restriction $\varphi > Tg_L$ ensures the measure of industries with patent protection is less than unity. According to (5.26), the fraction of industries with Northern quality leaders increases in the rate of population growth and the patent length for all values of parameter $\beta \in [-1, 1]$. Thus, the model generates the intuitive prediction that, *ceteris-paribus*, longer global patent protection (e.g. an increase in the patent length from 18 to 20 years) raises the number of Northern monopolies and results in a transfer of profits from Southern consumers to Northern firms.

Substituting (5.26) into (5.10) generates the steady-state solution for the rate of global innovation which equals the flow of patents:

$$\dot{A} = \left[\frac{g_L}{\varphi}\right]^{\frac{1}{(1+\beta)}} T^{\frac{-\beta}{(1+\beta)}} \tag{5.27}$$

The long-run rate of global innovation is increasing in the rate of population growth and decreasing in parameter φ. An increase in φ raises the level of R&D difficulty and reduces the rate of innovation. This result is standard in models of exogenous long-run Schumpeterian growth without scale effects.[12] What is novel about the present model is the role of parameter β in the long-run rate of innovation (and growth). If this parameter takes the value of

zero—which implies that there is no asymmetry across industries in the generation of knowledge spillovers—the rate of innovation is independent of the length of patent protection $T > 0$. Negative values of parameter β capture the "benevolent" case in which patents increase the rate of knowledge spillovers (decrease the rate of innovative R&D difficulty, $x(t)$, at each instant in time t—according to equation (5.8)). In this case, an increase in global patent protection raises the steady-state rate of innovation. However, strictly positive values of parameter β unveil the "malign" case of patent protection whereby stronger patent protection reduces the rate of global innovation, in addition to transferring income in the form of higher profits from Southern consumers to Northern firms.

The mathematical structure of the model is simple enough to render feasible the calculation of explicit steady-state solutions for all variables of interest. Denoting with a hat (^) the long-run equilibrium values of endogenous variables, we can combine the innovative and imitative-R&D conditions (5.14) and (5.20) to solve for the steady-state values of global per capita consumption expenditure and the North–South wage gap

$$\hat{c} = \frac{(1-\tau_A)(\rho - g_L)}{\alpha \bar{L}(1 - e^{-(\rho - g_L)T})} \frac{\lambda}{(\lambda - 1)} + \frac{(1-\tau_M)(\rho - g_L + \psi)}{\mu} \tag{5.28}$$

$$\hat{\omega} = 1 + \left[\frac{\alpha}{\mu} \frac{(1-\tau_M)(\lambda - 1)}{(1-\tau_A)\quad \lambda}\right]\left[1 + \frac{\psi}{(\rho - g_L)}\right]\bar{L}(1 - e^{-(\rho - g_L)T}) \tag{5.29}$$

where ψ is the risk of default for a Southern quality leader associated with the process of creative destruction

$$\psi = \frac{\dot{A}}{1 - v_P} = \frac{1}{T}\left\{\left(\frac{\varphi}{Tg_L}\right)^{\frac{1}{(1+\beta)}} - 1\right\}^{-1} \tag{5.30}$$

Equations (5.26) and (5.27) were used in the derivation of (5.30). The long-run North–South wage gap depends on virtually all parameters of the model. The wage of Northern workers exceeds that of Southern workers ($\hat{\omega} > 1$). Furthermore, the wage gap is increasing in factors that enhance the process of innovation—such as the productivity of labor in innovative R&D α, the subsidy to innovative R&D τ_A, and the magnitude of innovations λ. In contrast, parameters that encourage the transfer of technology from North to South—such as the labor productivity in imitative R&D μ, and subsidy to imitative R&D τ_M —have a positive impact on global wage inequality. The effect of patent protection on the North–South wage gap is ambiguous and depends on the relationship between the patent length and the steady-state rate of innovation described in (5.30).

Equations (5.28)–(5.30) will be used in the next section to analyze the consequences of globalization and the strength of intellectual property protection. We proceed by solving the model for the steady-state value of the rate of imitation.

Adding the Northern and Southern full-employment conditions (5.22) and (5.24) yields the following per capita full employment condition:

$$1 = \frac{g_L}{(\nu_P)^{\beta}\varphi\bar{L}} + \left[\frac{\nu_P}{\lambda} + (1-\nu_P)\right]\frac{c}{\omega} + \frac{\dot{M}}{\mu}$$

Substituting the endogenous variables c and ν_P in the above equation, and using the innovative-R&D condition (5.14) and equation (5.26), we may solve for the steady-state global imitation rate $d\hat{M}/dt$ to obtain

$$\frac{d\hat{M}}{dt} = \mu\left\{1 - \frac{(g_L/\varphi)^{\frac{1}{(1+\beta)}}}{\bar{L}(T)^{\frac{\beta}{(1+\beta)}}} - \left[1 - \frac{(\lambda-1)}{\lambda}\left(\frac{Tg_L}{\varphi}\right)^{\frac{1}{(1+\beta)}}\right]\frac{\lambda}{(\lambda-1)}\frac{(1-\tau_A)(\rho-g_L)}{\alpha\bar{L}(1-e^{-(\rho-g_L)T})}\right\} \tag{5.31}$$

where the term in square brackets is equal to $\nu_S+\nu_N+\nu_P/\lambda$, which is positive and less than unity. Substituting (5.31) and the innovative R&D condition (5.14) into the Southern full-employment condition (5.24), yields the steady-state value of the measure of industries with Southern quality leaders $\hat{\nu}_S$.

Globalization, intellectual property protection, and R&D subsidies

The model generates a steady-state equilibrium with several desirable properties: New higher-quality products are discovered in the North as a result of innovative R&D investments undertaken by Northern firms; each newly discovered product is produced in the North and enjoys global patent protection for a finite period of time, T; products whose patents have expired (generics) are produced in the North under conditions of perfect competition; the know-how of some generic products is transferred to the South (as a consequence of imitative R&D there) and these product are exported back to the North by Southern quality leaders until they are replaced by higher-quality products discovered in the North; other generic products are produced in the North until they are replaced by other newly discovered ones. This global process of Schumpeterian creative destruction generates product-cycle trade, long-run scale-invariant Schumpeterian growth, and a North–South wage gap.

Before proceeding with the comparative steady-state analysis, it is useful to derive an expression for the long-run growth rate of each consumer's utility function; in other words, we would like to derive an expression for long-run Schumpeterian growth. Since the measure of industries is normalized to unity and all industries are structurally identical, $A(t)$ denotes the economy-wide number of innovations at time t, as well as the "average" number of innovations per industry.

At each instant in time, there are v_P industries with Northern quality leaders. The average number of innovations in each of these industries is $j(\theta,t) = A(t)$ and the quantity produced by each Northern quality leader equals $q(\theta,t) = c/\lambda\omega$. Similarly, each of the remaining industries v_N+v_S is characterized by an average number of innovations $j(\theta,t) = A(t)$—the quantity produced equals $q(\theta,t) = c/\omega$ because every Southern quality leader and every competitive firm in the North producing generic products charges a price equal to the Northern wage $w_N = \omega$. Thus, the instantaneous utility of a typical household member at time t is

$$\ln u(t) = \int_{\nu_P} \ln\left[\lambda^{A(t)}\frac{c}{\lambda\omega}\right]d\theta + \int_{\nu_N+\nu_S} \ln\left[\lambda^{A(t)}\frac{c}{\lambda}\right]d\theta$$

Performing the integration yields the level of the instantaneous utility at time t

$$\ln u(t) = A(t)\ln\lambda + \nu_P \ln\left[\frac{c}{\lambda\omega}\right] + (\nu_N + \nu_S)\ln\left[\frac{c}{\omega}\right]$$

In the steady-state equilibrium, all variables of the right-hand side of the above expression are constant over time, except for the number of innovations $A(t) = \dot{A}t$. Differentiating the level of instantaneous utility with respect to time and using (5.27) yields

$$g_U = \frac{\dot{u}}{u} = \dot{A}\ln\lambda = \left(\frac{g_L}{\varphi}\right)^{\frac{1}{(1+\beta)}} T^{\frac{-\beta}{(1+\beta)}}\ln\lambda \tag{5.32}$$

The model generates endogenous long-run Schumpeterian growth without scale effects. As in a recent class of Schumpeterian growth models, growth is proportional to the rate of innovation, $\dot{A}$, which is equal to the steady-state flow of patents.[13] However, in the present model the rate of innovation depends on the patent length. The nature of this dependence is governed by parameter $\beta\in(-1, 1)$ which captures the structure of knowledge spillovers. We may summarize the properties of long-run Schumpeterian growth as follows:

Proposition 1 *Under the assumption that $T > 0$, the model's unique steady-state equilibrium is characterized by endogenous scale-invariant global Schumpeterian growth g_U with the following properties:*

- *a Long-run Schumpeterian growth is increasing in the rate of population growth g_L and the size of innovations λ, but decreasing in the difficulty of innovative-R&D φ.*
- *b The relationship between long-run Schumpeterian growth and patent length T depends on the structure of knowledge spillovers measured by parameter β: in the case of symmetric knowledge spillovers (i.e. $\beta = 0$), long-run growth is exogenous; if patents suppress the dissemination of useful knowledge*

(i.e. $\beta \in (0, 1)$), then an increase in patent protection (higher T) decreases long-run Schumpeterian growth; in contrast, if patents enhance the dissemination of useful knowledge (i.e. $\beta \in (-1, 0)$), an increase in patent protection increases long-run Schumpeterian growth.

Proof: See equation (5.32). QED.

We are now in a position to analyze the long-run affects of globalization and intellectual property protection in a post-TRIPs Schumpeterian North-South economy. As is well known, the process of globalization is multidimensional and its components include trade liberalization policies, the formation of multinational enterprises, international migration, so on. Following Dinopoulos and Segerstrom (2006), we envision a three-region global economy that consists of an open North, an open South, and a closed South, to address one dimension of globalization: An increase in the size of the South. Initially, free trade prevails between open South and open North, whereas the closed South follows an autarkic policy. Examples of closed South countries would include pre-1978 China, pre-1991 communist countries in Europe and Cuba, among others. In other words, one can think of the geographic dimension of globalization as a process by which countries in the closed South join the open South region by adopting free-trade policies with both open North and South. The long-run effects of geographic expansion of the market economy can be analyzed by considering the effects of permanent increase in the size of Southern population, $\overline{L}_S$, which implies an increase in the size of global population $\overline{L} = \overline{L}_S + \overline{L}_N$.

The following proposition summarizes the effects of Southern market-size expansion:

Proposition 2 *Globalization, viewed as a permanent expansion in the size of the South ($\overline{L}_S \uparrow$)*

- *a worsens the long-run wage-income distribution between North and South by raising the relative wage of Northern workers ($\omega \uparrow$);*
- *b permanently increases the rate of technology transfer from North to South ($M \uparrow$),*
- *c does not affect the long-run rates of innovation ($\dot{A} \leftrightarrow$) and Schumpeterian growth ($g_U \leftrightarrow$).*

Proof: The proof can be established with the help of equations (5.29), (5.31), and (5.32). Note that the term inside the square brackets of equation (5.31) equals $v_S + v_N + v_P/\lambda$ and it is positive. QED.

The long-run effects of patent protection can be analyzed with the help of equations (5.29), (5.31), and (5.32). Proposition 1(b) established the long-run growth effects of strengthening global protection of intellectual property by increasing the patent length $T > 0$.

The following proposition identifies the long-run effects of longer enforceable patents on wage income inequality and the rate of imitation:

Proposition 3 *A permanent increase in the global patent protection generated by an increase in the patent length ($T\uparrow$)*

a raises permanently the rate of technology transfer from North to South ($\dot{M}\uparrow$);
b permanently exacerbates the wage income inequality between North and South ($\omega\uparrow$) if the following sufficient, but hardly necessary, condition holds:

$$\hat{\nu}_P = \left(Tg_L/\varphi\right)^{\frac{1}{(1+\beta)}} \geq \beta/(1+\beta) \tag{5.33}$$

Proof: The proof follows from equations (5.31) and (5.29). Condition (5.33) is derived by assuming that $d\psi/dT \geq 0$, where the risk of default ψ is defined by equation (5.30). QED.

The economic intuition of part (a) is as follows: The innovative-R&D condition (5.14) implies that an increase in the patent length raises the value of innovation and requires a decrease in the ratio c/ω—the cost of innovation per dollar of consumption expenditure must increase— to ensure that the zero-profit condition is satisfied. An increase in T causes a permanent increase in the measure of industries with patent protection ν_P (see (5.26)). These two implications mean that longer patent protection shifts resources away from manufacturing and innovative R&D towards imitative R&D. This resource relocation effect is apparent from the per capita global resource condition

$$1 = \frac{g_L}{(\nu_P)^{\beta}\varphi\bar{L}} + \left[1 - \nu_P\frac{(\lambda-1)}{\lambda}\right]\frac{c}{\omega} + \frac{\dot{M}}{\mu}$$

where a decline in c/ω and an increase in ν_P imply a permanent increase in the rate of global imitation $\dot{M}$. This result highlights the general-equilibrium forces that could intensify North's demand for stronger intellectual property protection in the form of a longer patent length. Longer patents generate higher rates of imitation and shorter product cycles; in addition, they raise the rate of labor turnover in the global economy. The demand for stronger intellectual property protection might be even stronger if longer patents result in a slower rate of global innovations (see Proposition 1).

The effects of patent protection on the North–South wage gap, which are summarized in part (b) of Proposition 2, depend on whether longer patents increase or reduce the risk of default ψ for a Southern quality leader. From (5.30) it can be seen that the sign of $d\psi/dT$ is ambiguous. (This is so because an increase in T increases the measure of industries with patents and has an ambiguous effect on the rate of innovation.) The sufficient condition (5.33) guarantees that higher patent protection increases the risk of default ($d\psi/dT > 0$) for each Southern

quality leader. Condition (5.33) is satisfied if the measure of industries with patent protection is sufficiently high ($v_P \geq \beta/(1+\beta)$) relative to parameter $\beta \in (-1, 1)$. For instance, if longer patents increase the rate of innovation and growth (i.e. $\beta \in (-1, 0)$), or if the measure of industries with patent protection is more than 1/2 , then an increase in T raises the risk of default ψ. In all these cases, longer patent protection discourages imitation by increasing the risk of default for a successful Southern imitator and encourages innovation by increasing the duration of monopoly power enjoyed by Northern quality leaders. Equation (5.29) states that these two effects generate an unambiguous permanent increase in the North–South wage inequality.

We conclude this section by describing the effects of innovative and imitative R&D subsidies τ_A and τ_M, respectively. These effects are summarized in the following proposition and can be derived by inspecting equations (5.29) and (5.31):

Proposition 4 *A permanent increase in the innovative R&D subsidy* ($\tau A \uparrow$) *exacerbates wage-income inequality* ($\hat{\omega} \uparrow$) *and raises the rate of imitation* ($\dot{M} \uparrow$). *In contrast, a permanent increase in the imitative R&D subsidy* ($\tau M \uparrow$) *reduces the degree of North–South wage-income inequality* ($\hat{\omega} \downarrow$) *and does not affect the rate of imitation* ($\dot{M} \leftrightarrow$).

Proof: See equations (5.29) and (5.31). QED.

The apparent asymmetry between the impact of innovative and imitative R&D subsidies on the rate of imitation can be explained by their differential impact on the ratio c/ω. The innovative R&D condition (5.14) pins down the value of c/ω, which depends on the innovative R&D subsidy τ_A, but not the imitative R&D subsidy τ_M. Since τ_M does not affect the long-run values of v_P and ω, and does not appear explicitly either in the innovative or in the global-resource condition, it does not affect the long-run rate of imitation in this model. It is obvious then from the imitative R&D condition (5.20) that a permanent increase in τ_M reduces both global consumption per capita c and the wage gap ω without affecting their ratio.

Concluding remarks

We developed a North–South model of scale-invariant endogenous Schumpeterian growth, product-cycle trade and global patent protection. The model combines the deterministic R&D-based knowledge-creation function, introduced by the seminal works of Romer (1990) and Jones (1995), with the demand and market structures of quality-ladders Schumpeterian growth models. This combination renders possible the analysis of finite patents in a relatively simple analytical framework that permits the derivation of closed-form solutions.

The unique steady-state equilibrium of the model exhibits several novel properties: new higher-quality products are discovered in the innovating North through R&D investments; these products are manufactured in the North and exported to the South for a finite period of time which coincides with the length of a perfectly enforceable global patent; products whose patents have expired

(generics) are produced in the North initially under conditions of perfect competition; generic products are targeted by Southern firms that engage in imitative R&D and manage to copy a fraction of these products; Southern firms producing generics supply the global market until they are replaced by Northern firms that discover new higher quality products. This equilibrium generates endogenous product-cycle trade, an endogenous North–South wage gap, and endogenous rates of innovation and imitation.

The model highlights the role of patented products in the rate of knowledge spillovers that can enhance or suppress the introduction of new products. If products with active patents suppress the rate of knowledge spillovers more than products with expired patents, then an increase in patent length decreases the long-run rate of innovation and growth. If products with active patents enhance the rate of knowledge spillovers more than products with expired patents, higher patent length increases the long-run rate of innovation and growth.

The model generates several novel predictions for the long-run effects of globalization and patent protection: geographic expansion of the global market, caused by an increase in the size of the South worsens the North–South income distribution, does not affect the rate of innovation and growth, and increases the rate of imitation; an increase in the patent length (stricter global protection of intellectual property) increases the rate of imitation for a sufficiently large fraction of industries covered by patents, longer patents generate more global inequality; and in the case that patents suppress the rate of knowledge spillovers, longer patents decrease the long-run rate of innovation and growth. These results justify the concerns of many economists that North's demands for longer patents may worsen the disparities in world income and may suppress long-run Schumpeterian growth. The model is also consistent with a less pessimistic scenario; that is, even though longer patents worsen the wage-income distribution, they also reduce global poverty by boosting long-run global total factor productivity growth.

As usual, these results depend on the model's assumptions: We abstracted from the inherent uncertainty in innovative R&D; we did not analyze the case of asymmetric patent protection across the two regions; we did not analyze the stability and the welfare properties of the model due to the analytical complexity associated with dynamic North–South models; we assumed that patenting is the only mechanism that protects new inventions; and we did not deal with other important aspects of globalization such as trade liberalization, international labor migration and multinationals. We believe that the simple structure of the model allows a variety of generalizations that could relax these restrictive assumptions and generate new insight on the prospects and problems of globalization.

Notes

* This paper was prepared for the conference "Globalization: Prospects and Problems" in honor of Jagdish Bhagwati's 70th birthday at the University of Florida, January 28–30, 2005. We would like to thank our discussant Jonathan Eaton and the other conference participants for useful comments and suggestions. The authors are solely responsible for any remaining errors and omissions.

1 The opposite is true as well. Baldwin (1988) reports that US firms lose about $8 billion annually from international patent and copyright infringements.
2 Maskus (2000), among others, details the factors affecting intellectual property protection and the TRIPS agreement. Bhagwati (2004) argues that the TRIPS agreement is a prime example of regulatory capture of WTO by Northern multinationals. He expresses his opposition to it as follows:

> Clearly the rules sought by the pharmaceutical industries are unnecessarily harmful to poor countries. In particular, (1) TRIPs should not be in the WTO at all, (2) twenty-year patents at the WTO are excessive, and (3) access to generic drugs produced in developing countries, such as India and Brazil, that have manufacturing capacity should be freed for the poor countries, such as Botswana, that do not have such capacities but have medical emergencies such as AIDS, as certified by the World Health Organization, for example.
>
> (pp. 183–185)

In the same spirit, Stiglitz (2003) also highlights the importance of intellectual property rights and knowledge spillovers by echoing similar concerns

> No one denies the importance of intellectual property rights. But these rights need to balance out the rights and interests of producers with those of users—not only users in developing countries but producers in developed countries . . . after all, knowledge is the most important input into research, and stronger intellectual property rights might actually increase the price of this input.
>
> (p. 245)

3 Grossman and Helpman (1991), Grossman and Lai (2004) and Segerstrom *et al.* (1990) develop endogenous growth models with scale effects. Jones (1999), and Dinopoulos and Sener (2007) provide more details of the issues related to the scale-effects property of earlier endogenous growth models.
4 The literature on the economics of patents provides many examples that support this assumption. For instance, companies develop patent fences around a basic invention which prevent other firms from discovering similar products; and the threat of litigation on possible patent infringements by patent holders prevents challengers from exploiting fully possible knowledge spillovers associated with accessibility of the information contained in patent applications. Levin *et al.* (1987) provide ample evidence based on a survey of US manufacturing firms and many examples supporting a variety of mechanisms including patents that are used by firms to protect their intellectual property.
5 In this chapter, we assume that the patent protection is global. If the patent protection is not global and Southern firms can also target Northern state-of-the-art quality products whose patents have not expired, then a typical producer of a generic product can at most charge a price equal to $p_S^L = \lambda w_S$, since it only faces competition from the quality follower in the South. Since the product enjoys patent protection in the North, it cannot be exported; therefore, the product can be sold only to Southern consumers generating a flow of monopoly profits equal to $\pi_S^L = (\lambda w_S - w_S) q_S L_S$. If the same firm targets a Northern generic product whose patent has already expired, the successful Southern imitator would be able to sell the product to both Northern and Southern consumers and earn a higher flow of profits $\pi_S = (w_N - w_S)(q_N L_N + q_S L_S)$, provided that the Northern market is at least as large as the Southern market (i.e. $c_N L_N \geq c_S L_S$). Since the cost of imitation is the same across all industries, free entry in imitative R&D activities targeting generic Northern products would drive expected discounted profits down to zero and would imply negative expected discounted net benefits from targeting Northern products with active patents. In other words, Southern imitators do not have an incentive to target Northern products with active patents as long as there generic products produced by Northern firms at the steady-state equilibrium.

6 Howitt (1999) has introduced this trigger strategy in the context of a Schumpeterian growth model and has provided more details about it. In the absence of reentry costs, the Northern quality leader can always drive its Southern competitor out of the market by charging the limit price $p'_N = \lambda w_S < \lambda w_N$. The assumption of (large) market reentry costs in the South simplifies the exposition and algebra.

7 If the equilibrium wage is such that the right-hand side of the zero-profit condition exceeds its left-hand side, then innovative R&D is not profitable and the innovative activity stops.

8 This specification of imitative R&D difficulty corresponds to the permanent effects of growth (PEG) specification of the knowledge production process in Dinopoulos and Segerstrom (1999) and its purpose is to remove the scale-effects property inherent in all Schumpeterian growth models. Other specifications are also possible. For example, it is possible to use the level of Southern population, $L_S(t)$, the level of quality of the state-of-the-art generic product targeted for imitation, or the number of products discovered up to time t. We choose the simplest specification to minimize the complexity of algebra.

9 See Grossman and Helpman (1991, Chapter 11) for more details on this issue in a context of a growth model based on variety accumulation.

10 These three expression have been derived from (5.4), (5.16), and (5.7), respectively.

11 In the absence of patent protection (i.e. $T = 0$), the model has a steady-state equilibrium without growth. If $T = 0$ the reward to innovation becomes zero, and thus $v_P = 0$. This stops innovation activity (i.e. $\dot{A} = 0$) and firms have no demand for Northern researchers (*i.e.* $L_A = 0$). This effect pushes the Northern wage down and Southern labor-cost advantage diminishes. In addition, the probability $\dot{A}/(\nu_N + \nu_S)$ of replacing a Southern quality leader drops to zero. Southern quality leaders, safe from being replaced by further innovation, target all Northern industries that produce generics. Eventually all Northern products are copied by Southern firms, the relative wage of Northern workers becomes equal to the Southern wage, and the global economy reaches a steady-state equilibrium without trade, innovation, imitation and growth. For the remaining of the analysis, we focus on the steady-state equilibrium with strictly positive patent duration.

12 See, for example, Jones (1995), Segerstrom (1998) and Dinopoulos and Segerstrom (2006), among many others.

13 See Segerstrom (1998) and Dinopoulos and Segerstrom (1999, 2006), among others, for models of scale-invariant Schumpeterian growth with this property.

Bibliography

Baldwin, R. (1988) "Other Issues: North-South Trade, Intellectual Property Rights, and Subsidies," in R. Baldwin and D. Richardson (eds) *Issues in the Uruguay Round*, Cambridge, MA: NBER Conference Report.

Bhagwati, J. (2004) *In Defense of Globalization*, Oxford University Press.

Dinopoulos, E. and P. Segerstrom (1999) "A Schumpeterian Model of Protection and Relative Wages," *American Economic Review*, 89: 450–472.

—— (2006) " North-South Trade and Economic Growth," mimeo, Stockholm School of Economics, available at http://bear.cba.ufl.edu/dinopoulos/Research.html

Dinopoulos, E. and F. Sener (2007) "New Directions in Schumpeterian Growth Theory," in H. Hanusch and A. Pyka (eds) *Edgar Companion to Neo-Schumpeterian Economics*, Cheltenham: Edward Elgar.

Dinopoulos, E. and C. Syropoulos (2006) "Rent Protection as a Barrier to Innovation and Growth," *Economic Theory*, Forthcoming available at http://bear.cba.ufl.edu/dinopoulos/Research.html

Eaton, J. and S. Kortum (1996) "Measuring Technology Diffusion and the International Sources of Growth," *Eastern Economic Journal*, 22(4): 401–410.

—— (1999) "International Technology Diffusion: Theory and Measurement," *International Economic Review*, 40(3): 537–570.

Grossman, G. and E. Helpman (1991) *Innovation and Growth in the Global Economy*, The MIT Press.

Grossman, G. and E. Lai (2004) "International Protection of Intellectual Property," *American Economic Review*, 94(5): 1635–1653.

Howitt, P. (1999) "Steady Endogenous Growth with Population and R&D Inputs Growing," *Journal of Political Economy*, 107: 715–730.

Iwaisako, T. and K. Futagami (2003) "Patent Policy in an Endogenous Growth Model," *Journal of Economics*, 78(3): 239–258.

Jones, C. (1995) "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 103: 759–784.

—— (1999) "Growth: With or Without Scale Effects?," *American Economic Review Papers and Proceedings*, 89: 139–144.

Kortum, S. (1997) "Research, Patenting and Technological Change," *Econometrica*, 65: 1389–1419.

Levin, C., A. Klevorick, R. Nelson, and S. Winter (1987) "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3: 783–820.

Maskus, K. (2000) *Intellectual Property Rights in the Global Economy*, Washington, DC: Institute of International Economics.

Romer, P. (1990) "Endogenous Technological Change," *Journal of Political Economy*, 98(5): S71–S102.

Schumpeter, J. (1942) *Capitalism, Socialism and Democracy*, Harper and Row.

Segerstrom, P. (1998) "Endogenous Growth Without Scale Effects," *American Economic Review*, 88: 1290–1310.

Segerstrom, P., T. C. A. Anant, and E. Dinopoulos (1990) "A Schumpeterian Model of the Product Life Cycle," *American Economic Review*, 80: 1077–1092.

Sener, F. (2006) "Intellectual Property Rights and Rent Protection in a North-South Product-Cycle Model," Schenectady, NY: Union College, mimeo.

Stiglitz, J. (2003) *Globalization and its Discontents*, W. W. Norton.

6 Choosing between innovation and imitation in a model of international rivalry

Kenji Fujiwara, Koji Shimomura, and Kar-yiu Wong

Introduction

This chapter constructs a model of international rivalry to examine the competition in terms of technology and output between firms in two trading countries. We analyze how the less efficient firm chooses endogenously between innovation and imitation, and how its decision may affect the market equilibrium and profits of the firms. Conditions under which the less efficient firm is able to catch up or to surpass the technology of the advanced firm have been derived. The implications of removing the endogeneity of this decision of the less efficient firms are also examined.

Globalization in recent years has greatly enhanced trade and technology diffusion among countries. This phenomenon has brought to government planners and economists many contentious issues.[1] Among these issues is the one about international transfer of technology and the appropriate responses of governments. With firms producing similar products in various trading countries, the improvement of technologies and possible international spillover of technologies would affect the competitiveness of these firms and the welfare of their countries. Governments and economists are thus eager to find the right policies to improve the welfare of their own countries.

There is a large literature on the output-technology competition among firms in different countries. In terms of how firms compete in technology, papers in this literature can be divided into two categories: those that assume that all firms are choosing innovation optimally, and those that examine the competition between some firms that innovate and some firms (usually in other countries) that imitate.

One of the earlier papers that examines how international rival firms choose innovation is Spencer and Brander (1983). They also examine the roles of R&D subsidies provided by one or both governments. Bagwell and Staiger (1992, 1994) investigate the robustness of such R&D subsidies in the presence of uncertainty, and argue that Spencer and Brander's (1983) argument for the strategic R&D subsidy is not confined to a deterministic environment.

The innovation–imitation relations among firms have drawn more attention in the literature. For example, Segerstrom (1991) constructs a model of economic growth in which some firms innovate to improve the quality of their product and some others imitate. Using a dynamic North–South model, Grossman and

Helpman (1991) conclude that the North's R&D activity is enhanced by the existence of the imitating South. In a similar framework, Helpman (1993) discusses the effects of enforcement of intellectual property rights.[2] Dinopoulos and Syropoulos (1998), taking account of the more efficient firm's imitation-blocking and the less efficient firm's appropriating activities, identify the conditions under which these 'redistributive' expenditures become positive. Dinopoulos *et al.* (2006) construct a Schumpeterian growth model of international trade and examine the effect of patent protection on the rate of innovation and imitation. Liao and Wong (2004), also assuming an innovation–imitation relation between developed countries and developing countries, examine to what extent the governments of developing countries would be willing to provide intellectual property rights protection and investigate the roles and the shortcomings of the TRIPS agreement.

These papers all start with the presumption that less efficient firms are given a particular channel through which they improve their technologies: either innovation or imitation. This seems to be a serious limitation in many cases because in more general models the less efficient firms should have the option of choosing the optimal way to improve their technologies. In other words, the choice between innovation and imitation should be given to these firms, instead of arbitrarily assigning a technology-improving path to them. To see how serious this limitation is, we just have to note that many firms in many emerging countries started by imitating the technologies of firms in developed countries, but over time switched to developing their own technologies. In such a process, they gradually caught up with the advanced technologies and even surpassed them. The switch from imitation to innovation is missing in the papers mentioned above.

One very early exception that endogenizes the decision of a firm between innovation and imitation is Shimomura and Wong' (2001) study. Using a dynamic model of international rivalry, they examine the conditions for innovation and imitation by a less efficient firm, and investigate the circumstances under which the firm will choose to innovate or imitate. They show some interesting cases in which the less efficient firm chooses imitation initially but switches to innovation in the future. There are also cases presented in the paper in which the less efficient firm always chooses to imitate and never finds it profitable to innovate.

The purpose of this chapter is to offer another framework to examine the choice between innovation and imitation.[3] Using a simple model of oligopolistic competition, we endogenize the less efficient firm's choice between innovation and imitation. We also investigate the conditions under which the less efficient firm chooses to catch up or surpass the advanced firm in terms of technology. This model and the analysis allow us to answer one question: If the less efficient firm will choose to innovate, what happens if the firm is assumed to imitate instead? This question is important because many papers just assume the innovation–imitation relations between two firms without checking whether the less efficient may in fact choose to innovate instead, if it is allowed to. If one assumes that the less efficient imitates but prefers to innovate, what kind of error may one commit?

The rest of this chapter describes the model with two firms, one in each country. The firms, with different initial technology levels, attempt to improve

their technology and use the new technology to produce a homogeneous product to be exported to the rest of the world. Then we examine the case in which both firms innovate before producing the product. Thus they compete directly in terms of technology and output. Next we consider the case in which the more efficient firm innovates while the other one imitates. After which we examine the conditions for catching up and surpassing in the innovation–innovation case and in the innovation–imitation case. Then we endogenize the innovation/imitation decision of the initially less efficient firm. Finally we conclude.

The model

Consider a model consisting of three countries labeled Home, Foreign, and the Rest of the World (ROW). Both Home and Foreign have a firm producing a homogeneous product, which is exported to ROW. There is no demand for the product in Home or Foreign, and the demand in ROW can be represented by a linear function

$$p = a - q \tag{6.1}$$

where p is the market price and q the demand. The value of a is positive and sufficiently large so that the market can support both the Home firm and the Foreign firm. Initially the Home firm is able to produce the product using a marginal cost, c_0, which is independent of the output level. The corresponding initial marginal cost of the Foreign firm is c_0^*.[4] We assume that initially the Home firm has a more advanced technology.

Assumption 1 $c_0 < c_0^*$

If we denote the output of the Home (Foreign) firm by x (x^*), which is the same as the export of the product by the country, in equilibrium we have

$$q = x + x^* \tag{6.2}$$

The two firms compete in terms of technology and output. In terms of output, they compete in a Cournot–Nash fashion. In terms of technology, both of them are able to improve their technology, or to reduce their marginal costs, through one of the two channels: innovation and imitation.

Innovation requires the intentional expenditure on resources, which results in the improvement of technology. More specifically, we assume that after spending an amount of $eg^2/2$, $e > 0$ ($e^* g^{*2}/2$, $e^* > 0$), the Home (Foreign) firm is able to reduce its marginal cost by an amount of g (g^*). Variable e (e^*) can be interpreted as the cost of innovation.[5]

Define two variables:

$$z = \frac{a - c_0^*}{a - 2c_0^* + c_0} \tag{6.3}$$

$$z^* = \frac{a - c_0}{a - 2c_0 + c_0^*} \tag{6.4}$$

since $c_0^* > c_0$, equations (6.3) and (6.4) imply that $z > 1$ and $1 > z^* > 0$. To guarantee that if a firm chooses to innovate, the level of innovation is positive, we make the following assumption.

Assumption 2 $e > z, e^* > z^*$

If both firms choose to innovate, their new marginal costs become, respectively,

$$c_1 = c_0 - g$$
$$c_1^* = c_0^* - g^*$$

Imitation exists when a firm chooses to learn from the other firm. For example, if the Foreign firm imitates, its new marginal cost becomes

$$c_1^* = c_0^* - \theta(c_0^* - c_0) \tag{6.5}$$

where $\theta \in (0, 1)$.

Since initially the Home firm has a lower marginal cost, it has nothing to learn from the Foreign firm. This means that for the Home firm, imitation is the same as inaction, spending zero resources on innovation and thus getting zero innovation. We can thus reduce the present analysis to the following two cases: (A) Both firms innovate; and (B) The Home firm innovates while the Foreign firm imitates. These two cases are analyzed in the following sections.

Case (A): foreign firm innovating

In this case, both firms innovate. Denote the profits of the Home and Foreign firms by Π^v and Π^{*v}, respectively, where superscript v denotes the variables in this case. We have

$$\Pi^v = \left[a - (c_0 - g) - x - x^*\right] x - \frac{e}{2} g^2 \tag{6.6}$$

$$\Pi^{*v} = \left[a - (c_0^* - g^*) - x - x^*\right] x^* - \frac{e^*}{2} g^{*2} \tag{6.7}$$

Both firms choose the new technology level (marginal cost) first and then compete in terms of output. Since the present game never involves the problem of time inconsistency, we need not solve the game in a backward way; that is, the solution obtained by solving the game backward is equivalent to the solution derived in the first stage. Thus, we assume that the Home firm maximizes (6.6) with respect to x and g, whereas the Foreign firm (6.7) with respect to x^* and g^*,

both in a Cournot–Nash fashion. The corresponding first-order conditions for profit maximization are[6]

$$x\colon a - c_0 + g - 2x - x^* = 0 \tag{6.8}$$

$$g\colon x - eg = 0 \tag{6.9}$$

$$x^*\colon a - c_0^* + g^* - x - 2x^* = 0 \tag{6.10}$$

$$g^*\colon x^* - e^* g^* = 0 \tag{6.11}$$

This system is solved as follows. From (6.9) and (6.11), $x = eg$ and $x^* = eg^*$ follow. Then, substituting these into (6.8) and (6.10), we get

$$a - c_0 - (2e - 1)g - e^* g^* = 0$$

$$a - c_0^* - eg - (2e^* - 1)g^* = 0$$

Solving this system for g and g^* yields the optimal innovation levels:

$$g^v = g^v(e, e^*) = \frac{(e^* - 1)(a - c_0) + e^*(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*} \tag{6.12}$$

$$g^{*v} = g^{*v}(e, e^*) = \frac{(e - 1)(a - c_0^*) - e(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*} \tag{6.13}$$

Given Assumptions 1 and 2, $g^v > 0$ and $g^{*v} > 0$. Combining (6.12), (6.13), (6.9), and (6.11), we get the optimal outputs as follows:

$$x^v = x^v(e, e^*) = e\left[\frac{(e^* - 1)(a - c_0) + e^*(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*}\right] \tag{6.14}$$

$$x^{*v} = x^{*v}(e, e^*) = e^*\left[\frac{(e - 1)(a - c_0^*) - e(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*}\right] \tag{6.15}$$

Making use of (6.12)–(6.15), the profits of the firms as given by (6.6) and (6.7) reduce to

$$\Pi^v(e, e^*) = \frac{e(2e - 1)}{2}\left[\frac{(e^* - 1)(a - c_0) + e^*(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*}\right]^2 \tag{6.16}$$

$$\Pi^{*v}(e, e^*) = \frac{e^*(2e^* - 1)}{2}\left[\frac{(e - 1)(a - c_0^*) + e(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*}\right]^2 \tag{6.17}$$

The above analysis shows that the outputs, innovation levels, and profits are functions of the costs of innovation, e and e^*.

Lemma 1 *In the present case with both firms innovating, we have*

$$\frac{\partial g^v}{\partial e}<0,\quad \frac{\partial g^v}{\partial e^*}>0,\quad \frac{\partial g^{*v}}{\partial e}>0,\quad \frac{\partial g^{*v}}{\partial e^*}<0$$

$$\frac{\partial x^v}{\partial e}<0,\quad \frac{\partial x^v}{\partial e^*}>0,\quad \frac{\partial x^{*v}}{\partial e}>0,\quad \frac{\partial x^{*v}}{\partial e^*}<0$$

The proofs of all lemmas, propositions, and corollaries are given in the appendix.

Lemma 2 $\Pi_e^v<0, \Pi_{e^*}^v>0, \Pi_e^{*v}>0$ *and* $\Pi_{e^*}^{*v}<0$.[7]

Lemmas 1 and 2 make it clear why e^* and e^* are interpreted as the cost of innovation: An increase in its value a firm faces discourages the firm's innovation, output, and profit but encourages those of the other firm.

Case (B): foreign firm imitating

Denote the profits of the Home and Foreign firms by Π^m and Π^{*m}, respectively. Since imitation involves zero cost, the profits of the firms are

$$\Pi^m = [a-(c_0-g)-x-x^*]\,x-\frac{e}{2}g^2 \tag{6.18}$$

$$\Pi^{*m} = \left[a-c_1^{*m}-x-x^*\right]x^* \tag{6.19}$$

where $c_1^{*m}=c_0^*-\theta(c_0^*-c_0)$. Maximizing (6.18) with respect x to and g and (6.19) with respect to x^* gives the following optimality conditions:

$$x\text{: } a-c_0+g-2x-x^*=0 \tag{6.20}$$

$$g\text{: } x-eg=0 \tag{6.21}$$

$$x^*\text{: } a-c_1^{*m}-x-2x^*=0 \tag{6.22}$$

Solving this system immediately yields

$$x^m = x^m(e,\theta)=\frac{e[a-c_0+(1-\theta)(c_0^*-c_0)]}{3e-2} \tag{6.23}$$

$$x^{*m} = x^{*m}(e,\theta)=\frac{(2e-1)[a-c_0^*+\theta(c_0^*-c_0)]-e(a-c_0)}{3e-2} \tag{6.24}$$

$$g^m = g^m(e, \theta) = \frac{a - c_0 + (1 - \theta)(c_0^* - c_0)}{3e - 2} \tag{6.25}$$

where superscript "m" denotes the variables in the present case. The resulting profits of the firms are

$$\Pi^m(e, \theta) = \frac{e(2e - 1)}{2}\left[\frac{a - c_0 + (1 - \theta)(c_0^* - c_0)}{3e - 2}\right]^2 \tag{6.26}$$

$$\Pi^{*m}(e, \theta) = \left\{\frac{(2e - 1)[a - c_0^* + \theta(c_0^* - c_0)] - e(a - c_0)}{3e - 2}\right\}^2 \tag{6.27}$$

The dependence of the outputs, innovation level, and the profits of the firms are derived as follows.

Lemma 3 *In the present case with the foreign firm imitating, we have*

$$\frac{\partial x^m}{\partial e} < 0, \quad \frac{\partial x^m}{\partial \theta} < 0, \quad \frac{\partial x^{*m}}{\partial e} > 0, \quad \frac{\partial x^{*m}}{\partial \theta} > 0$$

$$\frac{\partial g^m}{\partial e} < 0, \quad \frac{\partial g^m}{\partial \theta} < 0$$

Lemma 4 $\Pi_e^m < 0, \Pi_\theta^m < 0, \Pi_e^{*m} > 0$ *and* $\Pi_\theta^{*m} > 0$.

As in the case of innovation by the less efficient firm, an increase in the cost of innovation faced by the advanced firm is detrimental to the firm's innovation, output, and profit but is good for the other firm. Not surprisingly, an increase in the degree of imitation, θ, has similar effects because it tends to help the less efficient firm but hurt the advanced firm.

Surpassing and catching up

Before we explain how the foreign firm chooses between innovation and imitation, we want to derive more properties of the two cases. In particular, we want to examine how in each of the cases surpassing or catching up is possible.

Definition (surpassing)

Surpassing by the Foreign firm occurs if and only if $c_1 > c_1^*$.

