Economics 471: Econometrics

University of Alabama

Department of Economics, Finance and Legal Studies

Fall 2015

Final Exam – Answers

1. (a) (1) $y = \alpha + \beta x + u$ (the model is correctly specified); (2) $E(u) = 0$ (mean zero error); (3) $E(u|x) = 0$ (errors are not correlated with the regressors); (4) no perfect collinearity; (5) $V(u|x) = \sigma^2$ (homoskedasticity); (6) $u \overset{iid}{\sim} N(0, \sigma^2)$

   (b) (1-4)

   (c) (1-5)

   (d) (1-6)

2. (a) $E(z_1) = E\left(\frac{x_1 - \overline{x}_1}{\sigma_{x_1}}\right) = \frac{1}{\sigma_{x_1}} E(x_1 - \overline{x}_1) = \frac{1}{\sigma_{x_1}}(\mu_1 - \mu_1) = 0$

   (b) $V(z_1) = V\left(\frac{x_1 - \overline{x}_1}{\sigma_{x_1}}\right) = \frac{1}{\sigma^2_{x_1}} V(x_1 - \overline{x}_1) = \frac{1}{\sigma^2_{x_1}} V(x_1) + \frac{1}{\sigma^2_{x_1}} V(\overline{x}_1) + \frac{1}{\sigma^2_{x_1}} 2{*}COV(x_1, \overline{x}_1) = \frac{1}{\sigma^2_{x_1}} V(x_1)$ $= \frac{1}{\sigma^2_{x_1}} \sigma^2_{x_1} = 1.$

   (c) Taking the average of each side of the regression equation give us

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} y_i &= \frac{1}{n}\sum_{i=1}^{n} \beta_0 + \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_{1i} + \beta_2 \frac{1}{n}\sum_{i=1}^{n} x_{2i} + \beta_3 \frac{1}{n}\sum_{i=1}^{n} x_{3i} + \frac{1}{n}\sum_{i=1}^{n} u_i \\ \overline{y} &= \beta_0 + \beta_1 \overline{x}_1 + \beta_2 \overline{x}_2 + \beta_3 \overline{x}_3 + \overline{u} \end{aligned}$$

   subtracting this equation from the original gives

$$(y_i - \overline{y}) = \beta_1 (x_{1i} - \overline{x}_1) + \beta_2 (x_{2i} - \overline{x}_2) + \beta_3 (x_{3i} - \overline{x}_3) + (u_i - \overline{u})$$

   In order to standardize the variables we must incorporate the standard deviations as

$$\begin{aligned} \frac{(y_i - \overline{y})}{\sigma_y} &= \beta_1 \frac{\sigma_{x_1}}{\sigma_y} \frac{(x_{1i} - \overline{x}_1)}{\sigma_{x_1}} + \beta_2 \frac{\sigma_{x_2}}{\sigma_y} \frac{(x_{2i} - \overline{x}_2)}{\sigma_{x_2}} + \beta_3 \frac{\sigma_{x_3}}{\sigma_y} \frac{(x_{3i} - \overline{x}_3)}{\sigma_{x_3}} + \frac{(u_i - \overline{u})}{\sigma_y} \\ z_y &= b_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 + e \end{aligned}$$

   (d) A one standard deviation change in x_1 will result in a b_1 standard deviation change in y.

   (e) $E(e) = E\left(\frac{u_i - \overline{u}}{\sigma_y}\right) = \frac{1}{\sigma_y} E(u_i - \overline{u}) = \frac{1}{\sigma_{x_1}}(0 - 0) = 0.$ $V(e) = V\left(\frac{u_i - \overline{u}}{\sigma_y}\right) = \frac{1}{\sigma^2_y} V(u - \overline{u}) = \frac{1}{\sigma^2_y} V(u) = \frac{1}{\sigma^2_y} \sigma^2_u = 1.$

3. (a) The omitted category is Protestant.

   (b) The omitted category is every one who is not Catholic.

