Econometrics 710 Problem Set #6

March 1, 2015

TA: Cristian Hernandez
cahernandez3@wisc.edu

Exercise 6.1

$$\begin{aligned}
\tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y \\
&= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + e) \\
&= \beta_1 + (\frac{1}{n}X_1'X_1)^{-1}(\frac{1}{n}X_1'X_2)\beta_2 + (\frac{1}{n}X_1'X_1)^{-1}(\frac{1}{n}X_1'e) \\
&\xrightarrow{p} \beta_1 + (\mathbb{E}x_{1i}x_{1i}')^{-1}(\mathbb{E}x_{1i}x_{2i}')\beta_2,
\end{aligned}$$

by the WLLN $n^{-1}\sum_i x_{1i}x_{1i}' \to_p \mathbb{E}x_{1i}x_{1i}', n^{-1}\sum_i x_{1i}x_{2i}' \to_p \mathbb{E}x_{1i}x_{2i}', n^{-1}\sum_i x_{1i}e_i \to_p \mathbb{E}x_{1i}e_i = 0$. In general, $\tilde{\beta}_1$ is not a consistent estimator for β_1 when $\mathbb{E}x_{1i}x_{2i}' \neq 0, \beta_2 \neq 0$. However, if $\mathbb{E}x_{1i}x_{2i}' = 0$, or $\beta_2 = 0$, then $\tilde{\beta}_1$ is a consistent estimator for β_1.

Exercise 6.2

$$\begin{aligned}
\hat{\beta} &= (X'X + \lambda I_k)^{-1}X'y \\
&= (X'X + \lambda I_k)^{-1}X'(X\beta + e) \\
&= \left(\frac{1}{n}X'X + \frac{1}{n}\lambda I_k\right)^{-1}\left(\frac{1}{n}X'X\beta + \frac{1}{n}X'e\right) \\
&\xrightarrow{p} Q_{xx}^{-1}(Q_{xx}\beta + 0) = \beta
\end{aligned}$$

by the WLLN, $\mathbb{E}x_ie_i = 0$, and $\frac{1}{n}\lambda I_k \to 0$ as $n \to \infty$ for fixed constant λ. So, $\hat{\beta}$ is a consistent estimator for β.

Exercise 6.3

When $\lambda_n = cn$

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{n}X'X + cI_k\right)^{-1}\left(\frac{1}{n}X'X\beta + \frac{1}{n}X'e\right) \\
&\xrightarrow{p} (Q_{xx} + cI_k)^{-1}(Q_{xx}\beta + 0) \\
&= (Q_{xx} + cI_k)^{-1}(Q_{xx} + cI_k - cI_k)\beta \\
&= \beta - c(Q_{xx} + cI_k)^{-1}\beta \neq \beta
\end{aligned}$$

Exercise 6.4

Note that $Pr(\mathbf{x}_{1i} = 1) = Pr(\mathbf{x}_{1i} = -1) = Pr(\mathbf{x}_{2i} = 1) = Pr(\mathbf{x}_{2i} = -1) = \frac{1}{2}$

(1) $\mathbb{E}\mathbf{x}_{1i} = 1/2 \cdot 1 + 1/2 \cdot (-1) = 0$
(2) $\mathbb{E}\mathbf{x}_{1i}^2 = 1/2 \cdot 1^2 + 1/2 \cdot 1^2 = 1$
(3) $\mathbb{E}\mathbf{x}_{1i}\mathbf{x}_{2i} = 3/8 \cdot 1^2 + 3/8 \cdot (-1)^2 + 2 \cdot 1/8 \cdot 1 \cdot (-1) = \frac{1}{2}$

