AAEC/ECON 5126 MIDTERM EXAM: SOLUTIONS

SPRING 2016 / INSTRUCTOR: KLAUS MOELTNER

This exam is open-book, open-notes, but please work strictly on your own. Please make sure your name is on every sheet you're handing in. You have 75 minutes to complete this exam. You can collect a maximum of 30 points. Each question is scored as indicated below. Vectors are given in lower-case boldface. Matrices are written in upper-case boldface.

1. Question I (6 points)

Consider the sample gradient for the classical liner regression model with sample size n:

$$g\left(\boldsymbol{\beta}, \sigma^2\right) = \begin{bmatrix} \frac{\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma^2} \\ -\frac{n}{2\sigma^2} + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^4} \end{bmatrix}$$

Solve for the MLE estimator $\hat{\sigma}^2$ in terms of residuals and sample size, conditional on having solved for $\hat{\boldsymbol{\beta}}$.

Solution:
Plugging in $\hat{\boldsymbol{\beta}}$ in the first derivative for σ^2 and setting to zero yields:

$$\begin{aligned} -\frac{n}{2\sigma^2} + \frac{\mathbf{e}'\mathbf{e}}{2\sigma^4} &= 0 \\ \frac{n}{2\sigma^2} &= \frac{\mathbf{e}'\mathbf{e}}{2\sigma^4} \\ \hat{\sigma}^2 &= \frac{\mathbf{e}'\mathbf{e}}{n} \end{aligned}$$

Date: March 17, 2016.

2. Question II (12 points)

Consider the CLRM for a full sample of n observations:

$$\mathbf{y} = \beta_0 \mathbf{i} + \beta_1 \mathbf{d}_1 + \beta_2 \mathbf{d}_2 + \beta_3 \mathbf{d}_3 + \beta_4 \mathbf{x} + \boldsymbol{\epsilon},$$

where all the usual assumptions hold, $\mathbf{i}$ is an n by 1 vector of "1"s, $\mathbf{d}_1$ is an indicator variable that takes the value of "1" for person i if that person is "female" (as opposed to "male"), and a value of "0" otherwise, $\mathbf{d}_2$ is an indicator variable that takes the value of "1" for person i if that person is "retired" (as opposed to "working"), and a value of "0" otherwise, and $\mathbf{x}_i$ is a continuous explanatory variable. The term $\mathbf{d}_3$ is an interaction term of $\mathbf{d}_1 * \mathbf{d}_2$ (in element-by-element multiplication).

Assume that n_1 of the n individuals are female, n_2 are retired, and n_3 are both (with $0 < n_k < n, \quad k = 1, 2, 3$).

(a) *(2 pts.)* For a single observation, show the expectation of y_i given $\mathbf{x}_i$ for all four possible cases. Is the marginal effect of x_i on $E(y_i|x_i)$ the same for all groups or not? How about the intercept?

(b) *(2 pts.)* Show that because of the inclusion of the interaction term the marginal effect of "female" on outcome y is allowed to differ for working and retired people. *(Hint: express the marginal effect as a difference in expectations.)*

(c) *(2 pts.)* Let $\mathbf{Z} = \begin{bmatrix} \mathbf{i} & \mathbf{d}_1 & \mathbf{d}_2 & \mathbf{d}_3 \end{bmatrix}$ and $\boldsymbol{\beta}_z = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}'$. Show the explicit form of $\mathbf{Z}'\boldsymbol{\epsilon}$.*(Hint: This should be a 4 by 1 vector of summation terms. State the condition for the elements over which the summation is taken underneath the summation sign.)*

(d) *(2 pts.)* Let b_4 be the OLS estimator for β_4. Write its solution in partitioned regression form, and show the explicit form of the residual maker matrix $\mathbf{M}_1$ in terms of $\mathbf{Z}$.

(e) *(4 pts.)* Show that b_4 is an unbiased estimator for β_4 (conditional on $\mathbf{x}, \mathbf{Z}$).

Solutions:

(a)

$$\begin{aligned}
E(y_i|d_{1i} = 0, d_{2i} = 0, x_i) &= \beta_0 + \beta_4 * x_i \\
E(y_i|d_{1i} = 1, d_{2i} = 0, x_i) &= (\beta_0 + \beta_1) + \beta_4 * x_i \\
E(y_i|d_{1i} = 0, d_{2i} = 1, x_i) &= (\beta_0 + \beta_2) + \beta_4 * x_i \\
E(y_i|d_{1i} = 1, d_{2i} = 1, x_i) &= (\beta_0 + \beta_1 + \beta_2 + \beta_3) + \beta_4 * x_i
\end{aligned}$$

The slopes (marginal effect of x_i) are the same for both groups, but the intercepts differ for all four cases.

(b)

$$E\left(y_i|d_{1i}=1,x_i\right)-E\left(y_i|d_{1i}=0,x_i\right)|d_{2i=0}=\beta_0+\beta_1+\beta_4*x_i-\left(\beta_0+\beta_4*x_i\right)=\beta_1$$
$$E\left(y_i|d_{1i}=1,x_i\right)-E\left(y_i|d_{1i}=0,x_i\right)|d_{2i=1}=\beta_0+\beta_1+\beta_2+\beta_3+\beta_4*x_i-\left(\beta_0+\beta_2+\beta_4*x_i\right)=\beta_1+\beta_3$$

Clearly, the two marginal effects differ by the interaction effect β_3.

