Financial Econometrics
MT Week 5 Assignment Answers

November 26, 2009

2.4 **Let $\hat{\mathbf{S}}$ be the sample covariance matrix of $\mathbf{z} = [\mathbf{y}\ \mathbf{X}]$, where X *does not include a constant***

$$\hat{\mathbf{S}} = n^{-1}\sum_{i=1}^{n}(\mathbf{z}_i - \bar{\mathbf{z}})'(\mathbf{z}_i - \bar{\mathbf{z}})$$

$$\hat{\mathbf{S}} = \begin{bmatrix} \hat{s}_{yy} & \hat{\mathbf{s}}'_{xy} \\ \hat{\mathbf{s}}_{xy} & \hat{\mathbf{S}}_{xx} \end{bmatrix}$$

and suppose n, the sample size, is known ($\hat{\mathbf{S}}$ is the sample covariance estimator). Under the small sample assumptions (including homoskedasticity and normality if needed), describe one method, using only $\hat{\mathbf{S}}$, $\bar{\mathbf{X}}$ (the 1 by $k-1$ sample mean of the matrix X, column-by-column), $\bar{y}$ and n, to

(a) **Estimate $\hat{\beta}_1, \ldots, \hat{\beta}_k$ from a model**

$$y_i = \beta_1 + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \epsilon_i$$

Since the model includes a constant, we know that $\sum_{i=1}^{n}\hat{\epsilon}_i = 0$ since it is one of the first order conditions. This then implies that $\sum_{i=1}^{n} y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2,i} - \ldots - \hat{\beta}_k x_{k,i} = 0$ and so must $\bar{y} - \hat{\beta}_1 - \hat{\beta}_2\bar{x}_1 - \ldots - \hat{\beta}_k\bar{x}_k = 0$, which is the first sum divided by n. Putting this together with the model evaluated at the OLS parameter estimates, we can conclude that

$$y_i - \bar{y} = \hat{\beta}_2\left(x_{2,i} - \bar{x}_2\right) + \ldots + \hat{\beta}_k\left(x_{k,i} - \bar{x}_k\right) + \hat{\epsilon}_i$$

This tells us that regression on demeaned data (known as regression through the origin) must have the same slope coefficients ($\beta_2, \ldots, \beta_k$) and an intercept of zero. Thus, when using demeaned data,

$$\tilde{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n}\right)^{-1}\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{y}}}{n} = \hat{\mathbf{S}}_{xx}^{-1}\hat{\mathbf{s}}_{xy}$$

where $\tilde{\mathbf{X}} = \mathbf{X} - \boldsymbol{\iota}\bar{\mathbf{X}}$ is the demeaned matrix of x data and $\tilde{\mathbf{y}} = \mathbf{y} - \bar{y}$ is the demeaned y data.[1] The intercept, β_1 can be estimated by

$$\hat{\beta}_1 = \bar{y} - \tilde{\boldsymbol{\beta}}'\bar{\mathbf{X}}$$

and so

$$\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_1\ \tilde{\boldsymbol{\beta}}'\right]'$$

Intuition: This problem highlights the strong link between the covariance of $\mathbf{x}_i$ and y_i when estimating linear regression coefficient. Aside from the intercept, all of the coefficients are normalized (by $\hat{\mathbf{S}}_{xx}^{-1}$) covariances ($\mathbf{s}_{xy}$). Moreover, tests that βs are zero can be directly interpreted as tests that covariances are zero.

[1] $\boldsymbol{\iota}$ is a n by 1 vector of 1s.

(b) **Estimate s, the standard error of the regression** By the identities used to derive the R^2,

$$TSS = SSE + RSS$$

and so

$$SSE = TSS - RSS = n\left(\hat{s}_{yy} - \tilde{\boldsymbol{\beta}}'\hat{\mathbf{S}}_{xx}\tilde{\boldsymbol{\beta}}\right)$$

and thus the standard error of the regression is

$$\hat{s} = \sqrt{\frac{SSE}{n-k}} = \sqrt{\frac{n\left(\hat{s}_{yy} - \tilde{\boldsymbol{\beta}}'\hat{\mathbf{S}}_{xx}\tilde{\boldsymbol{\beta}}\right)}{n-k}}$$