(4) $\mathbb{E}e_i^2 = \mathbb{E}(e_i^2|\mathbf{x}_{1i} = \mathbf{x}_{2i})Pr(\mathbf{x}_{1i} = \mathbf{x}_{2i}) + \mathbb{E}(e_i^2|\mathbf{x}_{1i} \neq \mathbf{x}_{2i})Pr(\mathbf{x}_{1i} \neq \mathbf{x}_{2i}) = 5/4 \cdot 3/4 + 1/4 \cdot 1/4 = 1$.
(5) $\mathbb{E}(\mathbf{x}_{1i}^2 e_i^2) = \mathbb{E}(\mathbf{x}_{1i}^2 e_i^2|\mathbf{x}_{1i} = \mathbf{x}_{2i})Pr(\mathbf{x}_{1i} = \mathbf{x}_{2i}) + \mathbb{E}(\mathbf{x}_{1i}^2 e_i^2|\mathbf{x}_{1i} \neq \mathbf{x}_{2i})Pr(\mathbf{x}_{1i} \neq \mathbf{x}_{2i}) = 1 \cdot 5/4 \cdot 3/4 + 1 \cdot 1/4 \cdot 1/4 = 1$
(6) Note that $\mathbb{E}(e_i^2|\mathbf{x}_{1i}, \mathbf{x}_{2i}) = 5/4$ if $\mathbf{x}_{1i} = \mathbf{x}_{2i}$ and $\mathbb{E}(e_i^2|\mathbf{x}_{1i}, \mathbf{x}_{2i}) = 1/4$ if $\mathbf{x}_{1i} \neq \mathbf{x}_{2i}$. Then, $\mathbb{E}(\mathbf{x}_{1i}\mathbf{x}_{2i}e_i^2) = \mathbb{E}(\mathbb{E}(\mathbf{x}_{1i}\mathbf{x}_{2i}e_i^2|\mathbf{x}_{1i}, \mathbf{x}_{2i})) = \mathbb{E}(\mathbf{x}_{1i}\mathbf{x}_{2i}\mathbb{E}(e_i^2|\mathbf{x}_{1i}, \mathbf{x}_{2i})) = 1 \cdot 1 \cdot (5/4) \cdot (3/4) + 1 \cdot (-1) \cdot (1/4) \cdot (1/4) = 7/8$.

Exercise 6.5

Note that

$$Q_{xx}^{-1} = \begin{bmatrix} Q_{11\cdot 2}^{-1} & -Q_{11\cdot 2}^{-1}Q_{12}Q_{22}^{-1} \\ -Q_{22\cdot 1}^{-1}Q_{21}Q_{11}^{-1} & Q_{22\cdot 1}^{-1} \end{bmatrix}, V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

$$\begin{aligned} V_{11} &= (Q_{11\cdot 2}^{-1}\Omega_{11} - Q_{11\cdot 2}^{-1}Q_{12}Q_{22}^{-1}\Omega_{21})Q_{11\cdot 2}^{-1} - (Q_{11\cdot 2}^{-1}\Omega_{12} - Q_{11\cdot 2}^{-1}Q_{12}Q_{22}^{-1}\Omega_{22})Q_{22\cdot 1}^{-1}Q_{21}Q_{11}^{-1} \\ &= Q_{11\cdot 2}^{-1}\Omega_{11}Q_{11\cdot 2}^{-1} - Q_{11\cdot 2}^{-1}Q_{12}Q_{22}^{-1}\Omega_{21}Q_{11\cdot 2}^{-1} - Q_{11\cdot 2}^{-1}\Omega_{12}Q_{22\cdot 1}^{-1}Q_{21}Q_{11}^{-1} + Q_{11\cdot 2}^{-1}Q_{12}Q_{22}^{-1}\Omega_{22}Q_{22\cdot 1}^{-1}Q_{21}Q_{11}^{-1} \\ &= Q_{11\cdot 2}^{-1}(\Omega_{11} - Q_{12}Q_{22}^{-1}\Omega_{21} - \Omega_{12}Q_{22}^{-1}Q_{21} + Q_{12}Q_{22}^{-1}\Omega_{22}Q_{22}^{-1}Q_{21})Q_{11\cdot 2}^{-1} \end{aligned}$$

Last equality comes from symmetry of Q_{xx}^{-1}, i.e., $(Q_{22\cdot 1}^{-1}Q_{21}Q_{11}^{-1}) = Q_{22}^{-1}Q_{21}Q_{11\cdot 2}^{-1}$. Similarly, V_{21}, V_{22} can be calculated as (6.22), (6.23).

Exercise 6.6

The moment equations are

$$\begin{gathered} \mathbb{E}(x_i e_i) = 0 \\ \mathbb{E}(x_i x_i' e_i^2 - \Omega) = 0 \end{gathered}$$

, or equivalently

$$\begin{gathered} \mathbb{E}(x_i(y_i - x_i'\beta)) = 0 \\ \mathbb{E}(x_i x_i'(y_i - x_i'\beta)^2 - \Omega) = 0 \end{gathered}$$

The method of moments estimators can be found by replacing the population moments with the sample moments, thus $(\hat{\beta}^{MM}, \hat{\Omega}^{MM})$ satisfy following equations;

$$\begin{gathered} \frac{1}{n}\sum_{i=1}^n x_i(y_i - x_i'\hat{\beta}) = 0 \\ \frac{1}{n}\sum_{i=1}^n (x_i x_i'(y_i - x_i'\hat{\beta})^2 - \hat{\Omega}) = 0 \end{gathered}$$

Then we get,

$$\begin{aligned} \hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i y_i\right) \\ \hat{\Omega} &= \frac{1}{n}\sum_{i=1}^n x_i x_i' \hat{e}_i^2 \end{aligned}$$

where $\hat{e}_i = y_i - \mathbf{x}_i'\hat{\beta}$.