Definition (catching up)

Catching up by the foreign firm occurs if and only if $c_0^* - c_0 > c_1^* - c_1 > 0$.

In other words, the foreign firm is said to have surpassed the home firm in terms of technology if its new marginal cost is lower than that of the home firm. The foreign firm is said to be catching up if it is closing the technology gap

between the two firms, but the home firm still has a more advanced technology. We want to derive conditions under which surpassing or catching up is possible.

Case (A)

We now examine the conditions for surpassing and catching up in case (A), in which the Foreign firm innovates (so does the Home firm).

Proposition 1 *Given Assumption 2, if the foreign firm chooses to innovate, it will surpass the home firm if and only if*

$$e > \frac{e^*(a - c_0^*)}{(a - c_0) - 3e^*(c_0^* - c_0)} \tag{6.28}$$

Corollary 1 *If the foreign firm has a cost of innovation not less than that of the home firm, $e^* \geq e$, then surpassing is not possible with both firms innovating.*

What the proposition and condition (6.28) mean is that if the home firm's cost of innovation is sufficiently higher than that of the foreign firm, the foreign firm will choose a new marginal cost lower than what the home firm chooses.

Surpassing can be illustrated graphically in Figure 6.1. Curve SP represents the following equation:

$$e = \frac{e^*(a - c_0^*)}{(a - c_0) - 3e^*(c_0^* - c_0)} \tag{6.29}$$

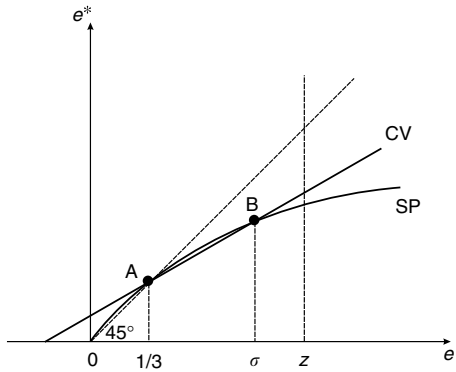

Figure 6.1 Curves SP and CV.

It passes through the origin. The slope of curve SP is equal to

$$\left.\frac{de^*}{de}\right|_{\mathrm{SP}} = \frac{[a - c_0 - 3e^*(c_0^* - c_0)]^2}{(a - c_0)(a - c_0^*)} > 0 \tag{6.30}$$

When $e^* \to 0$, the slope of the curve is greater than unity, implying that curve SP is above the 45° line in the small region near the origin (except at the origin). Equation (6.30) shows that the slope is monotonically decreasing with e^*. Curve SP cuts the 45° line when $e = e^*$, or, from (6.29), when $e = e^* = 1/3$. For $e > 1/3$, curve SP is entirely below the 45° line. For $e > z$ (Assumption 2), surpassing occurs in the region on the right-hand side of (below) curve SP. We now turn to catching up.

Proposition 2 *In the present case in which both sides choose to innovate, when given assumption 1, catching up occurs if and only if*

$$e > \frac{e^*(a - c_0) + (e^* - 1)(c_0^* - c_0)}{a - 2c_0^* + c_0} \tag{6.31}$$

Corollary 2 *If* $e^* \geq e$, *catching up is not possible*.

Catching up can be illustrated graphically. In Figure 6.1, curve CV represents equation

$$e = \frac{e^*(a - c_0) + (e^* - 1)(c_0^* - c_0)}{a - 2c_0^* + c_0} \tag{6.32}$$

The slope of curve CV is equal to

$$\left.\frac{de^*}{de}\right|_{\mathrm{CV}} = \frac{a - 2c_0^* + c_0}{a - 2c_0 + c_0^*} > 0 \tag{6.33}$$

Equation (6.33) shows that the slope of curve CV is constant and less than one. From (6.32), we see that curve CV cuts the horizontal axis at a point at which $e = -(c_0^* - c_0)/(a - 2\,c_0^* + c_0)$, and cuts the 45° line when $e = e^* = 1/3$.

It is interesting to compare curve SP and curve CV, as both of them are positively sloped.

Lemma 5 *Curves SP and CV intersect at two points, A* $(e = e^* = 1/3)$ *and B* $(e = e^* = \sigma,$ *where* $\sigma = -(a - c_0)/(a - 2\,c_0 + c_0^*) < 1)$. *For* $e \in (0, 1/3)$ *and* $e \in (\sigma, \infty)$, *curve CV is above curve SP. For* $e \in (1/3, \sigma)$, *curve CV is below curve SP*.

The above results are shown in Figure 6.1. Points A and B are the points of intersection between curve SP and curve CV. Since we focus on the area in which $e > z$, we can conclude that in this relevant region, curve SP is below and on the right-hand side of curve CV.

Case (B)

Let us now turn to the case in which the foreign firm chooses to imitate while the home firm innovates. We first note that imitation will never make the foreign firm as good as the home firm before home firm's innovation. In other words, if the foreign firm chooses to imitate, surpassing can never happen. This result is formally stated and proved below.

Proposition 3 *In case (B) with the foreign firm imitating, surpassing is not possible.*

Proposition 3 is quite intuitive. Surpassing (6.5) implies that at best the foreign firm can improve its technology to the initial technology level of the Home firm.[8]

Proposition 4 *In case (B) with the Foreign firm imitating, catching up occurs if and only if*

$$e > \frac{a - c_0 + (\theta + 1)(c_0^* - c_0)}{3\theta(c_0^* - c_0)} \tag{6.34}$$

Proposition 4 means that by imitating, the Foreign firm may make a bigger technology improvement than the home firm does. This occurs if the home firm's cost of innovation is high as compared with the degree of imitation. Catching up with the foreign firm imitating can also be illustrated graphically. Let us consider the following condition:

$$e = \frac{a - c_0 + (\theta + 1)(c_0^* - c_0)}{3\theta(c_0^* - c_0)} \tag{6.35}$$

which gives the threshold values of e and θ with which the technology gap between the two firms remains constant. The condition, when given a value of θ can be represented by a vertical line in the e–e^* space similar to Figure 6.1. The space to the right of the line represents values of e (when given θ) that lead to catching up.

Choosing between innovation and imitation

So far, we have focused the analysis on each of the two cases, assuming that the foreign firm either innovates or imitates. We now explain how the firm will choose between these two actions.

Since the foreign firm tries to maximize its profit, it innovates if and only if

$$\Pi^{*v}(e, e^*) > \Pi^{*m}(e, \theta) \tag{6.36}$$

To see the choice for the Foreign firm, let us consider the following condition:

$$\Pi^{*v}(e, e^*) = \Pi^{*m}(e, \theta) \tag{6.37}$$

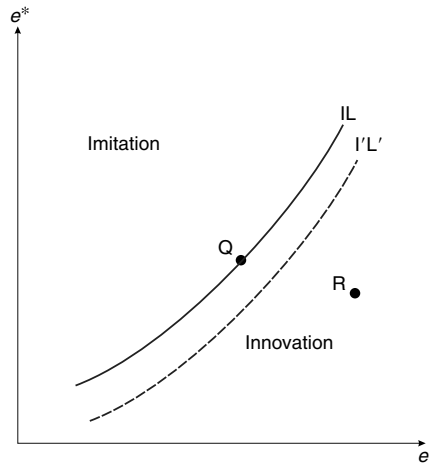

Figure 6.2 Curve IL.

Based on (6.36), (6.37) represents the combinations of e, e^*, and θ that will make the foreign firm indifferent between innovation and imitation. These combinations, when given θ, can be represented by the indifference locus, labeled IL, in Figure 6.2. Totally differentiate both sides of (6.37) to give

$$\Pi_e^{*v} de + \Pi_{e^*}^{*v} de^* = \Pi_e^{*m} de + \Pi_\theta^{*m} d\theta \tag{6.38}$$

From Lemmas 2 and 4, $\Pi_{e^*}^{*v} < 0$, while Π_e^{*m}, Π_e^{*v}, and Π_θ^{*m} are both positive. When given the value of θ, (6.38) is rearranged to give the slope of curve IL

$$\left.\frac{de^*}{de}\right|_{\text{IL}} = \frac{\Pi_e^{*m} - \Pi_e^{*v}}{\Pi_{e^*}^{*v}} \tag{6.39}$$

The sign of the slope is ambiguous. Thus, curve IL is positively sloped if and only if $\Pi_e^{*m} < \Pi_e^{*v}$, i.e. if and only if a rise in e will help the Foreign firm more when it imitates than when it innovates. This case is shown in Figure 6.2. The area on the right- (left-) hand side of curve IL represents the values of e and e^* that will lead to innovation (imitation) by the foreign firm.

Since $\Pi_\theta^{*m} > 0$, an increase in θ will shift curve IL down and to the right, say, I′L′. This means that the foreign firm will more likely choose to imitate. For example, point Q is on curve IL, meaning that the foreign firm is indifferent between imitation and innovation. If there is a rise in θ, the curve shifts to I′L′. Point Q now becomes a point above the curve, implying that the foreign firm now prefers to imitate.

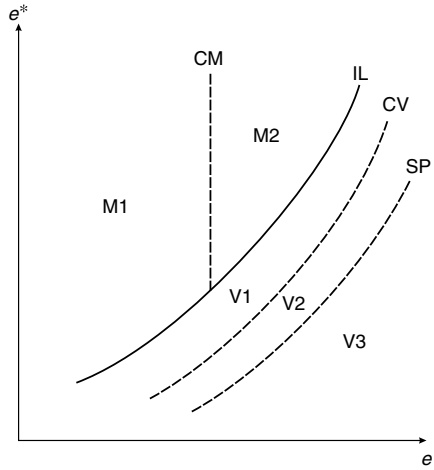

Figure 6.3 Innovation and imitation.

We can now add in curves SP and CV to the diagram. See Figure 6.3. By Lemma 5, for $e>z$, curve SP is below and to the right of curve CV. We can add in line CM, which gives the values of e (when given θ) that will help the foreign firm catch up when it chooses imitation. We can now identify the following regions in the diagram (Innovation or imitation refers to the decision of the foreign firm, as the home firm always innovates):

- V1: innovation with no catching up or surpassing;
- V2: innovation with catching up but no surpassing;
- V3: innovation with surpassing;
- M1: imitation with no catching up;
- M2: imitation with catching up.

So far, we have been focusing on the choice of the foreign firm and its profits. Now let us examine how the profit of the home firm may be different in different cases. In particular, we want to raise and answer one interesting question: Suppose that we do not endogenize the decision of the foreign firm, as many papers in the literature do, how would the profit of the home firm be affected? To see this question more clearly, consider the values of (e, e^*, θ) that are represented by point R in Figure 6.2. At this point, the foreign firm will choose to innovate because $\Pi^{*v}(e, e^*) > \Pi^{*m}(e, \theta)$. If it does, the home firm's profit is $\Pi^{v}(e, e^*)$. However, if we assume that the foreign firm imitates, the home firm's profit will

be $\Pi^m(e, \theta)$. How can these two profits of the home firm be compared? In other words, if one follows the practice of many papers in the literature and assumes that the Foreign firm imitates, then what kind of error may one commit?

Before we answer the above questions, let us first present the following two lemmas:

Lemma 6 *If the home firm always chooses its output and level of innovation optimally, its profit is negatively dependent on the foreign firm's output but independent of whether the foreign firm innovates or imitates*.

This lemma is just a simple application of the Envelope Theorem.

Lemma 7 *If* $\Pi^{*v}(e, e^*) > \Pi^{*m}(e, \theta)$, *then* $g^{*v} > \theta(c_0^* - c_0)$ *and* $c_1^{*v} < c_1^{*m}$.

Lemmas 6 and 7 are used to prove the following proposition, which answers the questions raised earlier.

Proposition 5 *If* $\Pi^{*v}(e, e^*) > \Pi^{*m}(e, \theta)$, *then* $\Pi^v(e, e^*) > \Pi^m(e, \theta)$.

The proposition implies that if the foreign firm prefers to innovate but is allowed to imitate only, its profit will drop while the profit of the home firm will rise. In other words, in the innovation–imitation models, if the foreign firm in fact prefers to innovate but is assumed to imitate, the profit of the foreign firm could be underestimated but that of the home firm be overestimated.

Concluding remarks

Using a simple model of international rivalry, we analyzed how two firms compete in terms of technology and output. We derived conditions for the less efficient firm to choose between innovation and imitation, and also conditions under which the less efficient firm is able to catch up or to surpass the technology of the other firm.

Endogenizing the innovation/imitation decision of the less efficient firm allows us to examine the potential risk of the artificial assumption that the less efficient firm always imitates. We argued that if the less efficient firm would choose to innovate, if allowed to, then assuming that it imitates will tend to underestimate its profit but overestimate the profit of the other firm.

In this chapter, we restricted our analysis to a static model in order to get more insights and results. It would be interesting to examine some dynamic issues such as switching between innovation and imitation over time using a dynamic model, as Shimomura and Wong (2001) do. Another approach to analyzing the issues of growth and innovation/imitation switching is to combine the present framework with a differential game of duopoly in which both of the two firms have two alternatives: to innovate and to imitate.[9] The latter could be a topic of future research.

Appendix

Proofs of Lemmas and Propositions stated in this paper:

Proof of Lemma 1

Differentiating (6.12) and (6.14) with respect to e and e^*, we get

$$\frac{\partial g^v}{\partial e} = \frac{-(3e^* - 2)[(e^* - 1)(a - c_0) + e^*(c_0^* - c_0)]}{[(2e - 1)(2e^* - 1) - ee^*]^2} < 0$$

$$\frac{\partial g^v}{\partial e^*} = \frac{(e - 1)(a - c_0^*) - e(c_0^* - c_0)}{[(2e - 1)(2e^* - 1) - ee^*]^2} > 0$$

$$\frac{\partial x^v}{\partial e} = \frac{-(2e^* - 1)[(e^* - 1)(a - c_0) + e^*(c_0^* - c_0)]}{[(2e - 1)(2e^* - 1) - ee^*]^2} < 0$$

$$\frac{\partial x^v}{\partial e^*} = e\frac{\partial g^v}{\partial e^*} > 0$$

The effects of e and e^* on $g^{*v}$and x^{*v}can be obtained in a symmetric way.

Proof of Lemma 2

Differentiating (6.16) and (6.17) with respect to e and e^*, respectively, yields

$$\Pi_e^v = -\frac{1}{2}\left[\frac{\psi^*}{(2e - 1)(2e^* - 1) - ee^*}\right]^2 < 0$$

$$\Pi_{e^*}^v = \frac{\psi\psi^* e(2e - 1)}{[(2e - 1)(2e^* - 1) - ee^*]^3} > 0$$

$$\Pi_e^{*v} = \frac{\psi\psi^* e^*(2e^* - 1)}{[(2e - 1)(2e^* - 1) - ee^*]^3} > 0$$

$$\Pi_{e^*}^{*v} = -\frac{1}{2}\left[\frac{\psi}{(2e - 1)(2e^* - 1) - ee^*}\right]^2 < 0$$

where the assumption that both g^v and g^{*v} are positive has been used and $\psi = (e - 1)(a - c_0^*) - e(c_0^* - c_0)$ and $\psi^* = (e^* - 1)(a - c_0) - e^*(c_0^* - c_0)$.

Proof of Lemma 3

The response of g^m on e and θ, and $\partial x^m/\partial\theta$ are trivially obtained, while the other results are derived as

$$\frac{\partial x^m}{\partial e} = \frac{-2\left[a - c_0 + (1-\theta)(c_0^* - c_0)\right]}{(3e-2)^2} < 0$$

$$\frac{\partial x^{*m}}{\partial e} = \frac{a - c_0 + (1-\theta)(c_0^* - c_0)}{(3e-2)^2} > 0$$

$$\frac{\partial x^{*m}}{\partial \theta} = \frac{(2e-1)(c_0^* - c_0)}{3e-2} > 0$$

This gives Lemma 3.

Proof of Lemma 4

Differentiating (6.26) and (6.27) with respect to e and θ, we have

$$\Pi_e^m = \frac{(2-5e)\zeta^2}{2(3e-2)^3} < 0$$

$$\Pi_\theta^m = \frac{-e(2e-1)\zeta(c_0^* - c_0)}{3e-2} < 0$$

$$\Pi_e^{*m} = \frac{2\zeta[(2e-1)\zeta^* - e(a-c_0)]}{(3e-2)^3} > 0$$

$$\Pi_\theta^{*m} = \frac{2(2e-1)(c_0^* - c_0)\left[\zeta^*(2e-1) - e(a-c_0)\right]}{(3e-2)^2} > 0$$

where $\zeta = a - c_0 + (1-\theta)(c_0^* - c_0)$ and $\zeta^* = a - c_0^* + \theta(c_0^* - c_0)$. The signs of the partial derivatives depend on the assumption that the outputs and innovation level are positive.

Proof of Lemma 5

Equations (6.29) and (6.32) imply that the horizontal gap between curves CV and SP is equal to

$$\frac{e^*(a-c_0) + (e^*-1)(c_0^* - c_0)}{a - c_0^* - (c_0^* - c_0)} - \frac{e^*(a - c_0^*)}{a - c_0 - 3e^*(c_0^* - c_0)}$$

$$= \frac{-(3e^* - 1)(c_0^* - c_0)[(e^* - 1)(a - c_0) + e^*(c_0^* - c_0)]}{[a - c_0^* - (c_0^* - c_0)][a - c_0 - 3e^*(c_0^* - c_0)]} \tag{6.40}$$

Setting the numerator in (6.40) to zero gives two roots: $e^* = 1/3$ and $e^* = \sigma < 1$. Comparison between (6.29) and (6.32) implies that curve CV is above curve SP when $e^* < 1/3$ and when $e^* > \sigma$. In between these two points, curve CV is below curve SP.

Proof of Lemma 6

We first note that equation (6.6) is identical to equation (6.18), meaning that the home firm's profit function is independent of whether the foreign firm innovates or imitates. Its decision depends only on what the Foreign firm produces. Its profit function can be written as $\Pi = \hat{\Pi}(x,g,x^*)$. If the firm always chooses x and g optimally, then

$$\frac{d\Pi}{dx^*} = \frac{\partial \Pi}{\partial x}\frac{dx}{dx^*} + \frac{\partial \Pi}{\partial g}\frac{dg}{dx^*} + \frac{\partial \Pi}{\partial x^*} = -x < 0$$

This proves the lemma.

Proof of Lemma 7

Suppose that $\Pi^{*v}(e,e^*) > \Pi^{*m}(e,\theta)$. If $g^{*v} \leq \theta(c_0^* - c_0)$ then innovation can never be a good choice because the Foreign firm can choose imitation instead, after paying nothing and getting a bigger or equal technology improvement. So we must have $g^{*v} > \theta(c_0^* - c_0)$, which implies that $c_1^{*v} < c_1^{*m}$.

Proof of Proposition 1

In case (A), with both firms innovating, the new gap between the firms' marginal costs is

$$c_1^{*v} - c_1^v = c_0^* - g^{*v} - c_0 + g^v$$

$$= \frac{e(3e^* - 1)(c_0^* - c_0) + (e^* - e)(a - c_0^*)}{(2e - 1)(2e^* - 1) - ee^*} \tag{6.41}$$

which in general can be both positive and negative. Thus, $c_1^{*v} - c_1^v < 0$ if and only if the numerator in (6.41) is negative, or if and only if

$$-e\left[a - c_0 - 3e^*(c_0^* - c_0)\right] + e^*(a - c_0^*) < 0 \tag{6.42}$$

Condition (6.42) is equivalent to (6.28).[10]

Proof of Proposition 2

To determine whether catching up exists, note that

$$\begin{aligned}(c_0^* - c_0) - (c_1^{*v} - c_1^v) &\equiv c_0^* - c_0 - (c_0^* - g^{*v} - c_0 + g^v) \\ &= g^{*v} - g^v \\ &= \frac{e(a - 2c_0^* + c_0) - e^*(a - c_0) - (e^* - 1)(c_0^* - c_0)}{(2e - 1)(2e^* - 1) - ee^*}\end{aligned}$$

From the definition, we can see that catching up exists if and only if the numerator is positive. Rearranging the terms will give the proposition.

Proof of Proposition 3

We note that

$$\begin{aligned}c_1^{*m} - c_1^m &= c_0^* - \theta(c_0^* - c_0) - c_0 + g^m \\ &= \frac{(a - c_0) + (3e - 1)(1 - \theta)(c_0^* - c_0)}{3e - 2} > 0\end{aligned}$$

This proves the proposition.

Proof of Proposition 4

With the foreign firm imitating, the change in the gap between the marginal costs of the firms is given by

$$\begin{aligned}(c_0^* - c_0) - (c_1^{*m} - c_1^m) &\equiv c_0^* - c_0 - [c_0^* - \theta(c_0^* - c_0) - c_0 + g^m] \\ &= \theta(c_0^* - c_0) - g^m \\ &= \frac{\theta(3e - 2)(c_0^* - c_0) - (a - c_0) - (1 - \theta)(c_0^* - c_0)}{3e - 2}\end{aligned}$$

The change in the gap is positive if and only if the numerator is positive. Rearranging the terms will give the proposition.

Proof of Proposition 5

Note first that the two reaction functions given by (6.8) and (6.9), or (6.20) and (6.21), of the home firm depend on x^* but not on whether the foreign firm innovates or imitates. Substituting (6.9) into (6.8) (or (6.21) into (6.20)), we get a function similar to the usual reaction function of a firm in a duopoly model.

If $\Pi^{*v}(e, e^*) > \Pi^{*m}(e, \theta)$ then by Lemma 7, $c_1^{*v} < c_1^{*m}$. This means that if the foreign firm chooses to innovate but if it is allowed to imitate only, its new marginal cost will be higher, no matter what the home firm chooses. It is well known in the regular duopoly model that a rise in the marginal cost of one of the firms will cause it to lower its output. By Lemma 6, the profit of the home firm will be higher.

Proof of Corollary 1

If $e^* \geq e$, then the numerator in (6.41) is positive, implying that $c_1^{*v} > c_1^v$.

Proof of Corollary 2

If $e^* \geq e$, the right-hand side of (6.31) can be rearranged to give

$$\frac{e^*(a - c_0) + (e^* - 1)(c_0^* - c_0)}{a - 2c_0^* + c_0} = e^* + \frac{(3e^* - 1)(c_0^* - c_0)}{a - 2c_0^* + c_0} > e^*$$

violating condition (6.31), implying that catching up is not possible.

Notes

1 Bhagwati (2004) makes an extensive and stimulating argument on globalization and welfare in the modern world.
2 Helpman's (1993) model is recently extended and reexamined by Mondal and Gupta (2006).
3 The present framework is based on that in Shimomura and Wong (2001), except that it is a static one. This allows us to investigate the rivalry between and strategies of the firms in more detail.
4 Unless stated otherwise, the variables of Foreign are distinguished by asterisks. Without loss of generality, we assume that the fixed costs of the firms are negligible.
5 Variable e is interpreted as the cost of innovation because $eg^2/2$ is the cost the home firm has to pay in order to reduce its marginal cost by g. As will be shown later, an increase in e will lower the firm's output, level of innovation, and profit.
6 Throughout the chapter, we assume interior equilibria.
7 Subindices are used to denote partial derivatives.
8 We do not consider the unrealistic case in which $\theta > 1$.
9 See, for example, Fershtman and Kamien (1987) and Tsutsui and Mino (1990) for recent work of analysis using differential game models of duopoly.
10 In deriving (6.42), we have assumed that $a - c_0 - 3e^*(c_0^* - c_0) > 0$.

Bibliography

Bagwell, K. and R. W. Staiger (1992) "The Sensitivity of Strategic and Corrective R&D Policy in Battles for Monopoly," *International Economic Review*, 33: 795–816.
Bagwell, K. and R. W. Staiger (1994) "The Sensitivity of Strategic and Corrective R&D Policy in Oligopolistic Industries," *Journal of International Economics*, 36: 133–150.
Bhagwati, J. N. (2004) *In Defense of Globalization*, New York: Oxford University Press.

Dinopoulos, E., A. Gungoraydinoglu and C. Syropoulos (2006) "Parent Protection and Global Schumpeterian Growth," in this volume.

Dinopoulos, E. and C. Syropoulos (1998) "International Diffusion and Appropriatability of Technological Expertise," in Baye, M. R. (ed.) *Advances in Applied Microeconomics*, Vol. 7, Amsterdam: Elsevier Science, 115–137.

Fershtman, C. and M. I. Kamien (1987) "Dynamic Duopolistic Competition with Sticky Prices," *Econometrica*, 55: 1151–1164.

Grossman, G. M. and E. Helpman (1991) *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.

Helpman, E. (1993) "Innovation, Imitation, and Intellectual Property Rights," *Econometrica*, 61: 1247–1280.

Liao, P. and K. Wong (2004) R&D Subsidy, Intellectual Property Rights Protection, and North-South Trade: How Good Is the TRIPS Agreement?, Mimeo, University of Washington

Mondal, D. and M. R. Gupta (2006) "Innovation, Imitation, and Intellectual Property Rights: A Note on Helpman's Model," *Journal of Economics*, 87: 29–53.

Segerstrom, P. S. (1991) "Innovation, Imitation, and Economic Growth," *Journal of Political Economy*, 99: 807–827.

Shimomura, K. and K. Wong (2001) Technology Adoption and Trade Policies, Mimeo, University of Washington.

Tsutsui, S. and K. Mino (1990) "Nonlinear Strategies in a Dynamic Duopolistic Competition with Sticky Prices,," *Journal of Economic Theory*, 52: 136–161.

7 Foreign manufacturing affiliates of US multinationals

Myths and realities in the globalization debate

James R. Markusen and Bridget Strand

Introduction

The globalization debate in general and the antiglobalization participants in particular often focus on the role of multinational firms in the globalization process. Multinationals are targeted in the debate and portrayed as the vehicles for spreading various ills to poor countries ranging from poor labor conditions to environmental degradation. Phrases like "social dumping" are commonplace, although this specific phrase is also heard within Europe, notably from the French and German governments.

Much of the anti-globalist position is based on carefully selected case studies and anecdotes and, on the basis of these, casual or even implicit theories about multinationals are advanced. In our complex world, one can probably find a case study to support virtually any preconceived idea if one searches hard enough. The rhetorical tactic is then to take this case study (of course more than one is preferred, but not necessary) and draw general inferences to the effect that "this is the way multinationals operate." The economics profession, in turn, is not always as helpful as it might be in clarifying the debate. While there have been many advances on the theory front in the last two decades and numerous empirical insights in the last eight years, textbooks and even some new research continue to advance ideas that are either contrary to modern theory and evidence or that may represent only a minor part of the actual role of multinationals in the world economy.

This chapter aims to provide an updated view of the role of multinationals, particularly manufacturing MNEs, in the world economy. Much of what we present here is empirical evidence, but that is in turn closely connected to and supportive of theoretical developments over the last two decades. In doing so, we will challenge the casual theorizing of the antiglobalists, and also the relevance of some more formal theory that continues to be advanced by the economics profession. We contend that, in certain respects, traditional economic theory and antiglobalists agree on points of positive theory, although they impose very different normative interpretations on these results. The approach of this chapter is to focus on the issue of positive economics. We argue that much traditional (but also some new) theory and antiglobalist arguments

are empirically misguided, or at least concentrate on only a small part of what multinational firms actually do in foreign manufacturing affiliates.

We offer one caveat before proceeding, which is that we are concerned about what activities go on within the ownership boundaries of multinational firms. To the extent that there are issues in clothing and footwear production; for example, foreign factories in these industries are typically not owned by North American and European multinationals, and it is incumbent on the antiglobalists to get this straight in their anecdotes.

Traditional theory, modern theory, and antiglobalists' implicit theory

Many international trade and finance textbooks continue to see direct investment as just part of the theory of capital movements, with capital moving from the high-income capital-abundant to the low-income capital-scarce countries. If the underlying paradigm is Hecksher–Ohlin, then the capital-scarce countries are also labor-abundant, low-wage countries. So, closely related is the assumption that multinationals are moving capital to low-wage countries, the flip-side of the usual Heckscher–Ohlin model of the world. This world-view is, of course, enormously appealing to international economists. It is in turn, closely consistent with the arguments of many of the antiglobalists: northern capitalists moving their resources to exploit cheap southern labor.

From here it is easy to substitute skilled labor for capital, or add skilled labor as a third factor. The factor–proportions theory then leads to the prediction that multinationals move capital to the unskilled-labor-abundant south, thus reducing the demand for and wages of unskilled-labor in the unskilled-labor scarce north. Of course, this is exactly the claim of the antiglobalists. The argument from the latter differs from the traditional capital-movement view not in positive theory, but in the normative interpretation of whether or not such a movement is good or bad.

The capital-movement view of foreign direct investment (FDI) has resurfaced in recent years in "vertical" or "resource seeking" models of multinational firms. An early model of this is by Helpman (1984), and it has resurfaced in recent years with a new vocabulary, principally "outsourcing" and "offshoring." These modes postulate that multinationals are firms that are able to geographically fragment the production process into stages with different factor intensities, and that stages are located in countries where the factors they use intensively are relatively cheap. Conceptually, this is very close to the factor-proportions world discussed in the previous two paragraphs. Firms will want, for example, to move an unskilled-labor-intensive phase of production to an unskilled-labor-abundant, low-wage country. This is pure Heckscher–Ohlin in new clothes and consistent with some antiglobalists' positions. It is important to realize the differences in the vocabulary used by media pundits and antiglobalists, offshoring refers only to the movement of an element of production abroad while outsourcing is done both domestically and abroad without any claim of ownership and management.

Another view of multinationals, although one much harder to document, is that multinationals are "creatures of market imperfections." The idea is that multinationals arise because of the existence of imperfections in markets and market structures. For example, multinationals may create their own channels of trade and information when markets are poorly developed, and/or are able to exert monopolistic market-power by virtue of size and efficiency. Trade which is generated by distortions and therefore may be harmful is, of course, an important contribution of Bhagwati and co-authors (e.g. Bhagwati, 1971; Bhagwati and Ramaswami, 1963; and Bhagwati *et al*., 1969). They show quite clearly how trade or factor movements in a second-best world of distortions may lower welfare.

At a superficial level, this seems quite consistent with several positions of the antiglobalists. However, we think that this is not the case, and that appealing to Bhagwati's theory of distortions to support the antiglobalist position is a stretch. The latter seems to assert that multinationals create the distortions, while Bhagwati's theory is about firms' responses to existing distortions. If markets are (exogenously) missing and multinationals create internal production and trade channels, that is very different from asserting that multinationals undermine existing indigenous markets and destroy competition. An additional example supporting Bhagwati's theory relates to the observation of prices in excess of marginal costs (an important subject of distortions theory) and concentration in equilibrium. We see these as a natural consequence of increasing-returns technologies and not something caused by the multinational firms *per se*. In fact, there may be procompetitive price effects from a multinational entering a previously distorted market.

Finally, another recent theory of multinationals is the "horizontal" or "market-seeking" view. Firms create firm-specific assets through investments in R&D and learning by doing. Once this knowledge capital is created, its services are essentially those of joint inputs or "public goods" within the firm, and can be applied over and over again to generate value in different locations. An early model of this is Markusen (1984), with more general-equilibrium structure added later by Horstmann and Markusen (1987, 1992), and Brainard (1993a). An important characteristic of these models is that they predict multinationals activity arising between similar countries, and indeed it is empirically found to arise more often between similar countries (Markusen and Venables, 1998, 2000). A factor-proportions basis for FDI is not needed and, in its absence, the asserted labor-market consequences of FDI mentioned above need not hold.

No one holds the view that these alternative theories of the multinational are mutually exclusive. Every scholar surely believes that these various motives and resulting investments occur in reality, and the only sensible question is the relative importance of these underlying causes in empirical reality. Just as we criticized the antiglobalists above for extrapolating from single (carefully selected) case studies to general theories, we should not rely too much on a theory that fits with only a small proportion of empirical observations. So this is the subject to which we now turn: what has a decade of empirical work taught us about multinational firms and which theories are most consistent with these results?

Finding 1: MNE investment flows primarily from–to high-income countries

There is now a good deal of evidence that the traditional Heckscher–Ohlin-based approaches, whether the old capital flow or new vertical modes, cannot account for the large bulk of what multinationals actually do. These models predict that FDI should increase with differences between countries, particularly with respect to factor endowments. Evidence accumulated over the last decade, starting with Brainard (1997) in particular, shows a different reality.

Table 7.1 shows that multinational investment flows primarily from *but also to* high-income, capital-abundant counties.

According to UNCTAD data, 69% of the world's *inward* FDI stock is found in developed countries. More than two-thirds of the total world FDI stock are firms from high-income countries investing in other high-income countries. This still leaves, of course, a large chunk of MNE activity flowing north–south, and Table 7.1 notes that 28% of the stock of inward investment is in developing countries. This is surely nontrivial, but further analysis shows that the overwhelming portion of this statistic goes into the best of the developing countries (e.g. Hong Kong, Singapore, Brazil, Chile). The least developed countries, a UN definition that includes 48 of the world's poorest countries, only accounts for one percent of the world's inward stock as noted in Table 7.1. Thus there is clearly not a strong case for relying on factor-proportions approaches to FDI to explain empirical results.

If one takes a gravity theory point of view instead, one could object that the developed countries also have a very large share of world income and so these statistics are not surprising. This position deserves examination, although we emphasize that it is not implied by any capital flow or vertical model of FDI. The lower part of Table 7.1 gives statistics that are ratios of two shares: the numerator is each group's share of the world's inward FDI stock and the denominator is that

Table 7.1 World FDI inward and outward stocks: 2003

FDI inward and outward stocks as shares in total world stocks in 2003		
Developed countries	Inward:	0.69
	Outward:	0.89
Developing countries	Inward:	0.28
	Outward:	0.10
Least developed countries	Inward:	0.01
	Outward:	0.00
Share of inward world FDI stock/Share of world GDP in 2003		
Developed countries		0.90
Developing countries		1.37
Least developed countries		1.07

Source: World Investment Report 2004, UNCTAD.

Note
Least developed countries is a UN definition consisting of 48 countries.

group's share of world GDP in 2003. It is true that the developed countries have a share ratio of less than one, suggesting that they are net outward investors, but the ratio is still rather large at 0.90. The fact that the developed countries have a share of inward investment equal to 90% of its share of world income is put into better perspective when per capita income is considered. On the contrary, the developing countries have a share greater than one, 1.37, but again we find that inward FDI is going to the better off of the developing countries. The least developed countries have a share very close to one, 1.07.

Again, it is not clear what to make of these last numbers in view of the capital flows and vertical models: the numbers are not predicted by those theories, they predict little or no relationship between FDI inflows and country size *per se*. If anything, they are more consistent with a simple gravity model approach based on some unknown underlying economic mechanism. They are also consistent with Markusen's "knowledge–capital approach" which integrates horizontal and vertical motives into one model which we discuss later (Markusen, 2002).

Finding 2: US outward FDI is primarily horizontal or "market seeking"

The UNCTAD data referred to above give the best geographic coverage. Unfortunately, they are limited to investment stocks and flows. This precludes us from understanding what the multinationals are actually doing, and thus dramatically limit our ability to discriminate among alternative theories of the MNE. The best alternative published source of data is the US BEA data, which gives sales of foreign affiliates and whether those sales are local, exported back to the parent, or exported to third countries. The drawback is that the data is only bilateral with the United States, so it only gives foreign affiliates of US companies or US affiliates of foreign companies. This is an important limitation but given the alternatives the BEA data allows a more thorough examination of the importance of alternative theories.

For the remainder of the chapter, we will use the Markusen–Maskus data set on foreign manufacturing affiliates of US multinationals using the published BEA data. This has been updated to include 39 host countries for the period 1984–2003 for which we have complete data. Before specifying a more formal model, Table 7.2 presents some summary statistics on the sales of the foreign manufacturing affiliates in 2000.

The first row of Table 7.2 notes that local sales in the host country account for 60% of manufacturing affiliate sales. Interestingly, only 12% of sales are back to the United States (and this figure would likely be a lot lower for services), 28% of sales are to third countries, sometimes referred to as "export platform production." Thus the bulk of foreign sales are local, lending a great deal of support to the horizontal or market-seeking motive. However, 40% export sales is certainly not a trivial figure and deserve closer analysis.

It is not very clear how to interpret this third-country figure in light of the theory and several of us are currently working on this problem. But the data in the subsequent lines of Table 7.2 give us some clue. Third-country exports as a share of total

Table 7.2 Distribution of sales by US affiliates between local sales, exports to the United States, exports to third countries

Sales of foreign affiliates of US multinationals: shares in total, 2000

	Local sales	*Export sales to the US*	*Export sales to third countries*	*Share of total export sales to third countries*
All countries in sample (39)	0.60	0.12	0.28	0.70
Ireland	0.13	0.16	0.71	0.82
Belgium	0.39	0.04	0.57	0.93
Greece	0.83	0.01	0.16	0.97
Holland	0.37	0.03	0.60	0.95
Portugal	0.80	0.00	0.20	0.98
Hong Kong	0.33	0.36	0.32	0.47
Indonesia	0.82	0.04	0.14	0.78
Malaysia	0.20	0.33	0.47	0.59
Philippines	0.44	0.19	0.37	0.66
Singapore	0.35	0.32	0.34	0.52
Canada	0.57	0.38	0.05	0.11
Mexico	0.53	0.39	0.08	0.17

sales are, perhaps surprisingly, highest in smaller EU countries, not some group of developing countries. A prevailing preliminary opinion is that these are basically horizontal or market–seeking investments made to serve the whole EU. But a particular country must be chosen (given plant-level scale economies), and naturally any choice is going to lead to a relatively large share of production exported to the rest of the EU. US manufacturing affiliates in Ireland, a favorite choice of American firms in the computer and pharmaceutical industries, export over 80% of their output to third countries, overwhelmingly the rest of the EU. These are really horizontal investments from the US parents point of view, with Ireland chosen as the specific location. A related comment most likely also applies to the Canada component of NAFTA. Much of Canadian affiliate production is exported to the United States; again, the motion is essentially a horizontal one with Canada chosen as the best location.

Contrast the Ireland figure with the sales of the Latin American group in the data. These five countries have traditionally had relatively high trade barriers, and much of the inward investment is for local sales (84%) generated by a "tariff jumping" motive. It is pretty clear that the Latin American affiliates in this group do not fit a vertical model of finding a low-cost location for producing for export to the United States, contrary to the claims of many antiglobalists.

The closest group to the traditional vertical model is the Southeast Asian group of five countries, but even in this case 43% of sales are local. We think that the bottom line here is that the conceptual view of multinationals as seekers of

low-wage labor, moving production abroad and exporting back to the US market is a very poor fit for the data.

Now we turn to more formal analysis of the data. We use an expanded and updated Markusen–Maskus dataset on manufacturing affiliates of US multinationals. Details of the theoretical foundations of the estimating equations are found in Carr *et al*. (2001), Markusen and Maskus (2001, 2002) and Markusen (2002) and in the Appendix to this chapter.

Dependent variables for each host/year observation:

Real sales of foreign manufacturing affiliates of US multinationals in country *j*, 1984–2003 for 39 host economies for which we have complete right-hand side variables. The whole sample contains 549 observations without any missing data depending on regression. Separate regressions for

A total sales	($RSALETOT_j$)
B local sales	($RSALEUS_j$)
C export sales back to the US	($RSALELOC_j$)
D export sales to third countries	($RSALEOTH_j$)

Summary statistics are given in Table 7.3.

Independent variables for each host/year observation are as follows:

Real US and real host-country *j* GDP (GDP_{us}, GDP_j).

Share of the labor force that is skilled for US and host countries *j* (SK_{us}, SK_j).

Index of investment costs/barriers for host countries J, compiled by Markusen and Maskus from the annual World Competitiveness Report data of the World Economic Forum ($INVC_j$).

Index of trade costs/barriers for the US and for the host countries, compiled by Markusen and Maskus from the annual World Competitiveness Report data of the World Economic Forum (TC_{us}, TC_j).

Distance variable from the US for each host country ($DIST_j$).

NAFTA: dummy variable = 0 for Canada at/after 1989 and Mexico at/after 1994, = 1 for all other country/year observations.

The specification for the central regression equation comes from Carr *et al*., (2001), Markusen (2002), and Markusen and Maskus (2001, 2002). It is based on the knowledge-capital model of Markusen (2002) and allows for both horizontal and vertical motives for foreign affiliate production. This specification is as follows:

$$\begin{aligned} RSALEST_{jt} = {} & \alpha + \beta_1 * (GDP_{us} + GDP_j) + \beta_2 * (GDP_{us} - GDP_j)^2 \\ & + \beta_3 * (SK_{us} - SK_j) + \beta_4 * (GDP_{us} - GDP_j) * (SK_{us} - SK_j) \\ & + \beta_5 * INVC_j * NAFTA_j + \beta_6 * TC_j * NAFTA_j \\ & + \beta_7 * TC_{us} * NAFTA_j + \beta_8 * DIST_j \end{aligned}$$

Table 7.3 Summary statistics

Variable	*Obervations*	*Mean**	*Std. Deviation*
SUMGDP	815	7858.96	1595.35
GDPDIFF	815	7032.68	1470.75
GDPDIFFSQ	815	51600000.00	20800000.00
GDPDIFF*SKLDIFF	743	680.53	659.37
SKLDIFF	746	0.10	0.09
SKLDIFFSQ	746	0.02	0.02
INVCJ*NAFTA	790	37.23	13.82
INVCI*NAFTA	819	24.64	5.29
TCJ*NAFTA	815	31.61	16.55
TCI*NAFTA	814	26.15	7.56
DIST	819	8545.92	3860.81
RSALETOT	761	18364.14	29892.00
RSALEUS	754	10906.42	18241.52
RSALELOC	737	2641.59	8059.81
RSALEOTH	744	5070.32	8955.44

Notes
*Units of Measure:
GDPJ: Billions of 1995 US$
GDPI: Billions of 1995 US$
SKJ: Share of host country's labor force that is skilled; 0.0–1.0
SKI: Share of US labor force that is skilled; 0.0–1.0
INVCJ: Index of qualitative impediments to investment; 0–100, 100 = highest costs
TCJ: Index of qualitative barriers to trade in host country; 0–100, 100 = most restrictive
TCI: Index of qualitative barriers to trade in US; 0–100, 100 = most restrictive
DIST: Kilometers from host country capital to Washigton DC.

