Econ 704 Econometrics I - Discussion 9

Professor: Jack Porter

TA: Emilio Cuilty Handout Desinged by: Jangsu Yoon

Exercise 1. (Consistent Asymptotic Variance Estimator) Assume that $y_i = X_i' \boldsymbol{\beta} + \varepsilon_i$, assumptions OLS1, OLS2 (or OLS2'), and OLS3' hold. Show that

$$A\hat{v}ar = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 X_i X_i'\right) \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \xrightarrow{p} \mathbb{E}\left(X_i X_i'\right)^{-1} \mathbb{E}\left(\varepsilon_i^2 X_i X_i'\right) \mathbb{E}\left(X_i X_i'\right)^{-1}.$$

Proof. This proof is borrowed from Bruce Hansen's textbook (p. 146) (BruceHansen). It is straightforward that $\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \mathbb{E}\left(X_i X_i'\right)$, and by continuous mapping theorem, $\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \xrightarrow{p} \mathbb{E}\left(X_i X_i'\right)^{-1}$. It remains to show that $\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{\varepsilon}_i^2 \xrightarrow{p} \mathbb{E}\left(X_i X_i' \varepsilon_i^2\right)$.

Observe that $\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \left(\hat{\varepsilon}_i^2 - \varepsilon_i^2\right) + \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \varepsilon_i^2$, and the second term converges in probability to $\mathbb{E}\left(X_i X_i' \varepsilon_i^2\right)$, so we need to show that the first term converges in probability to 0. First, use the triangle inequality:

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \left(\hat{\varepsilon}_i^2 - \varepsilon_i^2\right)\right\| &\leqslant \frac{1}{n}\sum_{i=1}^{n} \left\|X_i X_i' \left(\hat{\varepsilon}_i^2 - \varepsilon_i^2\right)\right\| \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\|X_i X_i'\right\| \cdot \left|\hat{\varepsilon}_i^2 - \varepsilon_i^2\right| \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\|X_i\right\|^2 \left|\hat{\varepsilon}_i^2 - \varepsilon_i^2\right|.
\end{aligned}$$

Note that we define a norm for a $k \times m$ matrix by $\left\|A\right\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m} a_{ij}^2}$.

Then $\left|\hat{\epsilon}_i{}^2 - \epsilon_i^2\right|$ is bounded by:

$$\begin{aligned}
\left|\hat{\varepsilon}_i^2 - \varepsilon_i^2\right| &= \left|\left(X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \varepsilon_i\right)^2 - \varepsilon_i^2\right| \\
&= \left|2\varepsilon_i X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \left[X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right]^2\right| \\
&\leqslant 2\left|\varepsilon_i X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right| + \left|X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right|^2 \\
&= 2\left|\varepsilon_i\right|\left|X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right| + \left|X_i'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right|^2 \\
&\leqslant 2\left|\varepsilon_i\right|\left\|X_i\right\|\left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right\| + \left\|X_i\right\|^2\left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right\|^2,
\end{aligned}$$

where the first equality follows from the triangle in equality and the second follows from the Cauchy-

Schwarz inequality.

Plugging into the triangle inequality for the summation, we find

$$\begin{aligned}\left\| \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \left(\hat{\varepsilon}_i^2 - \varepsilon_i^2 \right) \right\| &\leqslant 2 \left(\frac{1}{n} \sum_{i=1}^{n} |\varepsilon_i| \, \|X_i\|^3 \right) \left\| \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \right\| + \left(\frac{1}{n} \sum_{i=1}^{n} \|X_i\|^4 \right) \left\| \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \right\|^2 \\ &\xrightarrow{p} 0.\end{aligned}$$

□

Exercise 2. (Exercise 4.5. "Econometrics" by B. Hansen) Let (y_i, X_i) be a random sample with $E[y_i|X_i] = X_i'\beta$. Consider the Weighted Least Squares estimator of β

$$\tilde{\beta} = (X'WX)^{-1} (X'Wy)$$

where $W = diag(w_1, \ldots, w_n)$ and $w_i = x_{ji}^{-2}$, where x_{ji} is one of the X_i.

