PROBLEM SET # 2

ECONOMICS 240B -SPRING 2007
ANSWER KEY

Almost all of the answers are the work of Tomas Rau, who was the GSI in 2005. Tomas, I can't thank you enough!

1. Ruud's book questions

18.1 As we noticed in section, there were a small misprint in this question,

a)

$$\hat{\beta} = \frac{\sum_{n=1}^{2} w_n x_n y_n}{\sum_{n=1}^{2} w_n x_n^2} = \text{argmin}_{\beta} \sum_{n=1}^{2} w_n (y_n - \beta x_n)^2$$

For simplicity, I will work with the ratio of the 2 weights, defining $w = w_1/w_2$. Hence,

$$\mathbb{V}(\hat{\beta}|x) = \frac{w^2 x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2}{(w x_1^2 + x_2^2)^2}$$

$$\frac{\partial \mathbb{V}(\hat{\beta}|x)}{\partial w} = 2x_1^2 x_2^2 \frac{w\sigma_1^2 - \sigma_2^2}{(w x_1^2 + x_2^2)^3}$$

This FOC is equal to zero when $w = \sigma_2^2/\sigma_1^2$, therefore, one solution is $w_n = 1/\sigma_n^2$.

Note that we could have work with $\text{argmin}_{\beta} \sum_{n=1}^{2} w_n^2 (y_n - \beta x_n)^2$ and define $w = w_1^2/w_2^2$. It doesn't matter. In this case the data is weighted by $1/\sigma_n$.

b) Putting relatively more weight on the observation with smaller variance does not necessarily decrease the variance of an estimator with respect to LS. Note that, $\mathbb{V}(\beta_{LS}|x) = Var(\beta|x)$ when,

$$\frac{x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2}{(x_1^2 + x_2^2)^2} = \frac{w^2 x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2}{(w x_1^2 + x_2^2)^2}$$

solutions are: $w = 1$ (trivial case) and

$$w = \frac{2x_2^2 \sigma_2^2 + \sigma_2^2 x_1^2 - \sigma_1^2 x_2^2}{2x_1^2 \sigma_1^2 + \sigma_1^2 x_2^2 - \sigma_2^2 x_1^2}$$

c) This question is asking you to show that (away from the optimum, where $\mathbb{V}(\beta_{LS}|x) = \mathbb{V}(\hat{\beta}|x)$) there is a tradeoff between lower variance and higher x values. In any of these regressions, bigger x will provide a better estimation since $(x'x)^{-1}$ is dividing the variance term.

Define, $\lambda \equiv x_1/x_2$

Date: March 03 2007.

$$w = \frac{2\sigma_2^2 + \sigma_2^2\lambda^2 - \sigma_1^2}{2\lambda^2\sigma_1^2 + \sigma_1^2 - \sigma_2^2\lambda^2}$$

$$\frac{\partial w}{\partial \lambda} = \frac{4\lambda(\sigma_1 - \sigma_2)^2(\sigma_1 + \sigma_2)^2}{2\lambda^2\sigma_1^2 + \sigma_1^2 - \sigma_2^2\lambda^2} > 0 \quad \text{if} \quad \lambda > 0$$

The intuition is that, in addition to put relatively more weight on the observation with smaller conditional variance, a correction in the scale is necessary because the tradeoff just mentioned.

18.2 a)

$$\begin{aligned} \mathbb{V}(\hat{\mu}_{OLS}|X) &= \mathbb{V}(P_X y|X) = \mathbb{V}(X'(X'X)^{-1}X'y|X) \\ &= X(X'X)X'\mathbb{V}(y|X)X(X'X)^{-1}X' \\ &= X(X'X)X'\Omega_0 X(X'X)^{-1}X' = P_X\Omega_0 P_X = \mathbb{V}(X\hat{\beta}_{OLS}|X) \end{aligned}$$

b))

$$\begin{aligned} \mathbb{V}(y - \hat{\mu}_{OLS}|X) &= \mathbb{V}((I - P_X)y|X) = (I - P_X)\mathbb{V}(y|X)(I - P_X)' \\ &= (I - P_X)\Omega_0(I - P_X)' \end{aligned}$$

c)

$$\begin{aligned} \mathbb{C}(\hat{\mu}_{OLS}, y - \hat{\mu}_{OLS}|X) &= \mathbb{C}(P_X y, (I - P_X)y|X) \\ &= P_X\mathbb{V}(y|X)(I - P_X) = P_X\Omega_0(I - P_X) \end{aligned}$$

which is $\neq 0$ unless $\Omega_0 = I$, hence $\hat{\mu}_{OLS}$ and $y - \hat{\mu}_{OLS}$ are uncorrelated.

