Econometrics 710
Midterm Exam
March 22, 2012
Sample Answers

1.

(a) Nearly everyone missed this question. The references to efficiency and large samples are irrelevant. The residual variances are unrelated to the efficiency of the estimators of β. The solutions is algebraic and elementary. Least squares, by definition, minimizes the sum of squared errors. Therefore $\hat{\sigma}^2$ is smaller than any other residual variance constructed from any other estimator. Thus $\widehat{\sigma}^2 \leq \widetilde{\sigma}^2$ and $\widehat{\sigma}^2 \leq \overline{\sigma}^2$. Constrained least squares minimizes the sum of squared errors among all estimators which satisfy the restriction. Therefore $\widetilde{\sigma}^2$ is smaller than any other residual variance constructed from any other estimator satisfying the restriction, including efficient minimum distance. Thus $\widetilde{\sigma}^2 \leq \overline{\sigma}^2$. Thus the variance estimators algebraically satisfy $\hat{\sigma}^2 \leq \tilde{\sigma}^2 \leq \bar{\sigma}^2$, which is the **reverse** of the assertion.

(b) Since

$$\begin{aligned}\tilde{e}_i - \hat{e}_i &= \left(y_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}\right) - \left(y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right) \\ &= \boldsymbol{x}_i'\left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right)\end{aligned}$$

Then

$$\begin{aligned}\hat{\sigma}^2 T_n &= \sum_{i=1}^n \left(\tilde{e}_i - \hat{e}_i\right)^2 \\ &= \sum_{i=1}^n \left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right) \boldsymbol{x}_i \boldsymbol{x}_i' \left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right) \\ &= \left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right)' X'X \left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right)\end{aligned}$$

Also, recall that

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = \left(X'X\right)^{-1} R \left[R'(X'X)^{-1}R\right]^{-1} (R'\widehat{\boldsymbol{\beta}} - c).$$

Thus

$$\begin{aligned}\hat{\sigma}^2 T_n &= (R'\widehat{\boldsymbol{\beta}} - c)' \left[R'(X'X)^{-1}R\right]^{-1} R' \left(X'X\right)^{-1} X'X \left(X'X\right)^{-1} R \left[R'(X'X)^{-1}R\right]^{-1} (R'\widehat{\boldsymbol{\beta}} - c) \\ &= (R'\widehat{\boldsymbol{\beta}} - c)' \left[R'(X'X)^{-1}R\right]^{-1} R' \left(X'X\right)^{-1} R \left[R'(X'X)^{-1}R\right]^{-1} (R'\widehat{\boldsymbol{\beta}} - c) \\ &= (R'\widehat{\boldsymbol{\beta}} - c)' \left[R'(X'X)^{-1}R\right]^{-1} (R'\widehat{\boldsymbol{\beta}} - c) \\ &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' R \left[R'(X'X)^{-1}R\right]^{-1} R'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' R \left[R'(\frac{1}{n}X'X)^{-1}R\right]^{-1} R' \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\end{aligned}$$

Applying standard asymptotic theory,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d Z \sim N(0, V)$$

so

$$T_n \to_d \frac{Z'R\left[R'Q^{-1}R\right]^{-1} R'Z}{\sigma^2}$$

(c) Under homoskedasticity, $V = Q^{-1}\sigma^2$, so $R'Z \sim N(0, R'Q^{-1}R\sigma^2)$, and the distribution in the previous question is χ_q^2, where q is the dimension of R.

2.

