Student: _____	**Instructor:** Richeng Piao	**Assignment:** Practice Problem Set 10
Date: _____	**Course:** ECON 2560 - Applied Econometrics	

1. Consider a panel data set and the following regression model.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + u_{it}$$

What do subscripts *i* and *t* refer to?

- **A.** Subscripts *i* and *t* identify the entity and time period respectively.
- **B.** Subscripts *i* and *t* identify different individuals.
- **C.** Subscripts *i* and *t* identify different time periods.
- **D.** Subscripts *i* and *t* identify the time period and entity respectively.

Answer: A. Subscripts *i* and *t* identify the entity and time period respectively.

ID: Review Concept 10.1

2. A student aims to study the impact on atempted teen suicides (*AS*, number of attempted teen suicides in a state in a year), of the ease of access to mental health facilities (*MH*, number of operating mental health clinics in a state in a year). He obtains data from the 48 contiguous U.S. states for the years 2008 and 2016 and estimates the following regression equation for 2008:

$$\widehat{AS} = 40.52 - 0.35MH.$$
$$(2.98)\ (0.16)$$

Standard errors are given in parentheses. The student wants to find whether the effect of the ease of access to mental health clinics on attempted teen suicides is significant or not.

The test statistic associated with the 2008 study will be [ ].

(*Round your answer to two decimal places. Enter a minus sign if your answer is negative.*)

At the 5% significance level, the student (1) _____ the hypothesis that the effect of the ease of access to mental health clinics on attempted teen suicides is not significant.

Now, the student estimates the following regression equation from the data collected for 2016:

$$\widehat{AS} = 20.04 - 0.52MH.$$
$$(1.98)\ (0.21)$$

Standard errors are given in parentheses. The student wants to find whether the effect of the ease of access to mental health clinics on attempted teen suicides for the year 2016 is significant or not.

The test statistic associated with the 2016 study will be [ ].

(*Round your answer to two decimal places. Enter a minus sign if your answer is negative.*)

At the 5% significance level, the student (2) _____ the hypothesis that the effect of the ease of access to mental health clinics on attempted teen suicides is not significant.

Based on the given regression equations for 2008 and 2016, can we definitively conclude that the effect on attempted teen suicides of the ease of access to mental health clinics is accurately captured by the given regression equations?

- ○ **A.** Yes, because both the regression equations indicate a similar effect of the ease of access to mental health clinics.
- ○ **B.** Yes, because they suggest that the ease of access to mental health clinics inversely affects teen births.
- ○ **C.** No, because the extent to which this ease of access affects attempted teen suicides has increased from 2008 to 2016.
- ○ **D.** No, because these regression equations might have a substantial omitted variable bias.

(1) ○ fails to reject ○ rejects (2) ○ fails to reject ○ rejects

Answers – 2.19

(1) rejects

– 2.48

(2) rejects

D. No, because these regression equations might have a substantial omitted variable bias.

ID: Concept Exercise 10.1.1

3. Which of the following is an example of panel data?

- **A.** Data on the performance of Golden State Warriors, Cleveland Cavaliers, Chicago Bulls, New York Knicks, and Dallas Mavericks in the NBA playoffs for the year 2007.
- **B.** Data on the performance of Golden State Warriors, Cleveland Cavaliers, Chicago Bulls, New York Knicks, and Dallas Mavericks in the NBA playoffs for the years 2000 to 2015.
- **C.** Data on the performance of Golden State warriors in the NBA playoffs for the years 2000 to 2015.
- **D.** Data on the performance of Golden State Warriors, Cleveland Cavaliers and Chicago Bulls in the NBA playoffs for the year 2007, and of New York Knicks, and Dallas Mavericks in the NBA playoffs for the year 2010.

A panel in which the variables are studied for all '*n*' entities for all '*T*' time periods is (1) _____ panel.

When we have an omitted variable bias in a panel regression it is sometimes impossible to measure the omitted variables, although these omitted variables are sometimes constant across entities or through time.

Such omitted variable biases in panel regressions can be removed by using an OLS regression with (2) _____.

(1) ○ an unbalanced ○ a balanced (2) ○ fixed effects ○ data on omitted variables

Answers B.

Data on the performance of Golden State Warriors, Cleveland Cavaliers, Chicago Bulls, New York Knicks, and Dallas Mavericks in the NBA playoffs for the years 2000 to 2015.

(1) a balanced

(2) fixed effects

ID: Concept Exercise 10.1.2

---

4. Assume that for the $T = 2$ time periods case, you have estimated a simple regression in changes model and found a statistically significant positive intercept.

This implies:

- **A.** that the panel estimation approach is flawed since differencing the data eliminates the constant (intercept) in a regression.
- **B.** a negative mean change in the LHS variable in the absence of a change in the RHS variable since you subtract the earlier period from the later period.
- **C.** a positive mean change in the LHS variable in the absence of a change in the RHS variable.
- **D.** that the RHS variable changed between the two subperiods.

Answer: C. a positive mean change in the LHS variable in the absence of a change in the RHS variable.