18.5 There are at least 3 ways of presenting the solution. I will shortly present 3.

i) The easiest way of getting this is to *transform* the model into a Gauss-Markov one. Once done this we can apply the known formula of partition regression. Recall that the partition fit (Frisch-Waugh) for OLS is

$$\hat{\beta}_1 = (X_{1\perp 2}' X_{1\perp 2})^{-1} X_{1\perp 2}' y_{\perp 2}$$

So, let C be the cholesky decomposition matrix such that $\Omega_0 = CC'$. We know that pre multiplying the model by C^{-1} we got a transformed model with spherical variance matrix, hence the Frisch-Waugh formula works over the transformed model.
Let $X_{1\perp 2}^* = (I - P_{X_2^*})X_1$ and $y_{\perp 2}^* = (I - P_{X_2^*})y^*$
where $X_i^* = C^{-1}X_i$, $i = 1, 2$ and $y^* = C^{-1}y$. Hence, we have that the generalized partition regression formula for GLS is given by

$$\hat{\beta}_1^* = (X_{1\perp 2}^{*'} X_{1\perp 2}^*)^{-1} X_{1\perp 2}^{*'} y_{\perp 2}^*$$

ii) This method is a bit messier. We use cholesky decomposition to rewrite the GLS estimator in partition fashion. Let $\Omega_0 = CC'$

$$\hat{\beta}_{GLS} = (X'\Omega_0^{-1}X)^{-1}X'\Omega_0^{-1} = ((C^{-1}X)'(C^{-1}X))^{-1}(C^{-1}X)'(C^{-1}y)$$

$$\begin{pmatrix} \hat{\beta}_{1,GLS} \\ \hat{\beta}_{2,GLS} \end{pmatrix} \begin{pmatrix} (C^{-1}X_1)'(C^{-1}X_1) & (C^{-1}X_1)'(C^{-1}X_2) \\ (C^{-1}X_2)'(C^{-1}X_1) & (C^{-1}X_2)'(C^{-1}X_2) \end{pmatrix}^{-1} \begin{pmatrix} (C^{-1}X_1)'C^{-1}y \\ (C^{-1}X_2)'C^{-1}y \end{pmatrix}$$

Now, using the partition matrix inverse formula we can solve and get the same expression as method i).

iii) This solution is nice but requires more calculation.

$$\hat{\beta}_{GLS} = \text{argmin}_{b_1,b_2}(y - x_1b_1 - x_2b_2)'\Omega_0^{-1}(y - x_1b_1 - x_2b_2)$$

The Foc's for b_1 and b_2 will yield the same solution.

18.7 We know that,

$$S^2 = \frac{(y - X\hat{\beta}_{OLS})'(y - X\hat{\beta}_{OLS})}{N-K} = \frac{y'(I-P_X)'(I-P_X)y}{N-K}$$

$$\begin{aligned}
\mathbb{E}(S^2|X) &= tr(\mathbb{E}(S^2|X)) = \mathbb{E}(tr(S^2)|X) \\
&= \mathbb{E}[tr(y'(I-P_X)'(I-P_X)y|X]/(N-K) \\
&= \mathbb{E}[tr((I-P_X)yy'(I-P_X)')|X]/(N-K) \\
&= tr((I-P_X)\mathbb{E}[yy'|X](I-P_X)')/(N-K) \\
&= tr((I-P_X)(\Omega_0 + X\beta\beta'X')(I-P_X)')/(N-K) \\
&= tr((I-P_X)\Omega_0(I-P_X)')/(N-K) \\
&= tr((I-P_X)'(I-P_X)\Omega_0)/(N-K) \\
&= tr(\tfrac{(I-P_X)\Omega_0}{(N-K)})
\end{aligned}$$

where in the first and second equality we used the fact that $\mathbb{E}(S^2|X)$ is scalar and that $\mathbb{E}(\cdot)$ is a linear operator. In the fourth equality we used $tr(AB) = tr(BA)$.

19.3 As we saw in section,

$$\mathbb{C}(\epsilon_t, \epsilon_{t-s}) = \sigma_v^2\phi^s/(1-\phi^2)$$

which implies that,

$$\rho(s) = \phi^s$$

hence,

TABLE 1. up to 7 lags

$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(4)$	$\rho(5)$	...
-.498	.248	-.1235	.061	-.030	....

In pg. 457, the values obtained were $\hat{\rho}(1) = -.498$, $\hat{\rho}(2) = .093$, $\hat{\rho}(3) = -.093$. The difference is because these are sample correlations (no structure imposed) and the correlations we computed are from an AR(1) model which doesn't fit the data accurately.

19.5 Note that $LS \equiv GLS$ implies that,

$$(X'X)^{-1}X' = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \qquad (*)$$

Note that this is possible with $\Omega \neq \sigma^2 I$. See lemma 19.1 in Ruud's book. Anyway, note that the "true" variance of OLS is

$$\mathbb{V}(\hat{\beta}_{LS}|X) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Using $(*)$ we have that,

$$\begin{aligned}\mathbb{V}(\hat{\beta}_{LS}|X) \quad & (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\\ &= (X'\Omega^{-1}X)^{-1} = V(\hat{\beta}_{GLS})\end{aligned}$$

Now the question is asking whether *sampling variance* matrix from OLS software; $s^2(X'X)^{-1}$ can be used as an estimator of $(X'\Omega^{-1}X)^{-1}$. We showed that the "true" variance of OLS may be equal to the variance of GLS. So, the answer is no.