(c)

$$\mathbf{Z}'\boldsymbol{\epsilon}=\begin{bmatrix}\mathbf{i}'\boldsymbol{\epsilon}\\ \mathbf{d}_1'\boldsymbol{\epsilon}\\ \mathbf{d}_2'\boldsymbol{\epsilon}\\ \mathbf{d}_3'\boldsymbol{\epsilon}\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{n}\epsilon_i\\ \sum_{i|d_{1i}=1}\epsilon_i\\ \sum_{i|d_{2i}=1}\epsilon_i\\ \sum_{i|d_{1i}=1,d_{2i}=1}\epsilon_i\end{bmatrix}$$

(d)

$$b_4=\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\mathbf{y}$$
$$\mathbf{M}_1=\mathbf{I}-\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$$

where $\mathbf{I}$ is an n by n identity matrix.

(e) Plugging in the full regression model for $\mathbf{y}$ in the partitioned regression form of b_4 and taking expectations yields:

$$E\left(b_4|\mathbf{x},\mathbf{Z}\right)=$$
$$E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\mathbf{Z}\boldsymbol{\beta}_z\right)+$$
$$E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\mathbf{x}\beta_4\right)+$$
$$E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\boldsymbol{\epsilon}\right)=\beta_4$$

since $\mathbf{M}_1\mathbf{Z}=\mathbf{0}$ and:

$$E\left(\mathbf{x}'\mathbf{M}_1\boldsymbol{\epsilon}\right)=$$
$$E\left(\mathbf{x}'\boldsymbol{\epsilon}-\mathbf{x}'\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{\epsilon}\right)=$$
$$E\left(\mathbf{x}'\boldsymbol{\epsilon}\right)-\mathbf{x}'\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}E\left(\mathbf{Z}'\boldsymbol{\epsilon}\right)=0$$

by assumption (3) of the CLRM.

3. Question III (12 points)

(Continuation of Question II): Now assume the analyst ignores the interaction term $\mathbf{d}_3$, and instead estimates the following (incorrect) model:

$$\mathbf{y} = \beta_0 \mathbf{i} + \beta_1 \mathbf{d}_1 + \beta_2 \mathbf{d}_2 + \beta_4 \mathbf{x} + \boldsymbol{\nu}$$

(a) *(2 pts.)* Show the explicit form of $\boldsymbol{\nu}$.

(b) *(2 pts.)* Let $\tilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{i} & \mathbf{d}_1 & \mathbf{d}_2 \end{bmatrix}$ and $\boldsymbol{\beta}_{\tilde{z}} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 \end{bmatrix}'$. Show the explicit form of $\tilde{\mathbf{Z}}'\mathbf{d}_3$. *(Hint: This should be a 3 by 1 vector of summation terms that each equal a specific (known) constant.)*

(c) *(2 pts.)* Let $\hat{b}_4$ be the OLS estimator for β_4 for this model. Write its solution in partitioned regression form, and show the explicit form of the residual maker matrix $\mathbf{M}_1$ in this case.

(d) *(4 pts.)* Show that $\hat{b}_4$ is a *biased* estimator for β_4 (conditional on $\mathbf{x}, \tilde{\mathbf{Z}}$).

(e) *(2 pts.)* Describe under which (hypothetical) condition the bias would vanish.

Solutions:

(a)

$$\boldsymbol{\nu} = \beta_3 \mathbf{d}_3 + \boldsymbol{\epsilon}$$

(b)

$$\tilde{\mathbf{Z}}'\mathbf{d}_3 = \begin{bmatrix} \mathbf{i}'\mathbf{d}_3 \\ \mathbf{d}_1'\mathbf{d}_3 \\ \mathbf{d}_2'\mathbf{d}_3 \end{bmatrix} = \begin{bmatrix} \sum_{i|d_{1i}=1,d_{2,i}=1} \mathbf{i} \\ \sum_{i|d_{1i}=1,d_{2,i}=1} \mathbf{i} \\ \sum_{i|d_{1i}=1,d_{2,i}=1} \mathbf{i} \end{bmatrix} = \begin{bmatrix} n_3 \\ n_3 \\ n_3 \end{bmatrix}$$

In all three cases, the summation "picks up" only indicators for which the interaction equals one.

(c)

$$\hat{b}_4 = \left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1} \mathbf{x}'\mathbf{M}_1\mathbf{y}$$

$$\mathbf{M}_1 = \mathbf{I} - \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}}\right)^{-1} \tilde{\mathbf{Z}}'$$

(d) Plugging in the full (**TRUE!**) regression model for $\mathbf{y}$ in the partitioned regression form of $\hat{b}_4$ and taking expectations yields:

$$
\begin{aligned}
&E\left(b_4|\mathbf{x},\mathbf{Z}\right) = \\
&E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\tilde{\mathbf{Z}}\boldsymbol{\beta}_{\tilde{z}}\right) + \\
&E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\mathbf{x}\beta_4\right) + \\
&E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\mathbf{d}_3\beta_3\right) + \\
&E\left(\left(\mathbf{x}'\mathbf{M}_1\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_1\boldsymbol{\epsilon}\right)
\end{aligned}
$$

The first and last expectations go to zero based on the same arguments as above. However, the third expectation term does not go to zero. More explicitly:

$$
\mathbf{x}'\mathbf{M}_1\mathbf{d}_3 = \mathbf{x}'\mathbf{d}_3 - \mathbf{x}'\tilde{\mathbf{Z}}\left(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}'\mathbf{d}_3
$$

So the expectation of this term only goes to zero if $E\left(d_{3i}=0\right)\ \forall i$, i.e there are no retired females in the *population* (recall we are only conditioning this model on $\mathbf{x}$ and $\tilde{\mathbf{Z}}$, but not on $\mathbf{d}_3$, so this must hold for the population, not just for the sample).