AAEC/ECON 5126 final exam

Spring 2019 / Instructor: Klaus Moeltner

May 11, 2019

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 120 minutes to complete this exam. You can collect a maximum of 50 points. Each question is scored as indicated below. Vectors are given in lower-case boldface. Matrices are written in upper-case boldface.

Question I (18 points):

Consider the following linear regression model for observation i:

$$\begin{aligned} & y_i = \beta_0 + \beta_1 s_i + \mathbf{x}_i' \boldsymbol{\gamma} + \epsilon_i \quad \text{with} \\ & \epsilon_i \sim n\left(0, \sigma^2\right), \quad \forall i = 1 \dots n, \end{aligned} \tag{1}$$

where y_i is the sales price of a single-family residential home (in dollars), s_i is square footage, $\mathbf{x}_i$ includes a set of additional (exogenous) regressors, and ϵ_i is a typical error term with the usual CLRM properties, as shown in the second line of (1).

Part (a), 8 points

(a) Show $E\left(y_i | s_i, \mathbf{x}_i\right)$, where $E\left(.\right)$ is the expectation operator.

(b) What is the interpretation of β_1 with respect to y_i? Provide mathematical support for your answer.

(c) If one were to use $ln\, y_i$, where ln is the natural logarithm, instead of y_i in (1), how would that change the interpretation of β_1 with respect to y_i? Provide mathematical support for your answer.

(d) If, *in addition*, one were to use the log of square footage, $ln\, s_i$ instead of s_i in (1), how would that change the interpretation of β_1 with respect to y_i? Provide mathematical support for your answer.

Solution:

(a) Expected sales price is given as:

$$E\left(y_i | .\right) = \beta_0 + \beta_1 s_i + \mathbf{x}_i' \boldsymbol{\gamma}$$

(b) β_1 gives the linear effect of s_i on price, that is the change in price, in dollars, due to a unit change in square footage, ceteris paribus. Mathematically:

$$\frac{\partial y_i}{\partial s_i} = \beta_1$$

(c) We now have:

$$y_i = exp\left(\beta_0 + \beta_1 s_i + \mathbf{x}_i'\boldsymbol{\gamma} + \epsilon_i\right) \quad \text{and}$$
$$\frac{\partial y_i}{\partial s_i} = y_i * \beta_1, \quad \text{or} \quad \frac{\left(\frac{\partial y_i}{y_i}\right)}{\partial s_i} = \beta_1$$

Thus, β_1 can be interpreted as the proportional change in home price due to a 1-unit change in square footage, ceteris paribus.

(d) Using logs for both y_i and s_i gives:

$$y_i = exp\left(\beta_0 + \beta_1 ln\ s_i + \mathbf{x}_i'\boldsymbol{\gamma} + \epsilon_i\right) \quad \text{and}$$
$$\frac{\partial y_i}{\partial s_i} = y_i * \beta_1 * \frac{1}{s_i}, \quad \text{or} \quad \frac{\left(\frac{\partial y_i}{y_i}\right)}{\left(\frac{\partial s_i}{s_i}\right)} = \beta_1$$

Thus, β_1 can be interpreted as an elasticity, that is the percentage change in price due to a 1-percent change in square footage, ceteris paribus.

Part (b), 6 points

Now consider another model that uses price divided by square footage as the dependent variable, i.e.:

$$\begin{aligned} y_i^* &= \frac{y_i}{s_i} = \beta_0 + \beta_1 s_i + \mathbf{x}_i'\boldsymbol{\gamma} + \epsilon_i \quad \text{with} \\ \epsilon_i &\sim n\left(0, \sigma^2\right), \quad \forall i = 1 \dots n, \end{aligned} \tag{2}$$

(a) What is the new interpretation of β_1? Assuming diminishing marginal utility of housing space holds for the entire range of square footage found in the data, what would you expect its sign to be?

(b) Compute the direct effect of s_i on y_i for this model. How is it fundamentally different from all other effects of square footage on price derived in part (a) above?

(c) At what value of s_i (which may or may not be represented by the data) is this direct effect maximized? Under which additional condition is this indeed a maximum, and how does your answer relate to your argument regarding the expected sign of β_1 from above?

Solution:

(a) Now β_1 signifies the change in price per square foot due to a 1-unit change in square footage. Under diminishing utility over space holds, one would expect its sign to be negative.

(b) The direct effect of s_i on y_i can be derived as:

$$\begin{aligned} y_i &= s_i\beta_0 + \beta_1 s_i^2 + s_i\mathbf{x}_i'\boldsymbol{\gamma} + s_i\epsilon_i \quad \text{and} \\ \frac{\partial y_i}{\partial s_i} &= \beta_0 + 2\beta_1 * s_i + \mathbf{x}_i'\boldsymbol{\gamma} + \epsilon_i \end{aligned}$$

Thus, the 1-unit effect of s_i on price now changes over the range of s_i.