ID: Test B Ex 10.2.1

---

5. A student wants to study how the wheat production (P, measured in million tons) has changed from 2000 to 2010. To eliminate the effect of the unobserved variables that are constant over time, the student specifies the regression in changes. He randomly selects 145 rural counties in the country and estimates the following regression:

$$\underset{(0.43)\ (0.95)}{\left(\widehat{P_{2010} - P_{2000}}\right) = 2.54 + 1.35\left(M_{2010} - M_{2000}\right)},$$

where M denotes the moisture content of the soil. Standard errors are given in parentheses.

The student wants to test whether or not the change in the moisture content of the soil from 2000 to 2010 had a significant impact on the change in the wheat production from 2000 to 2010.

Suppose β_0 and β_1 denote the intercept and slope coefficient on the change in M, respectively.

The test statistic associated with the test the student wants to conduct (H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$) is [ ].

(*Round your answer to two decimal places.*)

At the 5% significance level, the researcher will (1) _____ the null hypothesis.

Suppose the student realizes that he has not included data on agricultural policies, which remains constant over the 10 year period, but is different for different rural counties. This non-inclusion (2) _____ to omitted variable bias.

(1) ○ reject ○ fail to reject

(2) ○ leads ○ does not lead

Answers 1.42

(1) fail to reject

(2) does not lead

ID: Concept Exercise 10.2.1

6. A researcher is using a panel data set on n = 1000 workers over T = 10 years (from 2001 through 2010) that contains the workers' earnings, gender, education, and age. The researcher is interested in the effect of education on earnings. Determine whether each of the following is an example of unobserved person-specific or time-specific variables that are correlated with both education and earnings.

1. Unobserved ability. (1) _____.

2. Unemployment level. (2) _____.

3. Unobserved motivation. (3) _____.

4. Unobserved household environment. (4) _____.

5. GDP growth. (5) _____.

How would you control for these person-specific and time-specific effects in a panel data regression?

- **A.** Include period-specific and time-specific variables in the regression.
- **B.** Subscripts *i* and *t* identify different time periods.
- **C.** Subscripts *i* and *t* identify different individuals.
- **D.** Subscripts *i* and *t* identify the time period and entity respectively.

(1) person-specific
time-specific

(2) time-specific
person-specific

(3) person-specific
time-specific

(4) person-specific
time-specific

(5) time-specific
person-specific

Answers (1) person-specific

(2) time-specific

(3) person-specific

(4) person-specific

(5) time-specific

A. Include period-specific and time-specific variables in the regression.

ID: Review Concept 10.2

---

7. A researcher is using a panel data set on n = 1000 workers over T = 10 years (from 2001 through 2010) that contains the workers' earnings, gender, education, and age. The researcher is interested in the effect of education on earnings. Suppose you run a regression of earnings on person-specific and time-specific control variables.

Can this regression be used to estimate the effect of gender on an individual's earnings or the effect of the national unemployment rate on an individual's earnings?

- **A.** It can be used to estimate the effect of both gender and the national unemployment rate on an individual's earnings
- **B.** Neither effect can be estimated using this regression.
- **C.** It can be used to estimate the effect of gender on an individual's earnings, but not the effect of the national unemployment rate on an individual's earnings.
- **D.** It can be used to estimate the effect of the national unemployment rate on an individual's earnings, but not the effect of gender on an individual's earnings.

Answer: B. Neither effect can be estimated using this regression.

ID: Review Concept 10.3

---

8. A researcher is using a panel data set on n = 1000 workers over T = 10 years (from 2001 through 2010) that contains the workers' earnings, gender, education, and age. The researcher is interested in the effect of education on earnings. Suppose you run a regression of earnings on person-specific and time-specific control variables.

   Why might the regression error for a given individual be serially correlated?

   - ☐ **A.** An unexpected natural disaster occurs in a particular individual's city.
   - ☐ **B.** An unexpected earnings increase that is persistent through some part of the sample period.
   - ☐ **C.** An individual wins $100 from a scratch lottery ticket.
   - ☐ **D.** A random increase in a particular individual's ability.

   Answer: A. An unexpected natural disaster occurs in a particular individual's city., B. An unexpected earnings increase that is persistent through some part of the sample period.

   ID: Review Concept 10.4

---

9. The fixed effects regression model:

   - ○ **A.** in a log-log model may include logs of the binary variables, which control for the fixed effects.
   - ○ **B.** has n different intercepts.
   - ○ **C.** the slope coefficients are allowed to differ across entities, but the intercept is "fixed" (remains unchanged).
   - ○ **D.** has "fixed" (repaired) the effect of heteroskedasticity.

   Answer: B. has n different intercepts.

   ID: Test A Ex 10.3.1

---

10. In the panel regression analysis of beer taxes on traffic deaths, the estimation period is 1982–1988 for the 48 contiguous U.S. states.

    To test for the significance of entity fixed effects, you should calculate the F-statistic and compare it to the critical value from your $F_{q,\,\infty}$ distribution, where q equals:

    - ○ **A.** 7.
    - ○ **B.** 54.
    - ○ **C.** 47.
    - ○ **D.** 48.

    Answer: C. 47.