26.7 This question was a bit tricky. We are asked to calculate the expectation of the elements of $\mathbb{E}_T[\epsilon_t(\hat{\beta}_{OLS})\epsilon_t(\hat{\beta}_{OLS})']$. Note that the stacking in here is a bit different (see pg. 704). We proceed element by element,

$$\epsilon_t(\hat{\beta}_{OLS}) = \begin{pmatrix} y_{1t} - x'_{1t}\hat{\beta}_1 \\ . \\ . \\ y_{Jt} - x'_{Jt}\hat{\beta}_J \end{pmatrix}$$

$$\mathbb{E}_T[\epsilon_t(\hat{\beta}_{OLS})\epsilon_t(\hat{\beta}_{OLS})'] =$$

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} (y_{1t} - x'_{1t}\hat{\beta}_1)^2 & ... & (y_{1t} - x'_{1t}\hat{\beta}_1)(y_{Jt} - x'_{Jt}\hat{\beta}_J) \\ . & ... & . \\ . & ... & . \\ (y_{Jt} - x'_{Jt}\hat{\beta}_J)(y_{1t} - x'_{1t}\hat{\beta}_1) & ... & (y_{Jt} - x'_{Jt}\hat{\beta}_J)^2 \end{pmatrix}$$

We can use the usual stacking notation. So the representative element of this matrix is like, $\hat{\epsilon}'_j\hat{\epsilon}_i$ where $\epsilon_j = y_j - X_j\hat{\beta}_j = (I_T - P_{X_j})y_j$ is a Tx1 vector. Hence we want to get $\mathbb{E}[\hat{\epsilon}'_j\hat{\epsilon}_i]$

$$\begin{aligned}\mathbb{E}[\hat{\epsilon}'_j\hat{\epsilon}_i]] \quad & \mathbb{E}[y'_j(I_T - P_{X_j})(I_T - P_{X_i})y_i]\\ &= \mathbb{E}[\epsilon'_j(I_T - P_{X_j})(I_T - P_{X_i})\epsilon_i]\\ &= tr(\mathbb{E}[\epsilon'_j(I_T - P_{X_j})(I_T - P_{X_i})\epsilon_i])\\ &= \mathbb{E}[tr((I_T - P_{X_j})(I_T - P_{X_i})\epsilon'_j\epsilon_i]\\ &= tr((I_T - P_{X_j})(I_T - P_{X_i})\mathbb{E}[\epsilon'_j\epsilon_i])\\ &= w_{0ji}tr(I - P_{X_j} - P_{X_i} + P_{X_j}P_{X_i})\\ &= w_{0ji}[T - K_j - K_i + tr(P_{X_j}P_{X_i})]\end{aligned}$$

So an unbiased estimator of w_{0ji} is

$$s_{ji} = \frac{\hat{\epsilon}'_j\hat{\epsilon}_i}{T - K_j - K_i + tr(P_{X_j}P_{X_i})}$$

2. Empirical Question

1) We want to test the proportionality hypothesis between consumption and income[1]. Note that the variables are in log. Consider the following: $y = \alpha x^{\beta}$ in order to have proportionality between y and x we need that β be equal to one. What happens if we take logs? $log(y) = log(\alpha) + \beta log(x)$, which is just the model we want to test. Some people omit the constant which impose the restriction $\alpha = 1$. Here is the output of my code. The code is at the end of the document.

```
part a)
              Bols     std. errors
    cons      0.5714   0.1379
    log(inc)  0.8519   0.0203

    Proportionality test:

    t-statistic  -7.3019
    critical value  1.9702

The null hyp of proportionality is rejected

part b)
    Proportionality test (with Eicker-White cov matrix):

    t-statistic    -5.8086
    critical value, t(233)  1.9702

The null hyp of proportionality is rejected

part c)
    Auxiliar regression of residuals

                B       std errors
    cons        0.8441    0.2714
    log(inc)   -0.2543    0.0789
    log(inc)2   0.0195    0.0057

T = N*R2 = 23.832 Critical value, Chi2(2): 5.9915

part d)
    Weighted Least Squares

                B       std errors
    cons        0.4627    0.1525
    log(inc)    0.8684    0.0227

t-statistic= -5.793209 critical value, t(233): 1.9702
```

---

[1] I'm considering a two-sided alternative

```
The null hyp of proportionality is rejected

part e)
    statistic -6.0485
critical value, t(233): 1.9702

The null hyp proportionality is rejected
```

E-mail address: `carolina@econ.berkeley.edu`