March, 2007

Answers to Problem Set 2

Maria Carolina Caetano, University of California, Berkeley
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,

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

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

\[ \mathbb{V}(\hat{\beta} | x) = \frac{w^2 x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2}{(wx_1^2 + x_2^2)^2} \]

\[ \frac{\partial \mathbb{V}(\hat{\beta} | x)}{\partial w} = 2w^2 x_2^2 \frac{w \sigma_1^2 - \sigma_2^2}{(wx_1^2 + x_2^2)^3} \]

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, \( \mathbb{V}(\beta_{LS} | x) = \text{Var}(\beta | x) \) when,

\[ \frac{x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2}{(x_1^2 + x_2^2)^2} = \frac{w^2 x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2}{(wx_1^2 + x_2^2)^2} \]

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

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

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, \( \lambda \equiv x_1/x_2 \)
There are at least 3 ways of presenting the solution. I will shortly present
the easy 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 formula for GLS is given by

\[
\begin{align*}
\hat{\beta}_{GLS} & = (X'X)'(X'X)^{-1}\hat{y} \\
& = X'(X'X)^{-1}y \quad \text{if } \Omega_0 = I, \text{ hence } \hat{\beta}_{GLS} \text{ and } y - \hat{\mu}_{OLS} \text{ are uncorrelated.}
\end{align*}
\]

18.2 a) 
\[
\mathbb{V}(\hat{\mu}_{OLS}|X) = \mathbb{V}(P_Xy|X) = \mathbb{V}(X'(X'X)^{-1}X'y|X) \\
= X(X'X)^{-1}P_X\mathbb{V}(y|X)(X'X)^{-1}X' \\
= X(X'X)^{-1}P_X\mathbb{V}(y|X)(X'X)^{-1}X' = P_X\Omega_0 P_X = \mathbb{V}(X\hat{\mu}_{OLS}|X)
\]

b) 
\[
\mathbb{V}(y - \hat{\mu}_{OLS}|X) = \mathbb{V}((I - P_X)y|X) = (I - P_X)\mathbb{V}(y|X)(I - P_X)' \\
= (I - P_X)I_0(I - P_X)'
\]

c) 
\[
\mathbb{C}(\hat{\mu}_{OLS}, y - \hat{\mu}_{OLS}|X) = \mathbb{C}(P_Xy, (I - P_X)y|X) \\
= P_X\mathbb{V}(y|X)(I - P_X) = P_XI_0(I - P_X) \\
\]

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

\[
\hat{\beta}_1 = (X'_{1*}X_{1*})^{-1}X'_{1*}y_{1*}
\]

So, let \(C\) be the cholesky decomposition matrix such that \(\Omega_0^{1/2}CC'\). 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 \(X_{i*} = (I - P_{X_i})X_i\) and \(y_{i*} = (I - P_{X_i})y^*\), 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

\[
\hat{\beta}_i = (X'_{i*}X_{i*})^{-1}X'_{i*}y_{i*}
\]

ii) This method is a bit messier. We use cholesky decomposition to rewrite the GLS estimator in partition fashion. Let \(\Omega_0^{1/2}CC'\)

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

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

\[ \hat{\beta}_{GLS} = \arg\min_{b_1,b_2} (y - x_1b_1 - x_2b_2)'\Omega_0^{-1}(y - x_1b_1 - x_2b_2) \]

The Foc’s for \( b_1 \) and \( b_2 \) will yield the same solution.

18.7 We know that,

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

\[ \mathbb{E}(S^2|X) = tr(\mathbb{E}(S^2|X)) = \mathbb{E}(tr(S^2|X)) = \mathbb{E}(tr(y'(I - P_X)'y|X)/(N - K)) = tr((I - P_X)\mathbb{E}(yy'|X)(I - P_X)')/(N - K) \]

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,

\[ \mathbb{C}(\epsilon_t, \epsilon_{t-s}) = \sigma^2\phi^s/(1 - \phi^2) \]

which implies that,

\[ \rho(s) = \phi^s \]

hence,

<table>
<thead>
<tr>
<th>Table 1. up to 7 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho(1) )</td>
</tr>
<tr>
<td>-0.498</td>
</tr>
</tbody>
</table>

In pg. 457, the values obtained were \( \hat{\rho}(1) = -0.498, \hat{\rho}(2) = 0.093, \hat{\rho}(3) = -0.093 \). 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,

\[ (X'X)^{-1}X' = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \quad (*) \]

Note that this is possible with \( \Omega \neq \sigma^2I \). See lemma 19.1 in Ruud’s book.

