Econometrics 710
Midterm Exam
March 11, 2010
Sample Answers

1. My friend is confused. The assumption that the observations are iid implies nothing about the relationship between y_i and $\boldsymbol{x}_i$. It does not imply that the linear equation $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i$ is a regression. (Calling it a regression does not make it one!) Even if $\boldsymbol{x}_i'\boldsymbol{\beta}$ is the conditional mean so that $E(e_i|\boldsymbol{x}_i) = 0$, the iid assumption does not imply anything about the conditional variance $E(e_i^2|\boldsymbol{x}_i)$. We say that a regression is "homoskedastic" if $E(e_i^2|\boldsymbol{x}_i) = \sigma^2$ is a constant independent of $\boldsymbol{x}_i$. This is not implied by the assumption that the observations are iid.

2. No, the estimates are different, except in the special case that the variances are uncorrelated in-sample: $X_1'X_2 = 0$. We can calculate that

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_1 &= \left(X_1'X_1\right)^{-1} X_1'Y \\
\tilde{u} &= M_1 Y \\
M_1 &= I - X_1\left(X_1'X_1\right)^{-1} X_1' \\
\tilde{\boldsymbol{\beta}}_2 &= \left(X_2'X_2\right)^{-1} X_2'\tilde{u} \\
&= \left(X_2'X_2\right)^{-1} X_2'M_1 Y
\end{aligned}$$

From the FWL theorem, we have the expression

$$\hat{\boldsymbol{\beta}}_2 = \left(X_2'M_1X_2\right)^{-1} X_2'M_1Y$$

These two expressions are different, except when

$$\begin{aligned}
X_2'X_2 &= X_2'M_1X_2 \\
&= X_2'X_2 - X_2'X_1\left(X_1'X_1\right)^{-1} X_1'X_2
\end{aligned}$$

which happens when $X_2'X_1 = 0$. Another way to see the difference is to take the expression for $\tilde{\boldsymbol{\beta}}_2$ and substitute in the OLS estimates on the full model

$$Y = X_1\hat{\boldsymbol{\beta}}_1 + X_2\hat{\boldsymbol{\beta}}_2 + \hat{e}$$

to find

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_2 &= \left(X_2'X_2\right)^{-1} X_2'M_1Y \\
&= \left(X_2'X_2\right)^{-1} X_2'M_1\left(X_1\hat{\boldsymbol{\beta}}_1 + X_2\hat{\boldsymbol{\beta}}_2 + \hat{e}\right) \\
&= \left(X_2'X_2\right)^{-1} X_2'M_1X_1\hat{\boldsymbol{\beta}}_1 + \left(X_2'X_2\right)^{-1} X_2'M_1X_2\hat{\boldsymbol{\beta}}_2 + \left(X_2'X_2\right)^{-1} X_2'M_1\hat{e} \\
&= \left(X_2'X_2\right)^{-1} X_2'M_1X_2\hat{\boldsymbol{\beta}}_2 \\
&= \hat{\boldsymbol{\beta}}_2 - \left(X_2'X_2\right)^{-1} X_2'X_1\left(X_1'X_1\right)^{-1} X_1'X_2\hat{\boldsymbol{\beta}}_2 \neq \hat{\boldsymbol{\beta}}_2
\end{aligned}$$

where the second-to-last equality uses $M_1X_1 = 0$ and $M_1\hat{e} = M_1My = 0$

3. Yes, it is consistent. There are two key steps to this observation. One, the observations are selected randomly. Two, the subsample has $N = n/2$ observations and $N \to \infty$ as $n \to \infty$. Since the observations are selected randomly from the the dataset, this is equivalently to taking a sample of size N from the population. Thus the slope coefficient is equivalent to estimation based on an iid sample of size N. As the slope coefficient is consistent for the population projection coefficient (for any random sample), then so is this estimator.