The NAFTA dummy on investment and trade costs is essentially making the extreme assumption that for US firms, investment costs and trade costs for US parents into Canada and Mexico fall to zero in 1989 and 1994, respectively. Not having this dummy makes the opposite extreme assumption, that US parents get no investment or trade-cost benefits from NAFTA. We estimated these equations both ways, and got much more plausible (and very different) results when using this NAFTA dummy. Results are shown in Tables 7.4–7.7, which are weighted least square with country fixed effects.

Since the formulation of the central estimating equation and the results have been discussed at length in previously published articles, we will instead use the material to formulate some comparative statistics on how host country characteristics affect US affiliate activity. Once we have the estimated coefficients, we take partial derivatives to estimate marginal effects (e.g. GDP variables appear three times, SK variables twice). Due to the interaction terms, comparative statisticcs on GDP and SK variables depend on the values of each other. We will try to clarify this as we go.

The comparative static results are presented in Table 7.8.

In the top panel, we present derived estimates (explained in the Appendix) of the elasticity of US affiliate sales with respect to the combined US and host

Table 7.4 Regression results for US outward foreign affiliate sales data

Dependent variable	RSALETOT
Total observations	703
Adjusted R^2	0.946

Variable	*Coefficient*	*T-Stat*	*P-Value*
SUMGDP	11.38	15.81	0.00
GDPDIFSQ	0.00	−7.13	0.00
SKDIFF	127467.30	4.61	0.00
GDPDIFF*SKDIFF	−16.99	−5.65	0.00
INVCJ*NAFTA	−230.48	−4.68	0.00
TCJ*NAFTA	69.46	1.89	0.06
TCI*NAFTA	−114.00	−2.03	0.04
DIST	−5.84	−30.24	0.00
INTERCEPT	56257.86	14.29	0.00

Table 7.5 Regression results for local sales of US foreign affiliate manufacturing

Dependent variable	RSALELOC
Total observations	698
Adjusted R^2	0.949

Variable	*Coefficient*	*T-Stat*	*P-Value*
SUMGDP	8.13	19.06	0.00
GDPDIFSQ	0.00	−10.83	0.00
SKDIFF	65817.97	4.00	0.00
GDPDIFF*SKDIFF	−7.56	−4.23	0.00
INVCJ*NAFTA	−69.91	−2.39	0.02
TCJ*NAFTA	25.24	1.16	0.25
TCI*NAFTA	−64.79	−1.94	0.05
DIST	−3.70	−32.30	0.00
INTERCEPT	26789.65	11.49	0.00

Table 7.6 Regression results for third country sales of US foreign affiliate manufacturing

Dependent variable	RSALEOTH
Total observations	690
Adjusted R^2	0.910

Variable	*Coefficient*	*T-Stat*	*P-Value*
SUMGDP	2.92	10.58	0.00
GDPDIFSQ	0.00	−3.75	0.00
SKDIFF	23486.69	2.20	0.03
GDPDIFF*SKDIFF	−4.81	−4.15	0.00
INVCJ*NAFTA	−29.80	−1.52	0.13
TCJ*NAFTA	12.10	0.86	0.39
TCI*NAFTA	20.06	0.91	0.36
DIST	−0.04	−0.27	0.79
INTERCEPT	−14380.89	−9.59	0.00

Table 7.7 Regression results for US sales of US foreign affiliate manufacturing

Dependent variable	RSALEUS
Total observations	685
Adjusted R^2	0.917

Variable	*Coefficient*	*T-Stat*	*P-Value*
SUMGDP	0.33	1.35	0.18
GDPDIFSQ	0.00	1.95	0.05
SKDIFF	30142.50	3.19	0.00
GDPDIFF*SKDIFF	−3.64	−3.54	0.00
INVCJ*NAFTA	−145.20	−8.39	0.00
TCJ*NAFTA	30.16	2.42	0.02
TCI*NAFTA	−56.92	−2.92	0.00
DIST	−2.56	−22.07	0.00
INTERCEPT	44476.42	33.62	0.00

Table 7.8 Economic interpretations of the econometric results

Panel A: Elasticities of US manufactuirng affiliate sales in country j with respect to SUMGPD (combined size of US and host country)

	Total sales	*Local sales*
No change in size difference	4.872	24.178
All income growth in host-country	13.110	8.900
All income growth in the United States	9.640	7.350

Panel B: Proportional change in US foreign affiliate sales given a one-percentage-point increase in the share of the host labor force that is skilled

	Total sales	*Local sales*	*Sales to US*	*Sales to third*
Evaluated at mean US-host size difference	−0.004	−0.040	−0.003	0.023
Evaluated as host size goes to zero	0.001	−0.032	−0.002	0.025

Panel C: Proportional change in affiliate sales given a one-standard deviation fall in trade and investment costs

	Investment costs	*Trade costs*	*Both*
Total sales	0.173	−0.063	0.111
Local sales	0.366	−0.158 n	0.208
US sales	0.184	−0.046	0.138
Third-country sales	0.081	−0.040 n	0.042

Note
n: not significant.

country market size, SUMGPD = GDP_{us} + GDP_j. The result depends on which country is growing so we present the results three ways: each country grows the same amount so that GDPDIFF = GDP_{us} − GDP_j is constant, all growth is in the host country, and all growth is in the United States. The elasticity calculations are done for both total sales and local sales.

The calculated elasticities are in the top panel of Table 7.8 and are quite large, adding further weight to the horizontal/market-seeking model of the multinational. Values in excess of one are predicted by Markusen's knowledge–capital model with plant-level scale economies. A growth in host-country market size expands multinational production at both the intensive and extensive margin. At the intensive margin, existing affiliates expand production roughly in proportion to market size, giving an elasticity of one, but pro-competitive, firm-size effects boost that figure. At the extensive margin, new affiliates in existing and new industries enter, finding it profitable to switch from exporting to local production. It must be emphasized again that such numbers are not predicted by any vertical or factor-proportions approach to the multinational.

Finding 3: US outward manufacturing affiliate production is not unskilled-labor seeking

The Heckscher–Ohlin approach outlined earlier tends to lump all types of labor together and assume labor and capital are the two factors of production in an economy, closed or open. If instead the two factors were skilled and unskilled-labor, then the HO approach would suggest that US affiliates should be drawn to unskilled-labor-abundant countries. This certainly also appears to be the assumption of the antiglobalists: that outward multinational activity is costing the US fewer skilled jobs and is taking advantage of lax labor laws in developing countries. However, this view is no longer universal insofar as there has been a lot written in the business press about the outsourcing (meaning in fact offshoring) of skilled jobs in information technology and business-process outsourcing.

Our estimates show that US manufacturing affiliate production has essentially a zero response to host-country skilled-labor measures. Pol Antras (2005) has found a different result using labor and capital: the more capital abundant the host country the higher the level of outward affiliate activity. Interestingly, we find a skilled-labor-seeking effect for affiliate production exported to third countries but not for local sale, but both effects are tiny. Once again, both some conventional economist's wisdom and some street wisdom that multinationals should be unskilled-labor seeking do not seem consistent with the evidence.

The effects of skilled-labor endowments are shown in the middle panel of Table 7.8. Since the skilled-labor variable is a share (0.00 to 1.00) of the labor force, we are reporting "quasi-elasticities," the proportional response of affiliate sales to a one-percentage-point increase in the skilled-labor share. Referring

back to the regression equation, SK_j appears twice in the equation, so these quasi-elasticities also depend on the size difference between the United States and the host country. We present two different results in Table 7.8 (middle panel), the first using the mean value of GDPDIFF = $(GDP_{us} - GPD_j)$ and the allowing the size of the host country to go to zero (GDPDIFF = GDP_{us}). Results allowing the host size to go to zero show that US manufacturing affiliate sales are skilled-labor-seeking, although this effect is near zero.

Overall, the results suggest that skill composition is less important for local sales than for export sales. This makes sense in that there are not good alternatives to producing locally other than exporting from the United States. For export sales on the other hand, host countries compete with other host countries and thus US affiliate location decisions become more sensitive to country characteristics such as skill composition. But again, all effects are tiny, essentially zero in economic terms.

Finding 4: effects of host-country liberalization on inward affiliate production

Finally, we turn to the effects of the host-country trade and investment cost variables. We find two results.

1 A fall in host-country investment costs raises US affiliate activity and a fall in host-country trade costs lowers US affiliate activity.
2 A one-standard-deviation fall in both host investment and trade costs has a significant positive effect on total US affiliate production, and this effect is much higher for local sales and exports to the United States than it is for third country exports.

Results for the effects of changes in investment and trade costs are presented in the bottom panel of Table 7.8. The trade and investment cost indices are from survey data, which we feel create two problems of quantification. First, it is hard to interpret percentage changes in these variables, since the index has no strong cardinal properties. Second, it is hard to compare the trade and investment cost data. What we do therefore, is consider the proportional changes in affiliate sales following a one-standard-deviation fall in each of them. We feel that this helps in comparing the effects of the two types of costs on US outward affiliate activity.

The results from these estimations indicate that an increase in investment costs lowers affiliate sales. While at first glance this may seem rather obvious, there is some theory that suggests it is not. In particular, we might expect that the scale of a multinational's involvements with a country might fall with higher investment costs, but there may also be a substitution effect toward using owned subsidiaries away from arm's length options like licensing or subcontracting in a more secure host environment. If this substitution effect outweighs the scale effect, then the coefficient on investment costs could be positive. Our results indicate that this is

not the case even with the assumption that NAFTA membership reduces the investment costs to zero.

Results for the effects of trade costs are presented in the bottom panel of Table 7.8 and indicate that a increase in trade costs raises affiliate activity. This is also not obvious from classical theory, yet is consistent with the knowledge capital model's theory concerning horizontal FDI. Indeed, vertical models of direct investment in which intermediates are imported and output exported back to the parent suggest that an increase in trade costs should decrease affiliate activity. The robust positive sign on host trade costs in our results gives one more indication that horizontal motives dominate in US outward manufacturing investment.

Now consider a *fall* in trade and investment costs. The right-hand column of the bottom panel of Table 7.8 adds the (negative of the) two coefficients together; that is, it gives the effect of a combined reduction in investment costs and trade costs of one standard deviation each. The coefficient for the changes in total sales, 0.111 is definitely significant in economics terms (e.g. a $3.8 bn increase in US affiliate sales in Brazil for 2003). There is evidence of a positive effect on overall affiliate sales from this combined liberalization due to the large increase in sales attributed to a reduction in investment costs.

Summary and conclusions

The purpose of the chapter is to review recent empirical findings on multinational production, particularly foreign manufacturing affiliates of US multinationals, and compare these results to ideas advanced in the globalization debate and also found in some traditional economic theory. We argue that some traditional theory and positions of antiglobalists are similar from the point of view of positive theory, but generate quite different normative interpretations from these two groups. Our chapter sticks to positive theory, and argues that both groups misrepresent the principal activities of US manufacturing affiliates abroad.

Four principal points are made. First, US multinationals' investments in manufacturing affiliates and indeed investments from all parent countries flow primarily to high-income countries. A traditional view that foreign direct investment is just a capital flow from capital-abundant to capital-scare countries receives little support in the data. That portion of FDI that does flow from high-income to developing countries flows primarily to the most advanced of the developing countries. These results are consistent with the lack of empirical support for a factor proportions trade model and additionally refute the antiglobalists claim that North–South trade is dominant, hurting US jobs and taking advantage of foreign workers.

Second, US outward FDI is primarily horizontal or "market-seeking." The view that multinationals are moving jobs abroad to produce cheaply to serve the US market is a very minor part of US affiliates production: in our data set affiliate sales back to the US account for only 12% of total sales. There is a significant portion of sales (28%) that goes to third countries. However, examination of this

indicates that much of this is exports of US affiliates located in one European country to the rest of the European Union. We view these as essentially horizontal investments to serve the EU market, but a particular location is chosen on the basis of various cost considerations such as labor unions, taxes, and government regulations among others.

Third, we find that US outward manufacturing affiliate production is not unskilled-labor seeking. In a Heckscher–Ohlin world view, the flip side of the idea that multinationals are moving capital to capital-scare countries is that they are moving capital to low-wage, labor-abundant countries. Of course, labor comes in many grades, but at very least our results refute the idea that multinationals are low-skilled labor seeking. Perhaps it is in the nature of what multinationals do that they are primarily involved in skilled-labor intensive activities, and some of these are located abroad to serve local markets.

Fourth, we find that a fall in host country investment costs raises US affiliate activity and a fall in host-country trade costs lowers US affiliate sales. The latter result is yet more evidence that outward FDI is primarily horizontal or market-seeking. The former result seems obvious, but in fact investment costs may cause a substitution of arm's length modes of entry to internalization within the firm via FDI. Our results indicate that to the extent this substitution effect occurs, it is dominated by a scale effect reducing investment.

In this same section, we note that a one-standard-deviation fall in both host investment and trade costs has a significant and positive effect on total US affiliate production. This is a finding that should be of interest to developing countries with high trade barriers that fear that liberalization will result in disinvestment by market-seeking firms. This was a worry in Canada prior to the US–Canada free-trade agreement and a worry in Mexico prior to NAFTA. These fears did not materialize, perhaps in part because in both countries liberalization went far beyond simple reductions in trade barriers.

We conclude again by acknowledging the caveat that we are concerned about activities that take place within the ownership boundaries of the firm. There may indeed be valid issues about the movement of low-skilled activities to third-world "sweatshops" in industries like clothing and footwear, but the antiglobalists need to understand that these are owned and managed either by local entrepreneurs or are subsidiaries of third-country firms which are generally not North American or European. Perhaps more progress can be made toward a better world if we begin with a proper understanding of the underlying reality.

Appendix 1

Data

Data for the estimation form a panel of cross-sectional observations for 39 countries over the period 1984–2003. The dependent variable of real sales volume of US nonbank manufacturing affiliates is used to indicate productive activity. These

sales data are reported annually by the Bureau of Economic Analysis and broken down into sales to the local market, exports back to the parent country, and exports to third countries. Annual sales are reported in millions of 1995 US$ and are converted from local currencies using exchange rates from the *International Financial Statistics* (IFS) publication of the International Monetary Fund. Real gross domestic product is measured in billions of 1995 US$ for each country in the panel. Therefore, the annual real GDP values in local currencies were converted into US$ using the IFS exchange rates and then the US GDP deflator for that year was applied in order to obtain the 1995 US$ figure.

The variables for skilled labor abundance are defined as the total sum of employment in each country in occupational categories 0/1 (professional, technical, and kindred workers) and two (administrative workers), as defined in the *Yearbook of Labor Statistics*, divided by total employment in each country. The *Yearbook of Labor Statistics* is published annually by the International Labor Organization. There were a few missing observations for this variable within the panel and in these cases the country's period average skilled-labor ratio was used. The skilled-labor ratios are given a value between 0.0 and 1.0 and are used to compute the Skill Difference variable (SKDIFF) for the estimation by taking the value for the skilled-labor ratio for the United States and subtracting the value for the skilled-labor ratio for the host country.

In the process of updating the data on foreign affiliate sales from 2000 to 2003 there were changes in the ILO classifications of occupations used to construct the skilled labor ratios.The ILO adjusted their occupation classification system from the 1968 categories which were the following:

1 Professional, technical, and related workers;
2 Administrative and managerial workers;
3 Clerical and related workers;
4 Sales workers;
5 Service workers;
6 Agriculture, animal husbandry and forestry workers, fishermen and hunters;
7/8/9 Production and related workers, transport equipment operators and laborers;
10 Workers not classifiable by occupation.

To a new set of classifications which are

1 Legislators, senior officials, and managers;
2 Professionals;
3 Technicians and associate professionals;
4 Clerks;
5 Service workers and shop and market sales workers;
6 Skilled agricultural and fishery workers;
7 Craft and related trade workers;

8 Plant and machine operators and assemblers;
9 Elementary occupations;
10 Armed forces.

In creating the skilled-labor share variables the initial dataset uses the ratio of employment in 0/1 and 2 classifications to total employment. In updating the data with the new classifications, subcategories were cross referenced in order to capture the same occupational categories. The skilled-labor variables for years after 2000 use the ratio of employment in categories 1, 2 and 3 to total employment.

The cost of investing in the host country is calculated by taking the average of indices of perceived barriers to investment that are reported annually in the *World Competitiveness Report* of the World Economic Forum (WEF). These indices are based on survey data and are therefore qualitative rather than quantitative. We feel the variables (survey questions) that represent impediments to investment include restrictions on the ability to acquire control in a domestic company, limitations on the ability to employ foreign skilled labor, restraints on negotiating joint ventures, rigid control over hiring and firing practices, market dominance by a small number of firms, an absence of fair judicial administration, difficulties in acquiring local bank credit, restrictions over access to local and foreign capital markets, and inadequate protection of intellectual property. These indices are summed and the average is taken to represent the cost of investment in each country. The resulting value is between zero and 100 with higher numbers representing higher costs of investing.

The indices for trade costs in the host country and for the United States are calculated in a similar manner. Indices on the efforts to prevent the importation of competitive products and perceived measures of national protection are taken from the *World Competitiveness Report* and the average is then used as the trade cost index used in this analysis. The values range from zero to 100, with a value of 100 representing a virtually closed economy with exorbitantly high trade costs. TCI, the trade cost variable for the United States only varies across time in the estimation. Due to the qualitative nature of these indices, the regression coefficients represent partial effects of a change in the perceived costs of investing and trading, as documented by the WEF's extensive surveys of multinational enterprises.

In the update of the data from 2000 to 2003 the World Competitiveness Report is now known as the Global Competitiveness Report and the survey questions have been greatly expanded from the previous versions, resulting in having to readjust the investment cost index and trade cost index.

For the investment cost index, it was necessary to use a combination of different survey questions from the Global Competitiveness Report (GCR) for each question used for 1984–2000. In many cases, 2001 was a transition year for the GCR and therefore the questions are not necessarily the same for 2001–2004. Each original question is listed below in bold and then the approximating questions for 2001–2004 are given with the weights and scaled values. There are some questions that did not change at all and those are also identified.

The original investment cost index is calculated from questions that measure: restrictions on the ability to acquire control in a domestic company; an absence of fair administration of justice; restrictions on access to foreign capital markets; limitations on the ability to employ foreign skilled labor; controls on hiring and firing practices; protection of intellectual property; restraints on negotiating joint ventures; restrictions on access to local capital markets; difficulties in acquiring local bank credit; market dominance by a small number of enterprises. The values for these are rescaled to a common 1–10 scale and are summed and then subtracted from 100 with higher values reflecting higher investment costs.

1 Competition laws do not prevent unfair competition? 1 = do prevent 10 = do not prevent

2001 Corporate activity is 1 = dominated by a few groups, 7 = spread among many firms. Competition in the local market in most industries is 1 = limited and price-cutting is rare, 7 = intense and market leadership changes over time.
Entry of new competitors, 1 = almost never occurs in the local market, 7 = is common in the local market.
Antimonopoly policy is 1 = lax, 7 = effectively promotes competition

2002 Corporate activity is 1 = dominated by a few groups, 7 = spread among many firms. Competition in the local market in most industries is 1 = limited and price-cutting is rare, 7 = intense and market leadership changes over time.
Antimonopoly policy is 1 = lax, 7 = effectively promotes competition

2003 Corporate activity is 1 = dominated by a few groups, 7 = spread among many firms. Competition in the local market in most industries is 1 = limited and price-cutting is rare, 7 = intense and market leadership changes over time.
Antimonopoly policy is 1 = lax, 7 = effectively promotes competition.

2004 Antimonopoly policy is 1 = lax, 7 = effectively promotes competition.
Market dominance by a few enterprises is 1 = common, 7 = rare.

For each year the values are weighted equally and summed, and then rescaled for a 1–10 scale.

2 Cross-border ventures can/cannot be negotiated without government imposed restraint. 1 = cannot, 10 = can

2001 Approximately how many permits would you need to start a new firm? (range: 2–10)
International distribution and marketing from your country, 1 = takes place through foreign companies, 7 = is owned and controlled by local companies.

2002 Number of administrative procedures required to register a business (range: 2–20)

Number of procedures required to resolve a contract dispute (range: 10–47)
International distribution and marketing from your country, 1 = takes place through foreign companies, 7 = is owned and controlled by local companies.

2003 Government intervention in corporate investment is 1 = distorting, 7 = nonexistent
Starting a new business is 1 = extremely difficult and time consuming, 7 = easy.
International distribution and marketing from your country, 1 = takes place through foreign companies, 7 = is owned and controlled by local companies.

2004 What is the impact of your country's rules governing FDI on your business, 1 = damaging, 7 = beneficial.
Starting a new business is 1 = extremely difficult and time consuming, 7 = easy.

International distribution and marketing from your country, 1 = takes place through foreign companies, 7 = is owned and controlled by local companies.

This was the most difficult question to try and capture. First, the answers are all rescaled to a 1–10 scale. For 2001 and 2002, the values are weighted equally and summed and subtracted from 10. For 2003 and 2004, the values are weighted equally but the third is subtracted from 10 before weighting and then the values are summed.

3 Credit does/does not flow easily from banks to business. 1 = does not, 10 = does flow easily

2001–2004 Obtaining credit for your company has become 1 = more difficult, 7 = easier, over the last year.
How easy is it to obtain a loan in your country with only a good business plan? 1 = impossible, 7 = easy.

The values are rescaled to a 1–10 scale, are equally weighted, and are summed.

4 Access to foreign capital markets is/is not restricted for domestic companies. 1 = is restricted, 10 = is not restricted

2001 Citizens who wish to invest in stocks and bonds and/or open bank accounts in other countries are 1 = prohibited from doing so, 7 = free to do so.

2002–2004 Components and parts are obtained by 1 = almost always imported, 7 = always available locally.
Process machinery is obtained by 1 = almost always imported, 7 = always available locally.

The values are rescaled to a 1–10 scale, are equally weighted, and are summed. For 2002–2004 the values are then subtracted from 10. This measure is far from perfect but because it is only one of nine criteria in the investment cost index it will have to do for now.

5 Access to local capital markets is/is not restricted for foreign companies. 1 = is restricted, 10 = is not restricted

2001 Foreign investors who wish to invest in stocks and bonds and/or open bank accounts in your country are 1 = prohibited from doing so, 7 = free to do so.

2002–2004 Venture capital is available to entrepreneurs with innovative projects. 1 = not true, 7 = true.
Raising money on the local stock market is 1 = nearly impossible, 7-quite possible for a good company.

The values are rescaled to a 1–10 scale, are equally weighted, and are summed.

6 There is no/full confidence in the fair administration of justice in society. 1 = no confidence, 10 = full confidence

2001–2004 The judiciary is independent and not subject to interference by the government and/or parties to disputes. 1 = not true, 7 = true.

The values are rescaled to a 1–10 scale.

7 Foreign investors may/may not acquire control in a domestic company. 1 = may not, 10 = are free to do so

2001 no relevant questions, used an average between the 2000 and 2002 values.

2002–2004 Foreign ownership is 1 = rare and prohibited in key sectors, 7 = prevalent and encouraged.

The values are rescaled to a 1–10 scale.

8 Immigration laws prevent/do not prevent employing foreign skills. 1 = prevent, 10 = do not prevent

2001 Scientists and engineers in your country are 1 = nonexistent or rare, 7 = widely available.
Scientific research institutions are 1 = nonexistent, 7-best in their field.

2002–2003 Specialized research and training services are 1 = not available, 7-available from world class organizations.
Scientific research institutions are 1 = nonexistent, 7-best in their field.

2004 Labor regulations in your country 1 = prevent your company from employing foreign labor, 7 = do not prevent your company from hiring foreign labor.

The values are rescaled to a 1–10 scale, are equally weighted, and are summed. For 2001–2003, the values are subtracted from 10.

9 The questions for intellectual property protection and hiring and firing practices did not change after 2000 in the GCR and therefore no rescaling was necessary.

Pre-2000 data for the trade index was based on one question from the World Competitiveness Report (now the Global Competitiveness Report). This question read: "National protectionism prevents/does not foreign products and services from being imported" and was measured on a scale of 1–10 with 1 being the most restrictive. This number was multiplied by 10 and then subtracted from 100 to get the value for the trade cost index, with 100 being the most restrictive.

The questions from year to year changed after 2000 for some categories and this affected the trade cost index and the investment cost index. The questions below were used to measure the trade costs for the years 2001–2004.

2001 Hidden trade barriers are an important problem (1) or not an important problem (7).
This number was rescaled to a 1–10 scale and then multiplied by 10 and subtracted from 100.

2002 Hidden trade barriers are an important problem (1) or not an important problem (7).
The cost of importing foreign equipment: fees, tariffs, etc. raise the cost by less than 10% (1), 11–20% (2), 21–30% (3), 31–40% (4), 41–50% (5), 51–60% (6), 61–70% (7), 71–80% (8), 81–100% (9).

2003 Hidden trade barriers are an important problem (1) or not an important problem (7).
The cost of importing foreign equipment: fees, tariffs, etc. raise the cost by less than 10% (1), 11–20% (2), 21–30% (3), 31–40% (4), 41–50% (5), 51–60% (6), 61–70% (7), 71–80% (8), 81–100% (9).

2004 Hidden trade barriers are an important problem (1) or not an important problem (7).
The cost of importing foreign equipment: fees, tariffs, etc. raise the cost by less than 10% (1), 11–20% (2), 21–30% (3), 31–40% (4), 41–50% (5), 51–60% (6), 61–70% (7), 71–80% (8), 81–100% (9).

For 2002–2004, both values for the questions are rescaled to 1–10 scales and then are equally weighted by .5 and summed before being multiplied by 10 and subtracted from 100. The distance variable is given as the number of kilometers from each country's capital to Washington DC and theory dictates that it will have a negative relationship with investment and trade costs. However, because the distance variable interacts with more that one variable it is difficult to determine if it is primarily capturing trade costs or investment costs. It is important to note here that the survey questions that make up the basis for these indices do not include any questions about distance and therefore correlation between these variables is not an issue.

Table 7.3 lists the summary statistics for all the variables used in the estimation and also gives the units of measure for each of the variables. After rows of data with missing observations are eliminated, the panel used in the regression estimations

contains 549 observations. The estimation results for each dependent variable are given in Tables 7.4–7.7. To contend with the issue of different units of measure for the variables, the estimations are done using weighted least squares (WLS). In addition, distinguishing the time series contributions to the results requires incorporating country specific fixed effects into the regression equations. The country fixed effects only apply to the host countries due to the fact that a dummy for the United States in the estimation would result in perfect collinearity. We do not report the coefficients on the country specific dummy variables but most are significant.

Appendix 2

Data and calculations

$$\begin{aligned}\text{RSALEST}_{jt} = \alpha &+ \beta_1{*}(\text{GDP}_{us} + \text{GDP}_j) + \beta_2{*}(\text{GDP}_{us} - \text{GDP}_j)^2 \\ &+ \beta_3{*}(\text{SK}_{us} - \text{SK}_j) + \beta_4{*}(\text{GDP}_{us} - \text{GDP}_j){*}(\text{SK}_{us} - \text{SK}_j) \\ &+ \beta_5{*}\text{INVC}_j{*}\text{NAFTA}_j + \beta_6{*}\text{TC}_j{*}\text{NAFTA}_j \\ &+ \beta_7{*}\text{TC}_{us}{*}\text{NAFTA}_j + \beta_8{*}\text{DIST}_{j.}\end{aligned}$$

Consider the elasticity of total sales (RSALEST) with respect to SUMGDP. We do this first under the assumption that both counties have the same absolute increase in GDP, so the change in GDPDIFF is 0.

(A) no change in GDPDIFF

$$\frac{\partial\ \text{RSALEST}_j}{\partial\ \text{SUMGDP}} = \beta_1 \quad \frac{\partial\ \text{RSALEST}_j/\text{RSALEST}_j}{\partial\ \text{SUMGDP}/\text{SUMGDP}} = \beta_1 \frac{\text{mean(SUMGDP)}}{\text{mean(RSALEST}_j)}$$

Second, assume that all the increase in SUMGDP is in the host-country's GDP.

(B) all increase in country *j*: the change in GDPDIFF w.r.t GDPJ equals −1.

$$\begin{aligned}\frac{\partial\ \text{RSALEST}_j}{\partial\ \text{SUMGDP}} &= \beta_1 + \beta_2{*}2{*}\text{mean(GDPDIFF)}{*}(-1) \\ &+ \beta_4{*}\text{mean(SKDIFF)}{*}(-1)\end{aligned}$$

Third, assume that all the increase in SUMGDP is in the United States's GDP.

(C) all increase in the US: the change in GDPDIFF w.r.t GDPUS is +1.

$$\begin{aligned}\frac{\partial\ \text{RSALEST}_j}{\partial\ \text{SUMGDP}} &= \beta_1 + \beta_2{*}2{*}\text{mean(GDPDIFF)}{*}(-1) \\ &+ \beta_4{*}\text{mean(SKDIFF)}{*}(+1)\end{aligned}$$

Now consider the elasticity of total sales with respect to the host-country's skilled-labor share, SKJ:

(A) at mean GDPDIFF

$$\frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{SK}_j} = \beta_3 + *(-1) + \beta_4 *\mathrm{mean(GDPDIFF)}*(-1)$$

The "quasi-elasticity" is defined as the proportional response sales to a one-percentage-point increae in SKJ. This is given by a Taylor's expansion around the initial value, but for a very small change in SKJ we can safely ignore the second-order terms (the latter involves the square of the change 0.01 change in SKJ, or about 0.0001).

$$\frac{\Delta\ \mathrm{RSALEST}_j/\mathrm{RSALEST}_j}{\Delta\ \mathrm{SKJ}}\Delta\mathrm{RSALEST}_j$$

$$= \frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{SK}_j}\Delta\ \mathrm{SKJ} + (\text{higher-order-terms})$$

$$\frac{\Delta\mathrm{RSALEST}_j}{\Delta\mathrm{RSALEST}_j} = \frac{\partial\mathrm{RSALEST}_j}{\partial\mathrm{SK}_j}\frac{(0.01)}{\mathrm{mean(RSALEST}_j)}\Delta\mathrm{SKJ} = 0.01$$

(B) as country *j* becomes small: we don't have mean(GDP_{us}) printed out, so I used the formula $\mathrm{GDP}_{us} = ((\mathrm{GDP}_{us} + \mathrm{GDP}_j) + (\mathrm{GDP}_{us} - \mathrm{GDP}_j))/2 = (\mathrm{SUMGDP} + \mathrm{GDPDIFF})/2$ thus the formula for the derivative replaces mean(GDPDIFF) with mean(GDP_{us}) as GDP_j goes to zero.

$$\frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{SK}_j} = \beta_3 + *(-1) + \beta_4 * \mathrm{mean(GDP}_{us})*(-1)$$

Proportional changes in affiliate sales with respect to a one standard-deviation fall in INVC_j and TC_j . These are more straightforward.

$$\frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{INVC}_j} = \beta_5$$

$$\Delta\mathrm{RSALEST}_j = \frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{INVC}_j}\Delta\mathrm{INVC}_j + (\text{higher-order-terms})$$

$$\frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{RSALEST}_j} = \frac{\partial\ \mathrm{RSALEST}_j}{\partial\ \mathrm{INVC}_j}\frac{-\ \mathrm{standev(INVC}_j)}{\mathrm{mean(RSALEST}_j)} = \beta_5\frac{-\ \mathrm{standev(INVC}_j)}{\mathrm{mean(RSALEST}_j)}$$

A similar proceed is used to calculate the effects of a one standard-deviation fall in TC_j . Derivatives of local sales, exports sales to the United States, and export sales to third countries are defined and derived in a similar fashion.

Bibliography

Antras, Pol (2005) "Property Rights and the International Organization of Production," *American Economic Review Papers and Proceedings*, 95: 25–32.

Bhagwati, Jagdish (1971) "The Generalized Theory of Distortions and Welfare" in *Trade, Balance of Payments and Growth: Essays in Honor of Charles P. Kindleberger*, Amsterdam: North Holland, 69–90.

Bhagwati, Jagdish, and V. K. Ramaswami (1963) "Domestic Distortions, Tariff, and the Theory of Optimum Subsidy," *Journal of Political Economy*, 71: 44–50.

Bhagwati, Jagdish, V. K. Ramaswami, and T. N. Srinivasan (1969) "Domestic Distortions, Tariffs, and the Theory of Optimum Subsidy," *Journal of Political Economy*, 77: 1005–1010.

Blonigen, Bruce A, Ronald B. Davies, and Keith Head (2003) "Estimating the Knowledge-Capital Model of the Multinational Enterprise: Comment," *American Economic Review*, 93: 980–994.

Blonigen, Bruce A, Ronald B. Davies, Glen R. Waddell, and Helen Naughton (2007) "FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment," *European Economic Review*, 51: 1303–1325.

Brainard, S. Lael (1993) "A Simple Theory of Multinational Corporations and Trade with a Trade-off between Proximity and Concentration," NBER working paper No. 4269.

Brainard, S. Lael (1997) "An Empirical Assessment of the Proximity-Concentration Tradeoff between Multinational Sales and Trade," *American Economic Review*, 87: 520–544.

Carr, David L., James R. Markusen and Keith E. Maskus (2001) "Estimating the Knowledge-Capital Model of the Multinational Enterprise," *American Economic Review*, 91: 693–708.

Ekholm, Karolina, Rikard Forslid and James R. Markusen (2007) "Export-Platform Foreign Direct Investment," *Journal of the European Economics Association*, 5: 776–795.

Helpman, Elhanan (1984) "A Simple Theory of Trade with Multinational Corporations," *Journal of Political Economy*, 92: 451–471.

Horstmann, Ignatius J. and James R. Markusen (1987) "Strategic Investments and the Development of Multinationals," *International Economic Review*, 28: 109–121.

Horstmann, Ignatius and James R. Markusen (1992) "Endogenous Market Structures in International Trade," *Journal of International Economics*, 20: 225–247.

Markusen, James R. (1984) "Multinationals, Multi-Plant Economies, and the Gains from Trade," *Journal of International Economics*, 16: 205–226.

Markusen, James R. (2002) *Multinational Firms and the Theory of International Trade*, Cambridge: MIT Press.

Markusen, James R. and Anthony J. Venables (1998) "Multinational Firms and the New Trade Theory," *Journal of International Economics*, 46: 183–203.

Markusen, James R. and Anthony J. Venables (2000) "The Theory of Endowment, Intra-Industry and Multinational Trade," *Journal of International Economics*, 52: 209–234.

Markusen, James R. and Keith E. Maskus (2001) "Multinational Firms: Reconciling Theory and Evidence," in Magnus Blomstrom and Linda Goldberg (eds), *Topics in Empirical International Economics: A Festschrift in Honor of Robert E. Lipsey*, Chicago: University of Chicago Press, 71–95.

Markusen, James R. and Keith E. Maskus (2002) "Discriminating among Alternative Theories of the Multinational Enterprise," *Review of International Economics*, 10: 694–707.

Shatz, Howard J. (2004) "US Multinational Affiliate Exports from Developing Countries," *Journal of Economic Geography*, 4: 323–344.

8 Who makes the rules of globalization?

Corporate influence in global and regional trade agreements

Alan V. Deardorff

Introduction

In this chapter I argue that profit maximizing firms, even though they contribute to social welfare when they compete in the market, may not do so when they influence the political process. In particular, I suggest, through several examples from both the real world and from economic theory, that corporations have played a significant role in the formulation of the rules of the international trading system. They did this in the formation of the World Trade Organization (WTO) where they were responsible for the expansion to cover both intellectual property and services. And they do this in preferential trading arrangements such as the North American Free Trade Agreement (NAFTA), where they inserted the notorious Chapter 11 and specified rules of origin for automotive products. All of this is quite consistent with economic theory, including the literature on the political economy of trade policy. I also use a simple duopoly model to illustrate a domestic firm's interest in setting rules of origin. The corporate influence on rules need not be bad, but there is no reason why it should be good either, as these examples illustrate.

When economists think of globalization, most of us think first of free trade and therefore the gains from trade that we have been studying and teaching for two centuries. But actual globalization takes many forms, including not only trade but also foreign direct investment (FDI), financial capital flows, and sometimes even migration of labor. More important, even the liberalization of trade has not been the simple move to zero trade barriers that we understand from our models, but has taken the form of trade liberalization that was only partial. Some of this is multilateral, but with barriers only partially eliminated. And increasingly it is "minilateral," with barriers removed more completely but only between pairs, or among small groups, of countries in free trade areas (FTAs) and the like.

Furthermore, in both of these cases the liberalization of trade typically is attached to a variety of other measures that may or may not generate the same sorts of benefits as reducing tariffs. Multilaterally, the World Trade Organization has expanded beyond the General Agreement on Tariffs and Trade (GATT) to include services and intellectual property, as well as to address policies relevant to trade other than trade barriers. Minilaterally, FTAs routinely address issues other than trade, including not only investment but also labor standards and environment, the

effects of which may be very different from tariffs. And even for trade, because FTAs are only partial liberalization, they have features that may reduce or even reverse the beneficial effects of free trade. The importance of all of this is that, while we have a solid understanding from economic theory and experience of the benefits of multilateral free trade, these actual modes of globalization are something else.

Many of those who oppose or are skeptical of globalization—who tend not to be economists—base their doubts fundamentally on suspicion of corporations. Free trade increases the ability for large corporations to operate across national borders, thus—the skeptics would say—increasing their power over the economy and over peoples and governments around the world. Because corporations pursue only their self-interest and not the social good, they exploit the world for their own profit. Thus Globalization ≡ Corporate Power ≡ Everybody Else Loses!

As economists we take issue with the second of these identities, and perhaps also with the first. Economic theory tells us, since Adam Smith, that firms, pursuing their self-interest but competing with one another, lead, under ideal conditions at least, to maximized well-being for society as a whole. And indeed their competition among themselves reduces and, again under ideal conditions, even eliminates their own profits. So the fact that corporations are controlling economic activity does not, as long as they compete with each other or with smaller firms, mean that society loses. On the contrary, it is the society that gains, and the corporations themselves find their profits competed away. Indeed, by this argument globalization itself may reduce corporate power by forcing large firms to compete with each other across borders, rather than allowing each to enjoy market power behind its home country's trade barriers.

This would be fine if in fact globalization did entail simply the move to free trade. But as I have said, actual globalization includes both more and less than free trade. One implication of this is that the welfare theorems of trade theory need not apply, since their conditions are not met. This is the meaning of the Theorem of the Second Best, where partial moves toward free trade may be welfare reducing if other markets remain distorted. I am not too worried about that, however, except perhaps in cases where the importance of particular distortions and their implications for gains from trade are well established. I certainly would not in general want to reject partial trade liberalization based on just the vague fear that it may not be welfare improving. Distortions, after all, do not only undermine the benefits of some moves toward liberalization; they also increase the benefits of others.[1]

My greater concern is that the particular ways that the world has embraced globalization have been selected for it not by obviously unbiased and high-minded academics, but rather by governments acting under the influence of special interests— often corporate interests. That is, governments have negotiated both multilateral trade treaties and minilateral preferential arrangements in response to, and in the presence of, pressures brought to bear upon them by domestic interests. These interests could take a variety of forms, and in principle they could represent very broad constituencies. But in practice it seems clear that corporate

interests have so far dominated. And while economic theory is reassuring about the effects of the profit motive when it drives behavior in the marketplace, there is no such reason to trust its effects in the political arena. In short, my concern in this chapter is that the influence of producers, especially of large corporations, has dominated the drafting of the texts of international economic agreements and has done so, understandably, so as to promote the interests of those producers.[2] Their interests are not necessarily harmful to society, but they are not necessarily beneficial either. And there is no reason to expect an invisible hand to guide governments that are under corporate influence to move toward desirable outcomes.

In the remainder of this chapter I will first elaborate on some particular examples of corporate influence that are suggested by casual observation. Most of these are ones that I mentioned when I first began to worry about this issue in Deardorff (2003). I will then turn to looking at what economic theory can suggest about this process in the context specifically of FTAs. I will first review the implications of some of the existing literature, then provide my own simple example of how producer interests might seek to determine rules of origin.

Episodes of corporate influence

The "rules" of globalization might be thought to take many forms at many levels, including import tariffs and the procedures for collecting them, the nontariff barriers that still restrict trade in many sectors, and the national laws that specify these restrictions. However, I will focus here only on the rules that are embodied in international trade agreements, and specifically on the WTO and NAFTA.

TRIPs

I will start with what to me is the most egregious example of corporate influence on international rulemaking: the Uruguay Round Agreement on Trade Related Intellectual Property Rights (TRIPs). When issues of intellectual property rights (IPRs) first began to be discussed as part of the Uruguay Round, the claim was that any new rules would be confined to their "trade related" aspects. That presumably referred to the entry of IPR-violating goods into international trade, especially the export of such goods into a country where the IPR was established and enforced. It is not clear that new rules were needed for that, but in any case the negotiations soon left that narrow problem behind. The resulting TRIPs agreement requires that all WTO member countries enact and enforce IPR laws that are roughly comparable to those in the developed countries. The effect of this, since most IPRs are owned by developed-country firms, has been to force developing countries to protect the IPRs of individuals and firms in the developed countries. And this, aside from any stimulus it might provide to the creation of intellectual property of particular use to developing countries, is a change that can only benefit rich-country owners of IPRs and harm poor-country consumers. A case can of course be made for IPR protection as a second-best means of stimulating innovation. But I have argued in Deardorff (1990, 1992) that, even on efficiency

grounds, extending IPR protection to the whole world is going too far. And developing countries—notably India—resisted this effort to bring IPRs into the rules of the world trading system.

