Rauli Susmel
Econometrics 1

Homework 3 – Solutions

1. Prove the result that the restricted least squares estimator never has a larger variance matrix than the unrestricted least squares estimator.

2. Prove the result that the R^2 associated with a restricted least squares estimator is never larger than that associated with the unrestricted least squares estimator. Conclude that imposing restrictions never improves the fit of the regression.

3. Reverse Regression. This and the next exercise continue the analysis of Exercise 10, Chapter 8. In the earlier exercise, interest centered on a particular dummy variable in which the regressors were accurately measured. Here, we consider the case in which the crucial regressor in the model is measured with error. The paper by Kamlich and Polachek (1982) is directed toward this issue.
Consider the simple errors in variables model, $y = \alpha + \beta x^* + \varepsilon$, $x = x^* + u$, where u and ε are uncorrelated, and x is the erroneously measured, observed counterpart to x*.
(a) Assume that x*, u, and ε are all normally distributed with means μ*, 0, and 0, variances σ_*^2, σ_u^2, and σ_ε^2 and zero covariances. Obtain the probability limits of the least squares estimates of α and β.
(b) As an alternative, consider regressing x on a constant and y, then computing the reciprocal of the estimate. Obtain the probability limit of this estimate.
(c) Do the `direct' and `reverse' estimators bound the true coefficient?

4. Reverse Regression - Continued: Suppose that we use the following model:

$$y = \beta x^* + \gamma d + \varepsilon,$$
$$x = x^* + u.$$

For convenience, we drop the constant term. Assume that x*, ε, and u are independent normally distributed with zero means. Suppose that d is a random variable which takes the values one and zero with probabilities π and 1-π in the population, and is independent of all other variables in the model. To put this in context, the preceding model (and variants of it) have appeared in the literature on discrimination. We view y as a "wage" variable, x* as "qualifications" and x as some imperfect measure such as education. The dummy variable, d, is membership (d=1) or nonmembership (d=0) in some protected class. The hypothesis of discrimination turns on $\gamma<0$ versus $\gamma=0$.

(a) What is the probability limit of c, the least squares estimator of γ, in the least squares regression of y on x and d? [**Hints:** The independence of x* and d is important. Also,

$$\text{plim } \mathbf{d}'\mathbf{d}/n = \text{Var}[d] + E^2[d] = \pi(1-\pi) + \pi^2 = \pi.$$

This minor modification does not effect the model substantively, but greatly simplifies the algebra.]
Now, suppose that x* and d are not independent. In particular, suppose $E[x^*|d=1] = \mu^1$ and $E[\mathbf{x}^*|d=0] = \mu^0$. Then, $\text{plim}[\mathbf{x}^{*\prime}\mathbf{d}/n]$ will equal $\pi\, \mu^1$. Repeat the derivation with this assumption.

(b) Consider, instead, a regression of x on y and d. What is the probability limit of the coefficient on d in this regression? Assume that x* and d are independent.
(c) Suppose that x* and d are not independent, but γ is, in fact, less than zero. Assuming that both preceding equations still hold, what is estimated by $\bar{y}|d=1 - \bar{y}|d=0$? What does this quantity estimate if γ does equal zero?

In the regression of **y** on **x** and **d**, if **d** and **x** are independent, we can invoke the familiar result for least squares regression. The results are the same as those obtained by two simple regressions. It is instructive to verify this.

$$plim\begin{bmatrix}\mathbf{x'x}/n & \mathbf{x'd}/n\\ \mathbf{d'x}/n & \mathbf{d'd}/n\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{x'y}/n\\ \mathbf{d'y}/n\end{pmatrix}=\begin{bmatrix}\sigma_*^2+\sigma_u^2 & 0\\ 0 & \pi\end{bmatrix}^{-1}\begin{pmatrix}\beta\sigma_*^2\\ \gamma\pi\end{pmatrix}=\begin{pmatrix}\beta/\left(1+\sigma_u^2/\sigma_*^2\right)\\ \gamma\end{pmatrix}$$