2.9 **Let y_i and x_i conform to the small sample assumptions and let $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$. Define another estimator**

$$\breve{\beta}_2 = \frac{\bar{y}_H - \bar{y}_L}{\bar{x}_H - \bar{x}_L}$$

where $\bar{x}_H$ is the average value of x_i given $x_i >$ median (x), and $\bar{y}_H$ is the average value of y_i for n such that $x_i >$ median (x). $\bar{x}_L$ is the average value of x_i given $x_i \leq$ median (x), and $\bar{y}_L$ is the average value of y_i for n such that $x_i \leq$ median (x) (both $\bar{x}$ and $\bar{y}$ depend on the order of x_i, and not y_i). For example suppose the x_i were ordered such that $x_1 < x_2 < x_3 < \ldots < x_i$ and n is even. Then,

$$\bar{x}_L = \frac{2}{n}\sum_{i=1}^{n/2} x_i$$

and

$$\bar{x}_H = \frac{2}{n}\sum_{i=n/2+1}^{n} x_i$$

(i) **Is $\breve{\beta}_2$ unbiased, conditional on X?**

$$\begin{aligned}
\breve{\beta}_2 =& \frac{\bar{y}_H - \bar{y}_L}{\bar{x}_H - \bar{x}_L} \\
=& \frac{\beta_1 + \beta_2\bar{x}_H + \bar{\epsilon}_H - \beta_1 - \beta_2\bar{x}_L - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L} \\
=& \frac{\beta_2\left(\bar{x}_H - \bar{x}_L\right) + \bar{\epsilon}_H - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L} \\
=& \frac{\beta_2\left(\bar{x}_H - \bar{x}_L\right)}{\bar{x}_H - \bar{x}_L} + \frac{\bar{\epsilon}_H - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L} \\
=& \beta_2 + \frac{\bar{\epsilon}_H - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L} \\
\mathrm{E}\left[\breve{\beta}_2 \middle| \mathbf{X}\right] =& \beta_2 + \mathrm{E}\left[\frac{\bar{\epsilon}_H - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L}\middle| \mathbf{X}\right] \\
=& \beta_2 + \frac{\mathrm{E}\left[\bar{\epsilon}_H \middle| \mathbf{X}\right] - \mathrm{E}\left[\bar{\epsilon}_L \middle| \mathbf{X}\right]}{\bar{x}_H - \bar{x}_L}
\end{aligned}$$

$$\begin{aligned} =&\beta_2 + \frac{0-0}{\bar{x}_H - \bar{x}_L} \\ =&\beta_2 \end{aligned}$$

so it is unbiased.

(ii) **Is $\breve{\beta}_2$ consistent? Are any additional assumptions needed beyond those of the small sample framework?**

$$\begin{aligned} \text{plim}\breve{\beta}_2 - \beta_2 =& \text{plim}\frac{\bar{\epsilon}_H - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L} \\ =& \frac{\text{plim}\bar{\epsilon}_H - \text{plim}\bar{\epsilon}_L}{\text{plim}\bar{x}_H - \text{plim}\bar{x}_L} \\ =& \frac{0-0}{\text{plim}\bar{x}_H - \text{plim}\bar{x}_L} \\ =& 0 \end{aligned}$$

as long as $\text{plim}\bar{x}_H - \text{plim}\bar{x}_L \neq 0$ which requires the x data to have some dispersion asymptotically, an additional assumption to the small sample framework. $\text{plim}\bar{\epsilon}_H = \text{plim}\bar{\epsilon}_L = 0$ since the ϵs are i.i.d. mean zero normals, and hence averages follow a law of large numbers.

(iii) **What is the variance of $\breve{\beta}_2$, conditional on X?**

$$\begin{aligned} \text{E}\left[\left(\breve{\beta}_2 - \beta_2\right)^2 \middle| \mathbf{X}\right] =& \text{E}\left[\left(\frac{\bar{\epsilon}_H - \bar{\epsilon}_L}{\bar{x}_H - \bar{x}_L}\right)^2 \middle| \mathbf{X}\right] \\ =& \left(\frac{1}{\bar{x}_H - \bar{x}_L}\right)^2 \text{E}\left[\bar{\epsilon}_H^2 + \bar{\epsilon}_L^2 - 2\bar{\epsilon}_H\bar{\epsilon}_L \middle| \mathbf{X}\right] \\ =& \left(\frac{1}{\bar{x}_H - \bar{x}_L}\right)^2 \left[\text{E}\left[\bar{\epsilon}_H^2 \middle| \mathbf{X}\right] + \text{E}\left[\bar{\epsilon}_L^2 \middle| \mathbf{X}\right] - \text{E}\left[2\bar{\epsilon}_H\bar{\epsilon}_L \middle| \mathbf{X}\right]\right] \\ =& \left(\frac{1}{\bar{x}_H - \bar{x}_L}\right)^2 \left[\frac{2\sigma^2}{n} + \frac{2\sigma^2}{n} + 0\right] \\ =& \frac{4\sigma^2}{n\left(\bar{x}_H - \bar{x}_L\right)^2} \end{aligned}$$

This provides an alternative method to show consistency since as long as $\bar{x}_H - \bar{x}_L$ does not converge to 0, the variance will converge to 0. Combining this result with the unbiasedness of $\breve{\beta}$ shows that the the estimator converge in mean square which implies consistency.