Solutions to the Midterm Exam

ECON 837

Prof. Simon Woodock, Spring 2008

1. We have $\mathbf{Z} = [\mathbf{y}\ \mathbf{X}]$, so that

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} \mathbf{y}' \\ \mathbf{X}' \end{bmatrix} \begin{bmatrix} \mathbf{y} & \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{y}'\mathbf{y} & \mathbf{y}'\mathbf{X} \\ \mathbf{X}'\mathbf{y} & \mathbf{X}'\mathbf{X} \end{bmatrix}$$

and hence

$$\mathbf{y}'\mathbf{y} = 150 \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 15 \\ 50 \end{bmatrix} \quad \mathbf{X}'\mathbf{X} = \begin{bmatrix} 25 & 0 \\ 0 & 100 \end{bmatrix}.$$

Since the model includes an intercept, we know the $(1,1)$ element of $\mathbf{X}'\mathbf{X}$ is n, and the first element of $\mathbf{X}'\mathbf{y}$ is $\sum y_i$. Therefore

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1/25 & 0 \\ 0 & 1/100 \end{bmatrix} \begin{bmatrix} 15 \\ 50 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 1/2 \end{bmatrix} \\
s^2 &= \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{1}{25-2}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \frac{1}{23}\left(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\right) \\
&= \frac{1}{23}\left(150 - 2\begin{bmatrix} 15 & 50 \end{bmatrix}\begin{bmatrix} 3/5 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 3/5 & 1/2 \end{bmatrix}\begin{bmatrix} 25 & 0 \\ 0 & 100 \end{bmatrix}\begin{bmatrix} 3/5 \\ 1/2 \end{bmatrix}\right) \\
&= \frac{1}{23}\left(150 - 2(9+25) + (9+25)\right) = 116/23 \\
R^2 &= 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{J}\mathbf{y}} = 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{y} - \frac{1}{n}\mathbf{y}'\mathbf{i}\mathbf{i}'\mathbf{y}} = 1 - \frac{116}{150 - \frac{1}{25}\left(\sum y_i\right)^2} = 1 - \frac{116}{150 - 15^2/25} \\
&= 1 - \frac{116}{141} = 25/141 \approx 0.177
\end{aligned}$$

So the cross-products matrix is sufficient for computing all of these, plus a few other things (test of $H_0 : \beta_2 = 0$, for example). However, there are other things we can/should do that we cannot using only the cross-products, such as plot residuals for evidence of heteroskedasticity, etc.

2. (a) We know

$$E[\hat{\mu}] = \mu \Leftrightarrow E\left[\sum_{i=1}^{n} c_i X_i\right] = \mu \Leftrightarrow \sum_{i=1}^{n} c_i E[X_i] = \mu \Leftrightarrow \mu \sum_{i=1}^{n} c_i = \mu \Leftrightarrow \sum_{i=1}^{n} c_i = 1.$$

(b) The sampling variance of any $\hat{\mu}$ in this class of estimators is

$$Var[\hat{\mu}] = Var\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i^2 Var[X_i] + 2\sum_{i=1}^{n}\sum_{j \neq i} c_i c_j Cov[X_i, X_j] = \sum_{i=1}^{n} c_i^2 Var[X_i] = \sigma^2 \sum_{i=1}^{n} i c_i^2.$$

Hence the best (minimum variance) unbiased estimator solves

$$\min_{c_i} \sigma^2 \sum_{i=1}^{n} i c_i^2 \ \text{ s.t. } \sum_{i=1}^{n} c_i = 1$$

which implies the Lagrangean

$$\mathcal{L} = \sigma^2 \sum_{i=1}^{n} i c_i^2 + \lambda\left[1 - \sum_{i=1}^{n} c_i\right]$$

with FOCs:

$$\begin{aligned}
2\sigma^2 i c_i - \lambda &= 0 \text{ for } i = 1, 2, ..., n \\
1 - \sum_{i=1}^{n} c_i &= 0.
\end{aligned}$$

We need to eliminate λ. From the first n FOCs we see that for each i,

$$\lambda = 2\sigma^2 i c_i \Rightarrow \lambda i^{-1} = 2\sigma^2 c_i \Rightarrow \lambda \sum i^{-1} = 2\sigma^2 \sum c_i = 2\sigma^2 \Rightarrow \lambda = \frac{2\sigma^2}{\sum i^{-1}}$$

upon substituting in the constraint$\sum c_i = 1$. Therefore, for each i we have

$$2\sigma^2 i c_i = \frac{2\sigma^2}{\sum i^{-1}} \Rightarrow c_i^* = \frac{i^{-1}}{\sum i^{-1}}$$

so that $\hat{\mu}^* = \sum_{i=1}^n c_i^* X_i$. This is fairly intuitive – it attaches a weight to each observation that is inversely proportional to its variance. That way, each observation contributes about the same information to $\hat{\mu}$.