So why did this effort succeed? Clearly because of corporate influence. A group of very large corporations especially in the pharmaceutical industry were the initiators of the push for TRIPs in these negotiations, and they played an active role in moving the negotiations along. It was they, after all, who depended most heavily on the defense of IPRs, since their products, once developed, can be imitated at very low cost. It was in their interest of course to prevent the importation into the United States and Europe of unlicensed generic drugs that might be manufactured in, say, India. But once they entered the international rule making arena, the advocates of IPRs did not stop there. They pushed to have the international trading system override the decision that many developing countries had made, that the short-term health of their populations was more important than the profits of the big drug companies. I say "short-term" here because in making this choice they were almost certainly giving up the chance for new drugs to be developed of particular interest to those same populations. Whether they understood that cost or not, I do not know, but it was the choice they had made. And the drug companies sought to rule that choice essentially illegal.

My point here is not that the TRIPs agreement is harmful, even though I may believe that it is. My point is that the TRIPs agreement was negotiated for reasons that had nothing to do with whether it was harmful or not, but only to do with whether it would increase the profits of some powerful corporations. And I am not saying either that these corporations were behaving badly. I happen to think that pharmaceutical companies have made a huge contribution to the well-being of modern society, and I do not generally begrudge them the large profits that they earn when one of their discoveries is a success. Their pursuit of profit through research and development as well as marketing has, I am quite willing to believe, made the world a better place.

The problem is that they were permitted to pursue profit not just in their industry but also by influencing international negotiations on the rules that they themselves would follow and benefit from. And not surprisingly, they used their power to design those rules to their own maximum advantage. They could hardly have been expected not to. The error was in letting them influence the negotiations in the first place. In their particular case, it would not have been hard to prevent this, since the relevance of their concerns for trade negotiations was always suspect. They could have been told to take their case to the World Intellectual Property Organization, rather than to GATT and the WTO.

GATS

I claim that there is no reason to expect corporate influence on rule making to be beneficial. But there is also no reason to expect it to be harmful. My second example, I believe, demonstrates this. Prior to the Uruguay Round, the GATT covered only trade in goods. Indeed, international transactions in services were not even regarded as trade. Since they typically required some sort of producer

presence in the country consuming the service, they were considered to be part of FDI, and so were subject to much weaker disciplines on restrictive policies than the rules governing trade policy under the GATT.

At some point during the 1980s, participants in certain service industries got the idea of arguing that what they did was trade after all, and that it should be covered under GATT. They made this case to the US government, which in turn placed services on the agenda of the Uruguay Round. The end result was not, as it turned out, to include services in the GATT *per se*, which continues to be the central pillar of the WTO, but instead to negotiate a new General Agreement on Trade in Services (GATS) to provide its own discipline to service transactions.

This made good sense, in two respects. First, many have argued that the gains from international service transactions are just as real as the gains from trade in goods, and indeed the economic causes and effects are very much the same. I made this case myself early on, in Deardorff (1985).

On the other hand, the policies that are applied to services are quite different from the policies that are applied to trade in goods, and they do not even include anything that is exactly comparable to the tariffs that were the original focus of the GATT. So the same rules could not apply directly, and it made sense to negotiate a new agreement to cover services.

The GATS itself did not accomplish much liberalization, if any, although some has come in subsequent agreements. And there are disagreements about how liberalization in these markets ought to proceed. But my point here is simply to note that these issues might never have been addressed had it not been for the influence of American corporate executives in the service industries who saw an opportunity and pressed their case. They did it, I presume, primarily to benefit themselves and their firms, but in this case economics is very much on their side and I think the world as a whole stands to benefit.

Not all would agree with this, of course. Precisely because services are interlinked with FDI, imports of services can be seen as even more threatening to a country's domestic interests than imports of goods. And furthermore, the perception that many service industries were dominated by large corporations from rich countries created the fear that, in these industries, developing countries stood only to lose from freer trade.

Of course this second concern is exactly what trade following comparative advantage is all about, and while competing industries may lose, their countries gain from services that are supplied to them more efficiently. Indeed, one can argue that the gains from trade are even larger here than in many goods industries, because so many services provide essential inputs, including infrastructure, for other industries. Access to world-class service providers might well allow some developing-country manufacturing industries to exploit a comparative advantage that would otherwise be undermined by the need to use expensive and low quality local services.

NAFTA Chapter 11

The North American Free Trade Agreement (NAFTA) was a massive document, and those of us who tried to analyze the likely effects of NAFTA tended to focus

on those few aspects of it that we best understood. Not surprisingly, there were pieces of it that we neglected, some that we failed even to know about. One of these that later drew a good deal of public attention was Chapter 11, on investment. Motivated by concern that a government might expropriate the property of a foreign direct investor, Chapter 11 requires that any such expropriation, direct or indirect, be accompanied by appropriate compensation. Furthermore, it entitles the victim of such expropriation to bring a case before a NAFTA tribunal against the national government of the country where the expropriation took place.

This, it turns out, may provide a powerful tool for corporations in one NAFTA country to use against governments (including state and local) in another NAFTA country, whenever a government uses a policy or regulation that reduces the corporation's profit. Or so it seems, at least from the cases that have been brought. All three NAFTA governments have been named in such cases. Opponents of NAFTA point especially to attempts at environmental regulation that have been challenged under Chapter 11 on the grounds that they reduce the profit of a foreign NAFTA firm. The most visible such case has been against the United States by the Canadian Methanex Corporation, seeking compensation for a California prohibition against a fuel additive that they supplied and that had been judged to be toxic.

This particular case was decided against the company. Other cases have been brought under Chapter 11, some decided in favor of the complainant, others against, while still others have been settled or withdrawn. I find it hard to judge the merits of these cases, since many who write about this issue do not even pretend to be unbiased. But my impression is that the panels deciding these cases have been far less expansive than the NAFTA opponents have feared.

However, my point here is not to argue the merits of Chapter 11 and the way it has been administered. Rather what I want to stress is the unprecedented nature of the agreement itself, in two respects. First, as I said above, Chapter 11 covers not just direct expropriation—as when a government takes over ownership of a foreign firm's property —but also indirect. What that actually means remains to be determined by the cases as they are brought and decided, but as the Methanex case suggests, some would argue that indirect expropriation includes policy changes that reduce the value of a foreign investment. If that interpretation stands, it could turn out to be an expansion of property rights beyond what most countries, including the United States, include in domestic law.

Second, and perhaps more importantly, Chapter 11 allows an unprecedented right of private action against the government of another country to be decided by an international tribunal. Prior to NAFTA—in the US–Canada FTA, for example, and also in the GATT—such tribunals were available in trade law only for disputes between governments. But NAFTA does not require that an aggrieved private party go through its own government; it files the case directly against the government of another country.

So why did NAFTA come to include such an unusually strong facility for private action? Consider who gains from Chapter 11. On the one hand, companies

from NAFTA countries that invest in other NAFTA countries gain, acquiring either leverage over policies that might harm them or substantial compensation, if their cases succeed or induce a settlement in their favor. On the other hand, the countries themselves may gain from a higher level of direct investment, as foreign investors are reassured by Chapter 11 against the risk of adverse policy decisions. But if the latter was the dominant factor in motivating Chapter 11, one has to wonder why governments eager to attract FDI could not have provided such reassurance more directly with their own laws. The likelihood therefore seems high that Chapter 11 was written expressly for the companies that stood to benefit from it. And the impression that this was the case seems heightened by the fact that nobody spoke of Chapter 11 when the NAFTA agreement was being debated.

So regardless of the merits of Chapter 11, its existence strikes me as fairly strong evidence, once again, of the power of corporate interests in setting the rules of the international trading system.

NAFTA rules of origin

Because it is a free trade area rather than a customs union, thus leaving external tariffs unchanged and usually different for imports of the same good into different member countries, the NAFTA must include rules of origin. That is, for a good to cross from one NAFTA country into another free of tariff, it must be regarded as having "originated" within NAFTA, and the word "originated" must be defined. What this definition should be is not obvious, and it can matter a great deal as I will illustrate later. Economists have examined the effects of various rules of origin, without any consensus emerging as to what they should be.[3]

Two plausible candidates would base origin either on the percentage of domestic value added or on changes in tariff classification between inputs and outputs. The NAFTA Chapter 4, on rules of origin, uses both of these criteria and more, in an elaborate series of provisions that I suspect only a team of lawyers and accountants together could understand. The complexity of these rules must itself discourage some producers from attempting to take advantage of the tariff preferences that NAFTA provides. I have no idea whether the complexity is a deliberate attempt to discourage trade, or is only the natural outcome of compensating the drafters of the agreement in proportion to the time they spend on it.

But there is one aspect of the NAFTA rules of origin whose purpose seems more explicit. The fourth article of Chapter 4 deals exclusively with Automotive Goods, spelling out rules of origin for these that are even more explicit, detailed, and (to me) opaque than for other goods. Are these more restrictive than the rest of NAFTA's rules of origin? I certainly cannot tell from reading the text. But I do happen to know that this text was drafted by a group that included at least one economist who was an employee of a US auto company. And that person's purpose was to draft rules that would permit his or her company to export tariff-free from Mexico to the United States while preventing its Japanese competitors from doing the same. A little later, I will give an example of how this might be possible, in case it is not obvious.

So it appears that, by having American firms at the table in drafting rules of origin, at least in the automotive sector, the NAFTA was explicitly designed to benefit certain US corporations. Was this the intent of the US government? Perhaps, but not necessarily. For all I know there may be a good reason why the automotive sector cannot be handled by the same rules as others, and, if so, then the expertise to write the rules would certainly not lie in the office of the United States Trade Representative (USTR), whose staff should be experts on trade, not automobiles. To whom should they turn to assist them if not experts from the auto industry, with US auto companies being the natural choice? But once in the room, those experts can hardly be expected to represent the interests of Japanese or European firms.

Other?

These four examples of the role that corporate influence has apparently played, two in the WTO and two in NAFTA, are just the examples that I happen to know about and think I understand. Given the extreme complexity of both of these agreements and the fact that corporate representatives certainly were involved during their negotiations, if not as parties to the negotiations then as advisors to the negotiators, I have to believe that there are many other perhaps smaller ways that corporate interests have been represented. It is of course not in their interest to draw attention to these.

What theory says

Political economy of trade policy

We trade economists have been led to look at political forces mainly by the otherwise inexplicable reluctance of governments to take our advice. After all, we have been very consistent for two centuries in advising all who will listen that governments should not interfere with trade. There are exceptions to our case for free trade, but we have always found credible counter-arguments to these. In the end, even those of us who have been most creative in constructing cases where protection might be beneficial have ultimately agreed that the economic interests of countries are best served by eliminating tariffs and other barriers to trade.[4] Yet, hardly any government in history seems to have listened. Given the choice between believing that we were wrong, or that they were pursuing something other than their countries' economic interests, we naturally chose the latter.

Hence the theories of political economy of trade have focused on explaining why import tariffs were positive. This literature, which has been surveyed frequently,[5] seems to have converged on the model of Grossman and Helpman (1994) as the standard. In this model, a specific factor in each industry benefits from a tariff on its output, and the owners of specific factors are thus motivated to contribute toward the election of incumbent politicians in return for protection. Politicians in turn maximize a combination of social welfare and these contributions, with the result that they provide positive protection.

The Grossman–Helpman model says rather directly that owners of capital (as a specific factor) exert influence on international trade policy through political contributions. The model has also received some empirical support from Goldberg and Maggi (1999), although with a higher weight on social welfare than many might have expected. It does not, however, address the making of the rules of trade policy, but only the setting of tariffs within rules that have already been made.

However, a subsequent application of the Grossman–Helpman model comes closer to addressing rule making. In Grossman and Helpman (1995), they use the same model to examine the formation of FTAs. After applying their framework in the first part of the paper just to the binary problem of whether or not to agree on an FTA with a given partner, they then turn to the question of exclusions. That is, actual FTAs do not always eliminate tariffs on absolutely every sector of the economy, and instead may exempt certain "sensitive" sectors. What are those sectors likely to be sensitive to? Politics, of course. Grossman and Helpman show how the selection of exempted sectors will depend, as do tariffs in their first application, on contributions from specific factors in exempted industries. These in turn depend on the extra returns that these specific factors stand to gain (or avoid losing) by being exempted. This, then, is an example of corporate influence over the rules of an FTA, if we equate corporate profits with returns to specific factors.

Another example is provided by Krishna (1998). He uses a model of imperfect competition and political economy to examine how the political process might favor one choice of FTA partner over another, and how both might compare to an alternative of multilateral liberalization. What he finds is that, the more likely is an FTA to be trade diverting and therefore welfare reducing, the more likely it is to be selected by the political process, over other potential FTAs and over multilateral liberalization. Thus, again, the rules of trade policy in the form of FTAs are influenced by the profits of those sectors that stand to gain. It should be said, however, that political economy models of issues like this do not necessarily rest on the influence of owners of capital. Levy (1997), for example, examines how FTAs influence the choice of multilateral liberalization, but in a median voter model.

Rules of origin—a theoretical example

I turn now to a theoretical example that is focused more explicitly on the determination of one of the rules of international trade—specifically, a rule of origin in an FTA. I will show a simple model in which a domestic firm stands to increase its profit by setting a rule of origin that favors it over its foreign competitor, thus giving it the incentive to influence the formulation of that rule in whatever way it can. As will be clear, the model attempts to formalize what seems to have been going on in the NAFTA rules of origin in automotive goods, as discussed above. I do not model the political process, which I presume could be similar to that of Grossman and Helpman, although I will conclude with a brief discussion of an alternative mechanism that might more plausibly have the same effect.

Consider a simple model of Bertrand duopoly. That is, there are two firms that are the only suppliers of a single differentiated product to the market in country U, which is segmented from other countries' markets. One of the firms is based in country U and the other in another country, J. Production requires an intermediate input that each can produce only in its own country, due to a specialized factor, perhaps expertise, that is available only there and that is needed to sustain the final product's differentiated characteristics. Their costs for this input are constants, b^U and b^J, respectively, for the quantity needed to produce one unit of the final good. The input can then be assembled at constant cost in either country U or country J, and also in another country, M, which will become the partner in an FTA with country U. These per-unit costs of assembly are the same for both firms, but differ across countries: c^U, c^J, and c^M, of which I will assume $c^U = c^J > c^M$. The firms face demand functions in country U, $D^U = D^U(p^U, p^J)$ and $D^J = D^J(p^J, p^U)$ respectively, where p^i is the price of firm i's product in country U.

The demand functions are linear, as I will specify below. Countries U and M levy specific tariffs on imports of the unassembled good, t^{Uu} and t^{Mu} respectively, which may but need not be different. Initially, country U also levies a specific tariff, t^{Uf}, on imports of the final good from either other country. I assume that this tariff is larger than the cost saving that either firm could secure by assembling in country M rather than country U, $t^{Uf} > \max[(c^U - c^M), (t^{Uu} - t^{Mu} + c^U - c^M)]$, so that both firms do assembly in country U.[6] Later I will have country U form an FTA with country M, eliminating this tariff on imports from there but subject to a rule of origin.

Formally, then, the model has two firms, A and B, each with constant unit costs of production k^i, $i = A, B$. They face demands for their products given by

$$D^i = a^i - p^i + sp^j, \quad i \neq j = A,B \tag{8.1}$$

where $a^i > 0$, $0 < s < 1$ are parameters.[7] That is, demand for each is downward sloping in its own price, but increases in the price of its competitor. Each firm's profit is therefore

$$\pi^i = \left(p^i - k^i\right)\left(a^i - p^i + sp^j\right) \tag{8.2}$$

With Bertrand competition, each firm selects its price to maximize this profit given the price of the other. First order conditions for this maximization imply their price-reaction functions,

$$p^i = (a^i + k^i + sp^j)/2 \tag{8.3}$$

which together yield equilibrium prices

$$p^i = \Omega\left(A^i + 2k^i + sk^j\right) \tag{8.4}$$

where

$$\Omega = 1/(4 - s^2), A^i = 2a^i + sa^j \tag{8.5}$$

Profits in equilibrium turn out to be

$$\pi^i = \Omega^2\left(A^i - (2 - s^2)k^i + sk^j\right)^2 \tag{8.6}$$

Applying this result to the firms U and J, before the FTA we have their input costs as b^U and $b^J + t^{Uu}$ respectively, but a common assembly cost in country U, c^U. Thus, in the notation of the model, $k^U = b^U + c^U$ and $k^J = b^J + t^{Uu} + c^U$. Using (8.6), their pre-FTA profits are

$$\begin{aligned}\pi^U &= \Omega^2\left(A^U - (2 - s^2)(b^U + c^U) + s(b^J + t^{Uu} + c^U)\right)^2 \\ \pi^J &= \Omega^2\left(A^J - (2 - s^2)(b^J + t^{Uu} + c^U) + s(b^U + c^U)\right)^2\end{aligned} \tag{8.7}$$

After the FTA, firm U can do assembly in the cheaper country M, since its exports from M back to U will easily satisfy any rule of origin, having been produced entirely in U and M. Thus, letting "~" denote the FTA, k^U becomes $\tilde{k}^U = b^U + c^M < k^U$. Firm J, on the other hand, since its input comes from outside the FTA, if it were to assemble in country M would fail to satisfy any rule of origin that, say, requires content from within the FTA countries of more than $c^M/(b^J + t^{Mu} + c^M)$ percent. Assuming this is the case, then firm J continues to assemble in U, and its cost continues to be $\tilde{k}^J = k^J = b^J + t^{Uu} + c^U$.

Applying these costs to (8.6), we find the FTA profits of the firms to be

$$\begin{aligned}\tilde{\pi}^U &= \Omega^2\left(A^U - (2 - s^2)(b^U + c^M) + s(b^J + t^{Uu} + c^U)\right)^2 \\ \tilde{\pi}^J &= \Omega^2\left(A^J - (2 - s^2)(b^J + t^{Uu} + c^U) + s(b^U + c^M)\right)^2\end{aligned} \tag{8.8}$$

Comparing (8.8) with (8.7),

$$\begin{aligned}\tilde{\pi}^U &= \left(\sqrt{\pi^U} + \Omega(2 - s^2)(c^U - c^M)\right)^2 > \pi^U \\ \tilde{\pi}^J &= \left(\sqrt{\pi^J} - \Omega s(c^U - c^M)\right)^2 < \pi^J\end{aligned} \tag{8.9}$$

Thus, a rule of origin, which can prevent firm J from taking advantage of the lower cost assembly in country M while firm U avails itself of it through the FTA, both raises the profit of firm U and lowers the profit of firm J. This is hardly surprising, and in the example here it can be achieved very simply: just specify as a rule of origin that imports from M to U will be duty free only if they have content from among the FTA countries of more than $\Gamma = c^M/(b^J + t^{Mu} + c^M)$ percent. That would easily be satisfied by an FTA content requirement of 100%, but since that might be too blatantly protectionist, the firm may be expected instead to use its knowledge of the industry to seek a content requirement that is lower, but that is high enough to achieve the same purpose, just more than Γ.

I have not attempted here to model the means by which a firm in this situation might influence the writing of the rule of origin. The Grossman–Helpman story,

I suppose, would be that the firm would offer a financial contribution of some sort to incumbent politicians in return for their setting a content requirement in this industry of more than Γ. Whether this would succeed or not would depend on how the politician weighs the financial contribution (which is limited by the gain in profit that is implicit in (8.9)) against the loss in economic welfare for the country as a whole. Realistically, it would also depend on whether the politician can manage to influence the negotiations on the FTA at the level that would make this possible.

A more plausible story for political influence in this context is suggested by Hall (2001).[8] In his description of lobbying, private interests do not in any sense pay money to politicians. Rather, they help them. That is, recognizing the limited time and resources that legislators have with which to accomplish their many objectives, only one of which is being reelected, they welcome assistance from lobbyists who assist them by doing research and drafting legal texts on issues of their common interest. Hall points out first that a great deal of lobbying takes place, not between lobbyists and the politicians whose minds they wish to change, but between lobbyists and politicians who are already in agreement—the politicians who because of ideology or their constituencies already wish to achieve the same objectives as the lobbyists. Such behavior makes no sense from the perspective of other models of political economy. For example, in Grossman and Helpman's story, a firm would never pay a politician to vote for a tariff that he or she already favored on grounds of social welfare. But it does make sense if, by helping the politician, the lobbyist can help to make progress toward their common objective. The result of this is that lobbyists and their staff members play an active role in formulating policy.

Hall's theory is intended to explain the interaction between lobbyists and legislators, but the idea seems to extend easily to the executive branch. The office of US Trade Representative, in particular, is surprisingly small in proportion to the size of its mission, and I would guess that it needs all the help it can get. In addition, the negotiations that it undertakes necessarily involve the intricacies of many individual industries, and even a much larger USTR staff would be unlikely to possess the expertise needed to deal with these. Therefore, it is natural for trade negotiators to enlist the assistance of willing participants from domestic industries who have this expertise. And that gives the industries access to the drafting of trade agreements. It is through this mechanism, it seems, that trade rules negotiated by the US government come to represent the interests of domestic industries.

Conclusion

In this chapter I have argued that profit maximizing firms, even though they contribute to social welfare when they compete in the market, may not do so when they influence the political process. In particular, I have suggested, through several examples from both the real world and from economic theory, that corporations have played a significant role in the formulation of the rules of the modern international trading system. They do this both in the multilateral system

overseen by the WTO, and within regional trade agreements such as NAFTA. Their influence on rules need not be bad, but there is no reason, either, why it should be good. We therefore need to give more thought to constraining not just trade policy, but also the making of the rules of trade, to try to assure that this process moves us to a better world.

Notes

1 For example,suppose we know that a domestic industry provides a positive externality that somehow cannot be internalized through a first-best production subsidy. Then while this means that a tariff may be beneficial if the industry competes with imports, it also means that a tariff (on another good) is even more harmful than usual if this industry exports.
2 It is, of course, not only corporations who influence policy, and that is true of the rules as much as of individual policy actions. Producers are also represented by labor unions and by producer organizations, including agricultural ones. Interests other than those of producers are also articulated by various nongovernmental organizations. Even consumers play a role, although many so-called consumer organizations seem to argue for protection that will raise prices to consumers. My focus here is on corporations because they seem to have played the greatest role in influencing rules of the WTO and NAFTA.
3 See Krishna (2005) for a recent survey of this literature.
4 See, for example, Krugman (1987).
5 See Rodrik (1995), Helpman (1999).
6 I do not require, though I permit, that U and M have different size tariffs on the unassembled good, even though that would be the normal justification for having a rule of origin at all.
7 The upper bound on s, $s<1$, is only to assure the plausible property that equal increases in both prices reduce demand, but only $s<2$ is needed for the model to be well behaved.
8 Published as Hall and Deardorff (2006). See also Deardorff and Hall (1997).

Bibliography

Deardorff, Alan V. (1985) "Comparative Advantage and International Trade and Investment in Services," in Robert M. Stern (ed.) *Trade and Investment in Services: Canada/US Perspectives*, Toronto: Ontario Economic Council, 39, 71.

——(1990) "Should Patent Protection Be Extended to All Developing Countries," *The World Economy*, 13, (December): 497–507.

——(1992) "Welfare Effects of Global Patent Protection," *Economica*, 59, (February): 35–51.

——(2003) "What Might Globalization's Critics Believe?" *The World Economy*, 26, (May): 639–658.

Deardorff, Alan V. and Hall, Richard L. (1997) "Explaining the Role of Interest Groups in United States Trade Policy," Research Seminar in International Economics Discussion Paper No. 415, University of Michigan.

Goldberg, Pinelopi Koujianou and Maggi, Giovanni (1999) "Protection for Sale: An Empirical Investigation," *American Economic Review*, 89, (December): 1135–1155.

Grossman, Gene M. and Helpman, Elhanan (1994) "Protection for Sale," *American Economic Review*, 84, (September): 833–850.

——(1995) "The Politics of Free-Trade Agreements," *American Economic Review*, 85(4), (September): 667–690.

Hall, Richard L. (2001) "Lobbying as Legislative Subsidy," Paper presented at the annual meetings of the American Political Science Association.

Hall, Richard L. and Deardorff, Alan V. (2006) "Lobbying as Legislative Subsidy," *American Political Science Review*, 100 (February): 69–84.

Helpman, Elhanan (1999) "Politics and Trade Policy," in Richard Baldwin, Daniel Cohen, Andre Sapir, and Anthony Venables (eds) *Market Integration, Regionalism, and the Global Economy*, Cambridge: Cambridge University Press, 86–112.

Krishna, Kala (2005) "Understanding Rules of Origin," NBER working paper No. 11150, (February).

Krishna, Pravin (1998) "Regionalism and Multilateralism: A Political Economy Approach," *Quarterly Journal of Economics*, 113 (February): 227–251.

Krugman, Paul R. (1987) "Is Free Trade Passé?" *Journal of Economic Perspectives*, 1, 131–144.

Levy, Philip I. (1997) "A Political-Economic Analysis of Free-Trade Agreements," *American Economic Review*, 87 (September): 506–519.

Rodrik, Dani (1995) "Political Economy of Trade Policy," in G. Grossman and K. Rogoff (eds) *Handbook of International Economics*, Vol. III, Amsterdam: North-Holland: 1457–1494.

Part 3

Policies and institutions

9 Preferential trading and welfare

The small-union case revisited

Arvind Panagariya[*]

Introduction

The welfare analyses of preferential trading arrangements have been characterized by generally inconclusive and messy results. In this chapter, I attempt to give order to the analysis of one important case: a union between two small countries. The analysis has two key advantages over the existing literature. First, the model employed is fully general in that it allows for goods that are exported and imported by both partners as well as those that are exported by one and imported by the other partner. Second, the results are derived for finite changes in tariff rates rather than being limited to infinitesimally small changes.

The main results can be summarized as follows: (1) Assuming all goods to be normal in consumption, if two small countries form a free trade area or exchange some tariff preferences, their joint welfare falls or rises as their joint output, valued at world prices, rises or falls. (2) If, in addition, the *numeraire* good uses only labor and all other goods use labor and a sector-specific factor, the exchange of preferences or free trade area necessarily lowers the union's joint welfare. (3) A union member is necessarily hurt by its own preferential liberalization. The higher are its external tariffs and the larger its imports from the partner, the more it loses from extending the preferences. (4) In the specific-factors case just mentioned, a union member necessarily benefits from the tariff preference it receives from the partner. The more it exports to the partner and the higher the latter's tariffs, the greater the gain. Finally, in the specific-factors case, an Free Trade Agreement (FTA) benefits a member more the larger its bilateral trade surplus with the partner and the lower its external tariffs relative to the partner.

Since the publication of Viner's (1950) classic work, *The Customs Union Issue*, a voluminous literature has come to exist on the welfare effects of preferential trade arrangements (PTAs). The literature has remained remarkable for its lack of robust and clear results, however. Much of the older, first-wave literature employs homogeneous goods models characterized by perfect competition. As Corden's (1984) survey of that literature testifies, the results of these models depend critically on the number of goods, the direction of trade and initial levels of various tariffs.[1] The new literature, triggered by what Bhagwati (1993) has called the Second Regionalism, has largely switched to imperfect competition models.

As far as the welfare effects of PTAs go, this switch has only added to the fragility of the results.[2]

Nevertheless, following an initial suggestion by Wonnacott and Lutz (1989) but much stronger advocacy by Krugman (1991) and Summers (1991), many analysts uncritically accepted the "natural trading partners" hypothesis whereby trade creation effects are likely to dominate the trade diversion effects if the potential union members are natural trading partners in the sense that they already trade intensively with each other and are geographically proximate. Bhagwati (1993) was quick to question the analytic basis of this idea while Bhagwati and Panagariya (1996) subsequently provided its systematic critique. This brought the literature on static welfare theory of PTAs back to square one: we still lacked robust results on the welfare effects of PTAs.

The purpose of this chapter is to revisit the old problem of a PTA between two or more small countries and give some measure of order to its analysis. Of course, the small-country context necessarily requires that goods be homogeneous. In differentiated goods models, firms, and countries necessarily have market power. Moreover, as long as trade restrictions take the form of nonprohibitive tariffs, which are the focus of this chapter and much of the literature, the small-country assumption necessarily implies perfectly competitive behavior in the goods markets. Thus, the natural context for analyzing the problem of a PTA between two small countries is that of homogeneous goods with perfect competition.

At the center of the literature on PTAs between small countries is the three-good Meade (1955) model, as analyzed originally by Lipsey (1958) and elaborated subsequently by Corden (1976), McMillan and McCann (1981), Lloyd (1982) and Panagariya (1997a, b).[3] As traditionally formulated and analyzed, this model is quite unrealistic, however. The analysis usually begins with the assumption that, of the three goods, each union member exports one and imports the other two. Of the latter, one good is imported exclusively from the partner and the other from the rest of the world. Prices in the rest of the world are fixed. Internal prices facing a union member differ from those in the rest of the world by tariff per unit. Initially, the *ad valorem* tariff is assumed to be the same on both imports. Preferential trade is then introduced by a small reduction in the tariff on the good imported from the partner.

There are four important limitations of this approach to analyzing the implications of preferential trading. First, there is no reason whatsoever in this model for countries to form a PTA. Given that the prices facing consumers and producers in each member country depend entirely on its own tariffs, the model permits no distinction between liberalization within the context of a PTA and voluntary reductions in tariffs by the country. This is because internal prices facing the country are altered solely by alterations in its own tariffs and not of the partner. This, in turn, implies that the country has no incentive to make its liberalization contingent on that of the partner. Its welfare is solely the outcome of its own tariff policy.

Second, assuming the good imported from outside is also produced by both PTA members, there is an internal inconsistency in the model.[4] The inconsistency

is explained most easily in the case when the PTA takes the form of an FTA, though it also applies to partial tariff preferences. Denoting the FTA members by A and B, suppose the tariff on the good imported from the outside country is higher in A. Then the price paid for this product by consumers in A will exceed that in B. Because goods produced in A and B are free to move anywhere within the union, all of within-union output will be sold in country A, which has the higher price. But the conventional analyses of the Meade model do not allow for this kind of arbitrage, assuming, instead, that the price received by producers in each member country continues to equal the price in the outside country plus its own tariff.[5] One way to get around the problem is to simply assume that the good imported from outside is not produced by union members at all, or at least the member with lower tariff on it. But such an assumption is quite restrictive and arbitrary.

Third, the analysis is carried out for small changes in the tariff on the partner country. Assuming substitutability, the model predicts that, as the tariff on the partner is lowered successively by small amounts, welfare first rises and then falls. This leaves uncertain the welfare outcome in the FTA equilibrium relative to that in the initial equilibrium. To date, we have do not have a general condition allowing comparison between pre- and post-FTA outcomes.

Finally, the results in the model are derived for a specific pattern of trade. As Lloyd (1982) shows systematically, altering this pattern of trade significantly influences the outcome. Thus, any conditions derived for the improvement in welfare even for a small change in the tariff preference depend on the assumed pattern of trade.

One or more of these criticisms also apply to other welfare analyses of PTAs between small countries. For instance, Berglas (1979) assumes a different pattern of trade and is able to deal with the first two of the above problems to some degree. But his analysis is also limited to small changes and relies on the chosen pattern of trade.

Here I present an analysis of PTAs between small countries that is quite general along many dimensions. First, I accommodate all possible patterns of trade into the model. This is done by considering a union between A and B that allows for four types of goods: those exported by both, those imported by both, those imported by A and exported by B and those exported by A and imported by B. Leaving aside nontraded goods, which are readily incorporated into the analysis following the technique in Dixit and Norman (1980: 208–211), these types of goods exhaust all possible patterns of trade. Second, I allow all goods to be produced and consumed everywhere and also for the full exploitation of arbitrage opportunities. This means that, in the presence of an FTA, within-union producers sell their output in the country where they get the highest price. Third, I allow for perfectly general preferences. The only restriction I impose is that all goods be normal in consumption and even this restriction is a sufficiency condition, not necessary. Finally, the analysis is not limited to infinitesimally small changes. Instead, it allows for finite changes in tariffs, permitting a direct comparison of pre- and post-FTA equilibria.

This level of generality is achieved with the aid of two important restrictive assumptions, however. First, I assume that the union member with higher tariff on a good continues to import that good after the formation of the PTA. This assumption is necessary to ensure that internal prices remain linked to the corresponding world prices via relevant tariff rates. In the analysis of tariff policy in a small country, this is a standard assumption. The traditional general-equilibrium models of PTAs, mentioned, for example, in note 3, also make this assumption. Nevertheless, there is a strand of the literature, principally diagrammatic and partial-equilibrium in its approach that relaxes this assumption. For example, Bhagwati and Panagriya (1996) consider the case in which a tariff preference eliminates the imports of higher-tariff member from the rest of the world and, thus, delinks its internal price of the good from the corresponding world price. The present chapter does not cover this case. Second, the chapter also assumes that there are no redundant tariffs. In particular, if a good is exported in the initial equilibrium, it is subject to a zero tariff. Theoretically, this is not a particularly strong assumption and it is commonly made in the literature. But, in practice, countries do have tariffs on goods they export. The assumption is nontrivial in the FTA analysis due to the fact that a good that is initially exported can also turn into an import good in the post-FTA equilibrium. That, in turn, can convert a redundant tariff into an effective one.[6]

A key advantage of my approach is that it links closely the analysis of PTAs to the more traditional tariff analysis in a small, open economy. Kowalczyk (1990) has argued in favor of bringing these two strands of the literature closer, offering his own version of how this can be done for infinitesimally small changes. Likewise, Neary (1998) has offered what he calls a "warehouse" model that also aims to build a bridge between customs union theory and tariff analysis in the presence of market power on the part of the customs union. His analysis also employs differential calculus and is, thus, confined to infinitesimally small changes.

The chapter is organized as follows. First I study the effect of an FTA on the joint welfare of the union. Next, I focus on individual welfare of a union member, distinguishing between the effects of its own liberalization and those of the liberalization by its partner. Then, I conclude. It should be borne in mind that though the analysis is presented for an FTA for sharpness, the extension to partial preferences is straightforward.

FTA and the joint welfare of the union

Let us begin with the notation to be used throughout. As usual, there are three countries, A, B, and C. The only variables of C, relevant to the analysis, are prices, which are fixed. By appropriate choice of units, we set them all equal to unity. Lower-case letters denote variables and functions relating to A and upper-case letters those relating to B. When there is no likelihood of confusion, I spell out explicitly the variables and functions relating to A only. Superscript 0 is used to denote the value of a function or variable in the initial equilibrium and superscript 1 that in the post-Free Trade Agreement (FTA) equilibrium.

There are four goods, 0, 1, 2, and 3. Both A and B export good 0 and import good 1. In addition, A imports good 2 and exports good 3 while B does the opposite. These four goods exhaust possible patterns of trade. We can accommodate many goods of each type but it will complicate the presentation without adding to the generality of the analysis. The interested reader can try out the extension on his own.

Country A imposes a tariff t_i on good i, which is positive if good i is imported and zero otherwise.[7] The price of good i is denoted p_i in country A and P_i in B. Correspondingly, p and P stand for the entire price vectors. Thus, in the initial equilibrium, the price vector in country A is written $p^0 = (p_0^0, p_1^0, p_2^0, p_3^0) = (1, 1 + t_1, 1 + t_2, 1)$ and that in B $P^0 = (P_0^0, P_1^0, P_2^0, P_3^0) = (1, 1+T_1, 1, 1+T_3)$. Letting u stand for utility level in country A the expenditure function there can be written $e(p, u)$ where $e(\cdot)$ is positive, concave, and linear homogeneous in p and increasing in u. The partial derivative of the expenditure function with respect to the ith price, denoted $e_i(p, u)$, gives the compensated demand curve for good i and is positive. On the supply side, we represent the revenue function in country A by $r(p)$, where $r(\cdot)$ is convex and linear homogeneous in p and its partial derivative with respect to the ith price, denoted $r_i(p)$, gives the output of good i. For brevity, we will write $e^0 = e(p^0, u^0)$, $e_i^0 = e_i(p^0, u^0)$, $r^0 = r(p^0)$ and $r_i^0 = r_i^0(p^0)$, where u^0 is endogenous. Analogous notation applies to post-FTA period distinguished by superscript 1.

Assuming that all tariff proceeds are rebated to consumers in a lump-sum fashion, in Country A, the initial equilibrium can be represented by the condition:

$$e^0 = r^0 + t_1(e_1^0 - r_1^0) + t_2(e_2^0 - r_2^0) \tag{9.1}$$

The only endogenous variable in this equation is u^0. Corresponding to (9.1), in B, the equilibrium is represented by

$$E^0 = R^0 + T_1(E_1^0 - R_1^0) + T_3(E_3^0 - R_3^0) \tag{9.2}$$

The only endogenous variable in (9.2) is U^0.

Suppose now that A and B form a free trade area such that the goods produced within the union are allowed to move between the countries free of duty. Without loss of generality and ruling out the customs unions case of $t_1 = T_1$, assume $t_1 > T_1$. As noted in the introduction, assume further that country A continues to import good 1 from C in the post-FTA equilibrium. Then the consumer price of good 1 in A in the post-FTA equilibrium is the same as in the initial equilibrium. We have $p_1^1 = p_1^0 = 1+t_1$. Because the good can be imported into B at $1+T_1$, which is less than $1+t_1$, the consumer price of the good there cannot rise above $1+T_1$. It then follows that all of good 1, produced in B (as also that produced in A), will be sold in A. The demand in B will be satisfied entirely by imports, with consumers there paying the price $P_1^1 = P_1^0 = 1+T_1$.

Good 2 is imported by A and exported by B. Again, by assumption, A continues to import the good from C in the post-FTA equilibrium. Therefore, the price in the country remains unchanged; $p_2^1 = p_2^0 = 1+t_2$, which is more than $P_2^1 = P_2^0 = 1$, the price prevailing in B. All of good 2 produced in B (as also that produced in A) is sold in A.

The outcome with respect to good 3 is opposite of that with respect to good 2. The price of this good in B is $P_3^1 = P_3^0 = 1+T_3$, which is higher than $p_3^1 = p_3^0 = 1$, the price in A. Therefore, all of A's output of this good will be sold in B and the domestic demand satisfied through imports.

The upshot of this discussion is that, in the post-FTA equilibrium, the prices facing consumers in each member will be the same as those in the initial equilibrium. But the producer prices, now equalized between members, are different from those prevailing in the initial equilibrium. The vector of producer prices is now given by $(1, 1+t_1, 1+t_2, 1+T_3)$. Note the asymmetry here: for the good imported by both countries in the initial equilibrium, the producer price is determined by the consumer price in the country with the higher tariff. The FTA gives all producers access to the highest price prevailing in the union.

My objective here is to study the effect of the FTA on the joint welfare of A and B. For this purpose, I will assume that A gives a "gift" to B in the amount G in terms of the numeraire good, good 0, such that it guarantees B the same level of welfare as in the initial equilibrium. We then calculate the welfare in A relative to its pre-FTA welfare. If it rises, we can conclude that the union as a whole benefits and if it falls, the union is hurt. If, in the absence of the gift, the FTA would have raised B's welfare, G takes a negative value.

We can now write the equilibrium conditions in the post-FTA equilibrium as follows:

$$e^1 = r^1 + t_1(e_1^1 - r_1^1 - R_1^1) + t_2(e_2^1 - r_2^1 - R_2^1) - G \tag{9.3}$$

$$E^1 = E^0 = R^1 + T_1E_1^0 + T_3(E_3^0 - R_3^1 - r_3^1) + G \tag{9.4}$$

Note that $e^1 = e(p^0, u^1)$, $r^1 = r(1, 1+t_1, 1+t_2, 1+T_3)$ and $R^1 = R(1, 1+t_1, 1+t_2, 1+T_3)$. Since the level of utility in B is held at its pre-FTA level through the lumps-sum transfer and the consumer prices are unchanged, the post-FTA expenditure there remains unchanged at E^0. Equations (9.3) and (9.4) have two endogenous variables, u^1 and G.

Subtracting (9.1) from (9.3) and (9.2) from (9.4), respectively, we obtain

$$\begin{aligned} e^1 - e^0 &= (r^1 - r^0) + t[(e_1^1 - r_1^1 - R_1^1) - (e_1^0 - r_1^0)] \\ &\quad + t_2[(e_2^1 - r_2^1 - R_2^1) - (e_2^0 - r_2^0)] - G \end{aligned} \tag{9.5}$$

$$0 = (R^1 - R^0) + T_1R_1^0 + T_3(R_3^0 - R_3^1 - r_3^1) + G \tag{9.6}$$

Adding (9.5) and (9.6) and rearranging, we obtain

$$\begin{aligned} &(e^1 - t_1e_1^0 - t_2e_2^0) - (e^0 - t_1e_1^0 - t_2e_2^0) \\ &\quad = [(r^1 - t_1r_1^1 - t_2r_2^1 - T_3r_3^1) - (r^0 - t_1r_1^0 - t_2r_2^0)] \\ &\quad\quad + [(R^1 - t_1R_1^1 - t_2R_2^1 - T_3R_3^1) - (R^0 - T_1R_1^0 - T_3R_3^0)] \end{aligned} \tag{9.7}$$

Recalling that expenditure and revenue functions are linear homogeneous in prices, this equation can be written as

$$\sum_{i=0}^{3} e_i^{\,1} - \sum_{i=0}^{3} e_i^{\,0} = \left[\sum_{i=0}^{3} r_i^{\,1} - \sum_{i=0}^{3} r_i^{\,0}\right] + \left[\sum_{i=0}^{3} R_i^{\,1} - \sum_{i=0}^{3} R_i^{\,0}\right] \tag{9.7′}$$

Or, rearranging slightly, we have

$$\sum_{i=0}^{3} [e_i^{\,1} - e_i^{\,0}] = \sum_{i=0}^{3} [(r_i^{\,1} + R_i^{\,1}) - (r_i^{\,0} + R_i^{\,0})] \tag{9.7″}$$

This is a neat expression. The left-hand side represents the difference between country A's post- and pre-FTA total expenditure *at world price* (recall that world prices are all set equal to 1 by the choice of units). The right-hand side represents the difference between post- and pre-FTA output of the union as a whole, also valued at world prices. If we hold country B's utility fixed, at world prices, the expenditure in country A can be increased by the increase in the value of the union-wide output.