    ID: Test B Ex 10.3.2

---

11. Which of the following are the features of an entity fixed effects regression model? (*Check all that apply*.)

- ☐ **A.** It is a method used to control omitted variables in panel data when they vary across entities as well as time.
- ☐ **B.** It has *n* different slope coefficients, one for each entity, but the intercept is same for each entity.
- ☐ **C.** It has *n* different intercepts, one for each entity, but the slope coefficient is same for each entity.
- ☐ **D.** It is a method used to control omitted variables in panel data when they vary across entities but not time.

Consider the following entity fixed effects regression model:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + \gamma_n D_{ni} + u_{it},$$

where β_0 is the intercept and $\beta_1, \gamma_2, \ldots, \gamma_n$ are the unknown coefficients to be estimated and $D_{2i} = 1$ when $i = 2$ and $D_{2i} = 0$ otherwise, and so forth.

In this fixed effect regression using binary variables, we cannot include *n* binary variables plus a common (1) _____ as it will lead to (2) _____.

(1)
- ○ slope
- ○ intercept

(2)
- ○ error-in-variable bias
- ○ perfect multicollinearity
- ○ functional form misspecification

Answers C. It has *n* different intercepts, one for each entity, but the slope coefficient is same for each entity., D. It is a method used to control omitted variables in panel data when they vary across entities but not time.

(1) intercept

(2) perfect multicollinearity

ID: Concept Exercise 10.3.1

---

12. In the time fixed effects regression model, you should exclude one of the binary variables for the time periods when an intercept is present in the equation:

- ○ **A.** to avoid perfect multicollinearity.
- ○ **B.** to allow for some changes between time periods to take place.
- ○ **C.** because there are already too many coefficients to estimate.
- ○ **D.** because the first time period must always be excluded from your data set.

Answer: A. to avoid perfect multicollinearity.

ID: Test B Ex 10.4.3

---

13. Which of the following statements most accurately describes a time fixed effects regression model?

- **A.** A regression model controlling for omitted variables that are constant across entities but evolve over time.
- **B.** A regression model controlling for omitted variables that are constant across entities and do not evolve over time.
- **C.** A regression model controlling for omitted variables that vary across entities and evolve over time.
- **D.** A regression model controlling for omitted variables that vary across entities but do not evolve over time.

Which of the following equations correctly describes a time fixed effects regression model using binary indicators?

- **A.** $Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_1 B1_i + \delta_2 B2_i + \delta_3 B3_i + ... + \delta_n Bn_i + u_i$.
- **B.** $Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \delta_3 B3_t + ... + \delta_T BT_t + u_{it}$.
- **C.** $Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_1 B1_t + \delta_2 B2_t + \delta_3 B3_t + ... + \delta_T BT_t + u_{it}$.
- **D.** $Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_i + \delta_3 B3_i + ... + \delta_n Bn_i + u_i$.

where δ_2,..., δ_T are unknown coefficients and where $B2_t = 1$ if $t = 2$ and $B2_t = 0$ otherwise, and so forth.

A researcher studying the effect on the number of farmers producing peanuts across 48 contiguous US. states, of expected peanut prices from 2000 to 2017, estimates a regression equation of the following form for the year 2005 for Nebraska:

$$\widehat{FP} = \hat{\beta}_0 + \hat{\beta}_1 PP + \hat{\beta}_2 EF + \hat{\beta}_3 PLC,$$

where $\widehat{FP}$ is the estimated number of farmers producing peanuts, *PP* is the average expected price of peanuts, *EF* is the extent of farming undertaken in a state, and *PLC* is the price loss coverage payout given by the government to farmers. This payout is the difference between the prevailing national market price of peanuts in a given year, and the price of peanuts set by the congress. The regression equation is of the given form for all the years from 2000 to 2017 and for all the 48 states. Since the equation is specific to Nebraska in 2005, the time and entity subscripts are not needed.

Assume the extent of farming undertaken in each of the 48 states studied, does not change throughout the course of the study, and any change in the price of peanuts set by the congress would apply to all the 48 states.

Therefore, the impact of a time fixed effect on the number of farmers producing peanuts in a particular year in a particular state is

captured by the estimate (1) _____, controlling for (2) _____.

Therefore, the impact of an entity fixed effect on the number of farmers producing peanuts in a particular year in a particular state is

captured by the estimate (3) _____, controlling for (4) _____.

(1) $\hat{\beta}_1$ $\hat{\beta}_2$ $\hat{\beta}_3$

(2) *PLC* *PP* *EF*

(3) $\hat{\beta}_1$ $\hat{\beta}_2$ $\hat{\beta}_3$

(4) *PLC* *PP* *EF*

Answers A. A regression model controlling for omitted variables that are constant across entities but evolve over time.

