Economics 471: Introductory Econometrics

Department of Economics, Finance and Legal Studies

University of Alabama

Fall 2019

Midterm II — Key

The exam consists of three questions on three pages. Each question is of equal value.

1. Consider a regression model through the origin: $y_i = \beta x_i + u_i$, $i = 1, 2, \ldots, n$, and the corresponding slope parameter estimator

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2},$$

but where the true data generating process (i.e., the truth) is $y_i = \beta x_i + \delta w_i + e_i$. We assume that e_i is mean zero, has a constant variance (σ^2) and is uncorrelated with both x_i and w_i. Given this information, answer the following:

   (a) What is the expected value of $\tilde{\beta}$?

   (b) Under what conditions is the estimator in part (a) unbiased?

   (c) Suppose we were to correctly specify the model ($y_i = \beta x_i + \delta w_i + u_i$); what would the estimator of β be? What would the estimator of δ be? Call these estimators $\widehat{\beta}$ and $\widehat{\delta}$, respectively.

   (d) Consider the estimators in part (c); what is the variance of $\widehat{\beta}$? What is the variance of $\widehat{\delta}$?

   (e) Without using formal proofs, is $\widehat{\beta}$ a consistent estimator of β? Is $\widehat{\delta}$ a consistent estimator of δ? How do you know?

(a) $\tilde{\beta} = \sum(\beta x_i + \delta w_i + e_i)x_i / \sum x_i^2$

$= \beta + \delta \sum x_i w_i / \sum x_i^2 + e_i / \sum x_i^2$

$E(\tilde{\beta}) = \beta + \delta \sum x_i w_i / \sum x_i^2$

(b) $E(\tilde{\beta}) = \beta$ if $\delta = 0$ or $\sum x_i w_i = 0$

(c) $\hat{\beta} = \dfrac{\sum \hat{r}_{1i} y_i}{\sum \hat{r}_{1i}}$ $\quad \hat{r}_{1i} = x_i - \hat{\gamma}_0 - \hat{\gamma}_1 w_i$ *

$\hat{\delta} = \dfrac{\sum \hat{r}_{2i} y_i}{\sum \hat{r}_{2i}}$ $\quad r_{2i} = w_i - \hat{\gamma}_0 - \hat{\gamma}_1 x_i$ **

(d) $V(\hat{\beta}) = \dfrac{\sigma^2}{\sum(x_i-\bar{x})^2(1-R_1^2)}$

R_1^2 is R^2 from reg *

$V(\hat{\delta}) = \dfrac{\sigma^2}{\sum(w_i-\bar{w})^2(1-R_2^2)}$

R_2^2 is R^2 from reg **

(e) $E(\hat{\beta}) = \beta$ & $V(\hat{\beta}) \to 0$ as $n \to \infty$

$E(\hat{\delta}) = \delta$ & $V(\hat{\delta}) \to 0$ as $n \to \infty$

2. Consider the population regression function $y = \alpha + u$. Assuming $\alpha > 0$, in the figure below, perform the following:

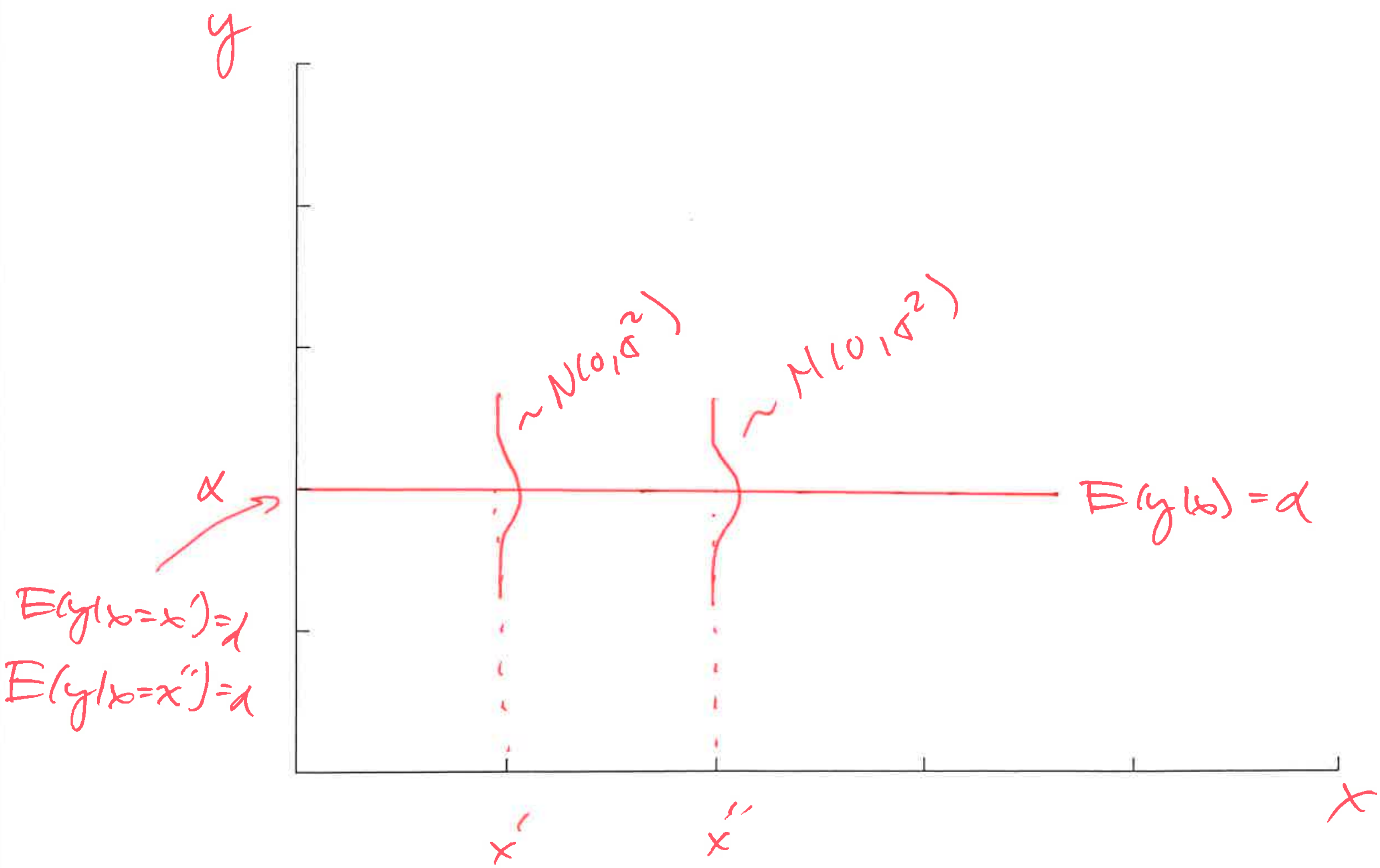
   (a) Label the axes

   (b) Plot and label the population regression curve.

   (c) Pick two values for x, plot their conditional expectations (i.e., $E(y|x)$).

   (d) For those two values of x in part (c), what is the marginal effect on $E(y|x)$ for each (i.e., $\partial E(y|x)/\partial x$)?

   (e) Assuming normally distributed, homoskedastic errors, plot and label the distribution of the error (u) for each of the points you listed in part (b).



(d) $$\frac{\partial E(y|x)}{\partial x} = 0 \quad \forall x$$

3. Consider the relationship between average monthly rent paid on rental units ($rent$) versus average city income ($avginc$), total city population (pop) and the percentage of students in the population ($pctstu$). Two gretl output files are below which correspond to two separate models. The univariate model is

Model 1: OLS, using observations 1–128
Dependent variable: lnrent

	Coefficient	Std. Error	t-ratio	p-value
const	−2.48821	0.435659	−5.711	0.0000
lnavginc	0.841260	0.0444821	18.91	0.0000

Mean dependent var	5.746195	S.D. dependent var	0.332707
Sum squared resid	3.662210	S.E. of regression	0.170485
R^2	0.739495	Adjusted R^2	0.737428
$F(1, 126)$	357.6765	P-value(F)	1.29e–38
Log-likelihood	45.82953	Akaike criterion	−87.65907
Schwarz criterion	−81.95501	Hannan–Quinn	−85.34148

and the multivariate model is

Model 2: OLS, using observations 1–128
Dependent variable: lnrent

	Coefficient	Std. Error	t-ratio	p-value
const	−3.36831	0.463944	−7.260	0.0000
lnavginc	0.877139	0.0413247	21.23	0.0000
lnpop	0.0313456	0.0270786	1.158	0.2493
pctstu	0.658487	0.120268	5.475	0.0000

Mean dependent var	5.746195	S.D. dependent var	0.332707
Sum squared resid	2.852256	S.E. of regression	0.151664
R^2	0.797110	Adjusted R^2	0.792201
$F(3, 124)$	162.3895	P-value(F)	8.98e–43
Log-likelihood	61.82675	Akaike criterion	−115.6535
Schwarz criterion	−104.2454	Hannan–Quinn	−111.0183

(a) Interpret the coefficient on $ln(avginc)$ in both Model 1 and Model 2.

(b) Interpret the coefficients on $ln(pop)$ and $pctstu$ in Model 2.

(c) Test the null hypothesis that $ln(avginc)$ is irrelevant in both Model 1 and Model 2.

(d) Test the null hypothesis that $ln(pop)$ and $pctstu$ are jointly irrelevant.

(e) Using (at least three of) the model selection criteria we discussed in class, what model has better predictive power?

(a) 1: if avginc ↑ by 1%, predicted rent ↑ 0.84%

2: holding lnpop & pctstu constant

a 1% ↑ avginc ⇒ predicted ↑ rent by 0.877%

(b) holding avginc & pctstu constant

a 1% ↑ in pop ⇒ predicted rent ↑ by 0.03%

holding avginc & pop constant

a 1% ↑ in students ⇒ predicted rent ↑ by 0.65%

(note: pctstu is in % terms already)

(c) 1: $H_0: \beta = 0$ vs. $H_1: \beta \neq 0$

$$t = \frac{0.841 - 0}{0.044} = 18.91 > 2 \Rightarrow \text{reject } H_0$$

2: $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

$$t = \frac{0.877 - 0}{0.041} = 21.23 > 2 \Rightarrow \text{reject } H_0$$

(d) $H_0: \beta_2 = \beta_3 = 0$ H_1: H_0 is not true

$$F = \frac{(SSR_R - SSR_u)/q}{SSR_u/(n-k-1)} = \frac{(R_u^2 - R_R^2)/q}{(1-R_u^2)/(n-k-1)}$$

$$= \frac{(3.66 - 2.85)/2}{2.85/(128-4)} = \frac{(0.797 - 0.739)/2}{(1-0.797)/(128-4)}$$

if $F > F_{2,124,0.25}$ ⇒ reject H_0

(e) model 1 model 2

R^2 $\hat{\sigma}$

$\bar{R}^2$ AIC

SSR SC

all smaller for model 2