Since β, Ω are continuously differentiable functions of population moments, the sample analogue estimators are semiparametrically efficient estimators for their population counterparts, in the sense that no semiparametric estimator can have a smaller asymptotic variance and their asymptotic variances equal to the semiparametric efficiency bound. To be more specifically, recall the proposition 5.13.2 (or proposition 6.19.1). Note that $\beta = Q_{xx}^{-1}Q_{xy}$, and $\Omega = \mathbb{E}(x_i x_i' y_i^2) - 2\mathbb{E}(x_i x_i' y_i x_i')\beta + \mathbb{E}(x_i x_i' \beta' x_i x_i' \beta) = \mathbb{E}(x_i x_i y_i^2) - 2\mathbb{E}(x_i x_i' y_i x_i')Q_{xx}^{-1}Q_{xy} + \mathbb{E}(x_i x_i' (Q_{xx}^{-1}Q_{xy})' x_i x_i' Q_{xx}^{-1} Q_{xy})$

Consider the class of distributions

$$L_8(\beta, \Omega) = \{F : \mathbb{E}y^8 < \infty, \mathbb{E}||x||^8 < \infty, \mathbb{E}x_i x_i' > 0\}$$

$(\hat{\beta}, \hat{\Omega})$ are semiparametrically efficient in the class of distributions $F \in \mathcal{L}_8(\beta, \Omega)$, if there is no further information about the model other than two moment conditions.

$(x_i x_i' y_i^2, x_i x_i' y_i x_i, x_i x_i', x_i y_i, x_i x_i' (Q_{xx}^{-1}Q_{xy})' x_i x_i')$ has finite variance if $F \in L_8(\beta, \Omega)$ by Cauchy-Schwartz inequality.

Exercise 6.7

(a) In the first two equations, β is defined as linear projection coefficient of y_i^* on x_i.

$$\beta = (\mathbb{E}x_i x_i')^{-1}(\mathbb{E}x_i y_i^*) = (\mathbb{E}x_i x_i')^{-1}(\mathbb{E}x_i(y_i - u_i)) = (\mathbb{E}x_i x_i')^{-1}(\mathbb{E}x_i y_i)$$

Last two equality holds because of measurement error assumption and $\mathbb{E}x_i u_i = 0$. Therefore, β is also coefficient from the linear projection of y_i on x_i

(b)

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}(X'y) = (X'X)^{-1}(X'(y^* + u)) \\
&= (\frac{1}{n}X'X)^{-1}(\frac{1}{n}X'(X\beta + u + e)) \\
&= \beta + (\frac{1}{n}\sum_{i=1}^{n} x_i x_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} x_i(u_i + e_i)) \xrightarrow{p} \beta + Q_{xx}^{-1} \cdot 0 = \beta
\end{aligned}$$

since $\mathbb{E}x(u_i + e_i) = 0$. $\hat{\beta}$ is consistent for β.

(c) $\sqrt{n}(\hat{\beta} - \beta) = (\frac{1}{n}\sum_{i=1}^{n} x_i x_i')^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i(e_i + u_i) \rightarrow_d N(\mathbf{0}, V)$, such that

$$V = Q_{xx}^{-1}\Omega Q_{xx}^{-1}$$

$$Q_{xx} = \mathbb{E}x_i x_i', \quad \Omega = \mathbb{E}(x_i x_i'(u_i + e_i)^2)$$

Exercise 6.8

From the equation (6.25)

$$\begin{aligned}\sqrt{n}(\hat{\sigma}^2-\sigma^2) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}e_i^2-\sigma^2\right)-2\left(\frac{1}{n}\sum_{i=1}^{n}e_ix_i'\right)\sqrt{n}(\hat{\beta}-\beta)+\sqrt{n}(\hat{\beta}-\beta)'\left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)(\hat{\beta}-\beta)\\ &= \sqrt{n}\left(\frac{1}{n}\sum_i e_i^2-\sigma^2\right)-2\cdot o_p(1)O_p(1)+O_p(1)O_p(1)o_p(1)\\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}e_i^2-\sigma^2\right)+o_p(1)\xrightarrow{d}N(0,\mathrm{var}(e_i^2))=N(0,\mathbb{E}e_i^4-\sigma^4)\end{aligned}$$

since $\frac{1}{n}\sum_{i=1}^{n}e_ix_i'=o_p(1)$, $\sqrt{n}(\hat{\beta}-\beta)=O_p(1)$, $\hat{\beta}-\beta=o_p(1)$, and $\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'=O_p(1)$.