B. $Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \delta_3 B3_t + ... + \delta_T BT_t + u_{it}$.

(1) $\hat{\beta}_3$

(2) *PLC*

(3) $\hat{\beta}_2$

(4) *EF*

ID: Concept Exercise 10.4.1

14. A researcher studying the effect on the number of farmers producing peanuts (*FP*) in state *i* in year *t,* across the 48 contiguous states in the U.S., of the average expected price of peanuts (*PP,* measured in dollars) in state *i* in year *t,* from 2000 to 2017, estimates the following regression equation based on all 17 years of data for all the 48 states:

$$\widehat{FP} = 0.98 + 2.12PP.$$
$$(0.23)\ (0.69)$$

Standard errors are given in parentheses. The researcher wants to find whether the effect of the expected peanut prices on the number of farmers producing peanuts is significant or not.

The test statistic associated with this study will be [ ].

(*Round your answer to two decimal places.*)

At the 5% significance level, the researcher (1) _____ the hypothesis that the effect of the average expected price of peanuts on the number of farmers producing peanuts is not significant.

After some deliberation, the researcher identifies that his estimated regression equation suffers from an omitted variable bias. The omitted variables being subsidies given to farmers by the government, price of peanuts set by the Congress, international demand for peanuts, etc. These variables affect the number of farmers producing peanuts and are correlated with the average expected price of peanuts.

The researcher identifies these omitted variables as having time fixed effects and remodels his regression equation as follows:

$$\widehat{FP} = 0.98 + 0.97PP + S,$$
$$(0.23)\ (0.83)$$

where *S* includes all the omitted variables that have time fixed effects that affect *FP* and are correlated with *PP.* Standard errors are given in parentheses. The coefficients on the time binary variables are not reported because they are not of primary interest.

After having incorporated time fixed effects, the researcher retests whether the effect of the average expected price of peanuts on the number of farmers producing peanuts is significant or not.

The test statistic associated with this study will be [ ].

(*Round your answer to two decimal places.*)

At the 5% significance level, the researcher (2) _____ the hypothesis that the effect of the average expected price of peanuts on the number of farmers producing peanuts is not significant.

A further analysis of the factors affecting the number of farmers producing peanuts, leads the researcher to identify another set of omitted variables such as the average farm size in different states, the average standard of living of farmers in different states, the opportunity cost of farming in different states, etc., These variables affect the number of farmers producing peanuts and are correlated with the average expected price of peanuts.

He identifies these omitted variables as having entity fixed effects and remodels his regression equation as follows:

$$\widehat{FP} = 0.98 + 1.75PP + Z + S,$$
$$(0.23)\ (0.98)$$

where *Z* includes all the omitted variables that have entity fixed effects that affect *FP* and are correlated with *PP*. Standard errors are given in parentheses. The coefficients on the time and state binary variables are not reported because they are not of primary interest.

After having incorporated time and entity fixed effects, the researcher retests whether the effect of the average expected price of peanuts on the number of farmers producing peanuts is significant or not.

The test statistic associated with this study will be [ ].

(*Round your answer to two decimal places.*)

At the 5% significance level, the researcher (3) _____ the hypothesis that the effect of the average expected price of peanuts on the number of farmers producing peanuts is not significant.

(1) ○ fails to reject ○ rejects

(2) ○ fails to reject ○ rejects

(3) ○ fails to reject ○ rejects

Answers 3.07

(1) rejects

1.17

(2) fails to reject

1.79

(3) fails to reject

ID: Concept Exercise 10.4.2

---

15. Consider the following binary variable version of the fixed effects model. Each regressor *Dj* is a binary variable that equals 1 when *i* = *j* and 0 otherwise.

$$Y_{it} = \beta_0 X_{0,\,it} + \beta_1 X_{it} + \gamma_1 D1_i + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + u_{it}$$

Suppose that $n = 3$ and that $X_{0,\,it} = 1$ for all i, t (the intercept, β_0, is the coefficient for the "constant" regressor). Show that the binary regressors and the "constant" regressor are perfectly multicollinear; that is, express one of the variables $D1_i$, $D2_i$, $D3_i$, and $X_{0,\,it}$ as a perfect linear function of the others.

When $n = 3$, the model reduces to

$$Y_{it} = \beta_0 X_{0,\,it} + \beta_1 X_{it} + \gamma_1 D1_i + \gamma_2 D2_i + \gamma_3 D3_i + u_{it}$$

So,

$$D1_i + D2_i + D3_i = \square$$

Since $X_{0,\,it} = 1$, then $X_{0,\,it}$ can be expressed as a lineal combination of $D1_i$, $D2_i$, and $D3_i$. Namely,

$$D1_i + D2_i + D3_i = X_{0,\,it} = 1$$

Show that this results hold for the general case (*n* binary regressors).

$$Y_{it} = \beta_0 X_{0,\,it} + \beta_1 X_{it} + \gamma_1 D1_i + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + u_{it}$$

So,

$$D1_i + D2_i + \ldots + Dn_i = \square$$

Similarly, since $X_{0,\,it} = 1$, $X_{0,\,it}$ can be expressed as a linear combination of $D1_i$, $D2_i$, ..., Dn_i. Namely,

$$D1_i + D2_i + \ldots + Dn_i = X_{0,\,it} = 1$$

What happens if you try to estimate the coefficients of the regression by OLS?