Don't forget to check the second order conditions as well to be sure we're at a minimum.

(c) From the variance expression we derived in part b,

$$Var[\hat{\mu}^*] = \sigma^2 \sum_{i=1}^n i (c_i^*)^2 = \sigma^2 \sum_{i=1}^n i \left(\frac{i^{-1}}{\sum_{j=1}^n j^{-1}} \right)^2 = \frac{\sigma^2}{\sum_{i=1}^n i^{-1}}.$$

Now note that $\bar{X}$ is also in the class of estimators we've denoted $\hat{\mu}$, with $c_i = 1/n$ for each i. Since $\hat{\mu}^*$ is the best estimator in this class, $\bar{X}$ can not have a smaller variance. But their variance could be the same. In fact, from our general expression for estimators in this class:

$$Var\left[\bar{X}\right] = \frac{\sigma^2}{n^2} \sum_{i=1}^n i = \frac{n(n+1)\sigma^2}{2n^2} = \frac{\sigma^2}{2} \frac{n+1}{n}.$$

Claim: $Var[\hat{\mu}^*] < Var\left[\bar{X}\right]$ for all $n \geq 2$..

Proof: For $n \geq 4$, this is easy. Notice that $Var\left[\bar{X}\right] > \sigma^2/2$ for all n. Now,

$$Var[\hat{\mu}^*] = \frac{\sigma^2}{1 + 2^{-1} + 3^{-1} + \cdots n^{-1}} < \frac{\sigma^2}{2}$$

if and only if $1 + 2^{-1} + 3^{-1} + \cdots + n^{-1} > 2$, which is true when $n \geq 4$. We need to check that $Var[\hat{\mu}^*] < Var\left[\bar{X}\right]$ for $n = 2$ and $n = 3$ also, but that is easy too:

$$\begin{aligned} Var[\hat{\mu}^*_{n=2}] &= \frac{2}{3}\sigma^2 < \frac{3}{4}\sigma^2 = Var\left[\bar{X}_{n=2}\right] \\ Var[\hat{\mu}^*_{n=3}] &= \frac{6}{11}\sigma^2 < \frac{2}{3}\sigma^2 = Var\left[\bar{X}_{n=3}\right] \end{aligned}$$

and hence $Var[\hat{\mu}^*] < Var\left[\bar{X}\right]$ for all $n \geq 2$.

(d) Notice that $\hat{\mu}^* \overset{m.s.}{\to} \mu$, because $E[\hat{\mu}^*] = \mu$ and

$$\lim_{n\to\infty} Var[\hat{\mu}^*] = \lim_{n\to\infty} \frac{\sigma^2}{\sum_{i=1}^n i^{-1}} = 0$$

(you were told $\lim_{n\to\infty} \sum_{i=1}^n i^{-1} = \infty$). It follows, therefore, that $\hat{\mu}^* \overset{p}{\to} \mu$, so that $\hat{\mu}^*$ is a consistent estimator of μ.

On the other hand, we also have $E\left[\bar{X}\right] = \mu$, but

$$\lim_{n\to\infty} Var\left[\bar{X}\right] = \lim_{n\to\infty} \frac{\sigma^2}{2} \frac{n+1}{n} = \frac{\sigma^2}{2} > 0$$

so that $\bar{X}$ does not converge to μ in mean square. Non-convergence in mean square does not necessarily imply non-convergence in probability. For the purpose of this exam, however, it was sufficient to establish non-convergence in mean square.

This result is pretty intuitive: $\bar{X}$ doesn't converge to μ as $n \to \infty$ because the X_i have larger and larger variance as $n \to \infty$. Hence values of X_i "far" from μ become more and more likely as $n \to \infty$, which keeps $\bar{X}$ from getting close to μ in probability.

3. (a) First, note that $\hat{y}_* = \mathbf{x}_*'\hat{\boldsymbol{\beta}} = \mathbf{x}_*' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$. We're asked whether $\hat{y}_*$ is an unbiased estimator of y_*. Since y_* is sampled from the same distribution as the original data, we know that $y_* = \mathbf{x}_*'\boldsymbol{\beta}+\varepsilon_*$, where ε_* is the (unobserved) error for the new observation. Consider:

$$\begin{aligned} E[\hat{y}_*] &= E\left[\mathbf{x}_*' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}\right] = \mathbf{x}_*' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E[\mathbf{y}] = \mathbf{x}_*' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E[\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}] \\ &= \mathbf{x}_*'\boldsymbol{\beta} + \mathbf{x}_*' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E[\boldsymbol{\varepsilon}] = \mathbf{x}_*'\boldsymbol{\beta} = E[y_*] \end{aligned}$$

This is not "unbiasedness" in the sense we are accustomed to. We have shown that $E[\hat{y}_*] = E[y_*]$, but we haven't shown $E[\hat{y}_*] = y_*$. However, $E[\hat{y}_*] = E[y_*]$ does imply that prediction errors have zero mean, $E[y_* - \hat{y}_*] = 0$, so the OLS predictor is usually called an "unbiased" predictor anyway.