Exercise 6.9

(a) We know that OLS estimator $\hat{\beta}$ is consistent.

$$\tilde{\beta} = \frac{1}{n}\sum_i\frac{x_i\beta+e_i}{x_i}=\beta+\frac{1}{n}\sum_i\frac{e_i}{x_i}\xrightarrow{p}\beta,$$

since $n^{-1}\sum_i\frac{e_i}{x_i}\to_p\mathbb{E}(\frac{e_i}{x_i})=\mathbb{E}(\frac{1}{x_i}\mathbb{E}(e_i|x_i))=0$ by the WLLN assuming that $\mathbb{E}|\frac{e_i}{x_i}|<\infty$. Thus, they are both consistent estimators for β.

(b) We also know that the OLS estimator $\hat{\beta}$ is semiparametrically efficient under the homoskedastic regression, i.e. $\mathbb{E}(e_i^2|x_i)=\sigma^2$.

Consider $\tilde{\beta}$, which is WLS using weights $w_i=x_i^{-2}$. In the heteroskedastic regression model, the GLS estimator is semiparametrically efficient, thus $\tilde{\beta}$ is efficient when $w_i=\sigma_i^{-2}$, i.e. $\mathbb{E}(e_i^2|x_i)=\sigma_i^2=x_i^2$.

Exercise 6.10

(a) A point forecast of y_{n+1} is $x_{n+1}'\widehat{\beta}$, where $\widehat{\beta}$ is computed with the sample $(y_i,x_i)_{i=1}^n$.

(b) The forecast error is $\widehat{e}_{n+1}=y_{n+1}-x'\widehat{\beta}$. Under the homoskedastic assumption, the variance of this forecast error is $\mathbb{E}\hat{e}_{n+1}=\sigma^2+x'V_{\hat{\beta}}x$ following from chapter 6.15. A natural estimator is $\hat{\sigma}^2+x'\hat{V}_{\hat{\beta}}x$

Exercise 6.11

(a)

$$\widehat{log(Wage)}=\underset{(0.0029)}{0.0904}Education+\underset{(0.0026)}{0.0354}Experience-\underset{(0.0053)}{0.0465}Experience^2/100+\underset{(0.046)}{1.185}$$

Here, I report Horn-Horn-Duncan robust standard errors.

(b)

$$\theta = r(\beta) = \frac{\beta_1}{\beta_2 + \frac{\beta_3}{50} experience}$$

Note that parameter of interest θ is also a function of a variable x, where $x = experience$. For given level of $experience = x$, $\hat{\theta}(x) = \frac{\hat{\beta}_1}{\hat{\beta}_2 + \frac{\hat{\beta}_3}{50} x}$, where $\hat{\beta}$ are OLS estimates from wage equation (a).

(c) For a given level of $experience = x$, the asymptotic standard error for $\hat{\theta}$ is

$$s(\hat{\theta}) = \sqrt{\hat{R}' \hat{V}_{\hat{\beta}} \hat{R}}$$

where $\hat{R}$ is a consistent estimator of $R = \frac{\partial r(\beta)}{\partial \beta} = \begin{pmatrix} \frac{1}{\beta_2 + \beta_3 x/50} \\ \frac{-\beta_1}{(\beta_2 + \beta_3 x/50)^2} \\ \frac{-\beta_1 x/50}{(\beta_2 + \beta_3 x/50)^2} \\ 0 \end{pmatrix}$ and $\hat{V}_{\hat{\beta}}$ is a consistent estimator of the covariance matrix of $\hat{\beta}$. Specifically, $\hat{R} = \begin{pmatrix} \frac{1}{\hat{\beta}_2 + \hat{\beta}_3 x/50} \\ \frac{-\hat{\beta}_1}{(\hat{\beta}_2 + \hat{\beta}_3 x/50)^2} \\ \frac{-\hat{\beta}_1 x/50}{(\hat{\beta}_2 + \hat{\beta}_3 x/50)^2} \\ 0 \end{pmatrix}$ with OLS estimates $\hat{\beta}$ and I use $\hat{V}_{\hat{\beta}} = \bar{V}_{\hat{\beta}}$, which is the Horn-Horn-Duncan robust covariance matrix estimator. (I also use the same covariance matrix estimator for the regression interval (e) and the forecast interval (f).)