- **A.** The inclusion of all binary regressors causes the OLS estimators to be biased.
- **B.** The inclusion of all binary regressors is likely to produce OLS estimators that are highly significant.
- **C.** The inclusion of all binary regressors causes the OLS estimators to be inconsistent.
- **D.** The inclusion of all binary regressors causes perfect multicollinearity, so the OLS estimators cannot be computed.

Answers 1

1

D. The inclusion of all binary regressors causes perfect multicollinearity, so the OLS estimators cannot be computed.

ID: Exercise 10.2

16. Consider the following binary variable version of the fixed effects model. Each regressor *Dj* is a binary variable that equals 1 when $i = j$ and 0 otherwise. Note that the binary variable $D1_i$ for the first group is arbitrarily omitted.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \ldots + \gamma_n Dn_i + u_{it}$$

Use the regression in the equation above and the tool palette to the right to answer the following questions.

What is the slope and intercept for entity 3 in time period 3?

The slope of entity 3 in time period 3 is [ ].
The intercept of entity 3 in time period 3 is [ ].

(Properly format your expressions using the tools in the palette. Hover over tools to see keyboard shortcuts. E.g., a subscript can be created with the _ character.)

What is the slope and intercept for entity 1 in time period 3?

The slope of entity 1 in time period 3 is [ ].
The intercept of entity 1 in time period 3 is [ ].

(Properly format your expressions using the tools in the palette.)

What is the slope and intercept for entity 2 in time period 3?

The slope of entity 2 in time period 3 is [ ].
The intercept of entity 2 in time period 3 is [ ].

(Properly format your expressions using the tools in the palette.)

Answers β_1

$\beta_0 + \gamma_3$

β_1

β_0

β_1

$\beta_0 + \gamma_2$

ID: Exercise 10.4

17. Do the fixed effects regression assumptions here imply that $\text{cov}\left(\hat{v}_{it}, \hat{v}_{is}\right) = 0$ for $t \neq s$ in the equation below?

Let $\hat{v}_{it} = \hat{X}_{it}\hat{u}_{it}$ and $\eta_i = \sqrt{\frac{1}{T}} \sum_{t=1}^{T} \hat{X}_{it}\hat{u}_{it}$. Then,

$$\begin{aligned}\text{var}(\eta_i) &= \text{var}\left(\sqrt{\frac{1}{T}} \sum_{t=1}^{T} \hat{v}_{it}\right) = \frac{1}{T}\text{var}\left(\hat{v}_{i1} + \hat{v}_{i2} + ... + \hat{v}_{iT}\right) \\ &= \frac{1}{T}\left[\text{var}\left(\hat{v}_{i1}\right) + \text{var}\left(\hat{v}_{i2}\right) + ... + \text{var}\left(\hat{v}_{iT}\right)\right. \\ &\quad \left. + 2\text{cov}\left(\hat{v}_{i1}, \hat{v}_{i2}\right) + ... + 2\text{cov}\left(\hat{v}_{iT-1}, \hat{v}_{iT}\right)\right]\end{aligned}$$

- **A.** No, because $\left(X_{i1}, X_{i2},..., X_{iT}, u_{i1}, u_{i2},..., u_{iT}\right)$, $i = 1,..., n$ are i.i.d. draws from their joint distribution.
- **B.** Yes, because the fixed effects regression assumptions do not allow for within-individual autocorrelation.
- **C.** No, because the fixed effects regression assumptions allow for within-individual autocorrelation.
- **D.** Yes, because $\left(X_{i1}, X_{i2},..., X_{iT}, u_{i1}, u_{i2},..., u_{iT}\right)$, $i = 1,..., n$ are i.i.d. draws from their joint distribution.

Answer: C. No, because the fixed effects regression assumptions allow for within-individual autocorrelation.

ID: Exercise 10.6

---

18. In the panel regression analysis of beer taxes on traffic deaths, the estimation period is 1982–1988 for the 48 contiguous U.S. states.

To test for the significance of time fixed effects, you should calculate the *F*-statistic and compare it to the critical value from your $F_{q,\,\infty}$ distribution, which equals (at the 5% level):

- **A.** 2.80.
- **B.** 2.64.
- **C.** 2.10.
- **D.** 2.01.

Answer: C. 2.10.

ID: Test A Ex 10.5.2

---

19. HAC standard errors and clustered standard errors are related as follows:

- **A.** they are the same.
- **B.** clustered standard errors are one type of HAC standard error.
- **C.** they are the same if the data are differenced.
- **D.** clustered standard errors are the square root of HAC standard errors.

Answer: B. clustered standard errors are one type of HAC standard error.

ID: Test A Ex 10.5.3

---

20. It is advisable to use clustered standard errors in panel regressions because:

- **A.** they are easier to calculate than homoskedasticity-only standard errors.
- **B.** the fixed effects estimator is asymptotically normally distributed when n is large.
- **C.** hypothesis testing can proceed in a standard way even if there are few entities (n is small).
- **D.** without clustered standard errors, the OLS estimator is biased.

Answer: B. the fixed effects estimator is asymptotically normally distributed when n is large.

ID: Test A Ex 10.5.4

---

21. In panel data, the regression error:

- **A.** should be calculated taking into account heteroskedasticity but not autocorrelation.
- **B.** only exists for the case of $T > 2$.
- **C.** is likely to be correlated over time within an entity.
- **D.** fits all of the three descriptions above.

