Econometrics 710
Final Exam
May 13, 2004
Sample Answers

1. The model is just-identified by the moment $E(x_i e_i) = 0$, so the appropriate (and asymptotically efficient) estimator is OLS $\hat{\beta} = (X'X)^{-1}(X'Y)$ which has the asymptotic distribution $\sqrt{n}\left(\hat{\beta} - \beta\right) \to_d N(0, V)$, $V = Q^{-1}\Omega Q^{-1}$ where $Q = Ex_i x_i'$ and $\Omega = Ex_i x_i' e_i^2$. We can estimate V by the White estimator

$$\hat{V} = \left(n^{-1}\sum_{i=1}^{n} x_i x_i\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} x_i x_i \hat{e}_i^2\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i\right)^{-1}.$$

An asymptotically efficient estimator of $\theta = \beta_1\beta_2$ is $\hat{\theta} = \hat{\beta}_1\hat{\beta}_2$ which has the asymptotic distribution $\sqrt{n}\left(\hat{\theta} - \theta\right) \to_d N\left(0, h'Vh\right)$ where

$$h = \frac{\partial}{\partial\beta}\left(\beta_1\beta_2\right) = \begin{pmatrix} \beta_2 \\ \beta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Thus an asymptotic standard error for $\hat{\theta}$ is $s(\hat{\theta}) = \sqrt{n^{-1}\hat{h}'\hat{V}\hat{h}}$ where $\hat{h} = \left(\hat{\beta}_2\ \hat{\beta}_1\ 0\ \cdots\ 0\right)'$

(a) An asymptotic 95% confidence interval for θ is $\hat{\theta} \pm 1.96 s(\hat{\theta})$.

(b) Draw B samples with replacement from the data and construct $\hat{\beta}_b^*$ and $\hat{\theta}_b^* = \hat{\beta}_1^*\hat{\beta}_2^*$ on each. Let $\hat{q}_1^*$ and $\hat{q}_2^*$ be the 2.5% and 97.5% sample quantiles of $\hat{\theta}_b^*$. These are estimates of q_1^* and q_2^*, the bootstrap quantiles of the distribution of $\hat{\theta}^*$. A percentile 95% confidence interval for θ is $[\hat{q}_1^*, \hat{q}_2^*]$.
Alternatively, another percentile interval can be constructed as follows. Let $\hat{q}_1^*$ and $\hat{q}_2^*$ be the 2.5% and 97.5% sample quantiles of $\hat{\theta}_b^* - \hat{\theta}$. A percentile 95% confidence interval for θ is $[\hat{\theta} - \hat{q}_2^*, \hat{\theta} - \hat{q}_1^*]$.

(c) Draw B samples with replacement from the data and construct $\hat{\beta}_b^*$, $\hat{\theta}_b^* = \hat{\beta}_1^*\hat{\beta}_2^*$, and $s(\hat{\theta}^*) = \sqrt{n^{-1}\hat{h}^{*\prime}\hat{V}^*\hat{h}^*}$ on each. Let $\hat{q}_1^*$ and $\hat{q}_2^*$ be the 2.5% and 97.5% sample quantiles of $T_b^* = \left(\hat{\theta}_b^* - \hat{\theta}\right)/s(\hat{\theta}^*)$. These are estimates of q_1^* and q_2^*, the bootstrap quantiles of the distribution of T^*. The equal-tailed percentile-t 95% confidence interval for θ is $[\hat{\theta} - s(\hat{\theta})\hat{q}_2^*, \hat{\theta} - s(\hat{\theta})\hat{q}_1^*]$.

2.

(a) Since the model is a regression, conditioning on a subset of the x_i's does not affect the validity of the regression. The OLS estimator $\hat{\beta}$ remains consistent. Algebraically, we can write the estimator as

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i' 1\left(x_{1i} > 0\right)\right)^{-1} \sum_{i=1}^{n} x_i y_i 1\left(x_{1i} > 0\right)$$

Thus

$$\begin{aligned}
\hat{\beta} - \beta &= \left(n^{-1}\sum_{i=1}^{n} x_i x_i' 1\left(x_{1i} > 0\right)\right)^{-1} n^{-1}\sum_{i=1}^{n} x_i e_i 1\left(x_{1i} > 0\right) \\
&\to_p \left(E\left(x_i x_i' 1\left(x_{1i} > 0\right)\right)\right)^{-1} E\left(x_i e_i 1\left(x_{1i} > 0\right)\right) \\
&= 0
\end{aligned}$$

since

$$E\left(x_i e_i 1\left(x_{1i} > 0\right)\right) = E\left(x_i 1\left(x_{1i} > 0\right) E\left(e_i \mid x_i\right)\right) = 0$$

by the law of iterated expectations. For this to work, it is critical that the model is a regresssion rather than a projection. (In the latter case, OLS will be inconsisent for the population projection coefficient β.)