Anyway, note that the “true” variance of OLS is
This question was a bit tricky. We are asked to calculate the expectation of this matrix is like, ˆ}

\[ \begin{pmatrix} y_{1t} - x_{1t}' \hat{\beta}_1 \\ \vdots \\ y_{Jt} - x_{Jt}' \hat{\beta}_J \end{pmatrix} \]

\[ \mathbb{E}_T[E_t(\hat{\beta}_{OLS})^t E_t(\hat{\beta}_{OLS})] = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} (y_{1t} - x_{1t}' \hat{\beta}_1)^2 & \ldots & (y_{1t} - x_{1t}' \hat{\beta}_1)(y_{Jt} - x_{Jt}' \hat{\beta}_J) \\ \vdots & \ddots & \vdots \\ (y_{Jt} - x_{Jt}' \hat{\beta}_J)(y_{1t} - x_{1t}' \hat{\beta}_1) & \ldots & (y_{Jt} - x_{Jt}' \hat{\beta}_J)^2 \end{pmatrix} \]

We can use the usual stacking notation. So the representative element of this matrix is like, \( \hat{\epsilon}_i^t \hat{\epsilon}_i \) 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 \( \mathbb{E}[\hat{\epsilon}_j^t \hat{\epsilon}_j] \)

\[ \begin{align*}
\mathbb{E}[\hat{\epsilon}_j^t \hat{\epsilon}_j] &= \mathbb{E}[y_j^t (I_T - P_{X_j})(I_T - P_{X_j}) y_j] \\
&= \mathbb{E}[\hat{\epsilon}_j^t (I_T - P_{X_j})(I_T - P_{X_j}) \hat{\epsilon}_j] \\
&= tr(\mathbb{E}[\hat{\epsilon}_j^t (I_T - P_{X_j})(I_T - P_{X_j}) \hat{\epsilon}_j]) \\
&= \mathbb{E}[tr((I_T - P_{X_j})(I_T - P_{X_j}) \hat{\epsilon}_j \hat{\epsilon}_j)] \\
&= tr((I_T - P_{X_j})(I_T - P_{X_j}) \mathbb{E}[\hat{\epsilon}_j \hat{\epsilon}_j]) \\
&= w_{0ji} tr(I - P_{X_j} - P_{X_i} + P_{X_j} P_{X_i}) \\
&= w_{0ji} [T - K_j - K_i + tr(P_{X_j} P_{X_i})]
\end{align*} \]

So an unbiased estimator of \( w_{0ji} \) is

\[ s_{ji} = \frac{\hat{\epsilon}_j^t \hat{\epsilon}_i}{T - K_j - K_i + tr(P_{X_j} P_{X_i})} \]
2. Empirical Question

1) We want to test the proportionality hypothesis between consumption and income. 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 \( \beta \) 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)

<table>
<thead>
<tr>
<th>B</th>
<th>std. errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>cons</td>
<td>0.5714</td>
</tr>
<tr>
<td>log(inc)</td>
<td>0.8519</td>
</tr>
</tbody>
</table>

Proportionality test:

\[ t\text{-statistic} = -7.3019 \]

<table>
<thead>
<tr>
<th>critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9702</td>
</tr>
</tbody>
</table>

The null hyp of proportionality is rejected

part b)

Proportionality test (with Eicker-White cov matrix):

\[ t\text{-statistic} = -5.8086 \]

<table>
<thead>
<tr>
<th>critical value, t(233)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9702</td>
</tr>
</tbody>
</table>

The null hyp of proportionality is rejected

part c)

Auxiliar regression of residuals

<table>
<thead>
<tr>
<th>B</th>
<th>std errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>cons</td>
<td>0.8441</td>
</tr>
<tr>
<td>log(inc)</td>
<td>-0.2543</td>
</tr>
<tr>
<td>log(inc)^2</td>
<td>0.0195</td>
</tr>
</tbody>
</table>

\[ T = N \times R^2 = 23.832 \]

Critical value, Chi2(2): 5.9915

part d)

Weighted Least Squares

<table>
<thead>
<tr>
<th>B</th>
<th>std errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>cons</td>
<td>0.4627</td>
</tr>
<tr>
<td>log(inc)</td>
<td>0.8684</td>
</tr>
</tbody>
</table>

\[ t\text{-statistic} = -5.793209 \]

<table>
<thead>
<tr>
<th>critical value, t(233)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9702</td>
</tr>
</tbody>
</table>

\[ ^1 \text{i'm considering a two-sided alternative} \]
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