Answer: C. is likely to be correlated over time within an entity.

ID: Text B Ex 10.5.4

---

22. If X_{it} is correlated with X_{is} for different values of s and t, then:

- **A.** X_{it} is said to be autocorrelated.
- **B.** the OLS estimator cannot be computed.
- **C.** statistical inference cannot proceed in a standard way even if clustered standard errors are used.
- **D.** this is not of practical importance since these correlations are typically weak in applications.

Answer: A. X_{it} is said to be autocorrelated.

ID: Test B Ex 10.5.5

---

23. Suppose the population regression is of the form:

$$Y_{it} = \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + \alpha_i + u_{it},\ i = 1,..,\ n,\ t = 1,..,\ T,$$

where $X_{1,it}$ is the value of the first regressor for entity i in time period t, $X_{2,it}$ is the value of the second regressor, and so forth; and $\alpha_i,..,\alpha_n$ are entity-specific intercepts.

Which of the following statements is true in describing the fixed effects regression assumptions?

- **A.** The variables for one entity are distributed identically to, but independently of, the variables for another entity; that is, the variables are i.i.d. across entities for $i = 1,..,\ n$.
- **B.** The variables for one entity are distributed identically to, but independently of, the variables for another entity and also within an entity; that is, the variables are i.i.d. across and within entities for $i = 1,..,\ n$, $t = 1,..,\ T$.
- **C.** The variables for one entity are distributed identically to, but independently of, the variables within an entity; that is, the variables are i.i.d. within entities for $t = 1,..,\ T$.
- **D.** The error term has a zero conditional mean across entities but a non-zero conditional mean within entities.

Which of the following statements is true in describing clustered standard errors?

- **A.** Clustered standard errors allow for heteroskedasticity and arbitrary autocorrelation within an entity, but treat the errors as uncorrelated across entities.
- **B.** Clustered standard errors allow for heteroskedasticity within an entity, but treat the errors as uncorrelated within and across entities.
- **C.** Clustered standard errors allow for arbitrary autocorrelation within an entity, but treat the errors as homoskedastic and uncorrelated across entites.
- **D.** Clustered standard errors allow for arbitrary autocorrelation within and across entities, but treat the errors as homoskedastic within and across entities.

Answers A.

The variables for one entity are distributed identically to, but independently of, the variables for another entity; that is, the variables are i.i.d. across entities for $i = 1,..,\ n$.

A.

Clustered standard errors allow for heteroskedasticity and arbitrary autocorrelation within an entity, but treat the errors as uncorrelated across entities.

ID: Concept Exercise 10.5.1

24. This exercise refers to the drunk driving panel data regression summarized below.

Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent variable: traffic fatility rate (deaths per 10,000).

Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Beer tax	0.41* (0.088)	− 0.78** (0.24)	− 0.77*** (0.36)	− 0.56 (0.35)	− 0.71** (0.32)	− 0.46 (0.33)	− 0.93* (0.32)
Drinking age 18				0.034 (0.074)	− 0.014 (0.084)		0.038 (0.108)
Drinking age 19				− 0.024 (0.057)	− 0.079 (0.066)		− 0.069 (0.103)
Drinking age 20				0.032 (0.054)	− 0.104*** (0.055)		− 0.118 (0.133)
Drinking age						− 0.003 (0.022)	
Mandatory jail or community service?				0.044 (0.101)	0.077 (0.113)	0.039 (0.109)	0.098 (0.165)
Average vehicle miles per driver				0.006 (0.005)	0.013 (0.012)	0.008 (0.008)	0.123 (0.053)
Unemployment rate				− 0.069* (0.015)		− 0.065* (0.013)	− 0.093* (0.023)
Real income per capita (logarithm)				1.78* (0.68)		1.54* (0.62)	1.04 (0.67)
Years	1982-88	1982-88	1982-88	1982-88	1982-88	1982-88	1982 & 1988 only
State Effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes
***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**							
Time effects = 0			4.97 (0.003)	10.62 (<0.001)	3.37 (0.004)	10.17 (<0.001)	37.06 (<0.001)
Drinking age coefficients = 0				0.38 (0.746)	1.52 (0.226)		0.45 (0.795)
Unemployment rate, income per capita = 0				29.38 (<0.001)		31.58 (<0.001)	25.16 (<0.001)
$\bar{R}^2$	0.097	0.818	0.894	0.995	0.841	0.931	0.857

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the ***10%, **5%, or *1% significance level.

New Jersey has a population of 6.5 million people. Suppose that New Jersey increased the tax on a case of beer by $1 (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year.

The predicted number of lives that would be saved over the next year is [ ]

(*Round your response to two decimal places*)

Construct a 95% confidence interval for your answer.

The 95% confidence interval for the number of lives that would be saved over the next year is [,]

(*Round your response to two decimal places*)

The drinking age in New Jersey is 21. Suppose that New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year.

The predicted (1) _____ in the number of traffic fatalities in the next year is [ ]

(*Round your response to two decimal places*)

Construct a 90% confidence interval for your answer.