It is difficult to write this down formally. One way to imagine this is to create a random variable u_i, independent of the observations, such that $P(u_i = 1) = 1/2$ and $P(u_i = 0) = 1/2$, so that $u_i = 1$ indicates that observation i has been selected into the estimation sample. Then you can write the estimator as

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' u_i\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i u_i\right)$$

Applying the WLLN we obtain

$$\hat{\boldsymbol{\beta}} \to_p E\left(\boldsymbol{x}_i \boldsymbol{x}_i' u_i\right)^{-1} E\left(\boldsymbol{x}_i y_i u_i\right)$$

Since u_i is independent of the observations, then $E\left(\boldsymbol{x}_i \boldsymbol{x}_i' u_i\right) = E\left(\boldsymbol{x}_i \boldsymbol{x}_i'\right) E\left(u_i\right) = E\left(\boldsymbol{x}_i \boldsymbol{x}_i'\right) \frac{1}{2}$ and $E\left(\boldsymbol{x}_i y_i u_i\right) = E\left(\boldsymbol{x}_i y_i\right) E\left(u_i\right) = E\left(\boldsymbol{x}_i y_i\right) \frac{1}{2}$ so the above expression simplifies to

$$\left(E\left(\boldsymbol{x}_i \boldsymbol{x}_i'\right) \frac{1}{2}\right)^{-1} E\left(\boldsymbol{x}_i y_i\right) \frac{1}{2} = \left(E\left(\boldsymbol{x}_i \boldsymbol{x}_i'\right)\right)^{-1} E\left(\boldsymbol{x}_i y_i\right)$$

which is the projection coefficient.

4. There are two possible interpretations of the question.

   (a) i. The advisor is asking you to exclude the single variable x_{2i} but keep the quadratic and interactions, or

   ii. The advisor is asking you to exclude x_{2i} completely from the regression

   A few of you interpreted the question as (i), but almost everyone interpreted the question as (ii). Many expressed some ambiguity, but then explained their reasoning for their choice. In fact, if you were to be asked this question, it would be quite appropriate to respond: Do you mean (i) or (ii)? If the advisor responds with "I meant (i)", then I believe it would be appropriate to respond: "Are you sure? Does it really make sense to exclude the linear variable x_{2i} but retain the quadratic and interaction terms? Why do you suggest this specification?" Perhaps the advisor has a good reason. If not, they might respond: "Oh I see, my suggestion was poorly thought. I should have meant (ii)." In grading this question, I let each of you interpret the question individually, and assessed your answer by your reasoning, justification, rigor and arguments.

   My answer interprets the question as (ii), as that had been my intention.

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + \beta_4 x_{2i}^2 + \beta_5 x_{1i} x_{2i} + e_i$$

   (b) The null and alternative are

$$H_0 \;:\; \begin{pmatrix} \beta_2 \\ \beta_4 \\ \beta_5 \end{pmatrix} = 0$$

$$H_1 \;:\; \begin{pmatrix} \beta_2 \\ \beta_4 \\ \beta_5 \end{pmatrix} \neq 0$$

   (c) The model does not state assumptions on the error. In this case it is appropriate to use the most general technique, which is a Wald statistic with a robust covariance matrix estimate. The Wald statistic is a good choice as the hypothesis is a linear restriction on the parameters. We can write the model as

$$y_i = z_i'\beta + e_i$$

where

$$z_i = \begin{pmatrix} 1 \\ x_{1i} \\ x_{2i} \\ x_{1i}^2 \\ x_{2i}^2 \\ x_{1i}x_{2i} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

The hypothesis can be written as

$$R'\beta = 0$$

where

$$R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The parameters are estimated by least-squares

$$\begin{aligned} \hat{\beta} &= \left(Z'Z\right)^{-1} Z'Y \\ \hat{e}_i &= y_i - z_i'\hat{\beta} \end{aligned}$$

The Wald statistic is

$$W_n = n\hat{\beta}' R \left(R'\hat{V}_\beta R\right)^{-1} R'\hat{\beta}$$

where my favorite estimate of the covariance matrix is

$$\begin{aligned} \hat{V}_\beta &= \hat{Q}^{-1} \\ \hat{Q} &= \frac{1}{n} Z'Z \\ \hat{\Omega} &= \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \hat{e}_i^2 \left(1 - h_{ii}\right)^{-1} \\ h_{ii} &= z_i' \left(Z'Z\right)^{-1} z_i \end{aligned}$$

(d) Under H_0, $W_n \to_d \chi_3^2$, a chi-square with 3 degrees of freedom.

(e) The rule is to reject H_0 in favor of H_1 if $W_n > c$ where c is the critical value, and to not reject H_0 if $W_n \leq c$. For a given significance level α, the critical value c is the upper α quantile of the χ_3^2 distribution, that is $P(\chi_3^2 > c) = \alpha$. (For the 5% significance level, it turns out that $c = 7.81$, but I wasn't expecting you to know that, even though you could have looked it up.) As 5% is the conventional significance level, you reject H_0 if $W_n > 7.81$.