Technically we also need the side condition $E\left(x_i x_i' 1\left(x_{1i} > 0\right)\right) > 0$. This is not automatic. A necessary condition is that x_{1i} have positive support on the region $(0, \infty)$. (If all x_{1i} are negative, then the available sample is empty.)

(b) If we observe the observation only if $y_i > 0$, this the sample is truncated, not censored (Tobit) model. (What is important is not the label, but to understand that this is not the same model as the one introduced in class.) However, the basic fact remains that truncation based on the dependent variable renders naive estimation methods inconsistent. Indeed

$$\begin{aligned}
\hat{\beta} - \beta &= \left(n^{-1}\sum_{i=1}^{n} x_i x_i' 1\left(y_i > 0\right)\right)^{-1} n^{-1}\sum_{i=1}^{n} x_i e_i 1\left(y_i > 0\right) \\
&\to_p \left(E\left(x_i x_i' 1\left(y_i > 0\right)\right)\right)^{-1} E\left(x_i e_i 1\left(y_i > 0\right)\right) \\
&\neq 0
\end{aligned}$$

Adding the assumption that e_i is independent of x_i and $N(0, \sigma^2)$, and letting $z_i = e_i/\sigma \sim N(0,1)$, we find

$$\begin{aligned}
E\left(e_i 1\left(y_i > 0\right) \mid x_i\right) &= E\left(e_i 1\left(e_i > -x_i'\beta\right) \mid x_i\right) \\
&= \sigma E\left(z_i 1\left(z_i > -x_i'\beta/\sigma\right) \mid x_i\right) \\
&= \sigma \lambda\left(-x_i'\beta/\sigma\right)
\end{aligned}$$

where $\lambda(s) = \phi(s)/\Phi(s)$. Thus

$$E\left(x_i e_i 1\left(y_i > 0\right)\right) = E\left(x_i E\left(e_i 1\left(y_i > 0\right) \mid x_i\right)\right) = \sigma E\left(x_i \lambda\left(-x_i'\beta/\sigma\right)\right).$$

And

$$E\left(x_i x_i' 1\left(y_i > 0\right)\right) = E\left(x_i x_i' E\left(1\left(z_i > -x_i'\beta/\sigma\right) \mid x_i\right)\right) = E\left(x_i x_i' \Phi\left(x_i'\beta/\sigma\right)\right).$$

Together

$$\hat{\beta} \to_p \beta + \sigma\left(E\left(x_i x_i' \Phi\left(x_i'\beta/\sigma\right)\right)\right)^{-1} E\left(x_i \lambda\left(-x_i'\beta/\sigma\right)\right) \neq \beta.$$

3.

(a) We know that $E\hat{\mu} = \mu$ and thus $\hat{\mu}$ is unbiased for μ. However, $\hat{\theta} = g(\hat{\mu})$ with $g(x) = 1/x$ is a nonlinear function of $\hat{\mu}$, so will be a biased estimator.

(b) The function $g(x)$ is strictly convex. Thus by Jensen's inequality, $E\hat{\theta} = Eg(\hat{\mu}) < g(E\hat{\mu}) = g(\mu) = \theta$. The inequality is strict since $Var(\hat{\mu}) > 0$. Thus $\hat{\theta}$ has upwards bias.

(c) A second-order Taylor series expansion is (in general)

$$g(\hat{\mu}) \simeq g(\mu) + g'(\mu)(\hat{\mu} - \mu) + \frac{1}{2}g''(\mu)(\hat{\mu} - \mu)^2$$

Since $g(x) = x^{-1}$, it follows that $g'(x) = -x^{-2}$ and $g''(x) = 2x^{-3}$. Thus the expansion can be written as

$$\hat{\theta} \simeq \theta - \mu^{-2}(\hat{\mu} - \mu) + \mu^{-3}(\hat{\mu} - \mu)^2$$

Taking expectations,

$$E\hat{\theta} - \theta \simeq -\mu^{-2}E(\hat{\mu} - \mu) + \mu^{-3}E(\hat{\mu} - \mu)^2 = \mu^{-3}Var(\hat{\mu} - \mu) = \frac{\sigma^2}{\mu^3 n}$$

since $\hat{\mu}$ is a sample mean. This is a positive number, suggesting a positive bias, and is consistent with the implication of Jensen's ienquality from part 2. This expression can be used to suggest the magnitude of the bias. Note that it depends on the mean and variance (μ and σ^2) as well as the sample size n.