(c) It is maximized at $s_i^* = -\frac{\beta_0 + \mathbf{x}_i'\boldsymbol{\gamma} + \epsilon_i}{2\beta_1}$. This is indeed a maximum if the second derivative of price w.r.t. square footage is negative, that is if $\beta_1 < 0$. This, of course, is expected as discussed above.

Part (c), 4 points
Somebody suggests using the following mathematically equivalent model to (2) and estimating it via OLS:

$$y_i = s_i\beta_0 + \beta_1 {s_i}^2 + s_i\mathbf{x}_i'\boldsymbol{\gamma} + s_i\epsilon_i \tag{3}$$

(a) How does this model violate CLRM assumptions?

(b) How could this be addressed econometrically to derive consistent estimates for *all* parameters? Show as much mathematical detail as possible.

Solution:

(a) The error variance is now given as $V(\epsilon_i) = si^2 * \sigma^2$ for each observation. Thus, errors become heteroskedastic, violating A4 of the CLRM.

(b) This could be fixed by using GLS or, more specifically, weighted LS (WLS) instead of OLS, with error variance-covariance matrix for the whole sample given as $\boldsymbol{\Omega} = diag\begin{bmatrix} s_1^2 & s_2^2 & \dots & s_n^2 \end{bmatrix}$. Then use $\boldsymbol{\beta}_{WLS} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$.

Question 2 (16 points)

Consider the following true relationship between nitrogen fertilizer (N) and yield (y) for a specific crop, for plot i:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 N_i + \epsilon_i \quad \text{if} \quad N_i < N^* \\ y_i &= P + \epsilon_i \quad \text{otherwise,} \quad \text{and} \\ \epsilon_i &\sim n\left(0, \sigma^2\right), \quad \forall i = 1 \ldots n, \end{aligned} \tag{4}$$

where P is often referred to as "plateau yield," and ϵ_i is a mean-zero normal error with equal variance σ^2 for all i, as shown in the second line of (1). N^* is the amount of fertilizer beyond which yield will simply get "stuck" at the plateau. Throughout this question assume that in an actual application N_i goes from zero to 200, and that N^*, while unknown, is located somewhere towards the middle of this range. Also assume no actual N_i exactly equals N^*.

Further assume $\beta_0 > 0$, $\beta_1 > 0$.

Part (a), 6 points

(a) Show $E(y_i|N_i)$ for both $N_i < N^*$, and $N_i > N^*$, respectively.

(b) Graph $E(y_i|N_i)$ for the entire range of N_i, with yield on the y-axis, and nitrogen on the x-axis. Add a few scattered dots around this line to symbolize the actual data points.

Solution:

(a)

$$\begin{aligned} E\left(y_i|N_i < N^*\right) &= \beta_0 + \beta_1 N_i \\ E\left(y_i|N_i > N^*\right) &= P \end{aligned}$$

(b) The graph should look something like this (ignore the dashed line for now):