Recall that, since the price vector facing the consumer before and after the FTA is the same, $e^0 = e(p^0, u^0)$ and $e^1 = e(p^0, u^1)$. It is then immediate that, assuming all goods to be normal, $e_i^{\,1} - e_i^{\,0}$ is positive or negative as $u^1 - u^0$ is positive or negative for all i. Therefore, from (9.7″), ruling out inferiority in consumption, the FTA increases or reduces the joint welfare of A and B as it increases or reduces the value of the union-wide output, at world prices. In the spirit of the standard tariff theory for a small, open economy, we have

Proposition 1 *Suppose two small countries form an FTA by eliminating tariffs between themselves but retaining them on the rest of the world at their original levels. All goods are normal in consumption and, after the union is formed, each good continues to be imported from the rest of the world into the union member with the higher tariff. Then the FTA increases the joint welfare of the union members if and only if it increases the value of the union's total output at world prices.*

Note that given the manner in which this proposition has been proved, it can be readily extended to more than two countries and many goods of each type. Indeed, it can also be proved readily for partial tariff preferences.

On the face of it, Proposition 1 gives a simple condition that allows us to determine the desirability of a specific PTA. But, despite its intuitive appeal, its practical application is limited for two reasons. First, it requires the knowledge of outputs in the post-PTA equilibrium, which is not available *ex ante*. As such, it can be applied only after the formation of the PTA. Second, and more importantly, even *ex post*, outputs can change due to many factors. For instance, technology may have changed for exogenous reasons or the external prices themselves may have shifted.

Given these limitations, we need to push the theoretical analysis a little further. Needless to say that this can only be accomplished by imposing further restrictions

on the structure of the model. It turns out that in at least one special albeit important case, we can derive a clear-cut result. This is the production model underlying the important contribution by Grossman and Helpman (1995) on the politics of FTAs and employed frequently in the recent international-trade literature including Panagariya and Rodrik (1993). This model assumes that the *numeraire* good uses only labor while all other goods use labor and a sector-specific factor.[8] Focusing on this model, I now proceed to demonstrate the following result.

Proposition 2 *In addition to the conditions stated in Proposition 1, if we assume that the numeraire good uses only labor and all other goods use labor and a sector-specific factor, a PTA between two small countries necessarily lowers the joint welfare of the union.*

To prove proposition 2, observe first that under the assumed production structure, the output of each non-numeraire good depends only on its own price with extra labor drawn from or absorbed by the numeraire good. Because the prices of goods 1 and 2 in country A and of good 3 in country B are the same before and after the PTA, we can set $r_1^1 = r_1^0$, $r_2^1 = r_2^0$ and $R_3^1 = R_3^0$ in equation (9.7″). This allows us to obtain

$$\sum_{i=0}^{3}(e_i^1 - e_i^0) = (r_0^1 + r_3^1) - (r_0^0 + r_3^0) + (R_0^1 + R_1^1 + R_2^1) - (R_0^0 + R_1^0 + R_2^0) \tag{9.8}$$

The first term in on the right-hand side represents the difference between the sums of the values of outputs of goods 0 and 3 at world prices after and before the PTA in country A. Likewise, the second term represents the difference between the sums of the values of outputs of goods 0, 1, and 2 at world prices after and before the PTA in country B. The remainder of the proof involves the demonstration that each of these terms is negative indicating that the value of output at world prices declines in each country in the post-PTA equilibrium.

Thus, consider the first term. We know that with the domestic prices of goods 1 and 2 being identical before and after the PTA in country A, labor employed in these sectors is the same in the two equilibriums. Therefore, the total supply of labor available for employment in sectors 0 and 3 before and after the PTA is the same. Let us represent this supply by O_0O_3 in Figure 9.1 and measure the employment of labor in sector 0 to the right from O_0 and in sector 3 to the left from O_3. Because the *numeraire* sector uses only labor, its value-of-marginal-product-of-labor (VMPL) curve, shown by line V_0V_0, is horizontal with the height of the line representing the wage rate (which equals 1 since good 0 is the numeraire good).

On the other hand, the VMPL curve of sector 3 slopes down as we increase labor employment in it and is shown by V_3V_3 in the pre-PTA equilibrium. The allocation of labor between sectors 0 and 3 is then determined by point E with labor O_0L_0 allocated to sector 0. Because the domestic prices of goods 0 and 3 in

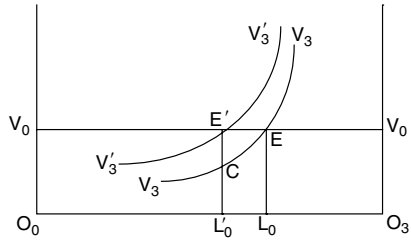

Figure 9.1 The value of outputs before and after the PTA.

country A coincide with the world prices in the pre-FTA equilibrium, the value of output of these goods at world prices is given by the area under the V_3V_3 curve up to point E plus rectangle $V_0EL_0O_0$.

Next, recall that the PTA raises the price of good 3 facing the producers of that good in country A from 1 to $1+T_3$. In turn, this shifts the V_3V_3 curve up to $V'_3V'_3$. The allocation of labor between goods 0 and 3 now moves to point E′ with $O_0L'_0$ labor employed in sector 0. The value of output of good 3 at world prices is now represented by the area under V_3V_3 (and not $V'_3V'_3$) up to point C and that of good 0 by rectangle $V_oE'L'_0O_0$. It is then immediate that the combined value of outputs of goods 0 and 3 at world prices declines by the triangular are EE′C in the post-PTA equilibrium relative to the pre-PTA equilibrium.

Using a construction similar to that in Figure 9.1 for country B, it can be shown that the PTA, which raises the producer prices of goods 1 and 2 above the respective world prices in that country, lowers the value of output at world prices. Thus, the proof of Proposition 2 is complete.

The key to Proposition 2 lies in the fact that, given the assumed production structure, non-*numeraire* goods are independent of one another in production (that is, neither substitutes nor complements). A change in the producer price of a good causes labor to move solely between that good and the *numeraire.* In country A, the only price change due to the FTA relates to the producer price of good 3, which increases from 1 to $1+T_3$. This change moves labor from good 0 to good 3. At the initial equilibrium, at world prices, the value of marginal product of labor in sector 3 is the same as in sector 0. Therefore, for the first unit of labor moved from good 0 to good 3, there is no net change in the value of output at world prices. But for subsequent units, due to the operation of diminishing returns, the value of marginal product at world prices declines in sector 3 but not in sector 0 since it is characterized by constant marginal product of labor. Therefore, the aggregate value of output at world prices declines.

A similar explanation applies to country B. There, the FTA leads to an increase in the producer price of good 1 from $1+T_1$ to $1+t_1$ (recall $t_1 > T_1$ by assumption) and of good 2 from 1 to $1+t_2$. Once again, labor is drawn from sector 0 where the value of marginal product at world prices exceeds that in good 1 and equals that

in good 2 at the initial equilibrium. Given constant marginal product in sector 0 and diminishing marginal product in sectors 1 and 2, the movement necessarily reduces the value of output at world prices.

Relaxation of two key assumptions, underlying Propositions 1 and 2, naturally gives rise to alternative possibilities. Thus, if the imports of a good from the outside country into the union member with the higher tariff cease altogether in the post-FTA equilibrium, the consumer price of that good falls. Under this scenario, we cannot determine the welfare effects of the FTA from the value of output alone and we cannot derive either Proposition 1 or Proposition 2. Alternatively, if the *numeraire* good also uses a specific factor or, more generally, all factors are used in all sectors, the effect of the FTA on the union's joint welfare is ambiguous due to the fact that resources may be drawn from the initially distorted sectors to initially undistorted sectors. In this case, Proposition 2 cannot be derived.

It is interesting to contrast Propositions 1 and 2 to the recent result of Panagariya and Krishna (1998) on welfare improving FTAs. In the spirit of Kemp and Wan (1976), these authors show that if two countries freeze their initial trade vectors with the rest of the world via appropriate external tariff vectors and form an FTA, their joint welfare necessarily rises. In contrast, this study has shown that if countries freeze their external tariffs rather than trade vectors, joint welfare falls. This contrast arises due to asymmetric implications of policy changes in the presence of price versus quantity distortions as explained in Krishna and Panagariya (1999).

Welfare of individual member countries

We were concerned above with the welfare of the union as a whole. Let us now briefly turn to the welfare of individual union members. Using partial-equilibrium models, I have emphasized in a number of recent writings (Panagariya, 1996, 1999a, b and Bhagwati and Panagariya, 1996) that, even in the small-union context, individual members stand to lose from preferential trade liberalization of their own and benefit from the partner's liberalization. This stands in sharp contrast to nondiscriminatory liberalization, which is beneficial to a small country undertaking such liberalization. The analysis below confirms the validity of this result in general equilibrium.

For the issue at hand, we rule out any direct transfers and set $G = 0$. Continuing to assume that the imports from the outside country do not cease, equation (9.5) remains valid with $G = 0$. Making use of linear homogeneity of $e(\cdot)$ in the prices, we can rewrite (9.5) as

$$\sum_{i=0}^{3}(e_i^1 - e_i^0) = (r^1 - r^0) - t_1[(r_1^1 - r_2^1) + R_1^1] - t_2[(r_2^1 - r_2^0) + R_2^1] \qquad (9.9)$$

As before, assuming all goods to be normal, the left-hand side is positive or negative as u^1 is larger or smaller than u^0. Therefore, the FTA raises or lowers country A's welfare as the right-hand side is positive or negative.

Suppose for a moment that all liberalization is done by country A and none by country B. Then, like the consumer prices, producer prices facing country A also remain unchanged in the post-FTA equilibrium. Therefore, we obtain $r^1 = r^0$ and $r_i^1 = r_i^0$ and equation (9.9) reduces to

$$\sum_{i=0}^{3} (e_i^1 - e_i^0) = - [t_1 R_1^1 + t_2 R_1^2] \tag{9.10}$$

The right-hand side being negative, we immediately obtain $u^1 < u^0$. Note that, as emphasized in the author's papers mentioned above, the loss to country A from its own liberalization is exactly equal to the tariff revenue it fails to collect on the imports from country B. The more it imports from country B, the greater the loss. We can state

Proposition 3 *Suppose a small country removes partially or wholly its tariffs on another small country but not the rest of the world. All goods are normal in consumption and the country continues to import each good from the rest of the world after the preferential liberalization. Then the country's welfare necessarily declines, with the loss in real income equaling the lost tariff revenue on the imports from the country receiving the tariff preference. The more the country imports from the partner and the greater the magnitude of tariff preference, the more it loses.*

Observe that this result does not rely on the special specific-factors model and applies to all technologies giving rise to the standard revenue functions. Also note that though the formal derivation in (9.10) assumes that tariffs on the partner are pushed all the way to zero, the extension to partial tariff preference is trivial. Hence, I have stated the above proposition to include both partial and full tariff preferences.

Next, consider the opposite case in which tariff preferences are extended exclusively by country B. In this case, the price of good 3 rises from 1 to $1+T_3$ and there is no loss of tariff revenue on the imports coming from country B. Equation (9.9) reduces to

$$\begin{aligned} \sum_{i=1}^{3} (e_i^1 - e_i^0) &= [(r^1 + r^0) - t_1(r_1^1 - r_2^1) - t_2(r_2^1 - r_2^0) \\ &= \sum_{i=0}^{3} (r_1^1 + r_i^0) + T_3 r_3^1 \end{aligned} \tag{9.11}$$

The first term on the right-hand side represents the change in the value of country A's output at world prices. The sign of this term is ambiguous in general. The second term represents the benefit reaped from the tariff-free access to country B's market; it equals precisely the tariff revenue due to country B. The more country A exports to country B and the greater the extent of tariff preference, the larger this term and more likely country A benefits from the preference.

Once again, in the special case considered previously, the ambiguity of the right-hand side can be eliminated. With the prices and outputs of goods 1 and 2 unchanged between the pre- and post-PTA equilibrium, we have $r_1^1 = r_1^0$, $r_2^1 = r_2^0$ allowing us to rewrite (9.11) as

$$\sum_{i=0}^{3}(e_i^1 - e_i^0) = [(r_0^1 + r_3^1) - (r_0^0 + r_3^0)] + T_3 r_3^1 \qquad (9.11')$$

We have already seen in the context of Proposition 2 that the term in the square brackets on the right-hand side of (9.11′) is negative and the area of triangle EE′C in Figure 9.1 represents its magnitude. Tariff revenue transfer to country A's exporters, shown by the second terms on the right-hand side above, equals the difference between the value of the output of good 3 at domestic and world prices. From Figure 9.1, the former is represented by the area under V′3V′3 up to E′ and the latter by the area under V3V3 up to point C. The difference between these two areas encompasses triangle EE′C and therefore necessarily exceeds the latter. The term on the right-hand side is necessarily positive, proving Proposition 4.

Proposition 4 *Suppose a small country receives a tariff preference on a good from another small country and the latter continues to import some quantity of the good from the rest of the world. Suppose further that the numeraire good uses only labor and other goods use labor and a sector-specific factor. Then the tariff preference necessarily benefits the country receiving the preference.*

We can combine Propositions 3 and 4 to obtain the following broader result.

Proposition 5 *Suppose the numeraire good uses only labor while other goods use labor and a sector-specific factor. Also assume that, following an exchange of tariff preferences, all goods continue to be imported from the rest of the world into the partner with higher tariff. Then, if trade between two PTA partners is approximately balanced in the post-PTA equilibrium, the country with higher external tariffs will lose and the one with lower external tariffs will benefit. Alternatively, if the countries have approximately equal tariff rates, the country with bilateral trade surplus will benefit while the other country will lose.*

Conclusion

Despite the passage of half a decade since the publication of Viner's (1950) seminal contribution to the theory of preferential trading, the welfare analyses of these arrangements have suffered from inconclusive and generally messy results. Whatever results have been derived relate to infinitesimally small tariff preferences. In this chapter, I have attempted to give order to the analysis of one important case: a union between two small countries.

My analysis has several advantages over the existing literature. First, the model employed is fully general in that it allows for goods that are exported and

imported by both partners as well as those that are exported by one and imported by the other partner. Second, the results are derived for finite changes in tariff rates rather than being limited to infinitesimally small changes. As such they can be applied to limited tariff preferences as well as a full free trade area. Finally, unlike the existing literature, the chapter offers clear and transparent results.

The main results of the chapter can be summarized as follows. First, assuming all goods to be normal in consumption, if two small countries form a free trade area or exchange some tariff preferences, their joint welfare falls or rises as their joint output, valued at world prices, rises or falls. Second, if, in addition, the *numeraire* good uses only labor and all other goods use labor and a sector-specific factor, the exchange of preferences or free trade area necessarily lowers the union's joint welfare. Third, a union member is necessarily hurt by its own preferential liberalization. The higher are its external tariffs and the larger its imports from the partner, the more it loses from extending the preferences. Fourth, in the specific-factors case just mentioned, a union member necessarily benefits from the tariff preference it receives from the partner. The more it exports to the partner and the higher the latter's tariffs, the greater the gain. Finally, in the specific-factors case, an FTA benefits a member more the larger its bilateral trade surplus with the partner and the lower its external tariffs relative to the partner.

The model in this chapter can be readily extended to make the decision to form the FTA endogenous. Even though the FTA reduces overall welfare of the union and therefore must hurt at least one member, if the political support function gives greater weight to producer profits than welfare as in Grossman and Helpman (1995) and Krishna (1998), it can be shown to be politically feasible.

Notes

* I am deeply indebted to Costas Syropoulos for numerous extremely helpful comments that led to a major revision of the chapter.

1 Some of the key contributions relying on two-good models are Gehrels (1956–1957), Lipsey (1957), and Bhagwati (1971). Among the analyses based on three-good models are Meade (1955), Mundell (1964), Corden (1976), Berglas (1979), Riezman (1979), McMillan and McCann (1981), Lloyd (1982), Kowalczyk (1990), and Panagariya (1997a, b). Some of the recent contributions employ these models to study political economy issues. Thus, Richardson (1993, 1994) and Levy (1997) use the two-good model and Panagariya and Findlay (1996) the three-good model to address the impact of PTAs on external protection.

2 Krugman (1991), Bond and Syropoulos (1996), Yi (1996), and Ethier (1998) use the differentiated goods model while Krishna (1998) employs a homogeneous-good, oligopoly model. See Panagariya (1999) for a survey of this and older literature.

3 Meade himself did not consider the small-union case. Instead, as discussed in detail in Panagariya (1997a), he allowed the terms of trade to vary and focused on the world welfare. In an important paper, Mundell (1964) later worked out explicitly the terms of trade effects of a PTA in this model.

4 This point was originally made by Richardson (1994) and incorporated into the production model, underlying their political economy analysis of FTAs, by Grossman and Helpman (1995). Panagariya and Krishna (1997) exploited it to establish a result similar to the Kemp–Ohyama–Kemp–Wan theorem for FTAs (Kemp and Wan, 1976).

5 As discussed in Panagariya (1997a), this problem also arises in a slightly different form in the case of a partial tariff preference.
6 Panagariya and Dutta-Gupta (1999) offer a partial-equilibrium analysis of this case. I am not aware of any general-equilibrium analysis of this case.
7 I remind the reader that this is not an entirely innocent assumption. Though a positive tariff on an exportable is redundant in the initial equilibrium, it may become effective in the post-FTA equilibrium. The implications of this possibility will be explained later in the paper.
8 The demand side in this paper is more general than in Grossman and Helpman (1995). They assume the utility function to be quasi linear. I only require that all goods be normal in consumption. On the other hand, they do not impose the assumption that each good continue to be imported after the FTA by the member with the higher tariff on it.

Bibliography

Berglas, Eitan (1979) "Preferential Trading: The n Commodity Case," *Journal of Political Economy*, 87(21): 315–331.

Bhagwati, Jagdish (1971) "Trade-Diverting Customs Unions and Welfare Improvement: A Clarification," *Economic Journal*, 81: 580–587.

Bhagwati, Jagdish (1993) "Regionalism and Multilateralism: An Overview," in Melo and Panagariya (eds) *New Dimensions in Regional Integration*, Cambridge, Great Britain: Cambridge University Press, 22–51. Also in Bhagwati, Krishna and Panagariya, 1999.

Bhagwati, Jagdish and Panagariya, Arvind (1996) "Preferential Trading Areas and Multilateralism: Strangers, Friends or Foes?" in Jagdish Bhagwati and Arvind Panagariya (eds) *The Economics of Preferential Trade Agreements*, AEI Press, Washington, DC. Also in Bhagwati, Krishna and Panagariya, 1999, pp. 1–78.

Bhagwati, Jagdish, Krishna, Pravin and Panagariya, Arvind (1999) *Trading Blocs: Alternative Approaches to Analyzing Preferential Trade Agreements*, Cambridge, MA: MIT Press, February.

Bond, Eric W. and Syropoulos, Constantinos (1996) "The Size of Trading Blocs, Market Power and World Welfare Effects," *Journal of International Economics*, 40: 411–437.

Corden, Max (1976) "Customs Unions Theory and the Nonuniformity of Tariffs," *Journal of International Economics*, 6: 99–106.

Corden, W. M. (1984) "Normative Theory of International Trade," in R. W. Jones and P. B. Kenen (eds) *Handbook of International Economics*, *Vol. I*, Amsterdam: North-Holland, chapter 2, pp. 63–130.

Ethier, W. (1998) "Regionalism in a Multilateral World," *Journal of Political Economy*, 106(6): 1214–1245.

Gehrels, Franz (1956–1957) "Customs Union from a Single-Country Viewpoint," *Review of Economic Studies*, 24(63): 61–64.

Grossman, Gene and Helpman, Elhanan (1995) "The Politics of Free Trade Agreements," *American Economic Review*, 85(4), September: 667–690. Also in Bhagwati, Krishna and Panagariya, 1999.

Kemp, M. and Wan, Henry Jr (1976) "An Elementary Proposition Concerning the Formation of Customs Unions," *Journal of International Economics*, 6, (February): 95–98. Also in Bhagwati, Krishna and Panagariya, 1999.

Kowalczyk, Carsten (1990) "Welfare and Customs Unions," NBER working paper No. 3476, October.

Krishna, Pravin (1998) "Regionalism and Multilateralism: A Political Economy Approach," *Quarterly Journal of Economics*, 113(1), (February): 227–251. Also in Bhagwati, Krishna and Panagariya, 1999.

Krishna, Pravin and Panagariya, Arvind (1999) A Unification of the Second Best Results in International Trade, Mimeo: University of Maryland.

Krugman, Paul (1991a) "The Move to Free Trade Zones," in *Policy Implications of Trade and Currency Zones*, Symposium Sponsored by the Federal Reserve Bank of Kansas City, 7–41.

——(1991b) "Is Bilateralism Bad?" in E. Helpman and A. Razin (eds) *International Trade and Trade Policy*, Cambridge, MA: MIT Press, pp. 9–23.

Levy, Philip (1997) "A Political-Economic Analysis of Free-Trade Agreements," *American Economic Review*, 87(4), (September): 506–519.

Lipsey, Richard (1957) "The Theory of Customs Unions: Trade Diversion and Welfare," *Economica*, 24, (February): 40–46.

——(1958) "The Theory of Customs Unions: A General Equilibrium Analysis," University of London, PhD thesis.

Lloyd, Peter J. (1982) "3x3 Theory of Customs Unions," *Journal of International Economics*, 12: 41–63.

McMillan, John and McCann, Ewen (1981) "Welfare Effects in Customs Union," *Economic Journal*, 91: 697–703.

Meade, James, E. (1955) *The Theory of Customs Unions*, Amsterdam: North-Holland.

Mundell, Robert A. (1964) "Tariff Preferences and the Terms of Trade," *Manchester School of Economic and Social Studies*, 32: 1–13. Also in Bhagwati, Krishna and Panagariya, 1999.

Panagariya, Arvind (1996) "The Free Trade Area of the Americas: Good for Latin America?" *World Economy*, 19(5): 485–515.

——(1997a) "Preferential Trading and the Myth of Natural Trading Partners," *Japan and the World Economy*, 9(4): 471–489.

——(1997b) "The Meade Model of Preferential Trading: History, Analytics and Policy Implications," in Cohen, B. J. (ed.) *International Trade and Finance: New Frontiers for Research. Essays in Honor of Peter B. Kenen*, New York: Cambridge University Press, 57–88. Also in Bhagwati, Krishna and Panagariya, 1999.

——(1999) *Preferential Trade Liberalization: The Traditional Theory and New Developments*, Mimeo: University of Maryland.

Panagariya, Arvind and Dutta-Gupta, Rupa (1999) Preferential Trading and Welfare: The Small-Union Case Revisited, Mimeo: University of Maryland.

Panagariya, Arvind and Pravin Krishna (1997) On the Existence of Necessarily Welfare-Enhancing Free Trade Areas, Mimeo: Center for International Economics, Department of Economics, University of Maryland.

Panagariya, Arvind and Rodrik, Dani (1993) "Political Economy Arguments for a Uniform Tariff," *International Economic Review*, 34(3): (August): 685–703.

Panagariya, Arvind and Ronald Findlay (1996) "A Political Economy Analysis of Free Trade Areas and Customs Unions," in Robert Feenstra, Douglas Irwin and Gene Grossman (eds) *The Political Economy of Trade Reform*, Essays in honor of Jagdish Bhagwati, Cambridge, MA: MIT Press, pp. 265–287. Also in Bhagwati, Krishna and Panagariya, 1999.

Richardson, Martin (1993) "Endogenous Protection and Trade Diversion," *Journal of International Economics*, 34(3–4): 309–24.

Richardson, M. (1994) "Why a Free Trade Area? The Tariff Also Rises," *Economics & Politics* 6(1), (March): 79–95. Also in Bhagwati, Krishna and Panagariya, 1999.

Riezman, Raymond (1979) "A 3x3 Model of Customs Unions," *Journal of International Economics*, 9(4): 341–354.

Summers, Lawrence (1991) "Regionalism and the World Trading System," *Policy Implications of Trade and Currency Zones*. Symposium Sponsored by the Federal Reserve Bank of Kansas City, 295–301.

Viner, Jacob (1950) *The Customs Union Issue*, New York: Carnegie Endowment for International Peace. Chapter 4 is reproduced in Bhagwati, Krishna and Panagariya, 1999.

Wonnacott, Paul and Mark Lutz (1989) "Is There a Case for Free Trade Areas?" *Free Trade Areas and U.S. Trade Policy.* Schott, Jeffrey, Washington, DC: Institute for International Economics, 59–84.

Yi, Sang-Seung (1996) "Endogenous Formation of Customs Unions under Imperfect Competition: Open Regionalism is Good," *Journal of International Economics*, 41: 153–77.

10 On the viability of conditional assistance programs

Wolfgang Mayer and Alex Mourmouras[*]

Introduction

The International Monetary Fund (IMF) and other International Financial Institutions (IFIs) provide financial and analytical assistance to member countries facing serious macroeconomic and structural imbalances. When such imbalances threaten a country, its government will approach (or will be approached by) an IFI to negotiate an agreement that specifies conditions under which assistance is made available. An increasingly important component of such conditional assistance programs is the implementation of structural reforms that aim to reduce economic policy distortions and adopt reforms that lead to a more efficient allocation of resources and higher economic growth.[1] As emphasized in the literature on transaction costs politics (Dixit, 2000, 2001), structural reforms have several key distinguishing features: they are far more difficult to monitor than macroeconomic reforms, have long gestation periods, and political-economy considerations play an important role in their implementation (see also Mussa and Savastano, 1999: 23).

This chapter examines the viability of IFI-supported assistance programs that incorporate structural economic reforms as conditions for receiving assistance.[2] It does so by highlighting the implications of the "difficult to monitor" and "political-economy" features of structural reforms. We describe a model in which an assistance-receiving government chooses policies under the influence of a special interest group and in which the assistance-providing IFI is unable to observe the actual implementation of structural reforms. The chapter's goal is to establish necessary conditions for the viability of a conditional assistance program; that is, to state conditions that must be met for the government not only to accept an IFI-designed assistance program but also to undertake the reforms called-for in the program, and to do so without endangering the political stability of the country and the financial stability of the IFI. If any one of these conditions is violated, the program is not viable.[3] The chapter also examines unconditional aid as an alternative to conditional aid, especially when the latter is not viable.

As mentioned, structural reforms are difficult to monitor. The IFI is frequently unable to observe the actual implementation and enforcement of reforms as distinct from the rather easily observable passage of reform laws and regulations.

The IFI's inability to observe the implementation of structural reforms makes conditional assistance programs susceptible to moral hazard; there is asymmetric information between assistance-receiving government and assistance-providing IFI concerning the implementation of conditions that form the core of the assistance agreement.[4] Unless the agreement itself creates sufficient incentives for reform, the government will be tempted to accept the IFI's assistance but not change its distorting economic policies.[5]

The government's policy choices are motivated by political considerations.[6] Its goal is to maximize the political support which it derives from special interest groups[7] and the general public.[8] Economic policies that distort the economy are typically chosen to placate special interests, even if they hurt the general public. In order to be viable, the combination of financing and adjustment offered by conditional assistance programs must take into account the interest group's influence. Specifically, the terms of the program that spell out the amount of the loan, its repayment conditions, and the type of reforms to be undertaken must be sufficiently attractive for the government to sacrifice some or all of the interest group's political support.[9]

A viable conditional assistance program requires the government not only to accept and implement the program, but to do so without endangering domestic political stability and without compromising the IFI's capacity to assist in the future. This chapter models a world economy in which every country's production is stochastic. The economy might perform well or poorly independent of the country's resource base. The economy's actual performance depends on its initial resource base, the amount of IFI assistance it receives to expand this base, the choice of economic policies, and unobservable stochastic production shocks. No matter how the economy performs, the general public is particularly sensitive to how externally dictated policies influence the outcome. A country's political stability becomes threatened when the public associates economic underperformance with IFI intervention; that is, when the public forms the belief that it would have been better off without IFI assistance and without reforms than with assistance and with reforms.

The program must take into consideration the threat of political unrest by the public by specifying a package of financing, repayment terms, and reforms that are consistent with the absence of strikes, revolutions, and other instability.[10] The program's expected repayment conditions—when both good and bad times are considered—must be sufficiently generous that the general public accepts the conditional assistance package as being worthwhile for the country and remains unorganized, thus safeguarding the country's political stability. To ensure continued political stability, the program's repayment conditions must also be sufficiently flexible, meaning that they must be more lenient during bad times than during good times. This flexibility is needed to ensure that that the public does not find it is worthwhile to organize against austerity related to excessively onerous program repayment terms, especially in recessions, crises or other bad economic times.[11]

Finally, IFIs are legally obligated to maintain the financial resources entrusted to them. Their charters contain the obligation to lend under adequate safeguards.

This means that the repayment conditions of each conditional assistance program must be such that, *ex ante*, the IFI expects, at the least, to break even. *Ex post*, the IFI gains from some programs and loses from others. But if production shocks are independently distributed among its large group of borrowers, IFI financial stability is secured if *ex ante* expected repayment on its loan portfolio is sufficiently large to cover the initial outlay.[12]

This chapter states the viability conditions for conditional assistance in a simple model with one aid-providing IFI, one aid-accepting foreign government, and one interest group that tries to influence the government's policies.[13] Given the economy's resources, its stochastic performance is limited to "good" and "bad." The probability of the performance being good, in turn, depends on whether the economy operates under policy distortions or is reformed. Assistance adds to the economy's resource base. How "good" and how "bad" the economy performs is influenced by the amount of assistance received.

This chapter provides three basic insights on the viability constraints of conditional assistance programs. The first one relates to programs that provide just enough political support to the government that it prefers joining the IFI at the negotiating table and signing an agreement to the alternative of going-it alone. It is easy to see why IFIs may not want to see their resources "wasted." In their effort to raise the welfare of their member countries, the IFIs will be tempted to expand assistance to as many of their members as they can reach. But since IFIs have limited financial resources, they will design assistance packages that are just barely attractive for the government to accept. The chapter shows that such "barely attractive" programs of reforms and assistance must always fail in their attempt to bring about economic reforms. In the presence of asymmetric information, such programs are attractive enough for the government to sign but do not contain enough incentives for it to reform and forego future political support from interest groups.

Second, the IFI's "aid cost" of reforms critically hinges on how dependent the government's political support is on the interest group prior to the implementation of economic reforms. The more successful the interest group is in its rent-seeking and the more of the rent it channels to support the government, the costlier it is politically for the government to reform and dislodge the interest group. On the other hand, the aid cost of reform is lowered when a government is more dependent on political support from the general public.

Third, we compare the effectiveness of conditional and unconditional assistance programs. When the IFI's objective is the maximization of the entire world's welfare, it is shown that viable conditional programs are always more effective than unconditional programs. However, unconditional programs offer an appropriate choice when conditional programs are no longer viable. While the viability of conditional programs critically hinges on how dependent the government's political support is on the interest group, the viability of unconditional programs is independent of the influence of the interest group. Accordingly, IFIs should not rigidly stick to programs that stipulate conditions as quid-pro-quos for providing assistance. The world is better off if IFIs offer unconditional assistance to countries in

which asymmetries of information regarding the true implementation of structural reforms are pervasive. IFI assistance to developing countries needs to be tailored to include a mix of conditional and unconditional programs, depending on how severe the information problems are in particular countries and sectors.

Political economy of the assistance-receiving country

There are three decision makers in this political-economy model of conditional aid provision. A domestic government chooses the assistance-receiving country's economic policies; a domestic interest group attempts to influence policy choices for its own benefit; and an IFI offers assistance that is contingent on reforming the economy. Assistance is provided through a loan that the country cannot obtain in the private market. Economic reforms entail the removal of policy distortions that favor the interest group. In addition to the three decision makers, there is the assistance-receiving country's general public. Its welfare, together with that of the interest group, critically shapes the government's policy choices.

The government's objective is to retain power which, in turn, depends on political support from the interest group and the general public. The interest group can appropriate economic rents through the adoption of economic policies which distort the economy and hurt the general public. The group tries to retain this privileged position by influencing the government through financial contributions, C. The latter are valuable to the government as it needs funding to shape the general public's perceptions of the government's performance. Interest group and general public have conflicting objectives. The general public benefits from economic reforms; the interest group is hurt by them. A completely reformed economy eliminates all rents to the interest group and, thereby, eliminates the group. The general public's welfare is measured by its own income which consists of national income adjusted for both loan repayments to the IFI and economic rents siphoned off by the interest group.

Let us define the country's national income by x and the loan repayment by z. The resulting net income, $(x-z)$, is split between general public and interest group. If an interest group exists, its appropriated rent is assumed to be a constant share, $0 \leq \gamma < 1$, of net income. When the interest group disappears, all of net income accrues to the general public. The government's political support function takes the usual form[14] of

$$G = C + a(1-\gamma)(x-z) \tag{10.1}$$

where G is a measure of political support and $0 < a < 1$ is a parameter that translates the public's net income, $(1-\gamma)(x-z)$, into political support for the government.[15] The value of a is less than one, meaning that a dollar of income received by the general public generates less political support than a dollar of financial contributions paid to the government.

An assistance agreement between the government and IFI spells out the amount of loan provided and the amount of repayments to be made. Repayments

are not fixed, but depend on the state of the economy at the time repayments are due. When the agreement is signed, neither government nor IFI knows how the economy will be performing, as the economy's production is stochastic. Production depends on the economy's resource base, the amount of assistance received, T, and the realization of the state of nature. Assuming constant resources, the country's aggregate production function can be written as

$$x = g(T)y \tag{10.2}$$

where y denotes national income in the absence of assistance, meaning that $g(0) = 1$. For $T > 0$, the assumption that $g(T) > 1$ expresses the magnifying influence of assistance on national income. It is assumed that there are diminishing returns to the amount of assistance received, such that $g'(T) > 0$ and $g''(T) < 0$ for all $T > 0$. For given resources, production is stochastic. National income in the absence of assistance, y, is either high, $\bar{y}$, or low, $\underline{y}$.The value of $\Delta y = (\bar{y} - \underline{y}) > 0$ serves as a measure of the economy's performance volatility.

All economic agents are able to observe and verify the value of actual national income. While none of them knows with certainty the cause for income being high or low, it is common knowledge that economic reforms—with the effect of removing interest-group-benefiting policy distortions—make the high-output state of nature more likely. More precisely, $0 < \pi_1 < 1$ is defined as the probability of attaining high output when the economy is reformed, whereas $0 < \pi_0 < 1$ is the probability of attaining high output when the economy is distorted. The value of $(\pi_1 - \pi_0) = \Delta\pi > 0$ serves as a measure of the effectiveness of economic reforms.

The IFI's aid agreement spells out the aid-receiving country's repayment obligations. In specifying repayment conditions, the IFI is concerned with the loan agreement's impact on the country's political stability. In particular, it fears that the general public will blame the conditional aid agreement for having made the country worse off than it would have been without this agreement. Political stability is seriously threatened when the general public views the aid agreement as an externally imposed burden on the country. Securing political stability requires flexible loan repayment conditions. Repayments when income is high, $\bar{z}$, and when income is low, $\underline{z}$, must be such that welfare of the general public is always at least as large under the aid agreement as it would be without it. For future reference, we will attach subscript $i = 1$ to the loan variable T_i and repayments z_i when the loan is conditional, while subscript $i = 0$ is employed to indicate unconditional aid.

In the early sections of this chapter, we are not trying to determine what kind of conditional assistance program is "optimal." In fact, the IFI's ultimate objective in extending a loan to a country's government is not explicitly stated. Instead, these sections focus on the viability of the instrument through which the IFI wants to operate. The instrument is conditional assistance, meaning assistance that is contingent on the implementation of economic reforms. Use of the conditional assistance process implies a set of constraints that must be satisfied to ensure the

loan-receiving country's implementation of economic reforms. The IFI's objective function is specified later, after the requirements for viable conditional assistance have been established; thereafter, the chapter compares the effectiveness of conditional aid with that of unconditional aid. The IFI's objective is assumed to be the maximization of world welfare, consisting of the combined welfare of aid-receiving and aid-providing countries.

The interest group as an obstacle to economic reforms

Prior to the IFI-government loan agreement, the country under consideration is saddled with significant economic policy distortions. The distortions are the result of rent-seeking by an interest group. The group makes financial contributions, C, to the policy-choosing government in return for rent-generating policies. As stated earlier, a successful interest group's rent is assumed to be a constant fraction, γ, of national income after loan repayments have been made. As long as there are no reforms, $\pi = \pi_0$, the interest group exerts influence. The group's expected net benefit from supporting a government that does not reform is

$$\gamma\left\{\pi_0\left[g(T_0)\bar{y} - \bar{z}_0\right] + (1 - \pi_0)\left[g(T_0)\underline{y} - \underline{z}_0\right]\right\} - C \tag{10.3}$$

where $T_0 > 0$ and $z_0 > 0$ when an unconditional loan is made and $T_0 = 0 = z_0$ when there is no loan agreement.

Prior to the IFI's loan offer—when $T_0 = \bar{z}_0 = \underline{z}_0 = 0$ in (10.3)—an interest group makes financial contributions to the government in return for rent-creating, distorting economic policies. The existence of such a group implies that expression (10.3) is nonnegative. If the group's expected net benefits were to become negative, its members would be better off if, instead of lobbying, they blended in as members of the general public. Accordingly, prior to receiving IFI assistance, there exists a maximum contribution level that an interest group is willing to make in order to avoid economic reforms. The maximum contribution equals the total rent generated by the distorting policies:

$$\bar{C}^0 = \gamma[\pi_0\bar{y} + (1 - \pi_0)\underline{y}] \tag{10.4}$$

where superscript 0 refers to this being the initial, pre-assistance period.

There also exists a minimum contribution level, $\underline{C}^0$, which the interest group must make to prevent the government from reforming the economy on its own, even if it did not receive any assistance. For the government to retain distorting policies, political support from the group and the public's diminished wellbeing must be at least as strong as it would be if the government reformed the economy, eliminated the privileged position of the interest group, and thereby enhanced the public's well-being.

Before the aid agreement, expected political support for a nonreforming government is $\{C^0 + a(1-\gamma)[\pi_0\,\bar{y} + (1-\pi_0)\,\underline{y}\}$, as follows from (10.1) and (10.2). If, on the other hand, the government were to enact reforms and lost

all interest group contributions, its expected political support would become $a[\pi_1\bar{y} + (1-\pi_1)\underline{y}]$. Public support is enlarged not only by a rising expected income—as the probability of high income increases—but also by getting all rather than only part of national income. On the other hand, the interest group's contribution has vanished. The minimum contribution the government must receive in order to work with the interest group, $\underline{C}^0$, is such that it makes the government indifferent between reforming and not reforming; that is,

$$\begin{aligned}\underline{C}^0 &= a[\pi_1\bar{y} + (1-\pi_1)\underline{y}] - a(1-\gamma)[\pi_0\bar{y} + (1-\pi_0)\underline{y}] \\ &= a\Delta\pi\Delta y + a\gamma[\pi_0\bar{y} + (1-\pi_0)\underline{y}]\end{aligned} \tag{10.5}$$

The actual level of contributions, C, made by the interest group prior to the IFI's loan offer, depends on the bargaining power of the group in dealing with the government. We are not explicitly modeling this bargaining process. Instead, we assume that the actual contribution exceeds the minimum the government must receive and that it is a fixed fraction, β, of the maximum amount the interest group is willing to contribute, such that

$$\underline{C}^0 < C = \beta\bar{C}^0 \leq \bar{C}^0 \tag{10.6}$$

The condition that $\underline{C}^0 < C = \beta\bar{C}^0$ is implied by our assumption that, prior to the IFI's arrival, an interest group and its government benefited from supporting each other in the design of economic policies. It follows from (10.6), after substitution of (10.4) and (10.5), that:

$$A = \{\gamma(\beta - a)[\pi_0\bar{y} + (1-\pi_0)\underline{y}] - a\Delta\pi\Delta y > 0\} \tag{10.7}$$

where A measures the interest group's actual political support of the government in excess of what is minimally required to prevent reforms. Note that this measure of the government's dependence on the interest group rises with both the share of income the group appropriates as rent, γ, and the bargaining strength of the government in dealing with the group, β; but this dependence declines with the public's influence on political support, as expressed by a.

Viability of conditional assistance programs

An IFI's conditional assistance offer consists of a loan that is contingent on the implementation of economic reforms. The loan must be repaid, and repayment conditions must be neither so harsh as to threaten the loan-receiving country's political stability nor so lenient as to compromise the IFI's financial stability. Most importantly, a conditional assistance program must contain incentives to implement reforms; it cannot simply be imposed by the IFI.