(a) We can rewrite the restriction as $\beta_1 = 2\beta_2$. Substituting this into the equation we find

$$y_i = \left(2x_{1i} + x_{2i}\right)\beta_2 + e_i$$

The CLS estimate of β_2 is the simple regression

$$\widetilde{\beta}_2 = \frac{\sum_{i=1}^n (2x_{1i} + x_{2i}) y_i}{\sum_{i=1}^n (2x_{1i} + x_{2i})^2}$$

and that for β_1 is

$$\widetilde{\beta}_1 = 2\widetilde{\beta}_2 = \frac{\sum_{i=1}^n (x_{1i} + x_{2i}/2) y_i}{\sum_{i=1}^n (x_{1i} + x_{2i}/2)^2}$$

(b) By the WLLN and CLT

$$\sqrt{n}\left(\widetilde{\beta}_1 - \beta_1\right) = 2\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n (2x_{1i} + x_2) e_i}{\frac{1}{n}\sum_{i=1}^n (2x_{1i} + x_2)^2} \rightarrow_d N\left(0, \frac{\mathbb{E}\left((2x_{1i} + x_2)^2 e_i^2\right)}{\left(\mathbb{E}(2x_{1i} + x_2)^2\right)^2}\right)$$

3.

(a) Since $\mathbb{E}(y_i \mid x_i) = (\gamma + \theta x_i)^{1/2}$ and $(\gamma + \theta x_i)^{1/2}$ is a function of x_i,

$$\mathbb{E}(u_i \mid x_i) = \mathbb{E}\left(y_i - (\gamma + \theta x_i)^{1/2} \mid x_i\right) = \mathbb{E}(y_i \mid x_i) - \mathbb{E}\left((\gamma + \theta x_i)^{1/2} \mid x_i\right) = (\gamma + \theta x_i)^{1/2} - (\gamma + \theta x_i)^{1/2} = 0$$

(b) Using $y_i = (\gamma + \theta x_i)^{1/2} + u_i$,

$$y_i^2 = \gamma + \theta x_i + (\gamma + \theta x_i)^{1/2} u_i + u_i^2$$

so

$$\begin{aligned}\mathbb{E}\left(y_i^2 \mid x_i\right) &= \gamma + \theta x_i + (\gamma + \theta x_i)^{1/2}\mathbb{E}(u_i \mid x_i) + \mathbb{E}\left(u_i^2 \mid x_i\right) \\ &= \gamma + \theta x_i + \mathbb{E}\left(u_i^2 \mid x_i\right)\end{aligned}$$

Equation (3) will be a regression if $\mathbb{E}\left(y_i^2 \mid x_i\right) = \gamma + \theta x_i$, which would only hold if $e_i = \mathbb{E}\left(u_i^2 \mid x_i\right)$. Since $\mathbb{E}\left(u_i^2 \mid x_i\right) \geq 0$ this is only possible if $e_i = \mathbb{E}\left(u_i^2 \mid x_i\right) = 0$, which means that the equation $y_i = (\gamma + \theta x_i)^{1/2}$ holds without error, which is unjustified given the stated situation. Can we make progress? The trouble is the extra component $\mathbb{E}\left(u_i^2 \mid x_i\right)$. If the error u_i is homoskedastic, $\mathbb{E}\left(u_i^2 \mid x_i\right) = \sigma^2$, then $\mathbb{E}\left(y_i^2 \mid x_i\right) = \gamma + \sigma^2 + \theta x_i$. It is just an intercept shift.

(c) Thus under homoskedasticity the least-squares estimates will be consistent for $(\gamma + \sigma^2, \theta)$ and thus θ can be recovered, but not γ. However, if the error v_i is heteroskedastic then neither the intercept nor slope coefficients can be consistently estimated. Another way of see this is to suppose for argument's sake that the conditional variance is linear in x_i, that is $\mathbb{E}\left(u_i^2 \mid x_i\right) = \sigma_0^2 + \rho x_i$. Then

$$\mathbb{E}\left(y_i^2 \mid x_i\right) = \gamma + \sigma_0^2 + (\theta + \rho) x_i$$

So the conditional variance changes both the intercept and slope. If you estimate the regression of y_i^2 on x_i, you don't know if you are estimating the desired equation or the conditional variance.

(d) Even if the slope coefficient is the only parameter of interest, homoskedasticity is an unreasonbly strong assumption in this context as its violation causes estimation inconsistency. Your friend's suggestion is not reasonable because the required assumptions lurking behing it are not justified by the assumptions of the economic model.