The 90% confidence interval for the predicted increase in the number of traffic fatalities in the next year is

[,]

(*Round your response to two decimal places*)

Suppose that real income per capita in New Jersey increases by 1% in the next year. Use the results in column (4) to predict the change in number of traffic fatalities in the next year.

The predicted (2) _____ in the number of traffic fatalities in the next year is

(*Round your response to two decimal places*)

Construct a 90% confidence interval for your answer.

The 90% confidence interval for the predicted increase in the number of traffic fatalities in the next year is

[,]

(*Round your response to two decimal places*)

Refer to the reported *F*-Statistics and *p*-values associated with testing for exclusion of group of variables. Should time effects be included ir the regression?

- **A.** Yes
- **B.** No.

A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis?

- **A.** I would include a binary variable *west* (=1 if the state is in the west and 0 otherwise). Then, I would test if the estimated coefficient for the binary variable is significant at a reasonable level.
- **B.** I would include a binary variable *west* (=1 if the state is in the west and 0 otherwise), and an interaction term *west* Unemployment rate* . Then, I would test if the estimated coefficient for the interaction term is significant at a reasonable level.
- **C.** I would compare the average number of fatalities in western states to the average number of fatalities in non-western states.
- **D.** I would include a binary variable *west* (=1 if the state is in the west and 0 otherwise), and an interaction term *west* Unemployment rate*. Then, I would test if the estimated coefficient for the binary variable *west* is significant at a reasonable level.

(1) increase / decrease (2) increase / decrease

Answers 364.00

– 81.90

809.90

(1) increase

22.10

– 57.02

101.22

(2) increase

11.57

4.30

18.84

A. Yes

B.
I would include a binary variable *west* (=1 if the state is in the west and 0 otherwise), and an interaction term *west*Unemployme*
rate . Then, I would test if the estimated coefficient for the interaction term is significant at a reasonable level.

ID: Exercise 10.1

25. This exercise refers to the drunk driving panel data regression summarized below.

Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent variable: traffic fatility rate (deaths per 10,000).

Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Beer tax	0.44* (0.011)	− 0.79** (0.32)	− 0.78*** (0.38)	− 0.54 (0.33)	− 0.71** (0.35)	− 0.46 (0.33)	− 0.96* (0.37)
Drinking age 18				0.029 (0.076)	− 0.015 (0.084)		0.043 (0.107)
Drinking age 19				− 0.018 (0.054)	− 0.081 (0.071)		− 0.061 (0.101)
Drinking age 20				0.032 (0.056)	− 0.109*** (0.051)		− 0.115 (0.128)
Drinking age						− 0.003 (0.021)	
Mandatory jail or community service?				0.032 (0.105)	0.082 (0.113)	0.044 (0.104)	0.097 (0.165)
Average vehicle miles per driver				0.008 (0.006)	0.014 (0.012)	0.007 (0.005)	0.123 (0.047)
Unemployment rate				− 0.068* (0.011)		− 0.065* (0.013)	− 0.091* (0.025)
Real income per capita (logarithm)				1.61* (0.62)		1.83* (0.62)	1.04 (0.65)
Years	1982-88	1982-88	1982-88	1982-88	1982-88	1982-88	1982 & 1988 only
State Effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes
***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**							
Time effects = 0			4.23 (0.001)	10.59 (<0.001)	3.44 (0.005)	10.83 (<0.001)	37.89 (<0.001)
Drinking age coefficients = 0				0.52 (0.777)	1.42 (0.244)		0.47 (0.769)
Unemployment rate, income per capita = 0				29.82 (<0.001)		31.91 (<0.001)	25.26 (<0.001)
$\bar{R}^2$	0.092	0.828	0.816	0.982	0.856	0.983	0.888

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the ***10%, **5%, or *1% significance level.

A researcher believes that traffic fatalities increase when roads are ice and so states with more snow will have more fatalities than other states. Comment on the following methods designed to estimate the effect of snow on fatalities.

The researcher collects data on the average snowfall for each state and adds this regressor ($AverageSnow_i$) to the regression given in the table above. Suppose snowfall does not vary over the time period. Should $AverageSnow_i$ be included in the regression?

- **A.** Yes, $AverageSnow_i$ will likely explain a significant part of traffic fatalities.
- **B.** No, because $AverageSnow_i$ is controlled for by time effects.
- **C.** No, because $AverageSnow_i$ will be perfectly collinear with the state fixed effects.
- **D.** Yes, because $AverageSnow_i$ is not controlled for by time effects.

The researcher collects data on the snowfall in each state for each year in the sample ($Snow_{it}$) and adds this regressor to the regression given in the table above. Should $Snow_{it}$ be included in the regression?

- **A.** Yes, $Snow_{it}$ varies with time, and so this method can be used along with state fixed effects.

- **B.** No, $Snow_{it}$ varies with time, and so this method cannot be used along with year effects.
- **C.** No, $Snow_{it}$ varies with time, and so this method can be used along with year effects but not state fixed effects.
- **D.** No, $Snow_{it}$ varies with time, and so this method cannot be used along with state fixed effects.