Viability requires the government not just to be willing to accept the loan and its repayment but also to implement the asked-for reforms. The IFI's offer must

create the necessary incentives to do so. It must be in the interest of the economic policy-setting government to reform the economy, even at the cost of losing support from a well-established interest group. In creating the right incentives, the IFI faces an important additional concern. The incentive structure (repayment conditions) must not permit outcomes that endanger the country's political stability. Consequently, four constraints must be satisfied to create the necessary incentives to reform without endangering political stability. First, at the time the IFI makes a loan offer, the government must consider it advantageous to accept the IFI-proposed conditional loan package. The government's expected political support with the loan and with economic reforms must be at least as strong as it would be without the loan and without economic reforms. This establishes a participation constraint. Second, the moral hazard problem of implementing economic reforms must be addressed. The government must be given an incentive to implement real reforms. Real rather than apparent reforms are required. It is not sufficient to pass laws and regulations that project the appearance of a reformed economy. The laws and regulations must also be fully implemented and enforced to make them effective.

While the IFI is able to observe the passage of economic reform laws and the economy's performance, it is unable to observe the actual implementation of reform measures. Only the domestic government and the domestic interest group know whether real reforms have been implemented. It is in the government's own interest to reform if expected political support is at least as strong when it accepts assistance and implements reforms as it would be with the same assistance and no implementation of reforms. In other words, there must be an *incentive constraint* to guarantee that it does not pay for the government to cheat on the assistance agreement.

The third and fourth constraints are imposed to secure political stability. A country's public inevitably views external interference in the government's economic policymaking with heightened suspicion. The adoption of external policy advice is rarely challenged if it seems to help the entire country. But the adoption of such advice becomes a major source of political instability if it seems to have hurt the country. Specifically, a threat to political stability emerges if, at any time during the assistance period, the public believes the country to have become worse off with the conditional loan than it would have been without such a loan. Different from the government's political support considerations, what matters here are not expectations about becoming better off or worse off through the acceptance of aid. What matters are actual outcomes. The IFI, therefore, must structure the loan repayment conditions in a way that, no matter what the prevailing state of the economy, the public can never blame the IFI as having made the country worse off. Hence, there are two political stability constraints, one when actual income is high and one when actual income is low.

For any given conditional loan, viability definitely requires that the participation constraint, incentive constraint, and two political stability constraints are satisfied. If any one of these constraints is violated, the conditional loan program is doomed to fail. Repayment conditions that satisfy the four constraints face one

additional hurdle. They also must be viable from the IFI's perspective. They must generate enough expected revenues to protect the IFI's own financial stability. In other words, the conditional aid package must be such that the expected repayment is at least as large as the initial loan value. The relevant measure for assessing IFI financial stability is expected rather than actual repayments; the IFI lends to many different countries and, depending on actual performance of these economies, earns a positive return on some loans and incurs a loss on other loans. As long as the IFI breaks even on average, it retains its financial strength.

We now provide detailed specifications of the constraint set that describes a viable conditional loan program. Starting with the participation constraint, the government is willing to accept a conditional aid package if the loan and its repayment obligations, combined with the impact of economic reforms, are such that political support for the government is not weakened. The alternative to accepting conditional assistance is the status quo, in which the interest group makes financial contributions to the government and the economy remains distorted. The government's expected political support under this *status quo* is expressed by the right-hand side (RHS) of (10.8), whereby we set $g(0) = 1$, the probability of high income is π_0, and only $(1-\gamma)$ of national income is left for the general public. The left-hand side (LHS) of (10.8) states the expected political support for the government when it receives a conditional loan in the amount of $T_1 > 0$ and it is obligated to pay back $\bar{z}_1$ when the economy performs well and $\underline{z}_1$ when it performs poorly. The implementation of economic reforms raises the probability of the economy performing well from π_0 to π_1, and disappearance of the interest group raises the general public's share of national income from $(1-\gamma)$ to 1. The government's participation constraint, therefore, is

$$a\left[\pi_1[g(T_1)\bar{y} - \bar{z}_1] + (1-\pi_1)[g(T_1)\underline{y} - \underline{z}_1]\right]$$
$$\geq a(1-\gamma)[\pi_0\bar{y} + (1-\pi_0)\underline{y}] + C^0 \tag{10.8}$$

The preceding section determined the interest group's maximum contribution, $\overline{C}^0$. We then assumed that bargaining between government and interest group resulted in an actual contribution of $C^0 = \beta\overline{C}^0$. Use of (10.4) and substitution for C^0 enables us to rewrite the participation constraint as

$$a\left[\pi_1[g(T_1)\bar{y} - \bar{z}_1] + (1-\pi_1)[g(T_1)\underline{y} - \underline{z}_1]\right]$$
$$\geq [a + \gamma(\beta - a)][\pi_0\bar{y} + (1-\pi_0)\underline{y}] \tag{10.8′}$$

Second, we introduce the *incentive constraint*. The government must have an incentive to implement and enforce the economic reform conditions of the accepted loan package. If the government accepts the loan and its repayment obligations, implements reforms, and loses interest group support, then $\pi = \pi_1$, $\gamma = 1$ and $C^1 = 0$ as expressed by the political-support expression on the LHS of (10.9). If, on the other hand, the government accepts assistance and its repayment

obligations, but fails to implement reforms and, thereby, retains interest group support, then $\pi = \pi_0$, $0 < \gamma < 1$ and $C^1 > 0$ as stated in the political-support expression on the RHS of (10.9).

The incentive constraint, which requires the government to be at least as well off accepting the loan and implementing reforms as it would be accepting the loan without implementing reforms, is then

$$a\{\pi_1[g(T_1)\bar{y} - \bar{z}_1] + (1 - \pi_1)[g(T_1)\underline{y} - \underline{z}_1]\} \geq a(1 - \gamma)\{\pi_0[g(T_1)\bar{y} - \bar{z}_1] + (1 - \pi_0)[g(T_1)\underline{y} - \underline{z}_1]\} + C^1 \tag{10.9}$$

where C^1 is the contribution level when assistance is received but there is no reform. The interest group's contribution, C^1, is again a fraction, β, of the maximum contribution the group is willing to make, $\overline{C}^1$. When the country did not receive any loans, this maximum contribution was described as $\overline{C}^0$ by (10.4). When the country now does receive a repayable loan, then the maximum contribution becomes

$$\overline{C}^1\ \gamma\{\pi_0[g(T_1)\bar{y} - \bar{z}_1] + (1 - \pi_0)[g(T_1)\underline{y} - \underline{z}_1]\} \tag{10.4′}$$

Substitution of $C^1 = \beta\overline{C}^1$ and of (10.4′) in (10.9) yields

$$a\{\pi_1[g(T_1)\bar{y} - \bar{z}_1] + (1 - \pi_1)[g(T_1)\underline{y} - \underline{z}_1]\} \geq [a + \gamma(\beta - a)]\{\pi_0[g(T_1)\bar{y} - \bar{z}_1] + (1 - \pi_0)[g(T_1)\underline{y} - \underline{z}_1]\} \tag{10.9′}$$

The third and fourth of our viability constraints require the loan package not to threaten the country's political stability. As stated previously, the loan repayment conditions must be such that, independent of the state of the economy, the public can never blame the IFI for having made the country worse off.[16] Hence, the country's actual income after loan repayment must be at least as large as its income would be if no loan had been received; and this must be so in good as well as in bad times. Accordingly, the political stability constraints are

$$g(T_1)\bar{y} - \bar{z}_1 \geq \bar{y} \tag{10.10}$$

$$g(T_1)\underline{y} - \underline{z}_1 \geq \underline{y} \tag{10.11}$$

Given a conditional loan of value $T_1 > 0$, constraints (10.8′), (10.9′), (10.10), and (10.11) can be used to solve for the highest attainable repayments, $\bar{z}_1$ and $\underline{z}_1$, that make the government accept the loan, induce the government to implement reforms, and not threaten the country's political stability. The IFI's choice is further narrowed by the condition that, given $T_1 > 0$, the solutions for $\bar{z}_1$ and $\underline{z}_1$ must be such that the IFI's *budget constraint* is satisfied; that is, the expected loan repayment must be at least as large as the actual amount of the conditional loan:

$$[\pi_1\bar{z}_1 + (1 - \pi_1)\underline{z}_1] \geq T_1 \tag{10.12}$$

A likely scenario for program failure

A typical IFI's general mission is to assist developing countries through loans that are not available in the private market. In pursuit of this mission, the IFI faces two major concerns. One is to bring as many governments of developing countries to the table and to offer them loan programs. The other is to specify the loans-repayment conditions in a way to meet its own budget constraint. This leads to a natural inclination on the part of IFI negotiators to offer as many loan agreements as possible and to make them just barely attractive enough for governments to accept them. In other words, there is pressure or at least an inclination to offer conditional loan packages that make the participation constraint "just binding." That way, repayments are as large as possible to build the pool for offering the largest possible number of loans. Of course, both government and IFI remain concerned with the country's political stability and, therefore, agree on repayment conditions that satisfy the two political stability constraints.

With pressure to have many governments sign conditional agreements and to waste as little as possible, the incentive constraint becomes a likely candidate for neglect. In addition, satisfying the incentive constraint encounters the following obstacle. In the real world, it is far more difficult to quantify the conditions under which a conditional loan package meets the incentive constraint than it is to figure out whether the participation constraint is satisfied. The participation constraint is met with the government's signature of the loan agreement. There is no such clear-cut signal for meeting the incentive constraint. An examination of the latter requires far more information. Not only must the IFI know how reforms enhance the prospect of a well-performing economy, but it also must know how entrenched the country's interest group is. In light of these information problems concerning the incentive constraint, it is highly likely that the IFI puts primary emphasis on the participation constraint.

Given the IFI's inclination to make the participation constraint binding, this section establishes the following implication for the fate of conditional loan programs:

Proposition 1 *A conditional loan offer inevitably results in reform failure if it is just barely attractive to the accepting government. When the participation constraint is binding and the political stability constraints are satisfied, the incentive constraint is always violated.*

Proposition 1 highlights the dangers of being preoccupied with signing loan agreements that preserve the IFI's financial and the recipient country's political stability. The temptation is to offer a deal that is just good enough to bring a government to the table. The government willingly accepts the loan and agrees to the stipulated repayments, but has no incentive to implement and enforce reforms. Its incentive is to cheat on the implementation of economic reforms. As shown in the next section, more lenient repayment conditions are called for to induce reforms. In other words, the IFI must allow for additional cost when it wants to induce

economic reforms. This additional cost weighs, of course, on the IFI's budget constraint. It is quite possible that there exists no loan package that meets the IFI's budget constraint when the loan is conditional even though the constraint could be met if the loan were unconditional. The comparison between conditional and unconditional loans will be taken up later.

To prove Proposition 1, we have the IFI offer a deal that is just good enough to be acceptable to the government; that is, the participation constraint of (10.8′) is binding. Furthermore, concern for political stability requires that constraints (10.10) and (10.11) are satisfied. To show that these three constraints necessarily imply violation of the incentive constraint (10.9′), we write (10.8′) as an equality and substitute the RHS of (10.8′) on the LHS of (10.9′) to obtain

$$[\pi_0\bar{y} + (1-\pi_0)\underline{y}] \geq \left\{\pi_0[g(T_1)\bar{y} - \bar{z}_1] + (1-\pi_0)[g(T_1)\underline{y} - \underline{z}_1]\right\}$$

After subtracting $[\pi_0\,\bar{y} + (1-\pi_0)\,\underline{y}]$ from both sides, the incentive constraint becomes:

$$0 \geq \left\{\pi_0[\bar{y}\Delta g_1 - \bar{z}_1] + (1-\pi_0)[\underline{y}\Delta g_1 - \underline{z}_1]\right\} \tag{10.9″}$$

where $\Delta g_1 = [g(T_1) - 1] > 0$. The political stability constraints of (10.10) and (10.11) reveal that both terms on the RHS of (10.9″) are nonnegative and that the incentive constraint of (10.9″) can be satisfied only if both (10.10) and (10.11) are binding. But this cannot be. The participation constraint of (10.8′), when written as equality, implies that it cannot be that both political stability constraints are binding. This can be seen after subtracting $a[\pi_1\bar{y} + (1-\pi_1))\underline{y}]$ from both the LHS and RHS of (10.8′), and using (10.7):

$$a\left\{\pi_1[\bar{y}\Delta g_1 - \bar{z}_1] + (1-\pi_1)[\underline{y}\Delta g_1 - \underline{z}_1]\right\} = A > 0 \tag{10.8″}$$

Hence, the fact that, prior to accepting assistance, an interest group influenced the government's policies requires that both $[\bar{y}\Delta g_1 - z]$ and $[\underline{y}\Delta g_1 - \underline{z}_1]$ are nonnegative and at least one of them is positive. This, in turn, implies that the incentive constraint of (10.9″) must be violated. The government has an incentive to cheat on the implementation of reforms.

Interest group strength and conditional loan viability

A conditional loan program is viable if the government has incentives to accept the loan and to implement economic reforms without threatening the country's political stability and the IFI's solvency. The main goal of this section is to relate conditional loan viability to the government's dependency on the interest group. The government's dependency rises with $0 \leq \gamma < 1$, which measures the fraction of the country's net income appropriated by the group as economic rent, and with β, which denotes the government's financial support share of this rent.

To highlight the various influences on conditional loan viability, we first solve for the loan repayment conditions, $\bar{z}_1$ and $\underline{z}_1$, that satisfy the participation constraint of (10.8′), the incentive constraint of (10.9′), and the political stability constraints of (10.10) and (10.11). As shown in the Appendix, the incentive constraint and the political stability constraint in the bad state of nature are binding, while the other two constraints are satisfied but not binding. Hence, we immediately know from (10.10) that

$$\underline{z}_1 = \underline{y}\Delta g_1 \tag{10.13}$$

where $\underline{y}\Delta g_1$ measures the loan-induced increase in national income in the bad state of nature. When income is low, the contractual repayment is exactly equal to the country's gain in income due to the loan.

The amount of repayment in the good-state of nature is obtained by going to the binding incentive constraint as restated in (A.1) of the Appendix, substitute (10.13) and (10.7), and solve for

$$\bar{z}_1 = \bar{y}\Delta g_1 - \frac{A}{B} \tag{10.14}$$

where $B = \{\pi_1 a - \pi_0[\beta\gamma + a(1-\gamma)]\}$. The first term on the RHS measures the gain in good-state of nature income made possible by the loan. $A > 0$ is, as explained earlier, a measure of the government's dependence on political support from the interest group. And the term B is positive since $\pi_1 a$ must exceed $\pi_0[\beta\gamma + a(1-\gamma)]$ when reforms are actually implemented.[17]

The government's political dependence on the interest group, as measured by the value of A, represents an important influence on the magnitude of the good-state of nature repayment. The more dependent the government is on interest group support, the smaller is its repayment requirement in the good state of nature. As mentioned already, this political dependency is directly related to the interest group's ability to capture rent, γ, and the government's ability to share in this rent, β. It is negatively related to the support influence of the public, as expressed by the value of a. More generally, one can state

Proposition 2 *The more successful the interest group is in appropriating rent, the more successful the government is in sharing this rent in form of financial contributions, and the less important the general public is in the political calculus of the government, the smaller is the aid-receiving government's repayment to the IFI and the more likely it is that the IFI's financial stability constraint is violated.*

The IFI's budget constraint was stated in (10.12). After substitution of (10.13) and (10.14), financial stability of the IFI requires that

$$E(z_1) = \pi_1\bar{z}_1 + (1-\pi_1)\underline{z}_1 = [\pi_1\bar{y} + (1-\pi_1)\underline{y}]\Delta g_1 - \pi_1 A/B \geq T_1 \tag{10.15}$$

To satisfy the budget constraint, the expected loan repayment must be at least as large as the value of the loan. Expected repayment, in turn, equals the expected gain in income from the loan when the economy is reformed minus a term whose value rises with the government's dependency on the interest group. More specifically, one can show that $\partial E(z_1)/\partial\gamma<0$ and $\partial E(z_1)/\partial\beta<0$, whereas $\partial E(z_1)/\partial a > 0$. Finally, we note that it is quite possible that the extra cost of creating incentives to reform the economy is sufficiently high for the IFI's budget constraint to be violated when the expected return on a conditional loan is rather small. This is particularly likely when there are strong bonds between interest group and government.

Is unconditional aid an alternative?

Offering a loan that is contingent on reforming the economy is not the only way of aiding a country. There exists the alternative of offering unconditional aid: the IFI provides assistance without conditioning it on the implementation of economic reforms. This section examines the effectiveness of unconditional aid relative to viable conditional aid, as well as the use of unconditional aid when conditional aid is no longer viable.

Comparing the effectiveness of alternative aid instruments requires specification of an objective function for the IFI. We assume here that the IFI's goal is to maximize expected income of the entire world; that is, of the aid-receiving country under consideration and the aid-providing rest of the world. Expected net income of the aid-receiving country is expected income made possible by aid minus expected repayments

$$g(T_i)[\pi_i\bar{y} + (1 - \pi_i)\underline{y}] - [\pi_i\bar{z}_i + (1 - \pi_i)\underline{z}_i] \tag{10.16}$$

where $i = 0$ when aid is unconditional and the economy is not reformed, whereas $i = 1$ when aid is conditional and the economy is reformed. Expected net income of the rest of the world is

$$E(Y^*) - T_i + [\pi_i\bar{z}_i + (1 - \pi_i)\underline{z}_i] \tag{10.17}$$

where $E(Y^*)$ denotes the rest of the world's expected income in the absence of any aid provision. The objective function of the IFI, therefore, is

$$W(T_i) = g(T_i)[\pi_i\bar{y} + (1 - \pi_i)\underline{y}] + E(Y^*) - T_i \tag{10.18}$$

where $W(T_i)$ denotes expected income of the world as a whole when aid is T_i. The IFI's goal is to maximize $W(T_i)$ with respect to T_i, without compromising the aid-receiving country's political and the IFI's financial stability. With conditional aid, there is the additional requirement that the repayment conditions contain sufficient incentives for the government to implement the called-for economic reforms.

For conditional aid, the relevant constraint set was already specified as (10.8′), (10.9′), (10.10), (10.11), and (10.12). After solving (10.8′)–(10.11) for $\bar{z}_1$ and $\underline{z}_1$, their substitution in (10.12) yielded (10.15). It is useful to restate this reduced-form financial stability constraint of the IFI as

$$E(Y_1)g(T_1) - [E(Y_1) + R + T_1] \geq 0 \tag{10.15′}$$

where $E(Y_1) = [\pi_1\bar{y} + (1-\pi_1)\underline{y}]$ denotes expected income of the reformed economy in the absence of aid, $[g(T_1)-1] = \Delta g_1$, and $R = \pi_1 A/B > 0$ is the expected limited liability rent.

The IFI maximizes its objective function of (10.18), setting $i = 1$, subject to the constraint of (10.15′). The Kuhn-Tucker conditions require that

$$E(Y_1)g'(T_1^*) \leq 1 \tag{10.19}$$

and that (10.15′) holds at T_1^*, where $T_1^* \geq 0$ denotes the IFI's world welfare-maximizing choice of conditional aid. If $E(Y_1)g'(0) > 1$, such that transferring a small amount of resources from the rest of the world to a reformed developing economy raises world welfare, then $T_1^* > 0$ as long as the IFI's financial stability constraint of (10.15′) is satisfied. If, however, the expected limited liability rent R is so high that the IFI's financial stability is compromised, then no conditional aid will be forthcoming.[18] The more rent the interest group appropriates in the absence of reforms, the more the government shares in this rent, and the less influential the general public is, the larger is R and the more likely it is that the IFI cannot afford to offer conditional aid.

For unconditional aid, (10.18) serves again as the objective function of the IFI, setting $i = 0$. The constraint set under unconditional aid, however, is quite different from the constraint set under conditional aid, since the IFI no longer is concerned with creating incentives for the government to reform. Consequently, the interest group remains active even after aid is given. There no longer is an incentive constraint, and the participation constraint becomes

$$\begin{aligned}&\left\{\pi_0\left[g(T_0)\bar{y} - \bar{z}_0\right] + (1-\pi_0)\left[g(T_0)\underline{y} - \underline{z}_0\right]\right\}\\ &\quad + \beta\left[\bar{C}^2 - \bar{C}^0\right]/\left[a(1-\gamma)\right] \geq [\pi_0\bar{y} + (1-\pi_0)\underline{y}]\end{aligned} \tag{10.20}$$

where $\bar{C}^2 = \gamma\{\pi_0[g(T_0)\bar{y} - \bar{z}_0] + (1-\pi_0)[g(T_0)\underline{y} - \underline{z}_0)]\}$ and $\bar{C}^0 = \gamma\{\pi_0\bar{y} + (1-\pi_0)\underline{y}\}$ are the maximum interest group contributions in the absence of reforms with and without an unconditional loan, respectively. Substitution of these contribution expressions reduces (10.20) to

$$\left\{\pi_0\left[\bar{y}\Delta g_0 - \bar{z}_0\right] + (1-\pi_0)\left[\underline{y}\Delta g_0 - \underline{z}_0\right]\right\} \geq 0 \tag{10.20′}$$

where $\Delta g_0 = [g(T_0)-1]$. In addition to (10.20′), the political stability constraints of (10.10) and (10.11), as well as the IFI's financial stability constraint of (10.12)

must be satisfied, with the only modification that loan repayments in the good and bad state of nature are now $\bar{z}_0$ and $\underline{z}_0$, respectively, rather than $\bar{z}_1$ and $\underline{z}_1$.

The Appendix shows that, for the amount of unconditional aid that maximizes the objective function of (10.18), T_0^*, there exists a nonempty set of loan repayment values, $\bar{z}_0$ and $\underline{z}_0$, such that participation, political stability, and IFI financial stability constraints are met. Accordingly, the IFI's world welfare-maximizing amount of unconditional aid is determined by the condition:

$$E(Y_0)g'(T_0^*) \leq 1 \tag{10.21}$$

where $T_0^* > 0$ if $E(Y_0)g'(0) > 1$.

The IFI's choice of conditional versus unconditional aid is stated as

Proposition 3

- *a* *If the expected net return on conditional aid is nonpositive at all aid levels, then neither conditional nor unconditional aid is provided; that is, if* $E(Y_1)g'(0) \leq 1$, *then* $T_1^* = T_0^* = 0$.
- *b* *If the expected net return on conditional aid is positive for some aid level and the limited liability rent is sufficiently small, then* $T_1^* > T_0^* \geq 0$ *and conditional aid yields higher world welfare than unconditional aid; that is, if* $E(Y_1)g'(0) > 1$ *and R is sufficiently small, then* $T_1^* > T_0^* \geq 0$ *and* $W(T_1^*) > W(T_0^*) \geq W(0)$.
- *c* *If the expected net return on unconditional aid is positive for some aid level and the limited liability rent under conditional aid is sufficiently large, then it is best for the IFI to give aid without attaching conditions; that is, if* $E(Y_0)g'(0) > 1$ *and R is sufficiently large, then* $T_1^* = 0 < T_0^*$.

Part (a) of Proposition 3 follows directly from (10.19) and (10.21) and the fact that $E(Y_1) > E(Y_0)$. Parts (b) and (c) can be explained with the help of Figure 10.1 which traces the $E(Y_i)g(T)$ functions for $i = 0, 1$. The choice of conditional aid is determined by (10.19), which requires that the slope of the $E(Y_1)g(T)$ function is equal to 1. The tangency point is marked by A_1, yielding conditional aid $T_1^* > 0$, provided the IFI's financial stability constraint (10.15′) is satisfied. Distance A_1C measures $E(Y_1)g(T_1^*)$, while distance BC measures the value of $E(Y_1)T_1^*$. If the limited liability rent is sufficiently small, such that R is less than distance A_1B, then $T_1^* > 0$ is the optimal choice. If, on the other hand, the limited liability rent is so large that R exceeds distance A_1B, then no conditional aid is provided. The choice of unconditional *aid* is determined by equation (10.21), which requires that the slope of the $E(Y_0)g(T)$ function is equal to 1. This happens at point A_0, yielding unconditional aid $T_0^* > 0$. Since $E(Y_0)g'(T) < E(Y_1)g'(T)$ for all T, it must be that $T_1^* > T_0^* > 0$ as long as $E(Y_0)g'(0) > 1$.

When the IFI's choice for both conditional and unconditional aid is positive, world welfare under conditional aid, $W(T_1^*)$, must always exceed world welfare

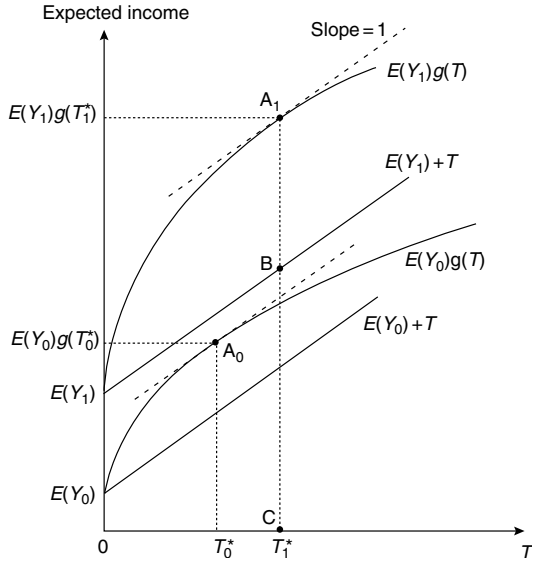

Figure 10.1 Contidional vs unconditional aid choices.

under unconditional aid, $W(T_0^*)$; that is,

$$E(Y_1)g(T_1^*) + E(Y^*) - T_1^* > E(Y_0)g(T_0^*) + E(Y^*) - T_0^* \tag{10.22}$$

To show this, we add and subtract $E(y_1)g(T_0^*)$ on the LHS of (10.22) and rewrite it as

$$\left[E(Y_1)g(T_1^*) - T_1^*\right] - \left[E(Y_1)g(T_0^*) - T_0^*\right] + \left[E(Y_1)g(T_0^*) - E(Y_0)g(T_0^*)\right] > 0 \quad (10.22')$$

Equation (10.22′) must hold since, evaluated at $E(Y_1)$, the first bracketed term on the LHS of (10.22′) must always exceed the second bracketed term. That is so, because the value of $[E(Y_1)g(T) - T]$ is maximized when T_1^* rather than T_0^* is chosen. Furthermore, the third bracketed term must be positive since $E(Y_1) > EY(_0)$.

Concluding remarks

Programs of economic adjustment and reform supported by international financial institutions have a mixed record of accomplishing their objectives. While IFIs almost always succeed in getting repaid, the policy programs supported by their

loans do not always succeed in their objectives of reforming the recipient countries' economies, improving efficiency, and raising economic growth. This chapter argued that asymmetries of information about the true implementation of reforms and the influence of special interests are at the root of such failures. It assessed the viability of IFI-supported programs using a model with three strategic actors: the government of an assistance-receiving country, a special interest group that influences the government's policies, and an aid-providing IFI that is willing to assist conditional on the implementation of economic reforms. In the absence of conditional assistance, the government has no incentive to adopt economic reforms. It is content receiving the interest group's political backing, in exchange for which it adopts distorted economic policies that allow the interest group to earn substantial economic rents. In order for IFI-supported programs to succeed in delivering economic reforms in this context, the interest group must lose this privileged position.

The model illustrates the difficult tradeoffs involved in designing IFI-supported programs and the consequences of ignoring viability constraints. The government must have an incentive to come to the IFI for support (the participation constraint) and also to implement the agreed-upon reforms (the incentive constraint). The program must also ensure that the entire country's political stability is maintained (the political stability constraints), and safeguard the IFI's resources (the IFI financial constraint). As Proposition 1 demonstrates, programs that are just attractive enough for a government to accept, without endangering the country's political stability and the IFI's financial solvency, always fail in the sense of resulting in no reforms. This type of nonviable conditional program clearly serves no one's interests: international assistance fails to catalyze reforms, which hurts the general public in assistance-receiving countries. The government maintains its close links with special interests and fails to deliver reforms despite program announcements. What's more, such programs do not serve the IFIs' own objectives of promoting efficiency-enhancing reforms and damage their credibility. The quality of IFI policy programs' seal of approval is lowered and the catalytic effect of IFI financing on private capital flows is reduced. Naturally, governments are reluctant to discontinue—or reduce the extent of—their privileged political relationship with influential interest groups, especially during noncrisis times. It may take an acute financial or political crisis for such a relationship to end. IFIs ought to pay special attention to incentive compatibility constraints in approving programs.

Programs that fail to ensure incentive compatibility may also lead to policy recidivism. Following the end of IFI loan disbursements, governments will face stronger incentives to continue their political relationship with interest groups. Unless reforms are not easily reversible (see below), governments will be tempted to go back to the bad old days and the policy gains and credibility earned during the period of IFI engagement will be short-lived. In such situations, IFIs will face pressures of their own to grant succession programs, especially if the economic situation deteriorates sufficiently. Such a cycle of stop-go program engagement and reform could well contribute to inappropriately prolonged use of IFI financing.

The extent to which policy recidivism is a problem clearly depends on the domestic political system. To a large extent, reform laws and regulations become a dead letter if they are resisted by implementing ministries and agencies in all countries. But reform and change may be particularly vulnerable to sabotage in many developing countries in which officials have broad discretionary powers of interpretation and enforcement of laws, rules, and regulations. Judicial and civil service reform to make bureaucracies in these countries more transparent and accountable could have important payoffs for the quality of governance in these countries and also lead to better implementation of IFI-supported programs. IFI-supported reforms could become less reversible if they are passed by parliament (rather than issued as government decrees), they are subject to public debate among stakeholders, and are highly publicized as prior actions or other conditions.

The chapter has also highlighted the need to respect a country's political stability constraint. Failure to do so is another cause for lack of program viability. In practice, IFI mission chiefs and country officials are well positioned to assess the immediate dangers to political stability during the period of program engagement. On the other hand, the parties' incentives after IFI disbursements end are less clear, especially given the turnover of elected officials and IFI mission teams. The paper highlights the need to asses the prospects for political stability at repayment time as well. The willingness of IFIs to restructure debts *ex post* (at repayment time) suggests that the political stability constraint is usually accounted for.

There have been (relatively few) cases in which programs fail to ensure the IFI's own financial integrity, presumably because IFIs are senior creditors and are also well aware of these constraints. IFIs have been exposed to significant financial risk in a few cases in which countries in social, political, and economic turmoil ran protracted arrears on their multilateral loans. Generally speaking though, failure to implement reforms does not result in stopping debt servicing to IFIs. Member countries continue making payments to IFIs even if they are under considerable financial and economic stress. But the incipient financial risk to IFIs is greater than suggested by the low frequency of protracted arrears. In a number of cases, IFIs avoided incurring arrears only by resorting to defensive lending—rolling over maturing loans without requiring additional reforms as a *quid pro quo*.

The difficulty in meeting all program viability constraints raises the question of whether there are ways to help a country avoid the pitfalls of conditional assistance. It is possible (Proposition 3) for the IFI to substitute unconditional loans for conditional loans when the strength of the interest group rises and conditional assistance fails to meet the IFI's budget constraint. A key question then is whether the IFI would want to lend to such a country. This depends on the IFI's objective function and the characteristics of borrowing countries. There could be a scenario for lending to systemically important countries (those whose policy externalities affect significantly other countries), as in the model presented by Mayer and Mourmouras (2002). As IFIs are largely rules-based institutions, uniformity of treatment considerations prevent

them from treating such borrowers preferentially. But uniformity of treatment and conditionality addressed to systemic borrowers could be watered down in fact, if not in principle.

The chapter's key finding is that the IFIs' information gap *vis-à-vis* borrowing governments regarding the true implementation of reforms is a key "fundamental" that should be considered in deciding the extent and nature of IFI assistance. If the information rent accruing to the interest group-influenced government becomes too high, then it is not worthwhile to provide assistance. IFIs must be selective in lending to opaque governments and undecipherable economic environments, a direction in which they have been moving in recent years. Improvements in information technology and international policy initiatives have greatly increased the transparency of economic data and policies during the last decade. This development is bound to reduce IFI informational disadvantages and limit the information rents accruing to assistance-receiving governments. Greater transparency could result in fewer nonviable conditional assistance programs being signed, improve the implementation rate of IFI-supported reforms, and help curtail the power of vested interests in these countries.

In practice, IFIs exert influence both through financial assistance and the provision of information and analytical support. Unlike the model presented in this chapter, in which information asymmetries are one-way and favor the assistance-receiving government, IFIs possess much specialized knowledge, including on the structural reforms needed to address specific weaknesses. The existence of such two-sided asymmetries of information may, to some extent, neutralize the government's rent and modify program viability constraints. More importantly though, IFIs need to take into account the pervasive *imperfections* in their own economic knowledge and the environment of Knightian uncertainty in which they operate. Like the economics profession at large, IFIs do not have the answers to many important questions about the nature, sequencing, and pacing of structural reforms. It is then an open question stating conditions under which IFIs should refrain from exerting their financial influence.

In deciding where to draw the line, IFIs should bear in mind the main point of this chapter: governments in assistance-receiving governments are in charge of reform implementation and necessarily maintain significant informational advantages *vis-à-vis* IFIs in this area; and the gains of democracy around the world notwithstanding, political systems in many developing countries remain highly imperfect. Many governments continue to obtain significant political support from interest groups that oppose efficiency-enhancing reforms. Information rents and special interest politics therefore give rise to incentives for politicians to sign programs with IFIs that they do not intend to implement fully. IFIs need to be appropriately selective and strive to ensure that approved programs contain sufficient incentives for assistance-receiving governments to adopt the envisaged reforms and avoid reform backtracking and reversals.

Appendix

Binding constraints when aid is conditional

We show that the binding *incentive constraint* of (10.9′) and the binding bad-economy political stability constraint of (10.11) imply that the participation constraint of (10.8′) and the good-economy political stability constraint of (10.10) are satisfied but not binding.

Stating (10.9′) as an equality and subtracting $a\pi_0\left[g(T_1)\bar{y} - \bar{z}_1\right] + (1 - \pi_0)$ $[g(T_1)\underline{y} - \underline{z}_1)$ from both sides, yields

$$a\Delta\pi\left\{\left[g(T_1)\bar{y} - \bar{z}_1\right] - \left[g(T_1)\underline{y} - \underline{z}_1\right]\right\}$$

$$= \gamma(\beta - a)\left\{\pi_0\left[g(T_1)\bar{y} - \bar{z}_1\right] + (1 - \pi_0)\left[g(T_1)\underline{y} - \underline{z}_1\right]\right\}$$

where $\Delta\pi = (\pi_1 - \pi_0) > 0$. Next, subtracting $a\Delta\pi(\bar{y} - \underline{y})$ from both sides of the above equation, and adding and subtracting $\gamma(\beta - a)\left[\pi_0\bar{y} + (1 - \pi_0)\underline{y}\right]$ on its RHS, results in

$$a\Delta\pi\left[(\bar{y}\Delta g_1 - \bar{z}_1) - (\underline{y}\Delta g_1 - \underline{z}_1)\right]$$

$$= A + \gamma(\beta - a)\left[\pi_0(\bar{y}\Delta g_1 - \bar{z}_1) - (1 - \pi_0)(\underline{y}\Delta g_1 - \underline{z}_1)\right] \tag{A.1}$$

where $A = -a\Delta\pi\Delta y + \gamma(\beta - a)\left[\pi_0\bar{y} + (1 - \pi_0)\underline{y}\right] > 0$, $\Delta y = (\bar{y} - \underline{y}) > 0$, and $\Delta g_1 = \left[g(T_1) - 1\right] > 0$.

The RHS of equation (A.1) must be positive since $A > 0$, whereas $(\underline{y}\Delta g_1 - \underline{z}_1) = 0$ by assumption and $(\bar{y}\Delta g_1 - \bar{z}_1) \geq 0$ from (10.10). It follows that the LHS of (A.1) must be positive and that the good-state political stability constraint is not binding; that is

$$(\bar{y}\Delta g_1 - \bar{z}_1) > 0 \tag{A.2}$$

Checking the participation constraint of (10.8′), we substitute the RHS of the binding incentive constraint of (10.9′) on the LHS of (10.8′), subtract $\left[a + \gamma(\beta - a)\right]\left[\pi_0\bar{y} + (1 - \pi_0)\underline{y}\right]$ from both sides of the restated participation constraint, and obtain

$$\left[a + \gamma(\beta - a)\right]\left\{\pi_0\left[\bar{y}\Delta g_1 - \bar{z}_1\right] + (1 - \pi_0)\left[\underline{y}\Delta g_1 - \underline{z}_1\right]\right\} \geq 0 \tag{A.3}$$

Since $(\bar{y}\Delta g_1 - \bar{z}_1) > 0$ and $(\underline{y}\Delta g_1 - \underline{z}_1) = 0$, the participation constraint is satisfied but not binding.

Constraint set when aid is unconditional

When aid is unconditional, the constraint set consists of the government's participation constraint of (A.4)—which simply restates (10.20′) of the main text—the

two political stability constraints of (A.5) and (A.6), as well as the IFI's financial stability constraint of (A.7)

$$\{\pi_0[\bar{y}\Delta g_0 - \bar{z}_0] + (1-\pi_0)[\underline{y}\Delta g_0 - \underline{z}_0]\} \geq 0 \tag{A.4}$$

$$g(T_0)\bar{y} - \bar{z}_0 \geq \bar{y} \tag{A.5}$$

$$g(T_0)\underline{y} - \underline{z}_0 \geq \underline{y} \tag{A.6}$$

$$[\pi_0\bar{z}_0 + (1-\pi_0)\underline{z}_0] \geq T_0 \tag{A.7}$$

First, one can see that (A.5) and (A.6) imply (A.4) since $[g(T_0-1] = \Delta g_0$. Second, one can show that (A.5) and (A.6), when binding at the optimal unconditional aid choice of T_0, imply that (A.7) is always satisfied but not binding. This can be seen after writing (A.5) and (A.6) as equalities, substituting for $\bar{z}_0$ and $\underline{z}_0$ in (A.7), and restating (A.7) as

$$\left[E(Y_0)g(T_0^*) - E(Y_0)\right]/T_0^* \geq 1 \tag{A.8}$$

The optimal choice of $T_0 > 0$, in turn, requires that $E(Y_0)g'(T_0^*) = 1$; and strict concavity of the $g(T)$ function implies that $\left[E(Y_0)g(T_0^*) - E(Y_0)\right]/T_0^* > E(Y_0)g'(T_0^*) = 1$, as can be seen from Figure 10.1. Consequently, it must be that, for the value of T_0^*, (A.7) is not binding for that combination of $(\bar{z}_0, \underline{z}_0)$ for which constraints (A.5) and (A.6) are binding.

The set of $(\bar{z}_0, \underline{z}_0)$ combinations which satisfy constraints (A.5), (A.6), and (A.7) is shown in Figure 10.2 as combinations in area ABC. Based on the discussion of the preceding paragraph, the IFI's financial stability constraint must lie to the left of point A since (A.7) is not binding at the point where (A.5) and (A.6) are binding.

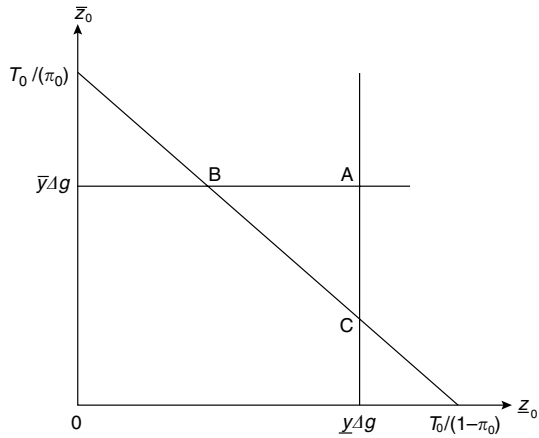

Figure 10.2 Repayment constraint set at the optimal unconditional aid choice.

Notes

* For useful comments on an earlier version of the paper we like to thank Burkhard Drees and Miguel Messmacher at the IMF, and Cem Karayalcin, Catherine Mann, T. N. Srinivasan, Devashish Mitra, and other participants in the Conference on "Globalization: Prospects and Problems" in honor of Jagdish Bhagwati's seventieth Birthday. Special gratitude goes to the University of Florida for hosting the event and to the conference organizers, especially Elias Dinopoulos.

1 Mussa and Savastano (1999: 19), in discussing the IMF's approach to economic stabilization, refer to the securing of sustainable external financing, the adoption of demand-restraining measures, and the implementation of structural reforms as the main prongs of the stabilization approach. Later they add that conditionality of structural reforms has become an increasingly important component of assistance programs, especially when designed by IFIs other than the IMF.

2 There has been a great deal of criticism of IFI-designed conditional assistance programs, especially when they have a strong structural reform component. Most prominent among the themes of criticism is that of a lack of "ownership" of reforms. As Mussa and Savastano (1999: 7) state, "IMF programs tend to perform best when their associated policies are most closely 'owned' by the national authorities in charge of implementing them." Ownership of reform policies, as the key to successful implementation of IFI-supported reforms, is also highlighted in IMF (2001) and Boughton and Mourmouras (2004).

3 Program viability is related to, but not the same as, the empirically tested "success" of conditional assistance programs. Dollar and Svensson (2000) consider a program to be a success or failure based on the assessment of the Operations Evaluation Department of the World Bank. The latter makes an overall judgment of whether the larger objectives of the reform have been attained. Hence, its focus is on whether reforms have been implemented. What impact the program has on the financial stability of the IFI is not being evaluated.

4 Some authors addressing information asymmetries in international markets have emphasized the screening and signaling role of IFIs (IMF, 2004b; Marchesi and Thomas, 1999; and Tirole, 2002).

5 Case studies suggest that failure to ensure incentive compatibility was a problem in several IFI-supported programs. In recent years, a number of borrowers misreported information to IFIs and used the proceeds of IFI loans for purposes that were contrary to the objectives of these programs. The IFIs have responded by strengthening their procedures to deal with mis- and nonreporting of information and by strengthening the audits of central banks of assistance-receiving countries (IMF, 2000, 2004a). These technical steps are in the right direction and will likely reduce the extent of the IFIs' information gap going forward.