(b) Note that $y_* - \hat{y}_* = \mathbf{x}_*'\boldsymbol{\beta} + \varepsilon_* - \mathbf{x}_*'\hat{\boldsymbol{\beta}} = \varepsilon_* + \mathbf{x}_*'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)$. Hence

$$\begin{aligned} Var[y_* - \hat{y}_*] &= E\left[\varepsilon_*^2\right] + 2E\left[\varepsilon_*\mathbf{x}_*'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right] + E\left[\mathbf{x}_*'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{x}_*\right] \\ &= \sigma^2 + 0 + E\left[\mathbf{x}_*'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}_*\right] \\ &= \sigma^2 + \mathbf{x}_*'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}_* \\ &= \sigma^2\left(1 + \mathbf{x}_*'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}_*\right) \end{aligned}$$

(c) The BLUE of y_* is $\mathbf{x}_*'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the least squares estimator. Technically, "best" means "minimum variance of prediction error" in this context.
How do we know that $\mathbf{x}_*'\hat{\boldsymbol{\beta}}$ is the BLUE of y_*? Straight from the GMT: we proved in lecture that the BLUE of $\mathbf{c}'\boldsymbol{\beta}$ is $\mathbf{c}'\hat{\boldsymbol{\beta}}$ for any $\mathbf{c}$. Hence the BLUE of $\mathbf{x}_*'\boldsymbol{\beta}$ is $\mathbf{x}_*'\hat{\boldsymbol{\beta}}$. This was a sufficient answer for full marks.
Now, it may not be obvious that the BLUE of y_* and the BLUE of $\mathbf{x}_*'\boldsymbol{\beta}$ are the same. Intuitively, it is is true because their difference, $y_* - \mathbf{x}_*'\boldsymbol{\beta} = \varepsilon_*$ is unpredictable (by definition). But suppose you wanted to show formally that $\mathbf{x}_*'\hat{\boldsymbol{\beta}}$ is the best (minimum variance of prediction error) linear unbiased estimator of y_*. Here's how you would proceed.
First, define the class of linear estimators $p = \mathbf{c}'\mathbf{y} = \mathbf{c}'\mathbf{X}\boldsymbol{\beta} + \mathbf{c}'\boldsymbol{\varepsilon}$ for arbitrary $\mathbf{c}$. The prediction errors are

$$p - y_* = \mathbf{c}'\mathbf{y} - y_* = \left(\mathbf{c}'\mathbf{X} - \mathbf{x}_*'\right)\boldsymbol{\beta} + \mathbf{c}'\boldsymbol{\varepsilon} - \varepsilon_*.$$

Unbiasedness is $E[p - y_*] = 0$, which requires $\mathbf{c}'\mathbf{X} = \mathbf{x}_*'$. In the unbiased case, the prediction error is $p - y_* = \mathbf{c}'\boldsymbol{\varepsilon} - \varepsilon_*$, with variance:

$$Var[p - y_*] = E\left[(\mathbf{c}'\boldsymbol{\varepsilon} - \varepsilon_*)^2\right] = E[\mathbf{c}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{c} + \varepsilon_* - 2\varepsilon_*\mathbf{c}'\boldsymbol{\varepsilon}] = \mathbf{c}'\sigma^2\mathbf{I}_n\mathbf{c} + \sigma^2 = \sigma^2(\mathbf{c}'\mathbf{c} + 1).$$

Note we used the spherical errors property, which implies the cross-product is zero in expectation. The best estimator in this class minimizes the variance of prediction error, subject to unbiasedness, by choice of $\mathbf{c}$. This gives the Lagrangean:

$$\mathcal{L} = \sigma^2(\mathbf{c}'\mathbf{c} + 1) - 2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{c} - \mathbf{x}_*)$$

with FOCs:

$$\begin{aligned} 2\sigma^2\mathbf{c} - 2\mathbf{X}\boldsymbol{\lambda} &= \mathbf{0} \\ 2(\mathbf{X}'\mathbf{c} - \mathbf{x}_*) &= \mathbf{0}. \end{aligned}$$

Solving the first set of equations for $\boldsymbol{\lambda}$, we get $\boldsymbol{\lambda} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{c} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*$ upon substituting in the constraint. We substitute this back into the first set of FOCs to eliminate $\boldsymbol{\lambda}$. Then we solve for $\mathbf{c} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*$, so that the BLUE is $\mathbf{c}'\mathbf{y} = \mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{x}_*'\hat{\boldsymbol{\beta}}$.