1. In which contexts would $\tilde{\beta}$ be a good estimator?

2. Using your intuition, in which situations would you expect that $\tilde{\beta}$ would perform better than OLS?

Proof. 1. First, $\tilde{\beta}$ is an unbiased estimator. That is,

$$E\left[\tilde{\beta}\right] = E\left[(X'WX)^{-1} (X'Wy)\right] = E\left[(X'WX)^{-1} (X'WE[y|X])\right] = \beta,$$

so the estimator is unbiased. Second, $\tilde{\beta}$ is the BLUE if $E[\epsilon\epsilon'|X] = W^{-1}$ by Gauss-Markov Theorem. To be specific, let $W^{1/2} = diag(\sqrt{w_1}, \ldots, \sqrt{w_n})$ and consider

$$\tilde{y} = \tilde{X}'\beta + \tilde{\epsilon}$$

where $\tilde{y} = W^{1/2}y$, $\tilde{X} = W^{1/2}X$, and $\tilde{\epsilon} = W^{1/2}\epsilon$. Then $E\left[\tilde{\epsilon}\tilde{\epsilon}'|\tilde{X}\right] = I_n$, where I_n is a $n \times n$ identity matrix. Then $\tilde{\beta}$ is the OLS estimator of $\tilde{y} = \tilde{X}'\beta + \tilde{\epsilon}$ satisfying (OLS 3), hence an efficient estimator.

2. If one of X_i (x_{ji}) is correlated with the variance of the error (where OLS 3 is violated), then this WLS would perform better than OLS. □

Exercise 3. Consider the linear model with two independent variables:

$$y = \alpha x + \beta z + \epsilon$$

where x, y, and z are $n \times 1$ observation vectors and ϵ is an error vector. Assume that ϵ has zero mean and covariance matrix $\sigma^2 I_n$ conditional on x and z. You are asked to investigate the properties of three alternative estimators for β : Let b denote the estimate resulting from a least-squares regression of y on x and z. Let b^* be the estimate obtained by regression y on z alone. Let b^{**} be the estimate resulting from the following step-wise procedure: First one regresses y on x alone, obtaining the estimate a^{**}; then one regresses the residual vector $y - a^{**}x$ on z, obtaining b^{**}.

1. Derive expressions for the mean and for the variance of each of the three estimators.

2. Show that $Var(b^{**}) \leq Var(b^*) \leq Var(b)$.

Proof. 1. Let's calculate the first estimator b. This is easily derived by Least square estimator formula.

$$b = (z'M_xz)^{-1}(z'M_xy)$$

where $y = (y_1, \dots, y_n)'$, $x = (x_1, \dots, x_n)'$, $z = (z_1, \dots, z_n)'$, $M_x = I_n - x(x'x)^{-1}x' = I_n - P_x$. Note that $I_n \in \mathfrak{M}_{n,n}(\mathbb{R})$ is a $n \times n$ matrix whose diagonal elements are 1 while other terms are zeros. $P_x := x(x'x)^{-1}x'$ is called as Projection matrix, which is symmetric and idempotent. Taking expectation to the estimator b,

$$E(b) = E\left[(z'M_xz)^{-1}(z'M_x(\alpha x + \beta z + \epsilon)\right] = \beta.$$

Why? The first term of $E(b)$ is $\alpha E\left[(z'M_xz)^{-1}(z'M_xx)\right]$ but $M_xx = (I_n - x(x'x)^{-1}x')x = x - x = 0$, since M_x is orthogonal to x. The second term is $E\left[(z'M_xz)^{-1}(z'M_xz)\beta\right] = \beta$. The last term is zero under $E\left[\epsilon|x,z\right] = 0$, because $E\left[(z'M_xz)^{-1}(z'M_xE\left[\epsilon|x,z\right])\right] = 0$. In the same sense, the conditional variance of this estimator is $\sigma^2(z'M_xz)^{-1}$, since

$$E\left[(z'M_xz)^{-1}(z'M_x\epsilon\epsilon'M_xz)(z'M_xz)^{-1}|x,z\right] = \sigma^2 E\left[(z'M_xz)^{-1}(z'M_xM_xz)(z'M_xz)^{-1}|x,z\right] = \sigma^2(z'M_xz)^{-1}$$

and M_x is an idempotent matrix.