6 Dollar and Svensson (2000) find that political economy variables have a strong influence on the success rate of conditional assistance programs. A similar conclusion is drawn by Ivanova *et al*. (2006).

7 One of the pioneers of studying the influence of interest groups on a country's welfare is Bhagwati (1980, 1982). Interest groups engage in "directly unproductive" (DUP), profit-seeking activities that lower welfare for the entire economy. The welfare consequences of these lobbying activities are further explored in Bhagwati and Srinivasan (1982).

8 Drazen (2002) and Mayer and Mourmouras (2002, 2005) have modeled the relationship between IFI and an interest group-influenced government in a setting of symmetric information.

9 This echoes the widely held belief that "ownership" of policies is essential for successful economic reforms.

10 Our political stability constraint is similar to Acemoglu and Robinson's (2006) revolution constraint. They work with a two-class, deterministic median-voter model and do

not consider external assistance. This is to be contrasted with our model of special interest politics under asymmetric information.

11 In reality, IFIs provide various relief measures to countries hit by bad economic conditions, including debt relief and roll-overs of liabilities to future periods. While they are usually negotiated *ex post*, assistance-receiving governments are aware of the availability of such relief measures at the time they sign the agreement. Hence, the availability of such measures has effects that are similar to a priori specified flexible repayment conditions.

12 We abstract from IFI operational expenses and the need for dividend payments to its creditor countries.

13 Increasing the number of groups would have a moderating influence on policy distortions, provided these groups have competing interests. To show this explicitly requires the formulation of a multisector economy and the discussion of what policy instrument raises a given group's welfare. This considerable complication would add little to enhance our understanding of program viability constraints in developing countries. In these countries political systems are highly imperfect and some groups (the military, the large land owners or crony industrialists) dominate the political influence process while other groups (the rural peasants) are not represented in the political system.

14 This form of the political support function was popularized by Grossman and Helpman (1994). A more general specification of the political support function was first introduced by Hillman (1982).

15 The general public of the loan-receiving country benefits from the assistance program in two ways: First, the IFI loan raises national income beyond what it would be under sole reliance on private capital markets. Second, the implementation of economic reforms raises expected national income and destroys (or at least weakens) the interest group, thereby enlarging the public's share of national income. The political-support-seeking government weighs this gain to the general public against the loss from vanishing interest group contributions.

16 As discussed in the Introduction, it is the *ex post* performance of the economy that matters for political stability of the country. In contrast, the government's decision to accept the assistance package is based on *ex ante* considerations.

17 The political stability constraint in the bad state of nature is always binding, and the aid-receiving country can raise its net income from assistance only when the good state of nature comes about. This good state comes about with probability of $\pi_1$when the economy is reformed and with probability of π_0 when there are no reforms. With reforms, the interest group collapses and all extra income goes to the public which, in turn, generates 'a' of political support from each extra unit of income gained. Hence, the expected gain in political support to the government from another unit of income gained by the reformed economy is $\pi_1 a$. With no reforms, the interest group remains in power and gains γ of each additional unit of income and contributes $\beta\gamma$ per income unit to the government. In addition, the government receives political support from the public equal to $a(1-\gamma)$ per unit of income gain. Hence, the expected gain in political support from another unit of income gained by the unreformed economy is $\pi_0[\beta\gamma + a(1-\gamma)]$. For the government to accept aid and implement reforms, it must be that $\pi_1 a > \pi_0[\beta\gamma + a(1-\gamma)]$.

18 Note that (10.15′) is always satisfied when $E(Y_1)g'(0) > 1$ and $R = 0$ since $g''(T) < 0$.

Bibliography

Acemoglu, D. and Robinson J. A. (2006) *Economic Origins of Dictatorship and Democracy*, New York: Cambridge (UK) University Press.

Bhagwati, J. N. (1980) "Lobbying and Welfare," *Journal of Public Economics*, 14: 355–364.

—— (1982) "Directly-Unproductive Profit-Seeking (DUP) Activities," *Journal of Political Economy*, 90: 988–1002.

Bhagwati, J. N. and Srinivasan, T. N. (1982) "The Welfare Consequences of Directly Unproductive Profit-Seeking (DUP) Lobbying Activities: Price Versus Quantity Distortions," *Journal of International Economics*, 13: 33–44.

Boughton, J. M. and Mourmouras A. (2004) "Whose Programme Is It? Policy Ownership and Conditional Lending," in D.Vines and C. L. Gilbert (eds) *The IMF and Its Critics: Reform of Global Financial Architecture*, Cambridge, UK: Cambridge University Press.

Dixit, A. (2000) IMF Programs as Incentive Mechanisms, mimeo, International Monetary Fund (June). Online. http://www.princeton.edu/~dixitak/home/incentive_schemes.doc

—— (2001) Some Lessons from Transaction-Cost Politics for Less-Developed Countries, Olson Memorial Lecture delivered at the University of Maryland, October 26. Online. http://www/bsos.umd.edu/umcc/olson_memorial_lecture_series_schedule.htm

Dollar, D. and Svensson J. (2000) "What Explains the Success or Failure of Bank-Supported Adjustment Programs," *Economic Journal*, 110: 894–917.

Drazen, A. (2002) Conditionality and Ownership in IMF Lending: A Political Economy Approach, *IMF Staff Papers*, 49: 36–67. Online. <http://www.imf.org/External/Pubs/FT/staffp/2001/00–00/pdf (accessed April 14, 2006).

Grossman, G. M. and Helpman, E. (1994) "Protection for Sale," *American Economic Review*, 84: 833–850.

Hillman, A. L. (1982) "Declining Industries and Political-Support Protectionist Motives,"*American Economic Review*, 72: 1180–1187.

IMF (2000) IMF Adopts Added Safeguards on Use of Its Resources, Public Information Notice No. 00/28 (April 4). Washington DC: International Monetary Fund. Online. <http://www.imf.org/ external/np/sec/pn/2000/pn0028.htm> (accessed April 14, 2006).

—— (2001) Strengthening Country Ownership of Fund-Supported Programs, (December 5) Washington DC: International Monetary Fund. Online. <http://www.imf.org/external/np/pdr/cond/2001/ eng/strength/120501.pdf> (accessed April 14, 2006).

—— (2004a) IMF Approves Decisions Strengthening the Effectiveness of the Legal Framework for the Provision of Information to the Fund, Public Information Notice No. 04/10 (February 23), Washington DC: International Monetary Fund.

—— (2004b) Signaling by the Fund—A Historical Review, Policy Development and Review Department (July 16), Washington DC: International Monetary Fund. Online. <http://www.imf.org/external/np/pdr/signal/2004/071604.pdf> (accessed April 14, 2006).

Ivanova, A., Mourmouras, A., Mayer, W. and Anayiotos, G. (2006) "What Determines the Implementation of IMF-Supported Programs?" Chapter 10, in A. Mody and A. Rebucci (eds), *IMF-Supported Programs: Recent Staff Research*, Washington DC: International Monetary Fund. Online. <http://www.imf.org/external/pubs/ft/wp/2003/wp0308.pdf> (accessed April 14, 2006).

Marchesi, S. and Thomas, J. P. (1999) "IMF Conditionality as a Screening Device," *Economic Journal*, 109: 111–125.

Mayer, W. and Mourmouras A. (2002) "Vested Interests in a Positive Theory of IFI Conditionality," IMF working paper 02/73, Washington, DC: International Monetary Fund. Online. <http://www.imf.org/external/pubs/ft/wp/2002/wp0273.pdf> (accessed April 14, 2006).

—— (2005) "The Political Economy of IMF Conditionality: A Common Agency Model," *Review of Development Economics*, 9: 449–466.

Mussa, M. and Savastano M. (1999) "The IMF Approach to Economic Stabilization," IMF working paper 99/104 (July), Washington, DC: International Monetary Fund. Online. <http://www.imf.org/external/pubs/ft/wp/1999/wp99104.pdf> (accessed April 14, 2006).

Tirole, J. (2002) *Financial Crises, Liquidity, and the International Monetary System*, Princeton, NJ: Princeton University Press.

11 Is the United States a large country in world trade?

Further evidence and implications for globalization

Stephen P. Magee, Kwang-Yeol Yoo, Nakgyoon Choi, and Hong Shik Lee

Introduction

The United States does not satisfy two necessary conditions for being a large country in world trade with monopsony power: sufficiently high market shares to control prices for nontransitory periods and the ability to change world prices using protection.

The market share evidence

The United States only has a 17% share of aggregate world imports; it has low market shares of world imports for 232 out of 236 three-digit SIC products and over 30% market shares for only four products (art, toys, cassette recorders, and pottery); the US does not have high protection for the most important products in world trade; and it does not have high protection for products with large percentages of the product exported to the United States.

The terms of trade evidence

United States protection has significantly affected the terms of trade for about 15% of products in world trade; except for Canada and Mexico, US protection has only affected the US terms of trade with 4% of world trade with its bilateral country trading partners; and aggregate US protection has not affected the US aggregate terms of trade since 1934.

Stephen Magee, one of the coauthors of this chapter, takes great pride in having worked under Jagdish Bhaqwati and finishing his PhD in 1969, being perhaps the first of Bhagwati's Ph.D. students out of MIT. Bhagwati's theoretical work on immizerization was empirically prophetic, explaining the 50% decline in the terms of trade of food producers that has occurred for the last century.[1] Magee learned the theory of distortions from Jagdish, and wrote his dissertation on factor market distortions. His career has been many times blessed by Jagdish's encouragement and advice.

A recent search for "Jagdish Bhagwati" on Google yielded over 345,000 hits. A search of Amazon yielded 82 hits, which is more books than most academics

write papers in their careers. While Jagdish is over the top in all that he does, he is a warm and congenial person, perhaps because of the gentle wisdom and patience of his brilliant wife, Padma Desai. Former US President Lyndon Johnson used to tell a story on himself that reminds us of Jagdish. Willy Brandt was Chancellor of Germany in the 1960s, and he did some background reading about Lyndon before meeting with him in the White House in Washington. Unfortunately, Brandt got Johnson's childhood mixed up with Abraham Lincoln's and asked Johnson if it was true that he had been born in a log cabin. Johnson replied, "No, I was born in a manger." While Jagdish did not have a Christian upnringing, like Johnson, he could have been born in a manger.

We present an ambitious chapter for this volume to pay appropriate homage to Jagdish—good guy, prophetic, and a brave leader of movements. One caveat: Chris Magee has warned the coauthors of this chapter that the line in his father's mind between the ambitious and the ridiculous is dangerously thin.

It is a widely held view in international economics that "the United States is a large country in world trade." The most common definition of a "large" country in international economics is that the country can control the world prices of its imports by changing its level of protection. Partly because of this, many thoughtful international economists for the last half century have complicated their models to incorporate large country effects. During the years we have worked on this paper, we have come to believe that the United States is not a large country in world trade.[2]

We advance two broad pieces of evidence that the United States is not a large country in world trade. The first deals with the small market shares of US imports in world trade. The second tests whether changes in US protection significantly improve the terms of trade of the US. If the United States is large, then an increase in its protection will generate a nontransitory improvement in the US terms of trade (the price of US exports over the price of US imports).[3]

The resistance of international economists to the possibility that the United States might be a small country has been explained by Magee and Magee (2005) as follows. All will agree that the United States, even with only a 17% share of world imports, is large enough to affect world prices by changing US protection. However, the standard in industrial organization for monopsony power in a market is that the United States must be able to control world prices. This is a much higher standard that requires high market shares and a sustained period of time required for prices to remain elevated. A country without monopsony power can increases its protection, affect world prices, and get transitory terms of trade welfare gains. However, it is not a large country because the welfare gains are quickly dissipated. Thus, we speculate that many international economists incorrectly believe that the United States is a large country because it can temporarily affect world prices.[4]

The US-large country hypothesis contradicts the historical fact that the United States supported free trade for nearly a half century after 1934.[5] If the United States is small, we have a simpler explanation of the free-trade bias of American foreign economic policy in the twentieth century than the convoluted political–economic tradeoffs required in the hegemonic literature.[6] The first work we did on this

chapter was a time-series test of whether US protection affected the US terms of trade. To our surprise, we found virtually no effects of US tariffs on the aggregate US terms of trade after 1934. This result was the first clue that the United States might not be a large country in world trade.[7] It is odd to us that the stylized fact of the United States being a large country in world trade has not been more explicitly tested.[8,9] We do not analyze the large country problem with respect to US exports.[10] The remainder of the chapter was done after the aggregate time series tests. We turn now to the three results.

Result I: the market share evidence

The US only has a 17% share of aggregate world imports

As Table 11.1 shows, the United States, the largest country in world imports, had only a 15% share of world imports in 1995. By 2003, that number had increased to 17% of world imports and 10% of world exports according to the CIA World Fact Book, 2004. In industrial organization, this is insufficient for a buyer to have monopsony power and the ability to control prices in a market. If the United States tried to lower world prices by raising it tariffs, foreign suppliers would substitute into the other 83% of the world market (over 200 other importing countries) to offset the attempted price control by the United States. Supplier substitutability to other national markets undermines the ability of a buyer with such a small market share to control prices.

A similar view of the low world concentration of world imports is provided by the Herfindahl–Hirshman Index (HHI) of market concentration used by industrial organization economists. The HHI is usually calculated to infer the likelihood of collusion among the largest sellers in markets. This measure is the sum of the market shares squared of suppliers in a market, with theoretical values ranging from 0 (e.g. farmers) to 10,000 (one seller). The measure does not prove the point we are making here, but it is instructive. We compare the concentration of world imports relative to domestic US product markets.

The US Justice Department and the US Federal Trade Commission have established the following cutoff points for the values of the HHI in evaluating mergers in markets: 0–1,000 as being unconcentrated, competitive markets; 1,000–1,800 as being markets with intermediate concentration; and above 1,800 as being highly concentrated markets (Carlton and Perloff, 1994, 805). The Herfindahl–Hirshman Index applied to importing countries indicates that world trade is highly unconcentrated. Table 11.2 shows that the HHI equals 543 for world imports and 497 for world exports. Both of these values are in the middle of the Justice Department-FTC's unconcentrated range of 0 to 1,000. Table 11.2 also shows how concentrated world imports are relative to 445 US four-digit SIC industries for which HHIs are calculated.

Note that 213 out of the 445 four-digit US SIC industries are more concentrated that world imports and 229 US industries are more concentrated that world exports. Nearly half of US industries are more concentrated than world imports.

Table 11.1 Shares of countries in world imports in 1995

	Rank of country in world imports	*Imports ($ billions)*	*%*
US	1	771.0	15.0
Germany	2	448.1	8.7
Japan	3	335.9	6.6
France	4	275.0	5.4
UK	5	263.8	5.1
Italy	6	204.1	4.0
Hong Kong	7	192.8	3.8
Canada	8	168.4	3.3
Belgium-Lux	9	154.2	3.0
S. Korea	10	135.1	2.6
China	11	129.1	2.5
Singapore	12	124.5	2.4
Spain	13	114.8	2.2
Taiwan	14	103.6	2.0
Africa		119.3	2.3
Middle East		137.3	2.7
South America		137.2	2.7
Other Asia		305.6	6.0
Eastern Europe		180.6	3.5
All others		825.4	16.1
Total		5,125.7	100.0

Source: *UNCTAD, Handbook of International Trade Statistics*, 1995.

Table 11.2 The concentration of world trade

HHI (Herfindahl–Hirshman Index) for world imports	543
HHI (Herfindahl–Hirshman Index) for world exports	497
US industries more concentrated than world imports	213
US industries less concentrated than world imports	229

Dividing the world import HHI into 10,000 gives 18 as the number of equal-sized importers in world trade. In short, the United States share of world imports is too low for world imports to be concentrated and for it to have monopsony power as a buyer in world markets.

The U.S. has low market shares of world imports for 232 out of 236 three-digit SIC products and over 30% market shares for only four products (art, toys, cassette recorders and pottery)

In industrial organization, we know that there are two necessary requirements for a country to exercise large-country power in world trade:[11] First, the country must have high enough product market shares to be able to increase their terms of trade (the price of exports over the price of imports)[12] by increasing protection. Second,

there must be barriers to exit so that foreign exporters exploited by higher US protection cannot shift their sales to other countries. Consider first the market share question.

The American industrial organization literature indicates that prices do not show elevated levels except in industries with four-firm concentration ratios above 50% (Carlton and Perloff, 1994, 358).[13] Work on the contestability of markets indicates individual firms with market shares over 70% may still be unable to control prices in the absence of entry and exit barriers. It is hard to find shares this high in markets without tacit or explicit collusion among the largest sellers. In world trade, large countries do not band together to collude in this way. In fact, the reverse is true: exporting countries have coordinated price increases through voluntary export restraint agreements. The United States has been a buyer rather than a seller of products subject to such agreements.

It must be true that the United States has market power in at least some products in world trade, that is, those in which it has the largest shares of world imports. In this section, we examine the US shares of three-digit SITC product markets applied to all of world trade. We then examine US product shares of world trade for only trade with the geographically closest US trading partners.

After over 100 years of American antitrust law, case law prohibits an antitrust case against defendants from proceeding past the initial summary judgment phase of a case unless the defendant firm (or group of conspiring firms) has a market share in excess of 30%. At time of trial, the antitrust hurdle is higher, requiring that large defendant firms have at least a 50% share of the market. While the United States is the world's largest importer, there are few products in which the United States has large shares.

Note in Table 11.3 that there are no products out of 236 three-digit SITC products in which imports into the United States exceed the 50% hurdle. In fact, there are only four products out of 236 in which the United States has more than 30% of world imports—art (40%), toys (31%), cassette recorders (30%), and pottery (30%). For those four products, the United States has insufficient shares of world imports to control prices on world markets.

Is there any evidence that the United States has attempted to exercise its buyer power in these product markets? If so, we would expect higher than average protection or possibly quantitative restrictions on these imports. Further, we would expect these products to be important in world trade, otherwise, why would the United States bother? Neither of these speculations are true. Art has a zero duty for US imports and the remaining products have low duties. As Table 11.3 also shows, these products are unimportant in world trade, with total trade in the world for these products constituting less than 1.5% of total world imports and 3.2% of US imports.

Another place to look for US monopsony market power would be for those products for which the United States had high tariff equivalents. These are products for which imports into the United States are subject to quantitative restrictions such as US quotas or voluntary export restraints by foreign countries. If the United States had sufficient monopsony power to control world prices, it would

Table 11.3 There are only 4 products out of 236 for which US shares of world imports are over 30%; US tariffs on these products are low

SITC product number	*US % of world imports*	*The % of this product in world imports*	*This product as a % of US imports*	*Rank of this product in world imports*	*US tariff rate (%)*	*SIC code of US tariff*	*Product*
	1995	*1991–1992*	*1995*	*1991–1992*	*1992*		
896	40.2	0.21	0.3	135	0		Art
894	30.5	0.65	1.8	40	4.6	3944	Toys
763	30.1	0.47	0.8	63	3.9	3651	Cassette recorders
666	30.0	0.13	0.2	171	7.3	326 **	Pottery
Total/ Average		1.46	3.2	102	4.5*		

Source: Tables 11A and 13A; source of the last two columns is Chris Magee, except for Art.

Notes
* This is the trade-weighted average tariff for all U.S. manufactured imports.
** This is the average tariff for five of the components of SIC 326.

Table 11.4 US shares of world imports for products with high US tariff-equivalents

	Tariff equivalent of the restriction (%)	*SITC number Rev 2*	*US % of world imports*	*Average rank of SITC product in world trade*	*SITC product*
Automobiles	11	781	29	1	Cars
Petroleum (after 1973)	96 (0)	333	24	2	Crude petroleum
Textiles and	30	651–658	8	96	Textiles
Apparel		842–848	24	78	Apparel
Carbon steel	16	671–679	12	86	Iron and steel

have a high share of world imports in these products. The only four such products are shown in Table 11.4.

None of the products has the required 50% of world imports required to satisfy the industrial organization standard. In fact, there are none with even at 30% share. Only automobiles, petroleum, and apparel are over 20%. We examine next whether adoption of the import restrictions listed in Table 11.4 were related to terms of trade motivations by the United States.

The US oil import quota (in effect only from 1959 to 1973) was adopted in the late 1950s because of a special-interest group—the oil producing states in the US Congress. Lyndon Johnson was the Majority Leader of the Senate and Sam Rayburn was Speaker of the House of Representatives. These two Texans and other politicians from the oil states pressured the Congress and Republican

President and former General Dwight Eisenhower to impose the quota. The action was justified using national defense arguments (the United States should reduce its dependence on foreign oil).

The Voluntary Export Restrictions (VER) on autos began in 1981. These restrictions were proposed by the Japanese, who convinced Detroit to support them because the VER would raise prices of cars in the United States. This VER was controlled by four of the largest Japanese exporters of cars and used to prevent smaller rival Japanese car manufacturers from entering and competing with them in the United States after 1981. Furthermore, the auto VER allowed the Japanese auto companies to capture the tariff-equivalent revenue, which normally would have been captured by the US Treasury. The terms of trade gains created by this trade restriction, if any, were captured by the Japanese, not the Americans.

The textile and apparel and the carbon steel restrictions were also voluntary export restraint agreements. Although these products are widely discussed in treatises on American protection, these products are quantitatively unimportant in world markets—neither product is in the top 75 products by value in world trade. There can be no large-country story associated with these products since the United States imports only 8% of the world's textiles and 12% of the world's iron and steel. In short, there is no evidence of monopsony or market power motivations in the four products in which the United States has the highest levels of protection—cars, oil, textiles and apparel, and carbon steel.

We just considered the US share of products in all of world trade. But countries far away from the United States might have small shares of their trade exported to the United States because of distance. We analyze now only exports of products from countries that are geographically closest to the United States and whose cumulative trade constitutes 80% of world exports. The following table shows a sample of the countries from our database.

The Figure 11.1 shows the relationship between distance and share of total exports going to the United States for the closest 95 countries (which constitute 83.7% of world exports), the last of which was Japan, that was 95th in distance. Note that only 15 countries have more than 50% of their exports going to the United States.

Table 11.5 Trade of the geographically closest trading partners to the United States in 1985

Country	*Rank in distance*	*Distance from US km*	*Exports to world ($ bill)*	*Exports to US ($ bill)*	*Exports to US % of total*
Canada	1	734	38.600	33.972	88.0
Mexico	8	3,039	16.719	9.569	57.2
UK	25	5,904	57.928	9.471	16.3
France	29	6,169	41.352	4.643	11.2
Germany	36	6,406	79.382	10.660	13.4
Italy	50	7,222	28.978	4.719	16.3
USSR	83	9,857	20.812	0.053	0.3
Japan	95	10,910	80.965	32.962	40.7

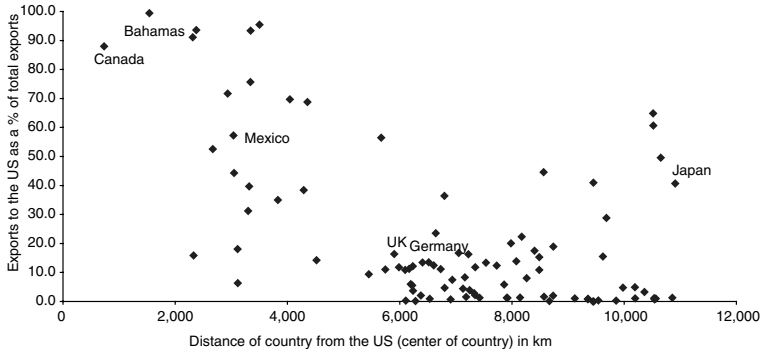

Figure 11.1 Percent of country exports to the United States in 1985 vs distance from the United States.

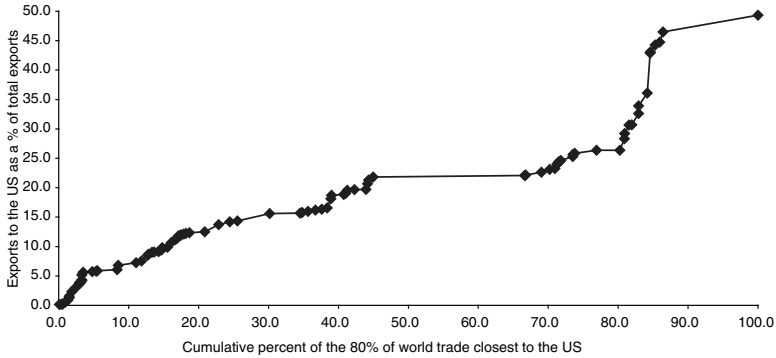

Figure 11.2 Percent of goods exported to the United States (Y) vs percent of total trade closest to the United States (X).

When we analyze the products shipped from these countries, we still see that the United States still does not have the high market shares we might have expected. Note in the Figure 11.2 that for this data set, the United States is less than 30% of world imports for over 80% of the data set's total trade. In addition, for none of these products does the United States import more than 50% of world exports.

The United States does not have high protection for the most important products in world trade

If the United States was exercising monopsony power in world trade, the greatest payoff from manipulating world prices would be in the biggest product markets

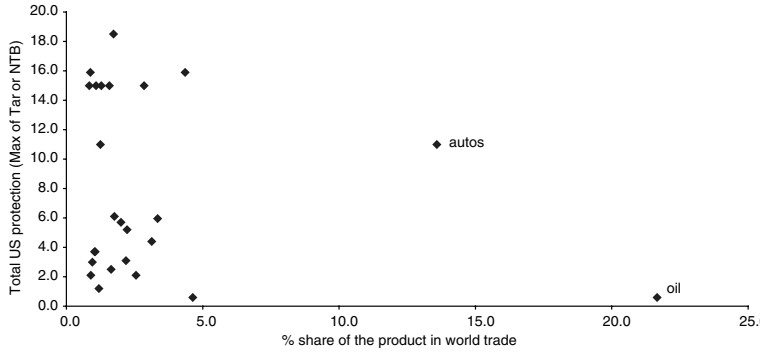

Figure 11.3 Total US protection (Y) vs size of the product in world trade (X).
Data source: Appendix Table 21.

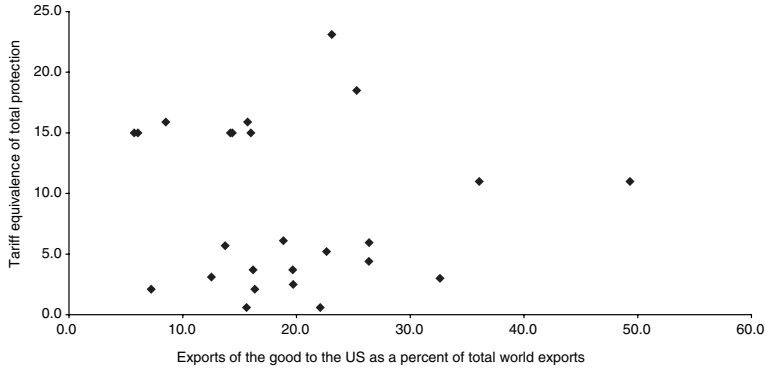

Figure 11.4 Percent of product exported to United States vs total US protection (tariff or NTB equivalent).

(e.g. lowering world prices for oil and autos since these two products are big US imports and they account for more than a third of world trade). If this were the case, there would be a positive correlation across products between US protection rates and the share of the good in world trade. We do not find this, as Figure 11.3 shows. The correlation between protection and the percent of the product in world trade is small, with or without the oil and auto outliers.

The United States does not have high protection for large percentages of the product exported to the United States

If the United States had power in world trade, it would get the greatest welfare gain in product markets dominated by US purchases. Given this, we would expect a

positive correlation between US tariffs and the percentage of the good in world trade shipped to the United States. We do not find this, as Figure 11.4 shows. The correlation is small between US protection and the percent of total world exports of the good that go to the United States.

Result II: The terms of trade evidence

US protection has significantly affected the US terms of trade for about 15% of products in world trade

We were unable to get long time series on the prices of US goods in international trade. The best we were able to do was get a very short nine year sample of annual data from 1980 through 1988 for 52 three-digit largely manufactured products.

Table 11.6 Summary of individual product regressions which display significantly positive effects of US protection on the US terms of trade (only the coefficients for US protection are shown)

Dependent variable: ln(Terms of trade)

SIC number	*Product*	*Regression 1 tariffs*	*Regression 2 tariff_Max*	*Regression 3 tariff_Sum*
209	Miscellaneous food preparations and kindred products	0.070124 (0.0141241)***		
242	Sawmills and planing mills	0.12314 (0.0390504)**		
286	Industrial organic chemicals	0.287411 (0.01067674)	0.005678 (0.001184)**	0.0040067 (0.0008417)**
354	Metalworking machinery and equipment	0.1873425 (0.0310276)**	0.1765342 (0.0526235)**	0.18235697 (0.0452368)**
357	Office, computing, and accounting machines	0.0318405 (0.0143225)*	0.047504 (0.0173125)*	0.0528341 (0.020126)*
362	Electrical industrial apparatus	0.0417789 (0.0035923)**		
363	Household appliances	0.0568673 (0.0081793)**		
365	Radio and television receiving equipment	0.0328697 (0.0476162)	0.0063289 (0.0019541)***	0.00628911 (0.0012356)***
386	Photographic equipment and supplies	0.1593024 (0.0322272)***		

Source: Appendix Table 15.

Notes
* significant at the 10% level
** significant at the 5% level
*** significant at the 1% level
tariff_Max = max{tariffs, NTBs}
tariff_Sum = tariff + NTBs
Standard errors in parenthesis.

We estimated individual terms of trade equations for all 52 of these products, with the dependent variable being the natural logarithm of the US terms of trade: that is, the price of US exports divided by the price of US imports of the product. The two independent variables were (1) the logarithm of the ratio of an index of US real GDP divided by a weighted index of OECD real GDP (the weights are the share of that country in US imports from the included OECD countries) and (2) three alternative measures of US tariff-equivalent protection. The three measures were

1 the tariff rate (duties as a percent of the value of US imports each year);
2 the maximum of the tariff rate or the tariff-equivalent of the nontariff barrier;
3 the sum of the tariff rate and the nontariff barrier.

Table 11.7 Individual product regressions each of which display significantly positive effects of US protection on the US terms of trade—the entire equation for regression 1 (tariffs only)

Dependent variable: ln(terms of trade)

SIC number	*Product*	*Constant*	*Tariff*	*ln(US/OECD weight GDP)*	R^2
209	Miscellaneous food preparations and kindred products	0.2240745 (0.2672353)	0.070124 (0.0141241)***	−5.899883 (3.767553)	0.5267
242	Sawmills and planing mills	−0.0939077 (0.0292747)	0.12314 (0.0390504)**	−2.881721 (1.130932)	0.8474
286	Industrial organic chemicals	−0.6563855 (0.08450087)	0.287411 (0.01067674)	−5.950235 (3.71536)	0.7840
354	Metalworking machinery and equipment	−0.5008067 (0.1110443)	0.1873425 (0.0310276)**	−23.99128 (11.39951)	0.8962
357	Office, Computing, and Accounting machines	−0.1451622 (0.056661)	0.0318405 (0.0143225)*	−1.114198 (1.966099)	0.6121
362	Electrical industrial apparatus	−0.1541141 (0.0205452)	0.0417789 (0.0035923)**	−4.276732 (0.912281)	0.9967
363	Household appliances	−0.1385046 (0.0292092)	0.0568673 (0.0081793)**	−7.886335 (2.45216)	0.9652
365	Radio and television receiving equipment	−0.2027454 (0.210275)	0.0328697 (0.0476162)	−0.9619763 (4.602891)	0.6589
386	Photographic equipment and supplies	−0.7181624 (0.143272)	0.1593024 (0.0322272)***	−5.063935 (2.226659)	0.9082

Source: Appendix Table 16.

Notes
* significant at the 10% level
** significant at the 5% level
*** significant at the 1% level
Standard errors in parenthesis.

Table 11.8 Products in which the United States appears to be a large country in world trade based on US tariffs and/or nontariff barriers having significant positive effects on the US terms of trade

SIC	*Product*	*US import value in millions (cif), 1985 ($)*	*Share % of US total imports %*	*US nontariff barrier % (A)*	*US tariffs % (B)*	*Max US barriers % (C) = Max A,B*	*Sum of barriers A+B (D)*
365	Radios and TVs	11,791	4.0	15.0	2.5	15.0	17.5
357	Office computers	10,909	3.7	0	3.8	3.8	3.8
286	Industrial organic Chemicals	4,298	1.5	15.0	5.6	15.0	20.6
354	Metalworking machines	3,651	1.2	15.9	4.8	15.9	20.7
242	Saw and planing mills	3,589	1.2	0	0.6	0.6	0.6
386	Photographic equipment supplies	3,356	1.1	0	4.3	4.3	4.3
363	Household appliances	2,126	0.7	0	3.2	3.2	3.2
362	Electrical Industrial Apparatus	2,064	0.7	0	3.9	3.9	3.9
209	Miscellaneous food and Preparations	1,524	0.5	0	3.6	3.6	3.6
	Total/Mean	43,308	14.8	5.1	3.6	7.3	8.7

Table 11.9 Determinants of US terms of trade: fixed effects model

Dependent variable: ln(terms of trade)			
Variable	*(1)*	*(2)*	*(3)*
	tariffs	*tariff_Max*	*tariff_Sum*
Constant	−0.0624924	−0.086324	−0.0324756
	(0.0207022)**	(0.0230853)**	(0.0084313)**
tar-1 (prods with sig pos TOT effects)	0.0145679	0.0545127	0.0845221
	(0.0048691)**	(0.0148574)**	(0.024877)**
tar-2 (prods with insig pos TOT effects)	0.0097932	0.0003607	0.0001317
	(0.0062725)	(0.0028465)	(0.0019809)
tar_3 (all other products)	−0.0071506	−0.0058232	−0.0048119
	(0.0025897)*	(0.0015225)**	(0.0014571)**
ln(US GDP_Index/ OECD GDP)	−3.460679	−3.707089	−4.039283
	(0.6994846)**	(0.5648427)***	(0.4286417)***
R^2	0.4601	0.3478	0.6121

Source: Observations (52 SIC industries for 9 years, 1980–1988) 468.

Notes
Standard errors in parenthesis
* : significant at the 10% level
** : significant at the 5% level
*** : significant at the 1% level
tar_1 = SIC209, SIC242, SIC286, SIC354, SIC357, SIC362, SIC363, SIC365, SIC386
tar_2 = SIC203, SIC206, SIC208, SIC333, SIC335, SIC349, SIC366, SIC367, SIC371
tar_3 = All other industries
tariff_Max = max {tariff, NTBs}
tariff_Sum = tariff + NTBs.

The results are shown in Table 11.8. Thus, we show the level of US protection for those products for which increases in US protection led to significant improvement in the US terms of trade over 1980–1988.

Table 11.8 shows that the United States appears able to alter its product terms of trade for about 15% of its imports. For the rest, US protection either had insignificantly positive effects on the product terms of trade or the wrong sign. Table 11.9 reports a fixed-effect panel regression over all 52 products. We group the tariff effects as shown.

These regressions and Table 11.10 estimated using the Feasible GLS procedure that accounts for heteroskedasticity gives a sense of the size of the US price effect. Using the tariff Max coefficient from both tables indicates that a one percentage point increase in US protection leads to a 5% improvement in the US terms of trade (the change in the logarithm is 0.05). The standard deviation is 5.7% for tariff_Max (which has the protection data for all 52 SIC products). Thus, a one standard deviation increase would increase the US terms of trade by 30% (5.7 times 0.05 = 0.289).

Table 11.10 Determinants of US terms of trade: feasible GLS

Dependent variable: ln(terms of trade)			
Variable	*(1)*	*(2)*	*(3)*
	tariffs	*tariff_Max*	*tariff_Sum*
Constant	−0.0751623	0.0002387	−0.1025638
	(0.0356254)**	(0.0085263)	(0.0452186)*
Tar_1	0.0258932	0.0508734	0.0752399
	(0.0058748)**	(0.0056251)***	(0.0086524)**
Tar_2	0.0048786	0.0052331	0.009272
	(0.0029724)	(0.008104)	(0.0098234)
Tar_3	−0.0033447	−0.0034503	−0.005366
	(0.0020197)*	(0.0009027)***	(0.0024614)**
ln(US GDP_Index/ OECD GDP_index)	−3.822343	−4.035127	−4.320947
	(2.144688)*	(2.128349)**	(2.333293)*

Source: Observations (52 SIC industries for 9 years, 1980–1988) 468.

Notes
Standard errors in parenthesis
*: significant at the 10% level
**: significant at the 5% level
***: significant at the 1% level
tar_1 = SIC209, SIC242, SIC286, SIC354, SIC357, SIC362, SIC363, SIC365, SIC386
tar_2 = SIC203, SIC206, SIC208, SIC333, SIC335, SIC349, SIC366, SIC367, SIC371
tar_3 = All other industries
tariff_Max = max {tariff, NTBs}
tariff_Sum = tariff + NTBs.

Except for Canada and Mexico, US protection has only affected the US terms of trade with 4% of world trade with its bilateral country trading partners

We examine now whether the United States can affect the US terms of trade *vis-à-vis* many of its bilateral trading partners. We showed above a figure of percentage of countries exports to the United States versus distance of the country from the United States. Table 11.11 shows the 20 countries with the highest percentages of their total exports going to the United States.

While it would appear that the data would support the United States having market power over these countries. However, an industrial organization view would be agnostic on this question. An industrial organization view would suggest that the United States could not exploit these countries by lowering the prices of their individual exports. The United States could not do this without pushing down individual product prices in the world markets for the country's major export products. We have shown above that the United States could do this for about 15% of the products in world trade. One limitation of this section is that we were unable to control for partner country trade policies.

International economists are reasonably drawn to the hypothesis that a large buyer will have market power over small sellers. However, the fallacy in this can

Table 11.11 Countries with the highest percentages of their exports going to the United States

Rank	*Country*	*Distance from US in km*	*Exports to US as % of total exports*
1	Bahamas	1,538	99.4
2	Trinidad and Tobago	3,501	95.4
3	Dominican Republic	2,376	93.6
4	Barbados	3,345	93.4
5	Haiti	2,312	91.1
6	Canada	734	88.0
7	Panama	3,341	75.6
8	Honduras	2,936	71.7
9	Guyana	4,043	69.7
10	Ecuador	4,357	68.8
11	Congo	10,515	64.9
12	Zaire	10,519	60.7
13	Mexico	3,039	57.2
14	Peru	5,671	56.5
15	Belize	2,670	52.6
16	Angola	10,653	49.5
17	Bermuda	8,564	44.6
18	El Salvador	3,049	44.3
19	Israel	9,452	41.0
20	Japan	10,910	40.7

be illustrated by the argument that a bank should have market power over each customer because each one is small relative to the size of the bank. The relevant consideration is the bank's market share and barriers to entry relative to the other banks. If that share is low, the bank does not have market power, irrespective of its size relative to its customers. Under GATT and now the WTO[14], country tariffs are levied on a product basis, not a country basis. This prevents countries from targeting individual country trading partners for exploitation.[15]

Aggregate US protection has not affected the aggregate US terms of trade since 1934

This section performs a time-series test of whether US protection affects the terms of trade with respect to the average product it imports. We test this by examining whether changes in US protection affect the aggregate terms of trade for total US trade, namely, the price of US exportables divided by the price of US importables. Correct specification required that we control for simultaneity and for the other determinants of both US protection and the US terms of trade.[16]

The results indicate that we can reject the null that the United States has no price effects for total trade, but only for the subperiod 1901–1929. This was a

Table 11.12 US protection displays price effects on the following countries' terms of trade (annual data, 1962–1996) (significantly negative effects of US protection on the countries' overall terms of trade)

Country	*Share of Country in Total Exports to the US, 1980 (%)*	*Total US protection (Tariffs + NTBs)*		*Share of Country in Total World Exports in 1980 (%)*
		coefficient	*t-value*	
Algeria	0.4	−0.0161361	−3.292*	0.726
Brazil	1.6	−0.0179384	−3.020*	1.054
Canada	19.6	−0.0016375	−2.990**	3.545
Colombia	0.2	−0.0048634	−3.743*	0.206
Dominican	0.4	−0.0060588	−1.795***	0.050
Egypt Republic	0.03	−0.0068867	−2.943**	0.159
Gabon	1.50E-04	−0.0125852	−3.056**	0.114
Iceland	0.009	−0.0042263	−4.238*	0.048
Mauritania	3.60E-05	−0.0024999	−2.055**	0.010
Mexico	2.7	—	—	0.815
Niger	—	−0.0056016	−1.805***	0.030
Nigeria	0.1	−0.0101437	−1.991***	1.359
Paraguay	0.02	−0.0083104	−2.624*	0.016
Trinidad and Tobago	0.6	−0.0216571	−11.06*	0.207
Total	25.7			8.3
Total excluding Canada and Mexico	3.3			4.0

Notes
* Significant at the 1% level.
** Significant at the 5% level.

period within which the United States was relatively larger than it is today. In all but one case, we cannot reject the null that the United States has no price effects after 1934. After 1934, there were essentially no long-run effects of US tariffs on the US terms of trade (because the US terms of trade were weakly exogenous using Johansen's cointegrating vector as an error correction term). The only rejection of the null was a short-run test indicating that US tariffs Granger cause the US terms of trade at a 5% level for the period 1934–1970. The result was not convincing, however, because impulse-response functions indicated that short-run changes in US tariffs had no effects on the terms of trade for the first twelve years.

Notes

1 E.R. Grilli and M.C. Yang, 1988, Primary Commodity Prices, Manufactured Goods Prices and the Terms of Trade of Developing Countries: What the Long Run Shows, *World Bank Economic Review* 2 (1), 1–44.