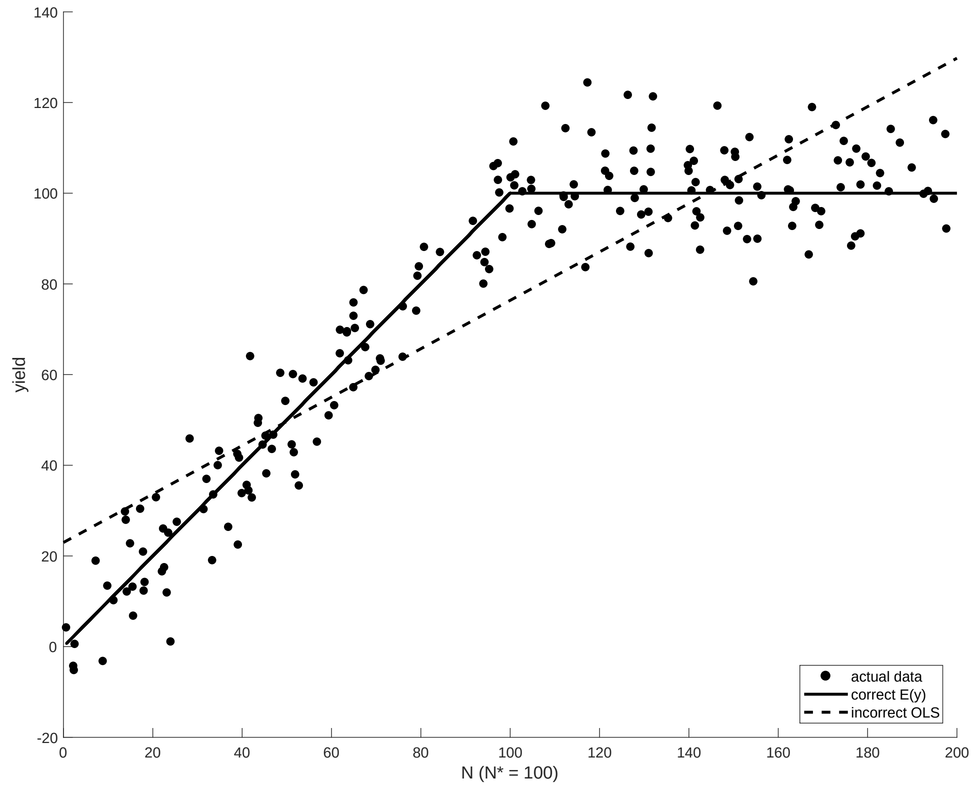

Figure 1: $E(y|N)$ with data points and incorrect OLS

Part (b), 6 points

Now assume a researcher is unaware of the true relationship between yield and nitrogen, and simply uses an OLS regression of y_i on a constant and N_i, using the *entire data*, to estimate β_0, β_1, and σ^2.

(a) Add your best guess for the estimated regression line (= predicted values for yield for the entire data range) to your graph.

(b) In which directions will the estimates for β_0 be biased? How about for β_1 and σ^2? *(Verbal answer is sufficient)*

Solution:

(a) The OLS regression line should look something like the dashed line in the graph above.

(b) The intercept will be biased upwards, the slope will be biased downwards, and the error variance will be biased upwards.

Part (c), 4 points
Now assume the researcher knows the general form of the true relationship in (1) as well as N^*.

(a) How could she use the subset of observations with $N_i < N^*$ and basic OLS to predict plateau yield P? Show some math.

(b) How would one derive a standard error for this prediction? Show some math. Assume that the estimated variance-covariance matrix for $\hat{\boldsymbol{\beta}} = \begin{bmatrix}\hat{\beta}_0 & \hat{\beta}_1\end{bmatrix}'$ is given as $\hat{V}_{\boldsymbol{\beta}}$.

Solution:

(a) Estimate β_0 and β_1 via OLS, then use: $\hat{P} = \hat{\beta}_0 + \hat{\beta}_1 * N^*$.

(b) $s.e.\left(\hat{P}\right) = \sqrt{\mathbf{x}'\hat{V}_{\boldsymbol{\beta}}\mathbf{x}}$, where $\mathbf{x} = \begin{bmatrix}1 & N^*\end{bmatrix}'$.

Question 3 (16 points)

Consider the CLRM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ (call it "Model 1"). Let $\mathbf{X}$ be partitioned into $\mathbf{X}_1$ and $\mathbf{X}_2$, which have dimensions n by k_1 and n by k_2, respectively. Let $k = k_1 + k_2$. Partition $\boldsymbol{\beta}$ accordingly into $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

Assume $\mathbf{X}_1$ and $\mathbf{X}_2$ are perfectly orthogonal in the sample and in the population. Furthermore, neither of them are correlated with the regression error, by the usual CLRM assumption.

Part (a), 6 points

(a) Using partitioned regression results, derive separate estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ (call them $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$).

(b) For the full model, express the residual vector $\mathbf{e}$ as a function of $\mathbf{y}$ and the projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, where $\mathbf{P}_j = \mathbf{X}_j \left(\mathbf{X}_j'\mathbf{X}_j\right)^{-1} \mathbf{X}_j' \quad j = 1, 2$.

(c) Show that $(\mathbf{P}_1 + \mathbf{P}_2)$ is idempotent under the model assumptions.

(d) Express the sum of squared residuals (SSR) as a function of $\mathbf{y}$ and the projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$. Call it SSR_1.

Solution:

$$\begin{aligned}
&\hat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}, \quad \text{since} \\
&\mathbf{X}_1'\mathbf{M}_2 = \mathbf{X}_1'\left(\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\right) = \\
&\mathbf{X}_1' - \mathbf{X}_1'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2' = \mathbf{X}_1' - \mathbf{0} = \mathbf{X}_1'
\end{aligned}$$

