Economics 471: Introductory Econometrics

Department of Economics, Finance and Legal Studies
University of Alabama
Spring 2022

Final Exam — Key

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

1. Consider a random sample of data $\{x_i, y_i\}_{i=1}^n$ and the model (without an intercept term) $y_i = \beta x_i + u_i$, where $E(u_i|x_i) = 0$ and $V(u_i|x_i) = \sigma_i^2$. With this information, answer the following questions:
   (a) Derive the method of moments estimator for β.
   (b) Derive the least-squares estimator for β
   (c) Suppose σ_i^2 is known, derive the conditional variance of $\widehat{\beta}$ given x (i.e., $V\left(\widehat{\beta}|x\right)$).
   (d) Suppose $\sigma_i^2 = \sigma^2$ for all i, simplify your conditional variance estimator from part (c).
   (e) Suppose σ^2 is unknown, give the estimator of σ^2 (i.e., $\widehat{\sigma}^2$). Be specific.

(a) $E(ux) = 0$

$$\frac{1}{n}\sum_{i=1}^{n}\left[(y_i - \hat{\beta}x_i)x_i\right] = 0 \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$

(b)

$$\sum_{i=1}^{n}\hat{u}_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}x_i)^2$$

$$\frac{\partial}{\partial\beta} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}x_i)x_i = 0 \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$

(c)

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^{n} u_i x_i}{\sum_{i=1}^{n} x_i^2}$$

$$V(\hat{\beta}|x) \overset{iid}{=} \frac{\sum_{i=1}^{n} x_i^2 V(u_i|x_i)}{\left[\sum_{i=1}^{n} x_i^2\right]^2} = \frac{\sum_{i=1}^{n} \sigma_i^2 x_i^2}{\left[\sum_{i=1}^{n} x_i^2\right]^2}$$

(d) if $\sigma_i^2 = \sigma^2 \ \forall i$

$$V(\hat{\beta}|x) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

(e)

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{\beta}x_i)^2$$

2. Consider a random sample of data $\{x_{1i}, x_{2i}, y_i\}_{i=1}^{n}$ and the model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, where $E(u_i|x_{1i}, x_{2i}) = 0$ and $V(u_i|x_{1i}, x_{2i}) = \sigma^2$. We know that an estimator of β_2 is

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^{n} \widehat{r}_{2i} y_i}{\sum_{i=1}^{n} \widehat{r}_{2i}^2}$$

and the conditional variance of that estimator is

$$\widehat{V}\left(\widehat{\beta}_2 | x_{1i}, x_{2i}\right) = \frac{\widehat{\sigma}^2}{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2 (1 - R_2^2)}.$$

With this information, answer the following questions:

(a) What model is used to estimate r_{2i}?

(b) For the model you listed in part (a), derive the least-squares estimator of the intercept parameter.

(c) For the model you listed in part (a), derive the least-squares estimator of the slope parameter.

(d) Suppose x_1 and x_2 are uncorrelated, what value should the slope estimator from part (c) take?

(e) Suppose the correlation between x_1 and x_2 approaches 1, what happens to the conditional variance of $\widehat{\beta}_2$ (i.e., $\widehat{V}\left(\widehat{\beta}_2 | x_{1i}, x_{2i}\right)$)?

(a) $x_{2i} = \gamma_0 + \gamma_1 x_{1i} + r_{2i}$

(b) & (c)

$$\sum_{i=1}^{n} \hat{r}_{2i}^2 = \sum_{i=1}^{n} (x_{2i} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{1i})^2$$

$$\frac{\partial}{\partial \alpha} = -2 \sum_{i=1}^{n} (x_{2i} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{1i}) = 0$$

$$\frac{\partial}{\partial \beta} = -2 \sum_{i=1}^{n} (x_{2i} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{1i}) x_{1i} = 0$$

$$\Rightarrow \quad \hat{\gamma}_0 = \bar{x}_2 - \hat{\gamma}_1 \bar{x}_1$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1)}{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2}$$

(d) if $\text{corr}(x_1, x_2) = 0$, $\gamma_1 = 0$

(e) as $\text{corr}(x_1, x_2) \to 1$, $R_2^2 \to 1 \Rightarrow V(\hat{\beta}_2 | x_1, x_2) \to \infty$

3. Consider a random sample of data $\{x_i, y_i\}_{i=1}^n$, where x is a continuous variable and y is a binary variable (i.e., only takes the value zero or one). Suppose our linear probability model is $y_i = \alpha + \beta x_i + u_i$, where $E(u_i|x_i) = 0$. With this information, in the figure below, perform the following:

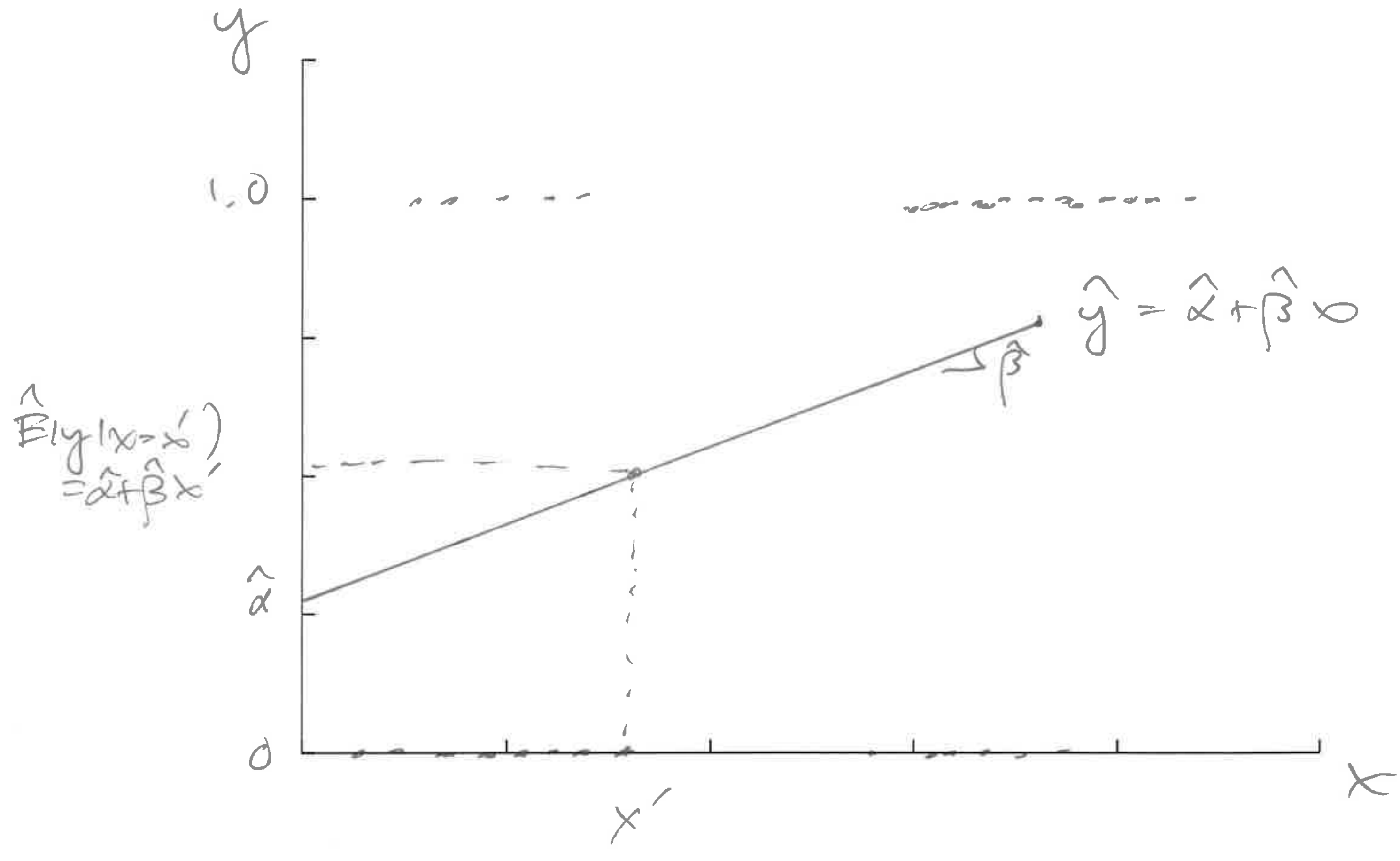
   (a) Label the axes.

   (b) Draw a feasible scatter plot of the data $\{x_i, y_i\}_{i=1}^n$.

   (c) Suppose α and β are positive, draw a feasible regression line $\widehat{E}(y|x)$.

   (d) Pick one value for x and plot its estimated conditional expectation. Interpret this value.

   (e) List at least one problem with the linear probability model that we discussed in class.



