BOSTON COLLEGE
Department of Economics
EC771: Econometrics
Spring 2004
Prof. Baum, Ms. Uysal
SOLUTION KEY FOR PROBLEM SET 7

1. A sample of 100 observations produces the following sample data

$$\bar{y_1} = 1,\ \bar{y_2} = 2,\ \mathbf{y_1'y_1} = 150,\ \mathbf{y_2'y_2} = 550,\ \mathbf{y_1'y_2} = 260$$

The underlying bivariate regression model is $y_1 = \mu + \epsilon_1,\ y_2 = \mu + \epsilon_2$
(a) Compute the ordinary least squares estimate of μ and estimate the sampling variance of this estimator.
(b) Compute the FGLS estimate of μ and the sampling variance of your estimator.

The model can be written as $\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} = \begin{bmatrix} \mathbf{i} \\ \mathbf{i} \end{bmatrix} \mu + \begin{bmatrix} \epsilon_\mathbf{1} \\ \epsilon_\mathbf{2} \end{bmatrix}$. Therefore, the OLS estimator is

$$m = (\mathbf{i'i} + \mathbf{i'i})^{-1}(\mathbf{i'y_1} + \mathbf{i'y_2}) = (n\bar{y_1} + n\bar{y_2})/(n+n) = (\bar{y_1} + \bar{y_2})/2 = 1.5$$

The sampling variance would be $var[m] = (1/2)^2\{var[\bar{y_1}] + var[\bar{y_2}] + 2cov[\bar{y_1}, \bar{y_2}]\}$ We would estimate the parts with

$$\begin{aligned} est.var[\bar{y_1}] &= s_{11}/n = ((150 - 100(1)^2)/99)/100 = .0051 \\ est.var[\bar{y_2}] &= s_{22}/n = ((550 - 100(1)^2)/99)/100 = .0152 \\ est.cov[\bar{y_1}, \bar{y_2}] &= s_{12}/n = ((260 - 100(1)(2))/99)/100 = .0061 \end{aligned}$$

Combining terms, est.var[m] = 0.0079. The GLS estimator would be

$$[(\sigma^{11}+\sigma^{12})\mathbf{i'y_1} + (\sigma^{22}+\sigma^{12})\mathbf{i'y_2}]/[(\sigma^{11}+\sigma^{12})\mathbf{i'i} + (\sigma^{22}+\sigma^{12})\mathbf{i'i}] = w\bar{y_1} + (1-w)\bar{y_2}$$

where $w = (\sigma^{11} + \sigma^{12})/(\sigma^{11} + \sigma^{22} + 2\sigma^{12})$.
Denoting $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$, $\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$.
The weight simplifies a little bit as the determinant appears in both the denominator and the numerator. Thus, $w = (\sigma_{22} - \sigma_{12})/(\sigma_{11} + \sigma_{22} - 2\sigma_{12})$. For our sample data, the two step estimator would be based on the variances computed above and $s_{11} = .5051$, $s_{22} = 1.5152$, $s_{12} = .6061$. Then, $w = 1.1250$. The FGLS estimate is $1.125(1) + (1 - 1.125)(2) = .875$. The sampling variance of this estimator is

$$w^2 var[\bar{y_1}] + (1-w)^2 var[\bar{y_2}] + 2w(1-w)cov[\bar{y_1}, \bar{y_2}] = .0050$$

as compared to .0079 for the OLS estimator.

2. For the model

$$\begin{aligned} y_1 &= \alpha_1 + \beta x + \epsilon_1 \\ y_2 &= \alpha_2 + \epsilon_2 \\ y_3 &= \alpha_3 + \epsilon_3 \end{aligned}$$

assume that $y_{i2}+y_{i3} = 1$ at every observation. Prove that the sample covariance matrix of the least squares residuals from the tree equations will be singular, thereby precluding computation of the FGLS estimator. How could you proceed in this case?

Once again, nothing is lost by assuming that $\bar{x} = 0$. Now, the OLS estimators are

$$a_1 = \bar{y_1},\ a_2 = \bar{y_2},\ a_3 = \bar{y_3}\ b = \mathbf{x'y_1}/\mathbf{x'x}.$$

The vector of residuals is

$$\begin{aligned} e_{i1} &= y_{i1} - \bar{y_1} + \beta x_i \\ e_{i2} &= y_{i2} - \bar{y_2} \\ e_{i3} &= y_{i3} - \bar{y_3} \end{aligned}$$

Now, if $y_{i2}+y_{i3} = 1$ at every observation, then $(1/n)\sum_i(y_{i2}+y_{i3}) = \bar{y_2}+\bar{y_3} = 1$ as well. Therefore, by just adding the two equations, we see that $e_{i2} + e_{i3} = 0$ for every observation. Let $\mathbf{e}_i$ be a 3x1 vector of residuals. Then $\mathbf{e}_i'\mathbf{c} = 0$, where $\mathbf{c}$=[0,1,1]'. The sample covariance matrix of the residuals is $\mathbf{S} = [(1/n)\sum_i \mathbf{e}_i\mathbf{e}_i']$. Then, $\mathbf{Sc} = [(1/n)\sum_i \mathbf{e}_i\mathbf{e}_i']\mathbf{c} = [(1/n)\sum_i \mathbf{e}_i\mathbf{e}_i'\mathbf{c}] = [(1/n)\sum_i \mathbf{e}_i x 0] = \mathbf{0}$, which means by definition, that $\mathbf{S}$ is singular.
We can proceed simply by dropping the third equation. The adding up condition implies that $\alpha_3 = 1-\alpha_2$. So, we can treat the first two equations as a seemingly unrelated regression model and estimate $\alpha_3$using the estimate of α_2.