(analogous for $\hat{\boldsymbol{\beta}}_2$)

$$\mathbf{e} = \mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{y} - \mathbf{P}_1\mathbf{y} - \mathbf{P}_2\mathbf{y}$$

$$\begin{aligned}
&(\mathbf{P}_1 + \mathbf{P}_2) * (\mathbf{P}_1 + \mathbf{P}_2) = \mathbf{P}_1\mathbf{P}_1 + 2\mathbf{P}_1\mathbf{P}_2 + \mathbf{P}_2\mathbf{P}_2 = \mathbf{P}_1 + \mathbf{P}_2, \quad \text{since} \\
&\mathbf{P}_1\mathbf{P}_2 = \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2 = \mathbf{0} \quad \text{by orthogonality}
\end{aligned}$$

$$\begin{aligned}
&SSR_1 = \mathbf{e}'\mathbf{e} = \left(\mathbf{y} - (\mathbf{P}_1 + \mathbf{P}_2)\,\mathbf{y}\right)'\left(\mathbf{y} - (\mathbf{P}_1 + \mathbf{P}_2)\,\mathbf{y}\right) = \\
&\mathbf{y}'\mathbf{y} - 2\mathbf{y}'(\mathbf{P}_1 + \mathbf{P}_2)\,\mathbf{y} + \mathbf{y}'(\mathbf{P}_1 + \mathbf{P}_2)\,\mathbf{y} = \\
&\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{P}_1\mathbf{y} - \mathbf{y}'\mathbf{P}_2\mathbf{y}
\end{aligned}$$

Part (b), 6 points

Now consider a second CLRM model ("Model 2") that regresses $\mathbf{y}$ only on $\mathbf{X}_1$, i.e. $\mathbf{y} = \mathbf{X}_1\boldsymbol{\gamma} + \boldsymbol{\nu}$.

(a) Write down the OLS solution for $\hat{\boldsymbol{\gamma}}$ and compare it to your estimator for $\boldsymbol{\beta}_1$ from part (a). Comment.

(b) For Model 2, express the residual vector $\mathbf{e}$, as well as the sum of squared residuals (SSR), as a function of $\mathbf{y}$ and the projection matrix $\mathbf{P}_1$. Call this sum SSR_2.

(c) Show the expression for the difference of the two SSRs, and argue that the SSR from Model 1 can be no larger than the SSR from Model 2. (*Hint: Recall that projection matrices are semipositive definite*).

Solution:

$$
\begin{aligned}
\hat{\gamma} &= \left(\mathbf{X}_1'\mathbf{X}\right)^{-1}\mathbf{X}_1'\mathbf{y} = \hat{\boldsymbol{\beta}}_1 \\
\mathbf{e} &= \mathbf{y} - \mathbf{X}_1\hat{\gamma} = \mathbf{y} - \mathbf{P}_1\mathbf{y} \\
SSR_2 &= \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{P}_1\mathbf{y} \\
SSR_2 - SSR_1 &= \mathbf{y}'\mathbf{P}_2\mathbf{y}
\end{aligned}
$$

This is a quadratic form that will be ≥ 0, since $\mathbf{P}_2$ is semi-positive definite. So the model with less information produces an SSR that is at least as large as the SSR for the more complete model.

Part (c), 4 points

What does this imply for the estimate of the (conditional) variance of $\hat{\boldsymbol{\beta}}_1$ compared to the (conditional) variance of $\hat{\boldsymbol{\gamma}}$ for both a finite sample of size n, and when $n \to \infty$?
(Hint: Take a close look at the expression for the estimated error variance, s^2, for each model.)

Solution:

$$
\begin{aligned}
\hat{V}\left(\hat{\boldsymbol{\beta}}_1\right) &= s_1^2\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}, \quad s_1^2 = \frac{SSR_1}{n-k} \\
\hat{V}\left(\hat{\gamma}\right) &= s_2^2\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}, \quad s_2^2 = \frac{SSR_2}{n-k_1}
\end{aligned}
$$

While $SSR_2 \geq SSR_1$, we have $k > k_1$, so the denominator of s_1^2 is smaller than the denominator of s_2^2. Thus, for a finite sample, the overall comparative magnitude of $\hat{V}\left(\hat{\boldsymbol{\beta}}_1\right)$ versus $\hat{V}\left(\hat{\boldsymbol{\gamma}}\right)$ is indeterminate. However, as n grows large the "denominator effect" vanishes, and, unambiguously, $\hat{V}\left(\hat{\boldsymbol{\beta}}_1\right) \leq \hat{V}\left(\hat{\gamma}\right)$. (Or, more technically correct, the difference between $\hat{V}\left(\hat{\boldsymbol{\beta}}_1\right)$ and $\hat{V}\left(\hat{\gamma}\right)$ becomes negative-semidefinite.)

Note: In retrospect, only the "finite n" part of this question really makes sense, since the ordinary limit of both variances goes to zero, as some of you pointed out (a gave full credit for that response), and the *plims* go to $\sigma_\epsilon^2 plim\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}$ and $\sigma_\nu^2 plim\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}$, respectively. Since no information on the relative magnitude of σ_ϵ^2 and σ_ν^2 is given, we can't make a clear statement on relative magnitude using *plims*.