(d) $\hat{y}' = \hat{E}(y|x=x') = \hat{P}(y=1|x=x')$

predicted prob for $y=1$ given $x=x'$

(e) possible for $\hat{y} < 0$ or > 1

heteroskedastic variance

4. Consider the gretl output below relating birthweight (bwght) in ounces to the price of cigarettes (cigprice), the tax on cigarettes (cigtax), a binary variable for male child (1 if male, 0 if not) and a binary variable for white child (1 if white, 0 if not). Subsequent models include squares of cigprice (cigprice2) and cigtax (cigtax2) and interactions between cigprice and male (cigpricemale) and cigtax and male (cigtaxmale). With the output from these three models, answer the questions on the following page:

Model 1: OLS, using observations 1–1388
Dependent variable: bwght

	Coefficient	Std. Error	t-ratio	p-value
const	104.596	11.9531	8.751	0.0000
cigprice	0.0571711	0.109470	0.5223	0.6016
cigtax	0.0115523	0.144071	0.08018	0.9361
male	3.00315	1.08391	2.771	0.0057
white	6.18002	1.32597	4.661	0.0000

Mean dependent var	118.6996	S.D. dependent var	20.35396
Sum squared resid	561505.6	S.E. of regression	20.14958
R^2	0.022809	Adjusted R^2	0.019982
$F(4, 1383)$	8.070141	P-value(F)	1.97e–06
Log-likelihood	−6135.401	Akaike criterion	12280.80
Schwarz criterion	12306.98	Hannan–Quinn	12290.59

Model 2: OLS, using observations 1–1388
Dependent variable: bwght

	Coefficient	Std. Error	t-ratio	p-value
const	8.71877	101.369	0.08601	0.9315
cigprice	1.63762	1.60836	1.018	0.3088
cigprice2	−0.00613186	0.00620244	−0.9886	0.3230
cigtax	−0.677358	0.431822	−1.569	0.1170
cigtax2	0.0184011	0.0109005	1.688	0.0916
male	3.08015	1.08463	2.840	0.0046
white	6.15143	1.32597	4.639	0.0000

Mean dependent var	118.6996	S.D. dependent var	20.35396
Sum squared resid	560179.4	S.E. of regression	20.14034
R^2	0.025117	Adjusted R^2	0.020881
$F(6, 1381)$	5.929936	P-value(F)	3.98e–06
Log-likelihood	−6133.760	Akaike criterion	12281.52
Schwarz criterion	12318.17	Hannan–Quinn	12295.23

Model 3: OLS, using observations 1–1388
Dependent variable: bwght

	Coefficient	Std. Error	t-ratio	p-value
const	102.861	17.4217	5.904	0.0000
cigprice	0.0453542	0.160162	0.2832	0.7771
cigtax	0.178820	0.209514	0.8535	0.3935
male	7.51115	23.8671	0.3147	0.7530
white	6.25203	1.32492	4.719	0.0000
cigpricemale	0.0156813	0.219219	0.07153	0.9430
cigtaxmale	−0.335240	0.288032	−1.164	0.2447

Mean dependent var	118.6996	S.D. dependent var	20.35396
Sum squared resid	559390.3	S.E. of regression	20.12615
R^2	0.026490	Adjusted R^2	0.022260
$F(6, 1381)$	6.263023	P-value(F)	1.66e–06
Log-likelihood	−6132.781	Akaike criterion	12279.56
Schwarz criterion	12316.21	Hannan–Quinn	12293.27

(a) Using three model selection criteria, pick the preferred model.

(b) Test the null hypothesis that cigarette tax does not impact birthweight in model 1.

(c) Test the null hypothesis that birthweight is a linear function of cigarette price and cigarette tax.

(d) List the marginal effect of cigarette price in each model.

(e) Test the null hypothesis that the marginal effects of cigarette prices and cigarette taxes do not depend on whether or not the baby is male.

(a)

	model
R^2	3
$\bar{R}^2$	3
$\hat{\sigma}$	3
LL	3
SC	1
AIC	3

(b) $t = \frac{0.01155 - 0}{0.14407} = 0.08018 < 2$

$\Rightarrow$ fail to reject

(c) R: model 1 $H_0: \beta_3 = \beta_5 = 0$
U: model 2

$$F = \frac{(SSR_R - SSR_u)/2}{SSR_u/(1388-7)}$$

$$= \frac{(561505.6 - 560179.4)/2}{560179.4/(1381)}$$

$$= 1.63 < F_{2,1381}$$

$\Rightarrow$ fail to reject

(d) $\frac{\partial \hat{y}}{\partial \text{cigprice}}$

model 1

$= 0.05717$

model 2

$= 1.63762 - 2 \cdot 0.00613 \text{ cigprice}$

model 3

$= 0.04535 + 0.01568 \cdot \text{male}$

(e) R: model 1 $\quad H_0: \beta_6 = \beta_7 = 0$

u: model 3

$$F = \frac{(561505.6 - 559390.3)/2}{559390.3/(1388-7)}$$

$= 2.611$

$< F_{2,1381}$

$\Rightarrow$ fail to reject H_0