Answers C. No, because $AverageSnow_i$ will be perfectly collinear with the state fixed effects.

A. Yes, $Snow_{it}$ varies with time, and so this method can be used along with state fixed effects.

ID: Exercise 10.7

---

26. When you add state fixed effects to a simple regression model for U.S. states over a certain time period, and the regression R^2 increases significantly, then it is safe to assume that:

- **A.** the coefficients on the other included explanatory variables will not change.
- **B.** time fixed effects are unimportant.
- **C.** the included explanatory variables, other than the state fixed effects, are unimportant.
- **D.** state fixed effects account for a large amount of the variation in the data.

Answer: D. state fixed effects account for a large amount of the variation in the data.

ID: Test A Ex 10.6.5

---

27. A researcher is interested in studying the factors that affected the monthly house rent (*P*, in hundred dollars) across City A for the time period 2014-2018. For her study, she selects the size of the house (*SZ*, measured in sq ft), the distance of the house from downtown (*D*, measured in kilometres), the average real income of households within a one km radius of the house (*I*, expressed in logarithmic terms), the inflation rate in the economy (π, measured in percentage) as regressors. She collects data across 100 randomly selected localities in the city for this time period. Here, the average real income of households within a one km radius of the house (*I*), and the inflation rate in the economy (π) are the economic variables.

The following table shows the regression results.

Dependent variable: Monthly house rent (in hundred dollars)					
Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
***Size of the house* (*SZ*)**	2.17** (0.75)	1.67* (0.84)	1.62^{+} (0.82)	1.51 (0.82)	1.62* (0.91)
***Distance of the house from downtown* (*D*)**				-2.09 (0.84)	-1.98 (0.76)
***Logarithm of average real income of households within a one km radius* (*I*)**				2.12** (0.98)	
***Inflation rate in the economy* (π)**				1.01** (0.79)	
Years	2014-2018	2014-2018	2014-2018	2014-2018	2014-2018
State effects?	no	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes
n	100	100	100	100	100
***F*-Statistics and *p*-values on Joint Hypotheses**					
Time effects = 0			4.22 (0.002)	10.12 (< 0.001)	3.48 (0.006)
***Real income of households within a one km radius, inflation rate in the economy* = 0**				25.20 (< 0.001)	
$\bar{R}^2$	0.091	0.789	0.791	0.799	0.788
Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under *F*-statistics. Individual coefficients are statistically significant at the $^{+}$10%, *5%, or **1% significance level.					

Let β_1 and $\beta_1 \Delta SZ$ denote the slope coefficient on the size of the house and the predicted change in the house rent associated with a small change in the size of the house (ΔSZ), respectively. Suppose the researcher wants to see the effect of an increase in size of house by 9 sq ft on house rent.

According to the regression in column (1), the 95% confidence interval for $\beta_1 \Delta SZ$ will be (,).

(*Round your answer to two decimal places.*)

Suppose that the researcher wants to test the hypothesis $\beta_1 \Delta SZ = 0$ using the calculated confidence interval. As the hypothesized value of $\beta_1 \Delta SZ$ (1) _____ in the calculated confidence interval for $\beta_1 \Delta SZ$, she would (2) _____ the hypothesis $\beta_1 \Delta SZ = 0$.

According to the regression in column (4), the 95% confidence interval for $\beta_1 \Delta SZ$ will be (,).

(*Round your answer to two decimal places. Enter a minus sign if your answer is negative.*)

The hypothesized value of $\beta_1 \Delta SZ$, 0, (3) _____ in the calculated confidence interval for $\beta_1 \Delta SZ$. So, the researcher would (4) _____ the hypothesis $\beta_1 \Delta SZ = 0$.

Which of the following statements are true regarding the specifications given in the above table? (*Check all that apply.*)

☐ **A.** The high value of $\bar{R}^2$ in column (3) implies that the time fixed effects account for a large amount of the variation in the data.

☐ **B.** The sensitivity of the estimated coefficient on *SZ* to including the economic variables, combined with the statistical significance of the coefficients on those variables in column (4), indicates that they should remain in the base specification.

☐ **C.** A $1 increase in the average real income of households within a 1 km radius of the house is associated with an increase in the price of the house by $2.12.

☐ **D.** The decrease in the value of the slope coefficient on *SZ* from column (1) to column (2) after including state effects was due to the elimination of the effect of omitted variables that differ across entities but are constant over time.

(1) ○ lies ○ does not lie (2) ○ fail to reject ○ reject (3) ○ lies ○ does not lie (4) ○ reject ○ fail to reject

Answers 6.30

32.76

(1) does not lie

(2) reject

– 0.87

28.05

(3) lies

(4) fail to reject

B.
The sensitivity of the estimated coefficient on *SZ* to including the economic variables, combined with the statistical significance o the coefficients on those variables in column (4), indicates that they should remain in the base specification.
, D.
The decrease in the value of the slope coefficient on *SZ* from column (1) to column (2) after including state effects was due to th elimination of the effect of omitted variables that differ across entities but are constant over time.

ID: Concept Exercise 10.6.1