(d) When $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right)$, we know that $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$. Hence

$$y_* - \hat{y}_* = \varepsilon_* + \mathbf{x}_*' \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) \sim N\left(0, \sigma^2 \left(1 + \mathbf{x}_*' \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \mathbf{x}_*\right)\right)$$

because linear combinations of normals are normal, and we derived the mean and variance previously. Consequently,

$$\frac{y_* - \hat{y}_*}{\sigma \sqrt{1 + \mathbf{x}_*' \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \mathbf{x}_*}} \sim N(0,1).$$

Of course σ is unknown, so this is not a useful test statistic. However, we know that $(n-k)\, s^2/\sigma^2 \sim \chi^2_{n-k}$ under normality, where $s^2 = \mathbf{e}'\mathbf{e}/(n-k)$. Furthermore we know s^2 is independent of $\hat{\boldsymbol{\beta}}$, and hence independent of $\hat{y}_*$, so that

$$\frac{\left(y_* - \hat{y}_*\right)/\sigma\sqrt{1 + \mathbf{x}_*' \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \mathbf{x}_*}}{\sqrt{\left((n-k)\, s^2/\sigma^2\right)/(n-k)}} = \frac{y_* - \hat{y}_*}{s\sqrt{1 + \mathbf{x}_*' \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \mathbf{x}_*}} \sim t_{n-k}$$

and we can use this to test hypotheses about y_*.

4. (a) The joint density of the data is $f(y_1, ..., y_n|\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda y_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n y_i\right)$, so that $\left(n, \sum_{i=1}^n y_i\right)$ is sufficient for λ by the factorization theorem.

(b) The MLE of λ maximizes $L(\lambda|y_1, ..., y_n) = f(y_1, ..., y_n|\lambda)$ with respect to λ. The FOC is:

$$n\lambda^{n-1} \exp\left(-\lambda \sum_{i=1}^{n} y_i\right) - \lambda^n \left[\left(\sum_{i=1}^{n} y_i\right) \exp\left(-\lambda \sum_{i=1}^{n} y_i\right)\right] = 0$$

with solution $\lambda_{MLE} = n/\sum_{i=1}^n y_i = 1/\bar{y}$ where $\bar{y}$ is the sample mean. Of course, you need to check the second order condition to ensure you're at a maximum.

(c) Now the joint density is $f(y_1, ..., y_n|x_1, ..., x_n, \beta) = \prod_{i=1}^{n} (x_i\beta)^{-1} \exp(-y_i/x_i\beta)$. The log-likelihood is easiest to work with:

$$l(\beta|y, x) = -\sum_{i=1}^{n} \ln(x_i\beta) - \sum_{i=1}^{n} (y_i/x_i\beta).$$

The MLE of β maximizes $l(\beta|y, x)$ with respect to β. The FOC is:

$$-\sum_{i=1}^{n} \frac{1}{\beta} + \sum_{i=1}^{n} \frac{y_i}{x_i\beta^2} = 0$$

with solution $\beta_{MLE} = n^{-1} \sum_{i=1}^n y_i/x_i$. Again, you need to check the SOC to ensure you're at a maximum.

(d) Here, the trick is to note that $E[y_i] = x_i\beta$. This means our linearity assumption is satisfied, and consequently that least squares regression of y_i on x_i yields an unbiased estimate of β. To see that $E[y_i] = x_i\beta$, return to the original parameterization for simplicity, where:

$$\begin{aligned}
E[y] &= \int_0^\infty y\lambda e^{-\lambda y} dy = \left[y e^{-\lambda y}\right]_0^\infty + \int_0^\infty e^{-\lambda y} dy && \text{(integration by parts)} \\
&= \left[y e^{-\lambda y}\right]_0^\infty + \frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda y} dy = \left[y e^{-\lambda y}\right]_0^\infty + \frac{1}{\lambda} && \text{(integrand is a pdf)} \\
&= \frac{1}{\lambda}
\end{aligned}$$

because

$$\begin{aligned} ye^{-\lambda y}|_{y=\infty} &= 1/\left(\lambda e^{\lambda y}\right)|_{y=\infty} = 0 \text{ by L'Hopital's rule, and} \\ ye^{-\lambda y}|_{y=0} &= 0 \text{ also.} \end{aligned}$$

So, $E[y_i] = \lambda^{-1} = x_i\beta$, and the linearity assumption is satisfied.