   (c) In the first equation δ_1 is exactly the difference between the means of $IdealChildren$ for Protestants (the reference group) and Catholics. In the second equations, δ_1 also associated with Catholics, but because the reference group has now changed, δ_1 rpresents the difference between the mean of $IdealChildren$ among Catholics and non-Catholics. The intercept term is the mean for the reference group category: Protestants in the first group, and non-Catholics in the second.

(d) Protestants: 2.775; Catholics: $2.775 + 0.134 = 2.909$; Jews: $2.775 + 4.405 = 3.180$; None: $2.775 - 0.150 = 2.625$; Other: $2.775 - 0.0854 = 2.6896$

(e) The $t-$test for the Catholic coefficient is not significant. I conclude that I should fail to reject the null that the slope is zero. Given that the slope is zero, in this setting, this simply reflects the difference in the two groups, I conclude that the two groups are the same with respect to the ideal family size.

(f) It cannot be determined from the data at hand. The coefficients are different, but there is no standard error for this difference given in the problem.

4. (a) In the linear model, the partial effect $\frac{\partial y}{\partial x} = \beta$, is the coefficient on x. In the quadratic model, the partial effect $\frac{\partial y}{\partial x} = \beta + 2\gamma x$ is a function of x. Plugging in the values from the table we see that $\widehat{\frac{\partial y}{\partial x}} = 4.0113$ in the linear model and $\widehat{\frac{\partial y}{\partial x}} = 9.5326 + 2\left(-1.6915\right)x = 9.5326 - 3.3832x$.

(b) In the linear model a one-unit increase in homework will bring about the same change in test score regardless of the value of homework. Given that we know the maximum score is 100, we need to find the value of homework that gives the predicted test score equal to 100. $100 = 49.838 + 4.0113 * homework^* \Rightarrow homework^* = \left(100 - 49.838\right)/4.0113 = 12.50$ (insane) hours per day.

(c) For the quadratic model with $\beta > 0$ and $\gamma < 0$ we know that the function achieves its maximum when $\frac{\partial y}{\partial x} = \beta + 2\gamma x = 0 \Rightarrow 9.5326 - 3.3832 * homework^* = 0 \Rightarrow homework^* = 9.5326/3.3832 = 2.8176 \Rightarrow$ the maximum test score is equal to $47.232 + 9.5326 * 2.8176 - 1.6916 * 2.8176^2 = 60.6616$.

(d) The easiest way to do this is to test the null: $H_0 : \gamma = 0$ in the quadratic model. The t-statistic for this test is -9.0473 and its respective $p - value = 0.0000$. Therefore we reject the model is linear. A more complicated test would be to take the RSS from both the linear and quadratic models, construct the F-statistic= $\frac{(332301.3 - 325165.7)/1}{325165.7/(3733-2-1)} = 81.835$ and compare them to the critical value from the F-table= 3.84 which would also suggest to reject the null. Note that $t^2 = F$ in this case of a single restriction.

(e) In the linear model: a test for the validity of the regression is the null: $H_0 : \beta = 0$. Note that the F-statistic is not avaiable in this EViews output and thus we must use the R^2 version of the test, noting that the resticted model which has no regressors will have a $R^2 = 0$. $F = \frac{R^2/q}{(1-R^2)/(n-k-1)} = \frac{0.0271/1}{(1-0.0271)/(3733-1-1)} = 103.9265 > 3.84$ and thus we reject the null. An alternative would be to note that this is a single restrict and $t^2 = 10.19134^2 = 103.9265$. For the quadratic model $F = \frac{R^2/q}{(1-R^2)/(n-k-1)} = \frac{0.0479/2}{(1-0.0479)/(3733-2-1)} = 93.8279 > 3.00$ (critical value) and thus we reject the null. No t alternative is possible in this case.

(f) Here we list all six selection criteria available from the EViews output. Note that the quadratic model performs better in each. It has a higher value for each of the first two selection criteria and a smaller value for the remaining selection criteria.

	Selection Criteria				
	R^2	$\overline{R}^2$	$\widehat{\sigma}$	AIC	SC
linear model	0.0271	0.0268	9.4374	7.3278	7.3311
quadratic model	0.0479	0.0475	9.3368	7.3066	7.3116