Next estimator is b^*. This can be obtained by regression on z only. So by OLS estimator formula,

$$b^* = (z'z)^{-1}z'y.$$

$E(b^*|x,z) = E\left[(z'z)^{-1}(z'(\alpha x + \beta z + \epsilon))|x,z\right] = \alpha(z'z)^{-1}z'x + \beta$ so that

$$E\left(b^*\right) = \alpha E\left((z'z)^{-1}z'x\right) + \beta,$$

hence b^* is a biased estimator. The conditional variance is much simpler, by $\sigma^2(z'z)^{-1}$ since

$$E\left[(z'z)^{-1}z'uu'z(z'z)^{-1}|x,z\right] = \sigma^2 E\left[(z'z)^{-1}(z'z)(z'z)^{-1}|x,z\right] = \sigma^2(z'z)^{-1}.$$

The last step is about b^{**}. The step follows: first one regresses y on x alone, obtaining the estimate a^{**}, then one regresses the residual vector $y - a^{**}x$ on z, obtaining b^{**}. So $a^{**} = (x'x)^{-1}x'y$, hence $y - a^{**}x = y - [(x'x)^{-1}x'y]x = y - x(x'x)^{-1}x'y$. Thus $y - a^{**}x = y - P_xy = M_xy$. The second step is to regress M_xy on z, hence $b^{**} = (z'z)^{-1}z'M_xy$. Hence

$$\begin{aligned} E(b^{**}|x,z) &= E\left[(z'z)^{-1}(z'M_x(\alpha x + \beta z + \epsilon))|x,z\right] \\ &= \alpha(z'z)^{-1}z'M_xx + \beta(z'z)^{-1}z'M_xz = \beta(z'z)^{-1}z'M_xz, \end{aligned}$$

so that b^{**} is also a biased estimator. The conditional variance is easy, by $\sigma^2(z'z)^{-1}(z'M_xz)(z'z)^{-1}$ since

$$E\left[(z'z)^{-1}z'M_xuu'M_xz(z'z)^{-1}|x,z\right] = \sigma^2 E\left[(z'z)^{-1}(z'M_xz)(z'z)^{-1}|x,z\right] = \sigma^2(z'z)^{-1}(z'M_xz)(z'z)^{-1}.$$

2. Show $Var(b) \geq Var(b^*)$. It holds if and only if $\sigma^2(z'M_xz)^{-1} \geq \sigma^2(z'z)^{-1}$. This is the same inequality of $z'z \geq z'M_xz$ or $z'(I_n - M_x)z = z'P_xz \geq 0$. To be sure, P_x is projection matrix, so that a symmetric and idempotent matrix, hence a positive semi-definite matrix. Therefore $z'P_xz \geq 0$ holds. This provides

$Var(b) \geq Var(b^*)$.

Next, we show $Var(b^*) \geq Var(b^{**})$. It is the same as showing $\sigma^2(z'z)^{-1} \geq \sigma^2(z'z)^{-1}(z'M_xz)(z'z)^{-1}$. This is modified by $I_n - (z'M_xz)(z'z)^{-1} \geq 0$. Since $(z'M_xz)(z'z)^{-1} = (z'z - z'P_xz)(z'z)^{-1} = I_n - z'P_xz(z'z)^{-1}$, so $I_n - (z'M_xz)(z'z)^{-1} = (z'P_xz)(z'z)^{-1}$. Since P_x is a positive semi-definite matrix, $z'P_xz \geq 0$. At last, $(z'z)^{-1} \geq 0$ by definition. Hence $Var(b^*) \geq Var(b^{**})$. □