Notes on Model Specification

To go with Gujarati, Basic Econometrics, Chapter 13

Attributes of a Good Econometric Model

A.C. Harvey: (in Gujarati, p. 453-4)

- **Parsimony:** Explain alot with a little.
- **Identifiability:** For a set of data, unique values exist for the parameters.
- **Goodness of Fit**
- **Theoretical Consistency**
- **Predictive Power**
  1. **Outside** the sample.
  2. The model is based on quantities observed *prior* to the time the prediction is needed.

Types of Specification Errors

1. Omitted Variables.
2. Including an irrelevant variable.
3. Incorrect functional form.

Consequences of Omitted Variable:

- Everything is wrong.

Consequences of Including an Irrelevant Variable:

1. Estimates of the parameters are unbiased and consistent.
2. σ^2 is correctly estimated.
3. Hypotheses tests are valid.
4. The estimated standard errors are generally inefficient (i.e., the variances are larger than those of $\hat{\beta}$).
   - This violates the BLUE condition: the estimated parameters are not the *best* linear unbiased estimators.

Tests of Specification Errors

Detecting the Presence of an Unnecesary Variable

Note: Gujarati says to use a t-test or F-test. This is not a recomended procedure.

We know that $\sigma_{\hat{\beta}_j}$ increases as $\text{Var}(x_j)$ decreases. We do not want to infer that x_j does not influence y simply because we don't have good observations on x_j.

Detecting Omitted Variables and Incorrect Functional Form

Examine the residuals: specification errors show up as patterns in the plots.

Suggested Plots to look at:

1. Y versus $\hat{Y}$
2. e versus X_j for some independent variable of interest.
3. Y versus $\hat{Y}$ versus X_j for some independent variable of interest.

Ramsey RESET Test:

1. Get $\hat{Y}$.
2. Add variants of $\hat{Y}$ ($\hat{Y}^2$, $\hat{Y}^3$, etc) to the RHS of the model.
3. Estimate the new model.
4. If the $\hat{Y}$ terms are significant (use an F-test to test for joint-significance of all the $\hat{Y}$ terms), then you have specification error.

From this, we do not get information on alternatives or corrections.

But, it is a good test because we do not need to know what Xs are causing the problem.

Other Tests:

1. Likelihood Ratio Test
2. Wald Test
3. Lagrange Multiplier Test
4. Hausman Test

Lagrange Multiplier Test for Adding a Variable

- Estimate a restricted regression (some set of variables is omitted), compute the residuals e_i.
- Regress e_i on all the regressors.
- The quantity nR^2 will be distributed as a χ^2_m, where m is the number of restrictions.
- If nR^2 exceeds the critical value, we can reject the restricted model.

This test is similar to the Ramsey RESET test. But, this test could be accomplished with a standard F-test.

Measurement Error

Error in Measuring Y

Say the truth is:

$$Y_i^* \ = \ \beta_0 \ + \ \beta_1 X_i \ + \ u_i$$

We measure Y^* with error:

$$Y_i \ = \ Y_i^* \ + \ \epsilon_i$$

Now we estimate the model based on observed data:

$$\begin{aligned}
(Y_i \ + \ \epsilon_i) \ &= \ \beta_0 \ + \ \beta_1 X_i \ + \ u_i \\
Y_i \ &= \ \beta_0 \ + \ \beta_1 X_i \ + \ (u_i \ - \ \epsilon_i) \\
Y_i \ &= \ \beta_0 \ + \ \beta_1 X_i \ + \ \gamma_i
\end{aligned}$$

If we assume that ϵ_i is very stochastic and unrelated to u_i:

$$\begin{aligned}
\mathbf{E}[u_i] \ &= \ \mathbf{E}[\epsilon_i] \ = \ 0 \\
cov(X_i, u_i) \ &= \ cov(\epsilon_i, X_i) \ = \ 0 \\
cov(\epsilon, u_i) \ &= \ 0
\end{aligned}$$

Then we will have unbiased estimates. But, the variance will be increased.

With no measurement error:

$$\text{Var}(\hat{\beta}) \;=\; \frac{\sigma_u^2}{\Sigma(X_i \;-\; \bar{X})^2} \tag{1}$$

With measurement error:

$$\text{Var}(\hat{\beta}) \;=\; \frac{\sigma_\gamma^2}{\Sigma(X_i \;-\; \bar{X})^2} \;=\; \frac{\sigma_u^2 \;+\; \sigma_\epsilon^2}{\Sigma(X_i \;-\; \bar{X})^2} \tag{2}$$

Measurement Error in X

Say the truth is:

$$Y_i \;=\; \beta_0 \;+\; \beta_1 X_i^* \;+\; u_i$$

We measure X^* with error:

$$X_i \;=\; X_i^* \;+\; \epsilon_i$$

Now we estimate the model based on observed data:

$$\begin{aligned}
Y_i &= \beta_0 \;+\; \beta_1(X_i^* \;+\; \epsilon_i) \;+\; u_i \\
Y_i &= \beta_0 \;+\; \beta_1 X_i^* \;+\; (u_i \;+\; \beta_1\,\epsilon_i) \\
Y_i &= \beta_0 \;+\; \beta_1 X_i^* \;+\; \gamma_i
\end{aligned}$$

Now we **cannot** assume that $\mathrm{cor}(X_i, \gamma_i) = 0$. So, our OLS estimates will be biased.

Leamer on Model Selection

Reasons for Model specification tests (From Leamer (1978, as adapted in Gujarati)):

1. Hypothesis Testing: Choose a true model.
2. Interpretive: Interpret data involving several correlated variables.
3. Simplification: construct a "fruitful" model.
4. Proxy: Choose between measures that purport to measure the same variable.
5. Data Selection: Select the appropriate data for estimation and prediction.
6. Post Data Model Construction: Improve an existing model.

Hendry on Model Selection

From Hendry and Richard 1983, as reported in Gujarati:

1. Be data admissable. The predictions made from the model must be logically possible.
2. Be consistent with theory.
3. Have weakly exogenous regressors.
4. Exhibit parameter constancy.
5. Exhibit data coherency. The residuals must be stochastic.
6. Be encompassing.

Nested models: tests are straightforward via F-tests.

Non-Nested Models:

1) Goodness of fit comparisons.

2) Estimate a model encompassing the two competing models.

Model A:

$$Y_i = \alpha_{10} + \beta_{11} X_i + \epsilon_i$$

Model B:

$$Y_i = \alpha_{20} + \beta_{21} Z_i + v_i$$

Create an encompassing model which the competing models are nested within:

Model C:

$$Y_i = \alpha_{30} + \beta_{31} X_i + \beta_{32} Z_i + \gamma_i$$

If β_{31} is significant, but β_{32} is not significant, then we would prefer model A over model B. But, we cannot actually reject model B - we have no statistical test to do this.