(d) Yes, the nonparametric bootstrap can be used here. Sampling from the observations with replacement, on each sample estimate $\hat{\mu}^*$ and $\hat{\theta}^* = 1/\hat{\mu}$. The bootstrap estimate of bias is $B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^* - \hat{\theta}$.

4. The three statistics are numerically identical, so the dispute is an illusion. The GMM Distance statistic equals the Wald statistic in linear models with linear restrictions. The GMM Distance statistic is the difference between the GMM criterion evaluated under the null and alternative hypotheses, e.g. $D = J_0 - J_1$. Since the model is just-identified, $J_1 = 0$. Thus $D = J_0$. The statistic J_0 is the test for overidentifying restrictions for the null model. Thus the three statistics are numerically identical.

5. Let $P = X(X'X)^{-1}X'$ and $PZ = \hat{Z}$. The 2SLS estimator is

$$\begin{aligned}
\hat{\beta} &= \left(Z'X(X'X)^{-1}X'Z\right)^{-1}\left(Z'X(X'X)^{-1}X'Y\right) \\
&= \left(Z'PZ\right)^{-1}\left(Z'PY\right) \\
&= \left(Z'PPZ\right)^{-1}\left(Z'PY\right) \\
&= \left(\hat{Z}'\hat{Z}\right)^{-1}\left(\hat{Z}'Y\right)
\end{aligned}$$

so the OLS regression of Y on $\hat{Z}$ yields the 2SLS slopes $\hat{\beta}$. However, the residuals from the latter regression are

$$\begin{aligned}
\tilde{e} &= Y - \hat{Z}\hat{\beta} \\
&= Y - PZ\hat{\beta} \\
&\neq Y - Z\hat{\beta} \\
&= \hat{e}
\end{aligned}$$

The correct 2SLS standard errors are calculated using the residuals $\hat{e}$, not $\tilde{e}$. Thus the "two-stage" procedure produces the correct estimates but not the correct standard errors.

This result does not depend upon whether or not the model is just identified or how homoskedasticity is treated.

6. The Monte Carlo procedure (as described) appears correct, but the conclusion is incomplete. (Side note: this is a Monte Carlo experiment, not a bootstrap procedure.) Note that the stated *conclusion* is that the test is oversized. This is a concrete statement about the true probability of Type I error. Specifically, let $p = P(T_n > 7.815)$. This can be any number. The test is properly sized if $p = .05$, undersized if $p < .05$ and oversized if $p > .05$. The stated conclusion is therefore equivalent to the rejection of the hypothesis that $p = .05$.

   The evidence in favor of this conclusion is that $\hat{p} = .07$. This is a point estimate, and has a sampling distribution. By the CLT, $\sqrt{B}(\hat{p} - p) \rightarrow_d N(0, p(1-p))$ as $B \rightarrow \infty$ Furthermore, under the null hypothesis of $p = .05$, $\sqrt{B}(\hat{p} - .05) \rightarrow_d N(0, (.05)(.95))$. Thus an appropriate test of $H_0 : p = .05$ is to reject for large values of the t-ratio

$$t = \frac{\hat{p} - .05}{\sqrt{(.05)(.95)/B}}$$

   In this particular case, $B = 200$ and $\sqrt{(.05)(.95)/200} = .015$, so $t = 1.33$. Alternatively, we can calculate standard errors for $\hat{p}$ using the formula $s(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/B} = .018$, yielding the similar t-ratio 1.11. In either case, we cannot reject H_0 at conventional significance levels. Based on this reasoning, it is incorrect to claim that the test must be oversized. A constructive recommendation would be to increase B.

   Alternatively, we can construct a confidence interval for the true unknown p. A 95% interval is $\hat{p} \pm 1.96 s(\hat{p}) = .07 \pm 1.96(.018) = [0.035, 0.105]$. The true p lies in this interval with 95% probability. Since the set includes .05, it is incorrect to conclude that the test is oversized.

   Another question might be: "If $B = 200$ is too small, how large should it be?". One simple answer is to ask how large should B be to reject at the 5% level the hypothesis that $p = .05$ when we observe $\hat{p} = .07$. This is equivalent to finding a B so that

$$\frac{.07 - .05}{\sqrt{(.05)(.95)/B}} > 1.96$$

   or

$$B > \left(\frac{1.96}{.02}\right)^2 (.05)(.95) = 457$$

   This is the smallest B for which $\hat{p} = .07$ allows us to reject $p = .05$