2 Chris Magee and Stephen Magee (2004) in a paper written after this one have also shown that the United States is not a large country in world trade using a different approach. See Chris Magee and Stephen Magee (2006) for the data used in this paper.

3 An intuitive explanation of the large country effect comes straight from microeconomics. Any tax imposed in a market with a positively sloped supply curve reduces the price received by suppliers. A small country in the world market faces a horizontal long-run supply curve and could gain no welfare advantage by imposing protection (i.e. rasing a tax) on its imports. But if the United States is large, it would face a positively sloped long-run supply curve and could obtain a welfare gain by increasing its protection and pushing down the price which foreign suppliers receive.

4 This literature showed that a country such as the United States should increase its tariff so long as the welfare gain from reducing prices paid to foreigners exceeded the welfare losses from the domestic price distortions until an "optimum tariff" was reached. See Corden (1984) for a survey of this literature. While our findings indicate that the United States is not large enough to control world prices, until Backus and Crucini (2000), we believed that the United States was large enough to affect world income. In fact, 33% of the world's income was earned in just three countries in 2003: the United States (21.4%), Japan (7%) and Germany (4.6%). Interestingly, both Japan and the integrated Germany are each geographically smaller than the state of Illinois (World Bank, 1993). And, the same three countries have 31% of world imports—that is, the United States (15%), Japan (7%) and Germany (9%). In addition to world income being more concentrated, many econometric models such as the Federal Reserve macro-international model, DRI, Klein, and so on all show that when income rises in the United States, foreign income rises as well. We accept these results as accurate and believe that the United States is certainly large enough to affect world income. We do not dispute Kindleberger's (1981, 1986) thesis that the United States is a hegemon in international politics. The United States has certainly played a large country political role for much of the twentieth century, eliminating a tragedy of the commons vacuum and providing the collective good of world leadership. Political scientists have mistakenly attempted to explain away the contradiction between the United States being large and the free-trade bias of the United States by saying that the United States gave up large country economic benefits (the right to exercise monopsony power by raising protection to get terms of trade welfare gains) in exchange for political benefits (greater US security through NATO, military bases, etc). As noted above, these complicated explanations are no longer required to explain US behavior over protection. In short, the United States can be large politically and have significant income effects without being a large country in world trade.

5 In a rational world of the United States and many small countries, we would expect the United States to have higher protection than everyone else. In fact, the United States has lower protection than most of its smaller trading partners and has accepted more than proportional cuts in its protection since 1934 to champion free trade. There have been a number of complicated theories advanced to explain this inconsistent behavior of a large country. But, the inconsistency disappears if the United States is a small country, since most of its optimum tariffs are near zero. US welfare will unilaterally increase as it cuts its protection. It is true that US welfare rises even more if foreign protection is also reduced. The latter explains the historical US role of encouraging freer trade through GATT and now the WTO.

6 This apparent irrationality of the United States favoring freer trade than its partners is discussed in the international relations literature on the US as a hegemon. See, for example, Krasner (1976) and Keohane (1984).

7 We started the research for this paper after Magee *et al*. (1989), Bohara and Kaempfer (1991) and other papers found that aggregate US tariffs were endogenous using time-series data. Bohara and Kaempfer found that increases in the US terms of trade led US tariffs to rise. Following the endogenous tariff literature, they reasoned that when import prices fall, import-competing firms are harmed and substitute out of economic activity and into political activity. The resulting increase in protectionist lobbying leads to a higher tariff. An alternative mechanism which produces the same result is Irwin's(1998) finding that the US *ad valorem* equivalent of specific duties rises when import prices fall. Both mechanisms generate a negative relationship between import prices and tariffs, implying positive causation from the US terms of trade (which has import prices in the denominator) to US tariffs. The thing which troubled us about this literature and the time-series tests (and Irwin's results) was that they failed to test for simultaneity. Given the stylized fact the United States is probably a large country in world trade, increases in US tariffs should increase the US terms of trade. This generates the same positive relationship between the terms of trade and tariffs as the endogenous tariff tests. But how could we know that the endogenous tariff results indicating a positive relationship between the US terms of trade and US tariffs were not driven by reverse causation—i.e. that the United States was a large country? If the United States was large, then an increase in its protection would improve the US terms of trade (the price of US exports over the price of US imports), explaining the positive relationship. Thus, our initial goal in writing this paper was to explain how much of the positive empirical relationship between tariffs and the terms of trade for the United States was driven by causation in each direction. That is, how big were large country effects (causation from tariffs to the terms of trade) versus endogenous tariff and import-price effects (causation from the terms of trade and import prices to tariffs) for the United States? Our initial research strategy with Kwang-Yeol Yoo was to measure how US protection and other control variables affected the US terms of trade over various subperiods from 1901–1988, while simultaneously accounting for the effect of the terms of trade and other endogenous protection determinants on US protection. Like Bohara and Kaempfer (1991), we estimated a vector autoregression model (VAR). In testing for how US tariffs affected the US terms of trade, we extended their methodology in several ways. We used Dickey-Fuller tests (DF) and augmented Dickey–Fuller tests (ADF) to see if each variable had a unit root. We estimate cointegrated systems of variables, following Johansen's (1988) maximum likelihood procedure. In subperiods in which we did not observe cointegration in the levels of the variables, we use the VAR model and standard Granger causality tests. We also estimated the error correction model (ECM), following to Engle and Granger (1987), which allowed a separation of long-run and short-run effects. Like Bohara and Kaempfer (1991), we find a regime change in US tariffs in the early 1930s, and hence performed all of our tests using two subperiods.

8 Alan Deardorff has stated in an e-mail that he knows of no formal empirical estimation of the optimum tariff for the United States outside of the CGE (computable general equilibrium) literature. He has never done this himself, although he (Deardorff, 1978) has a paper on the effects *of* the terms of trade, no effects *on* the terms of trade. For attempts to estimate the effects of tariffs on the terms of trade (?) see Whalley (1986,1992). For references to terms of trade effects in regional trade battles, see Perroni, Carlo and Whalley, John (1996).

9 There is support for a small-country view of the United States in other markets. Empirical work by Ruffin and Rassekh(1986) showed that the United States was not a large country in international capital markets.

10 There is a reason for not analyzing the country-size question with respect to US exports. With respect to average trade, the United States looks even less like a large country with respect to exports than imports (the United States only has a 12% share of world exports). International economists do not analyze the large-country problem with respect to a country's exports. The major export policy by which a country could

improve its welfare at the expense of foreigners would be to tax exports. This case is almost never driven by terms of trade considerations. We ignore US exports in this chapter because of a virtual absence of US export taxes and because the few which exist did not appear to be imposed because of terms of trade considerations. The latter is required for them to have a large-country motivation. Several generations of international trade economists, including ourselves, have mistakenly cited the United States as an example of a large country, in the sense that the United States could affect world prices. We have long taught our students the "optimum tariff" argument (see, *inter alia*, Johnson (1954) and Corden(1984)); that is, that self-interested large countries can exploit their monopsony power in world trade via protection.

11 Another consideration in the exercise of monopsony power is whether there is retaliation when the United States raises its protection. If all foreign countries would respond to US protection by raising their tariffs, then the US could affect little change in its terms of trade. Casually, the last century of tariff negotiations and tariff monitoring by both GATT and the WTO provides little evidence that the United States has been able to prevent foreign retaliation. For the pervasiveness of tariff retaliation by product and by country, both historically and in this century, see Conybeare's (1987) book on trade wars. This is another constraint on the United States being a large country.

12 The welfare of a country increases when the price of what it is selling increases relative to the price at which it is buying being a large country.

13 This percentage is biased downward for present purposes because industrial prices and rates of return are elevated above concentrations of 50% because firms frequently collude to raise price. Collusion is not an issue since buyer power by a large country would be exercised unilaterally (the country itself would raise its protection and push down world prices).

14 GATT was the General Agreement on Tariffs and Trade which regulated world protection until 1997. It has since been superseded by the World Trade Organization.

15 The strongest support for the price effects hypothesis was found in this section. We examined total country terms of trade and US protection data for the 58 countries for which we had data for all of the variables over the 35 year time period. Table 11.12 shows the 14 countries out of 58 countries for which US protection had the expected negative effect on the country terms of trade required for the United States to have price effects. These countries include Canada and Mexico, our two large contiguous trading partners. These 14 countries account for 26% of total US imports and 8% of world exports. These countries accounted for 67% of US imports back in 1980. Thus, of the countries tested, the United States had price effects with respect to 39% of US imports by value and 12% of world trade weighted by value. But, excluding the two closest trading partners, Canada and Mexico, the United States demonstrated the ability to affect the terms of trade for only 3% of US imports and 7% of world exports for the other 56 countries tested.

16 We use Magee, Brock, and Young's (1989) data for US tariffs, the country capital/labor endowment ratio, the terms of trade, and the real exchange rate and follow their specification for the determinants of endogenous protection. We used Hufbauer, Berliner and Elliott's (1986) calculations for the tariff equivalents of all US nontariff protection for the period 1922–1986 and assumed that the 1986 numbers for nontariff protection also apply to 1987 and 1988. We examine cointegration relationships among US protection, the labor-capital ratio, the terms of trade, unemployment, inflation, and the real dollar exchange rate. In the text, we use only the US tariff rate to measure US protection. In Magee and Magee, *et al.* (2006), we find that the qualitative results for US tariffs reported here also hold for total US protection (tariff plus the ad valorem equivalent of all nontariff barriers). We take logarithms of the terms of trade and capital per worker. We do not take logarithms of unemployment, the rates of protection, and inflation, since they already have a logarithmic interpretation. DF and ADF unit root tests are performed on each variable and results are reported in Table 11.13.

The tariff rate, terms of trade and capital per labor have unit roots in their ADF tests. Since first-differenced values of those variables no longer have unit roots, they are integrated of order one. However, unemployment turns out to have no unit root after adjusting for outliers in Great Depression period (1932–1935). Inflation also has no unit root. Thus, we include the tariff rate, capital per labor, and the terms of trade only in the cointegration analysis. Here are the definitions of the variables we are using:

TAR = US tariff rate (tariff duties divided by the value of total imports)
LTT = log(US terms of trade) = log(price of exports/price of imports),
UNE = US unemployment rate,
LKL = log(real value of US capital per labor in 1972 dollars) and
INF = US inflation rate of the producer price index.

As the first step to test cointegration, we perform a static OLS regression of TAR on LTT, LKL from 1901 till 1988 and residuals from that regression are used to test for the unit root. If the residuals have unit roots from ADF tests, the null hypothesis of no cointegration is not rejected. For the original work, see Yoo (1997).

Table 11.13 Dickey–Fuller and augmented Dickey–Fuller statistics for the entire sample, 1901–1988 (the same qualitative results hold for 1934–1988)

Variable	*DF/ADF statistics*	*Number of lags*	*Variable*	*ADF stats*	*Number of lags*
TAR	–2.83	1	TAR	–5.986	0
	(0.199)			(0.00)	
LTT	–2.12	0	LTT	–7.79	
	(0.586)			(0.00)	
LKL	–1.35	1	LKL	–3.96	0
	(0.904)			(0.00)	
UNE1	–3.1	1	—		
	(0.114)				
UNE2	–4.14	2	—		
(w/dummy 32–35)	(0.004)*				
INF	–6.299 *	0			
	(0.000)				

Notes
The definition of variables are given above.
P-values are given in parenthesis.
This is the number of lags that eliminates the serial correlation in the ADF regressions.
We get the same result whether we include the trend term as a regressor in ADF regressions or not. Above reports are ADF test result when the trend term is included in the ADF regression.

Table 11.14 Cointegration test by static OLS regression

Estimated cointegrating relationship	*Sample coverage*	*ADF statistics*	*Number of lags*
(1) TAR = 0.2080 + 0.2030*LTT – 0.117*LKL	1901–1988	–2.986	1
(2) TAR = 0.0977 + 0.1481*LTT – 0.078*LKL	1934–1988	–3.746*	0

Note
* Significant at 5% significance level in the Dickey–Fuller Table.

We do not reject the null of no cointegration when the whole sample is used. So, we split the sample into two subperiods. Table 11.14 reports the test results. The resulting cointegrating relationship is not particularly robust to the break point. For example, there is no cointegration in 1947–1988 but there is cointegration from most of 1930s and the early 1940s through 1988. A detailed discussion of break points the related tests is contained in Magee and Magee (2006). We report here the results from two subperiods: 1901–1933 and 1934–1988. This is the most reasonable break because it parallels Bohara and Kaempfer's (1991) finding of a regime change in 1934 and because the US Congress delegated broad powers of over US trade policy from itself to the Administration in the Trade Agreements Act of 1934.

We cannot reject the null of no cointegration in the period 1901–1933 but we do reject the null of no cointegration during 1934–1988. Explanations can be given to this phenomenon. Gardner and Kimbrough (1989) asserted that the tariff was a major source to the US government before income tax was introduced in 1913 and continued so until at least 1918. If this is true, then the interest-group model of endogenous tariffs that underlies our specification and which works well in the period 1934–1988, may not apply in the first subperiod. Equation (2) in Table 11.14 demonstrates the signs on the terms of trade and capital per laborer are those predicted by endogenous tariff theory. That is, an increase in the US terms of trade means that the relative price of imports are lower, which produces political pressure on Congress to raise tariffs and the US customs to collect more revenue. Increases in the capital-labor ratio means a decline in the economic importance and the political power of labor, the factor largely protected by tariffs, and hence lower tariffs. With respect to the subperiod 1934–1988, we use Johansen's maximum likelihood procedure of testing for the numbers of cointegrating vectors, to make sure that there is only one cointegrating relationship among these variables (i.e. $r = 1$). Uniqueness is required for validity of the ECM tests. Table 11.15 reports two test statistics—maximum eigenvalue (λ-max) and trace test—for various null and alternative hypothesis.

Table 11.15 Johansen's test for the number of cointegrating vectors (r), 1934–1988

Null hypothesis about rank r	*TAR LTT LKL*			
	Trace test statistics	*5% Critical value*	*λ-max test statistics*	*5% (10%) Critical value*
$r \leq 2$	1.970	8.803	1.970	8.803(6.691)
$r \leq 1$	13.608	17.844	11.638	14.595(12.783)
$r = 0$	32.403*	31.256	18.795**	21.279(18.959)

Notes
r: the rank or number of cointegrating vector.
* significant at the 5% level and ** significant at 10% level.
The critical values are from Table A2 in Johansen and Juselius(1990).
For example, in λ-max test, H1: $r = 2$ and in trace test, H1: $r \leq 2$ in H: 0: $r \leq 1$.

Bibliography

Backus, David K. and Crucini, Mario J. (2000) "Oil Prices and the Terms of Trade," *Journal of International Economics*, 50(1): 185–213.

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1993) *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*, Oxford.

Bohara, A. K. and Kaempfer, W. H. (1991) "A Test of Tariff Endogeneity in the United States," *American Economic Review*, 81: 952–960.

Carlton, D. and Perloff, J. M. (1994) *Modern Industrial Organization*, New York: Harper.
Conybeare, John A. C. (1987) *Trade Wars*, New York: Columbia University Press.
Corden, Max (1984) "The Normative Theory of International Trade," *Handbook of International Economics*, Amsterdam: North-Holland.
Crucini, Mario J. (1994) "Sources of Variation in Real Tariff Rates: The United States, 1900–1940," *American Economic Review*, 84: 732–743.
Das, Sanghamitra and Das, Satya P. (1994) "Quatitative Assessment of Tariff Endogeneity," *Economic Letters*, 44: 139–146.
Deardorff, Alan (1978) "The Terms-of-Trade Effect on Expenditure: Some Evidence From Econometric Models," *Journal of International Economics*, August.
Engle, Robert F. and Granger, C. W. J. (1987) "Cointegration and Error Correction: Representation, Estimation, and the Testing," *Econometrica* 55: 251–276.
Gardner, Grant W. and Kimbrough, Kent P. (1989) "The Behavior of U.S. Tariff Rates," *American Economic Review*. March, 211–218.
Hamilton, James D. (1994) *Time Series Analysis*, Princeton: Princeton University Press.
Henrique, Irene and Sadorsky, Perry (1994) "The Determinants and Persistence of Canadian Tariff Rates," *Canadian Journal of Economics*, Feburary, 112–128.
Hodrick, Robert J. (1989) "U.S. International Capital Flows: Perspectives From Rational Maximizing Models," *Carnegie-Rochester Conference Series on Public Policy*, 30: 231–288.
Hufbauer, Gary C. Berliner D. T. and Elliott, K. A. (1986) *Trade Protection in the U.S.: 31 Case Studies*, Washington DC: Institute for International Economics.
Irwin, Douglas A. (1998) "Changes in U.S. Tariffs: The Role of Import Prices and Commercial Policies," *American Economic Review*, 88: 1015–1026.
Johansen, Soren "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12: 231–254.
Johansen, Soren and Juselius, Kararina "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics* 52: 169–210.
Johnson, Harry G. (1954) "Optimum Tariffs and Retaliation," *Review of Economic Studies*, 21: 142–153.
Keohane, Robert O. (1984) *After Hegemony*, Princeton: Princeton University Press.
Kindleberger, Charles P. (1981) "Dominance and Leadership in the International Economy," *International Studies Quarterly*, 25: 242–254.
——(1986) "International Public Goods Without International Government," *American Economic Review*, 76: 1–13.
Krasner, Stephen D. (1976) "State Power and the Structure of International Trade," *World Politics*, 28, (April).
Magee, Stephen P., Brock, William A. and Young, Leslie (1989) *Black Hole Tariffs and Endogenous Policy Theory: Political Economy in General Equilibrium*, Cambridge: Cambridge University Press, Chapter 13.
Magee, Christopher S. P. and Magee, Stephen P. (2004) "Is the United States a Large Country in World Trade?" working paper, University of Texas at Austin.
Magee, Christopher S.P. and Magee, Stephen P., *et al.* (2006) *The United States is a Small Country in World Trade*, In draft.
Marshall, Alfred (1923) *Money, Credit and Commerce*, London: Macmillan, 213, 344.
Ostry, Jonathan D. (1988) "The Balance of Trade, Terms of Trade and the Real Exchange Rate," *IMF Staff Papers*, 35: 541–573.

Perroni, Carlo and Whalley, John (1996) "How Severe is Global Retaliation Risk under Increasing Regionalism?" *American Economic Review*,. 86(2): 57–61.

Persson, Torsten and Lars E. O. Svensson (1985) "Current Account Dynamics and the Terms of Trade: Harberger-Laursen-Metzler Two Generations Later," *Journal of Political Economy*, 93(1): 43–65.

Phillips, Peter C. B. and Ouliaris, S. (1990) "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58: 165–193.

Ruffin, Roy J. and Rassekh, Farhad (1986) "The Role of Foreign Direct Investment in U.S. Capital Outlfows," *American Economic Review*, 76: 1126–1130.

Sadorsky, Perry (1994) "The Behavior of U.S. Tariff Rates: Comment," *American Economic Review*, 84: 1097–1103.

Sen, Partha and Turnovsky, Stephen J. (1989) "Tariffs, Capital Accumulation, and the Current Account in a Small Open Economy," *International Economic Review*. 30: 811–831.

Stockman, Alan C. and Dellas, Harris (1989) "International Portfolio Nondiversification and Exchange Rate Variability," *Journal of International Economics*, 26(3/4): 271–290.

Stockman, Alan C. and Svensson, Lars E. O. (1987) "Capital Flows, Investment and Exchange Rates," *Journal of Monetary Economics*, 19(2): 171–201.

Stockman, Alan C. and Tesar, Linda L. (1995) "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements," *American Economic Review*, 85: 168–185.

Whalley, John (1986) *General equilibrium trade policy modeling*, Cambridge.

——(1992) *Applying general equilibrium*, Cambridge University Press.

World Bank (1993) *World Development Report, 1993*, New York: Oxford University. Press.

Yoo, Kwang-Yeol (1997) *Empirical Essays on Applied Rent Seeking*. PhD dissertation in Economics, University of Texas at Austin.

12 Taxing the brain drain

A reassessment of the Bhagwati proposal

John Douglas Wilson

Introduction

It has now been more than thirty years since Jagdish Bhagwati issued his controversial proposal for "taxing the brain drain." In its various forms, this proposal would enable developing countries to receive revenue from taxes levied on emigrants residing in developed countries. His original proposal (Bhagwati, 1972) was followed by two edited volumes (Bhagwati,1976a, b) that investigated the economic and legal issues that it raised. Another edited volume (Bhagwati and Wilson, 1989) looked at migration and tax policy more generally, including the problems that emigration creates for the source country's tax policy when it is unable to tax emigrants. The literature on the brain drain has continued to grow in recent years, and there has been an outpouring of research on the issues associated with taxation in the presence of international factor mobility. In addition, Bhagwati's views have also evolved since the original proposal, including a greater recognition of the political and administrative problems involved in extending a country's tax system to include emigrant incomes. But his basic argument for taxing the brain drain has remained remarkably constant. Recently, he has written, "Enhancing these good effects [of skilled emigration] requires that countries such as India and Taiwan adopt the diaspora model, extending a warmer embrace to their nationals abroad . . . However, the diaspora approach is incomplete unless the benefits are balanced by some obligations, such as the taxation of citizens living abroad . . . Estimates made by scholars . . . demonstrate that even a slight tax on Indian nationals abroad would substantially raise Indian government revenues" (Bhagwati, 2004: 215).

In this chapter, I reassess Bhagwati's proposal, using theoretical developments on international taxation that came after it was first put forth. I find that the proposal withstands some important arguments that have been made against it. The proposal remains remarkably valid after more than thirty years. I begin by describing Bhagwati's proposal and the reasoning behind it. Then I argue that the proposal is supported by the literature on optimal income taxation in an open economy. This literature assumes that the goal of government decision-makers is to maximize the well-being of their citizens (which may include emigrants). Since violations of this assumption have been a major concern, I demonstrate that a brain-drain tax may still be desirable in the case of "non-benevolent governments."

The proposal

An initial form of the proposal, as described in Bhagwati and Dellalfar (1973), calls for a tax on the incomes of "professional emigrants" from developing countries into developed countries. In its preferred form, the tax would be collected "under UN auspices," with the aid of the tax authorities in the host developed countries. It would be levied for only a "reasonable period," such as an emigrant's first ten years in the host country. The UN would route the revenue to the countries of origin, except that "corrupt and dictatorial" countries would not receive revenue. Instead, it would be dispersed to developed countries using the usual criteria for developmental spending.

It became apparent that enormous practical difficulties would be encountered in the implementation of such a tax, including the likelihood that attempts to involve the US tax authorities in its collection would result in its being declared unconstitutional because it discriminated against aliens. In later writings, Bhagwati responds by arguing that the tax could be collected by developing countries, using the type of "global tax system" employed by the United States and a few other countries, whereby the incomes of citizens at home and abroad are both taxed (see, e.g. Bhagwati, 1979). With the tax collected in this manner, its rationale rests on a comparison of global tax and scheduler tax systems, where the latter is based on residence, rather than nationality. In general, there exist considerable opportunities for individuals to engage in tax-avoidance and evasion activities that reduce or eliminate their tax payments on income earned abroad. The use of a global tax system to tax the brain drain still requires some cooperation of developed countries, including sharing of tax information. Such cooperation has proved difficult to obtain.

Bhagwati (1979: 62–63) emphasizes the equity aspects of a brain-drain tax. The basic idea is that skilled migrants typically earn economic rents, given the stiff immigration restrictions in developed countries, coupled with the relatively high returns to skill available there. The rationale for taxing these rents then follows from the usual equity arguments behind progressive taxation—arguments that are consistent with the welfare calculus of modern welfare economics. Of course, there remains the issue of why it should be the country of origin that subjects emigrants to its progressive tax system, rather than only the destination country. Here, Bhagwati (1979: 63) emphasizes the increased mobility of highly skilled individuals, coupled with their tendency to retain their national status and associated rights, often including the right to vote. If emigrants are not taxed, then we have "representation without taxation" for the source country.

While Bhagwati and Dellafar (1973) put forth this equity rationale for a brain-drain tax, they also stress the use of the tax to compensate developing countries for the losses experienced by those natives left behind. Bhagwati and Hamada (1974, 1975) develop theories of how these losses might occur, using models with unemployment. As an alternative to a tax collected from emigrants, Bhagwati has also discussed methods for transferring income from developed countries to developing countries to compensate the latter for losses caused by the brain drain

(e.g. Bhagwati, 1976, chapter 1). But he has long recognized that the magnitude and existence of such losses are highly uncertain, and therefore views the gains that developed countries experience from the brain drain as a better justification for brain-drain-related transfers (Bhagwati, 1979). Recent literature has emphasized the important gains for developing countries from high-skilled migration, which may arise from the resulting remittances and increased technology transfers, along with increased incentives for residents of developing countries to invest in human capital (Stark and Wang, 2002). Bhagwati (2004) emphasizes differences across countries in the degree to which they benefit or lose from the brain drain. Relatively small countries in Africa, from which a large fraction of skilled natives emigrate, can be expected to lose, whereas large developing countries are in a position to gain. In any case, Bhagwati has been careful not to let his proposal rest on the existence of losses to developing countries.

A central issue in the assessment of a brain-drain tax is the behavior of developing-country governments. An important argument against a brain-drain tax has been that non-benevolent governments would use the tax as another source of revenue for wasteful expenditures. I discuss this issue below, but first I discuss the theoretical arguments for the tax under the assumption that governments are benevolent.

Case for a tax on emigrant incomes

The literature on optimal income taxation in an open economy has built a case for home-country taxation of skilled emigrants by analyzing the difficulties in collecting a progressive income tax when emigrants cannot be taxed. In the standard model of income taxation in a closed economy, as developed by Mirrlees (1971), an income tax is used to redistribute income from high-skilled workers to low-skilled workers. The goal is to maximize a social welfare function that includes as arguments the utilities of all workers. There exists an equity-efficiency tradeoff, because the income tax distorts the supply of labor. Bhagwati and Hamada (1982) and Wilson (1980, 1982) extend the Mirrlees model to an open economy, where emigration is possible, but unlike Mirrlees, they restrict the income tax to be linear.[1] Their common message is that that the emigration by high-skilled workers leads to a less egalitarian tax system. The basic idea is that increasing the tax burden on skilled residents causes some of them to emigrate, and the resulting revenue loss represents a social cost from the source country's viewpoint. To limit this revenue loss, the country should reduce its taxation of skilled residents, leading to a less egalitarian tax structure. A similar message occurs from Wilson's (1992) analysis of optimal nonlinear income taxation under emigration. For all of these comparisons, the weight that a citizen receives in the measurement of social welfare is independent of whether the individual chooses to emigrate. It is the desire to avoid losing the tax revenue from high-income taxpayers that leads governments to implement a less egalitarian tax structure.

The extent to which high-skilled workers should be taxed in an open economy depends on how elastic emigration is with respect to changes in source-country

taxes. In the limit, where the elasticity is infinite, the only tax collected from the mobile-skilled should equal the cost of providing an additional skilled resident with public goods and services (see Wilson, 2005a, for additional discussion). In other words, countries would be unable to use taxation as a redistributive device if labor were perfectly mobile. A similar conclusion has emerged from the literature on the "Tiebout hypothesis": the mobility of households leads to an efficient provision of public goods by local governments, but it prevents these governments from redistributing income.

Allowing the home country to tax high-skilled emigrants enables it to increase its taxation of those high-skilled natives remaining at home without causing a revenue-reducing outflow of these individuals. Fewer leave because they are taxed abroad, and those that do leave continue to contribute to the home country's treasury. Thus, brain-drain taxes can be justified not just as a means of directly redistributing income from high-income emigrants to lower-income residents, but also as a way of increasing the ability of a country to tax high-income residents.

The case for a brain-drain tax is further strengthened by the analysis by Mirrlees (1982) of an optimal nonlinear income tax model in which both residents and emigrants are taxed. In this model, individuals have an opportunity to exit the tax system, interpreted as giving up their citizenship after emigrating. Nevertheless, Mirrlees finds that emigrants should still be taxed at relatively high rates. Indeed, he presents calculations where, comparing an emigrant and resident who are equally well off, the emigrant's tax payments to the home country should be at least as high as the resident's tax payments.

The Mirrlees model is particularly noteworthy as a theoretical basis for a brain-drain tax, because it includes the economic rents that some migrants receive. This is done by allowing some individuals to have access to high incomes abroad (net of monetary and psychic migration costs), relative to their incomes at home. In line with Bhagwati's reasoning, such rents help lead to high optimal taxes on emigrants, even though these rents vary widely across emigrants, and although individuals also possess heterogeneous abilities to opt out of the tax system.

One limitation of Mirrlees' analysis is the absence of an explicit model of host country restrictions on immigration, which Bhagwati cites as a major reason for such rents. It would be useful to construct a model of this type. By including the costly activities that potential immigrants undertake to enter their host countries (e.g. more costly or less-suitable forms of education, or unemployment while waiting for entry), actual rents might be lower than initially perceived.

As described above, recent literature on the brain drain suggests that it benefits those left behind in developing countries. But the demonstration that such benefits exist does not imply that unfettered migration maximizes their value. Thus, whereas Stark and Wang (2002) demonstrate that emigration opportunities stimulate human capital investment in the emigrants' home countries, they also find that these countries would benefit from restrictions on emigration.[2] A brain-drain tax would have a similar effect, by reducing incentives to migrate. Moreover, regardless of the responsiveness of migration rates to a brain-drain tax, its justification as a redistributive device would remain. Thus, this more recent

literature should not be viewed as providing a convincing argument against taxing the brain drain.

Nonbenevolent governments

A major concern about a brain-drain tax has been the possibility that some developing countries would abuse it. At the time the tax was proposed, there were concerns with "exit taxes" imposed by communist countries. On more philosophical grounds, Harry Johnson objected to requiring individuals who chose to leave a developing country to continue to pay taxes to this country, particularly since many developing countries were dictatorships.[3] But Bhagwati (1979) remarks that versions of the tax calling for it to be collected only from those emigrants retaining citizenship are not subject to such criticisms. These individuals could opt out of paying the tax by severing their ties with their home countries.

Related criticisms have been raised by Hufbauer (1989). He cites Tiebout's theory, where the opportunity to "vote with one's feet" causes local governments to efficiently respond to the preferences of their residents. According to Hufbauer, "The ability of an individual to emigrate affords both an escape hatch and a defense against political and economic oppression at the national level" (1989: 87). In his view, taxes on emigrant incomes represent restrictions on emigration, albeit mild forms. To the extent that a tax on the brain drain would inhibit emigration, it would reduce the ability of emigration to limit the behavior of governments.[4]

The possibility that potential emigration improves government decision-making may be addressed using a model in Bhagwati and Wilson (1989, Chapter 1). We sketch a model in which government officials in the home country care not only about the well-being of its citizens, measured by their "real income" (including emigrant income, net of emigration costs), but also about their own real income, given by the excess of tax revenue over public expenditures that increase the well-being of residents. To increase their income, officials must either raise taxes or reduce productive public expenditures, both of which reduce citizen income. But then some citizens will emigrate, causing a drop in tax revenue if the home country cannot tax their incomes abroad, and this lost revenue will lead to a further decline in citizen income. In other words, potential emigration increases the effective price of the officials' income, measured in units of foregone citizen income. As a result, these officials face incentives to substitute citizen income for their own income, which may be interpreted as less "waste" in government. If the home country could tax emigrant incomes, then the revenue loss from emigration would decline, thereby reducing or eliminating this desirable substitution effect. However, this tax would also provide an additional source of income for financing both citizen income and the officials' income. In other words, taxing emigrants creates an income effect, which may offset the substitution effect.

The basic idea here is that in the absence of a tax on emigrants, countries are forced to "compete" for skilled residents by reducing "waste" in government. This competition may then be desirable, a possibility that has been pursued further in papers by Gordon and Wilson (2003) and Wilson (2005b) on the

desirability of tax and expenditure competition when there is waste in government. As argued by Bhagwati and Wilson (1989), the desirability of this competition may be offset by the revenue losses resulting from emigration. Thus, the welfare effects of a brain-drain tax can go either way.

In practice, this revenue loss can be substantial. Desai, Kapur and McHale (2003) estimate the net fiscal impact of the migration of high-skilled emigrants from India to the United States. In 2001, there were a little more than one million Indian-born residents in the United States, with more than three-quarters of the working age population possessing a bachelor's degree or better. The foregone income tax revenue associated with these residents equaled about one-third of current individual income tax receipts. Taking into account the estimated savings in public expenditures from this emigration, estimates of the net fiscal loss range from 0.24% to 0.58% of Indian GDP in 2001.

The extent to which such revenue losses are offset by less waste in government is also unclear. Li and McHale (2006) find negative effects of skilled emigration on various measures of the "ability of government to formulate and implement sound policies." But they also find that skilled emigration has positive impacts on measures of "political institutions," including "the extent to which citizens of a country are able to participate in the selection of governments," referred to as "voice and accountability."

It is interesting to note here that Bhagwati's early view that the revenue from a brain-drain tax could be distributed by the UN to the host developing countries has the potential to both preserve beneficial competition for residents and provide these countries with income generated by taxes on emigrants. The trick would be to design the criteria for distributing the tax revenue so that an increase in any given country's emigration rate did not significantly increase the revenue transferred to that country (see Bhagwati and Wilson, 1989, Chapter 1). In other words, the pattern of revenue disbursements could be based on "initial" emigration rates, but countries could not perceive that more emigration would result in more revenue. In a world with changing migration patterns, such a system might be difficult to maintain, unless the revenue transfers were undertaken for developmental needs that were largely independent of these changes in migration patterns.

But there is an additional argument in favor of Bhagwati's proposal for implementing the brain-drain tax on citizens residing abroad. Here the taxation of emigrants replaces incentives for the countries to compete for residents with incentives to now compete for citizens. Specifically, the ability of a citizen to avoid the tax by giving up citizenship in the home country provides the country's government with incentives to engage in policies that are agreeable to citizens residing abroad. For example, if the taxes collected from these emigrants are largely wasted, then the citizen will be less likely to plan to return home for extended periods of time, in which case citizenship becomes less important.

There is a trade-off here. Allowing emigrants to avoid taxes by surrendering their citizenship limits the ability of the home government to tax these individuals. On the other hand, such avoidance activity limits the power of nonbenevolent

governments to some extent. As a result, incorporating this tax-avoidance feature into a brain-drain tax may increase its desirability.

In fact, Bhagwati's main rationale for a brain-drain tax—as a method of taxing the "rents" received by emigrants—suggests that this form of the tax should be used when developing-country governments are not benevolent. As noted above, he has argued that these rents result from the tight restrictions that developed countries impose on immigration from developing countries. Given such restrictions, the relevant migration elasticities are likely to be quite low with respect to changes in developing-country tax and public expenditure policies. Low elasticities reduce the degree to which a developing country needs to compete for residents through low taxes and desirable public expenditures. On the other hand, the elasticity of citizenship with respect to these developing-country policies can remain sizable, though migration elasticities are low. Given this combination of elasticities, a brain-drain tax will actually enhance incentives for developing countries to engage in welfare-improving tax and expenditure competition.

To conclude, Bhagwati's proposal to tax only citizens living abroad seems to be well-suited for addressing concerns about the functioning of developing-country governments.

Concluding remarks

I have argued that Bhagwati's proposal for a brain-drain tax survives major objections that have been levied against it, including the view that it would eliminate desirable competition for residents. Note, in addition, that governments compete not just for residents, but also for capital investment, and the latter form of competition is not directly affected by a brain-drain tax.

Recent research has presented new arguments for the absence of a "brain-drain problem," but this development does not eliminate the desirability of a brain-drain tax, which has been based more on the existence of the "brain-drain phenomenon."

A major stumbling block in the implementation of such a tax continues to be the administrative problems associated with collecting it. But these problems exist whenever taxes are levied on foreign-source income. Without the cooperation of host country governments, it is most difficult for the source country to collect these taxes. A large element of such cooperation would include information-sharing among governments, a topic that is now being actively researched (e.g. Keen and Ligthart, 2006).

In addition, there is the issue of the extent to which a brain-drain tax should be integrated into the existing source-country tax systems. Recall Bhagwati's argument that a brain-drain tax could be justified as part of a country's "global tax system." As commonly defined, such a system would treat identically the real incomes earned at home and abroad, except that credits or deductions would be needed to avoid double taxation of income by host and source countries, and perhaps to offset additional costs involved in earning income abroad. But if host-country tax rates exceeded those in source rates, then a credit for foreign taxes

paid would eliminate all tax payments to the source country. In the reverse case, taxes owed to the home country might exceed those envisioned by Bhagwati's original proposals, resulting in widespread tax evasion. Moreover, developing countries often encounter great difficulties in collecting income taxes from their own residents, suggesting that extending the same tax system to nonresidents would prove to be infeasible. Recognizing the practical problems involved in collecting a brain-drain tax, perhaps the best approach would be to levy a small tax at a flat rate on the incomes of citizens living abroad. At the cost of some additional complexity, progressivity in average tax payments could be introduced by taxing only incomes above a given exemption level. In their discussion of the various methods for taxing emigrants, Desai, Kapur and McHale (2004) note that if only 100,000 of the more than one million Indian citizens in the United States paid $5,000 each, the resulting $500 million annual revenue stream would still be substantial compared to India's individual income tax base of $5.84 billion and tertiary education budget of $2.7 billion.

Similar to other important policy initiatives, Bhagwati's proposal is not only well-grounded in existing economic theory, but also points to holes in the theory that need to be filled. Research inspired by this proposal should continue to further our understanding of how to design tax policies for an increasingly integrated world economy.

Notes

1 These papers, along with the paper by Mirrlees (1982) that is discussed below, are reprinted in Bhagwati and Wilson (1989). See Wilson (2005a) for a more extensive review of these and related papers.
2 See Stark *et al*. (2005) for further analysis.
3 Reported by Bhagwati (1979: 67).
4 Hufbauer also objects to such taxes because emigrants are not receiving the public goods financed by these taxes, and they have severed their cultural connections. The latter assertion is increasingly questionable, however, and emigrants do benefit from the "option value" of the right to return to their home countries and receive public good benefits.

Bibliography

Bhagwati, J. N. (1972) "The United States in the Nixon Era: The End of Innocence," *Daedalus*, 25–47.

—— (ed.) (1976a) *Taxing the Brain Drain, Vol. 1: A Proposal*, Amsterdam: North Holland.

—— (ed.) (1976b) *The Brain Drain and Taxation, Vol. 2: Theory and Empirical Analysis*, Amsterdam: North Holland.

—— (1979) "International Migration of the Highly Skilled: Economics, Ethics and Taxes," *Third World Quarterly*, 1: 17–30.

—— (2004) *In Defense of Globalizaton*, New York: Oxford University Press.

Bhagwati, J. N. and Dellafar, W. (1973) "The Brain Drain And Income Taxation," *World Development*, 1: 94–100.

Bhagwati, J. N. and Hamada, K. (1974) "The Brain Drain, International Integration Of Markets For Professionals And Unemployment: A Theoretical Analysis," *Journal of Development Economics*, 1: 19–24.

Bhagwati, J. (1975) "Domestic Distortions, Imperfect Information and the Brain Drain," *Journal of Development Economics*, 2: 139–153.

—— (1982) "Tax policy in the Presence of Emigration," *Journal of Public Economics*, 18: 291–317.

Bhagwati, J. N. and Wilson, J. D. (eds) (1989) *Income Taxation and International Mobility*, Cambridge, MA: MIT. Press.

Desai, M. A., Kapur, D. and McHale, J. (2003) "The Fiscal Impact of High Skilled Emigration: Flows Of Indians to the U.S.," WCFIA Working paper No. 03–01, Harvard University.

—— (2004) "Sharing the Spoils: Taxing International Human Capital Flows," *International Tax and Public Finance*, 11: 663–693.

Hufbauer, G. C. (1989) "The State, the Individual, and the Taxation of Economic Migration," in J. N. Bhagwati and J. D. Wilson (eds) *Income Taxation and International Mobility*, Cambridge, MA: MIT. Press.

Keen, M. and Ligthart, J. (2006) "Information Sharing and International Taxation: A Primer," *International Tax and Public Finance*, 13: 81–110.

Li, X. and J. McHale (2006) "Does Brain Drain Lead to Institutional Gain? A Cross Country Empirical Investigation," unpublished manuscript, University of British Columbia.

Mirrlees, J. A. (1971) "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38: 175–208.

—— (1982) "Migration and Optimal Income Taxes," *Journal of Public Economics*, 18: 319–341.

Stark, O. and Wang, Y. (2002) "Inducing Human Capital Formation: Migration as a Substitute for Subsidies," *Journal of Public Economics*, 86: 29–46.

Stark, O., Casarico, A., Devillanova, C. and Uebelmesser, S. (2005) "The New Economics of the Brain Drain: Mapping the Gains," unpublished manuscript.

Wilson, J. D. (1980) "The Effect of Potential Emigration on the Optimal Linear Income Tax," *Journal of Public Economics*, 14: 339–353.

—— (1982) "Optimal Linear Income Taxation in the Presence of Emigration," *Journal of Public Economics*, 18: 363–379.

—— (1992) "Optimal Income Taxation and International Personal Mobility," *American Economic Review*, *Papers and Proceedings*, 82: 191–196.

—— (2005a) "Income Taxation and Skilled Migration: The Analytical Issues," unpublished manuscript, Michigan State University.

—— (2005b) "Welfare-Improving Competition for Mobile Capital," *Journal of Urban Economics*, 57: 1–18.

Wilson, J. D. and Gordon, R. (2003) "Expenditure Competition," *Journal of Public Economic Theory*, 5: 399–417.

Index

Note: Page references in italics refer to tables, figures.