For example, for $experience = 10$, $\widehat{s}(\widehat{\theta}) = 0.2053$.

(d) For a given level of $experience = x$, a 90% asymptotic confidence interval for θ from the estimated model is

$$\hat{\theta} \pm 1.645 s(\hat{\theta})$$

where $\hat{\theta}$ is the estimate and $s(\hat{\theta})$ is the standard error from parts (b) and (c). Figure 1 shows the results.

(e) Let $z = (12, 20, 20^2/100, 1)'$. A 95% confidence interval for the regression function at this point is

$$z'\hat{\beta} \pm 1.96 \sqrt{z' \hat{V}_{\hat{\beta}} z} = [2.769, 2.815]$$

(f) Let $z = (16, 5, 5^2/100, 1)'$. A 80% forecast interval for log(wage) is

$$z'\hat{\beta} \pm 1.28 \sqrt{\hat{\sigma}^2 + z' \hat{V}_{\hat{\beta}} z} = [2.063, 3.533]$$

where $\hat{\sigma}^2$ is an estimator of σ^2. I use $s^2 = \frac{1}{n-k} \sum_{i=1}^{n} \hat{e}_i^2$. The forecast interval for the wage is obtained by taking the exponential function to the endpoints of above interval, which is $[7.867, 34.218]$

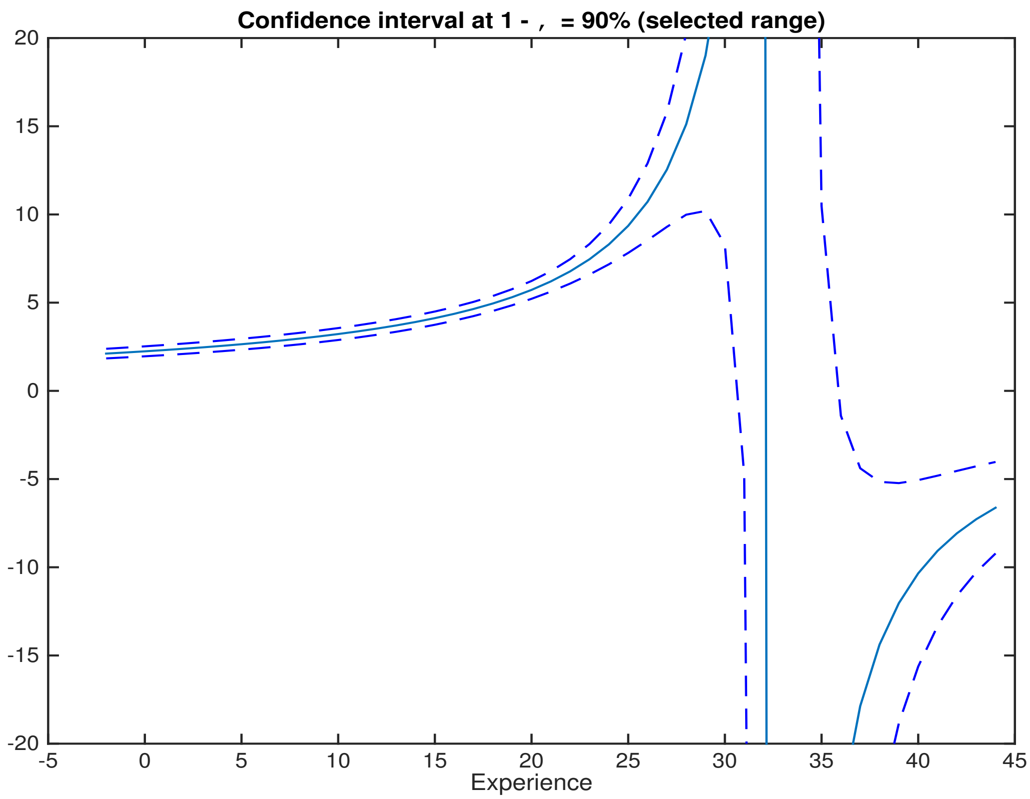

Figure 1: Estimates of the ratio of the returns to education to the returns to experience ($\hat{\theta}$) as function of experience with with 90% confidence interval at each point.