Therefore, although the coefficient on **x** is distorted, the effect of interest, namely, γ, is correctly measured. Now consider what happens if and d are not independent. With the second assumption, we must replace the off diagonal zero above with plim(**x′d**/n). Since u and d are still uncorrelated, this equals Cov[x*,d]. This is

$$\text{Cov}[x^*, d] = E[x^*d] = \pi\, E[x^* d|\, d=1] + (1-\pi)E[x^*d|\, d=0] = \pi\, \mu^1.$$

Also, plim[**y′d**/n] is now

$$\beta\text{Cov}[\mathbf{x}^*,d] + \gamma\text{plim}(\mathbf{d'd}/n) = \beta\pi\mu^1 + \gamma\pi$$

and

$$\text{plim}[\mathbf{y'x}^*/n] = \beta\text{plim}[\mathbf{x}^{*\prime}\mathbf{x}^*/n] + \gamma\text{plim}[\mathbf{x}^{*\prime}\mathbf{d}/n] = \beta\,\sigma_*^2 + \gamma\pi\,\mu^1.$$

Then, the probability limits of the least squares coefficient estimators is

$$plim\begin{bmatrix}\mathbf{x'x}/n & \mathbf{x'd}/n\\ \mathbf{d'x}/n & \mathbf{d'd}/n\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{x'y}/n\\ \mathbf{d'y}/n\end{pmatrix}=\begin{bmatrix}\sigma_*^2+\sigma_u^2 & \pi\mu^1\\ \pi\mu^1 & \pi\end{bmatrix}^{-1}\begin{pmatrix}\beta\sigma_*^2+\gamma\pi\mu^1\\ \beta\pi\mu^1+\gamma\pi\end{pmatrix}=\begin{pmatrix}\beta/\left(1+\sigma_u^2/\sigma_*^2\right)\\ \gamma\end{pmatrix}$$

$$=\frac{1}{\pi(\sigma_*^2+\sigma_u^2)+\pi^2(\mu^1)^2}\begin{bmatrix}\pi & -\pi\mu^1\\ -\pi\mu^1 & \sigma_*^2+\sigma_u^2\end{bmatrix}\begin{pmatrix}\beta\sigma_*^2+\gamma\pi\mu^1\\ \beta\pi\mu^1+\gamma\pi\end{pmatrix}$$

$$=\frac{1}{\pi(\sigma_*^2+\sigma_u^2)+\pi^2(\mu^1)^2}\begin{pmatrix}\beta(\pi\sigma_*^2+\pi^2(\mu^1)^2)\\ \gamma(\pi(\sigma_*^2+\sigma_u^2)+\pi^2(\mu^1)^2)+\beta\pi\sigma_u^2\end{pmatrix}.$$

The second expression does reduce to

$$\text{plim } c = \gamma + \beta\pi\,\mu^1\,\sigma_u^2/[\pi(\sigma_*^2 + \sigma_u^2) - \pi^2(\mu^1)^2\,],$$

but the upshot is that in the presence of measurement error, the two estimators become an unredeemable hash of the underlying parameters. Note that both expressions reduce to the true parameters if σ_u^2 equals zero.
Finally, the two means are estimators of

$$E[y|d=1] = \beta E[x^*|d=1] + \gamma = \beta\mu^1 + \gamma$$

and

$$E[y|d=0] = \beta E[x^*|d=0] = \beta\mu^0,$$

so the difference is $\beta(\mu^1 - \mu^0) + \gamma$, which is a mixture of two effects. Which one will be larger is entirely indeterminate, so it is reasonable to conclude that this is not a good way to analyze the problem. If γ equals zero, this difference will merely reflect the differences in the values of x*, which may be entirely unrelated to the issue under examination here. (This is, unfortunately, what is usually reported in the popular press.)