previous year using fixed effects regression. Will this regression give reliable estimates of the effects of the regressors (age, education, union status, and previous year's earnings) on earnings? Explain. (*Hint:* Check the fixed effects regression assumptions in Section 10.5.)

Empirical Exercises

E10.1 Some U.S. states have enacted laws that allow citizens to carry concealed weapons. These laws are known as "shall-issue" laws because they instruct local authorities to issue a concealed weapons permit to all applicants who are citizens, are mentally competent, and have not been convicted of a felony (some states have some additional restrictions). Proponents argue that, if more people carry concealed weapons, crime will decline because criminals are deterred from attacking other people. Opponents argue that crime will increase because of accidental or spontaneous use of the weapon. In this exercise, you will analyze the effect of concealed weapons laws on violent crimes. On the textbook Web site **www.aw-bc.com/stock_watson** you will find a data file **Guns** that contains a balanced panel of data from 50 U.S. states, plus the District of Columbia, for the years 1977–1999.[3] A detailed description is given in **Guns_Description**, available on the Web site.

a. Estimate (1) a regression of ln(*vio*) against *shall* and (2) a regression of ln(*vio*) against *shall, incarc_rate, density, avginc, pop, pb1064, pw1064,* and *pm1029*.

i. Interpret the coefficient on *shall* in regression (2). Is this estimate large or small in a "real-world" sense?

ii. Does adding the control variables in regression (2) change the estimated effect of a shall-carry law in regression (1), as measured by statistical significance? As measured by the "real-world" significance of the estimated coefficient?

iii. Suggest a variable that varies across states but plausibly varies little—or not at all—over time, and that could cause omitted variable bias in regression (2).

b. Do the results change when you add fixed state effects? If so, which set of regression results is more credible, and why?

[3]These data were provided by Professor John Donohue of Stanford University and were used in his paper with Ian Ayres, "Shooting Down the 'More Guns Less Crime' Hypothesis," *Stanford Law Review*, 2003; 55: 1193–1312.

c. Do the results change when you add fixed time effects? If so, which set of regression results is more credible, and why?

d. Repeat the analysis using ln(*rob*) and ln(*mur*) in place of ln(*vio*).

e. In your view, what are the most important remaining threats to the internal validity of this regression analysis?

f. Based on your analysis, what conclusions would you draw about the effects of concealed-weapon laws on these crime rates?

E10.2 Traffic crashes are the leading cause of death for Americans between the ages of 5 and 32. Through various spending policies, the federal government has encouraged states to institute mandatory seat belt laws to reduce the number of fatalities and serious injuries. In this exercise you will investigate how effective these laws are in increasing seat belt use and reducing fatalities. On the textbook Web site **www.aw-bc.com/stock_watson** you will find a data file **Seatbelts** that contains a panel of data from 50 U.S. states, plus the District of Columbia, for the years 1983–1997.[4] A detailed description is given in **Seatbelts_Description**, available on the Web site.

a. Estimate the effect of seat belt use on fatalities by regressing *FatalityRate* on *sb_useage*, *speed65*, *speed70*, *ba08*, *drinkage21*, ln(*income*), and *age*. Does the estimated regression suggest that increased seat belt use reduces fatalities?

b. Do the results change when you add state fixed effects? Provide an intuitive explanation for why the results changed.

c. Do the results change when you add time fixed effects plus state fixed effects?

d. Which regression specification—(a), (b), or (c)—is most reliable? Explain why.

e. Using the results in (c), discuss the size of the coefficient on *sb_useage*. Is it large? Small? How many lives would be saved if seat belt use increased from 52% to 90%?

f. There are two ways that mandatory seat-belt laws are enforced: "Primary" enforcement means that a police officer can stop a car and ticket the driver if the officer observes an occupant not wearing a seat belt; "secondary" enforcement means that a police officer can write a ticket if an occupant is not wearing a selt belt, but must have another

[4]These data were provided by Professor Liran Einav of Stanford University and were used in his paper with Alma Cohen, "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," *The Review of Economics and Statistics* 2003; 85(4): 828–843.

reason to stop the car. In the data set, *primary* is a binary variable for primary enforcement and *secondary* is a binary variable for secondary enforcement. Run a regression of *sb_useage* on *primary*, *secondary*, *speed65*, *speed70*, *ba08*, *drinkage21*, ln(*income*), and *age*, including fixed state and time effects in the regression. Does primary enforcement lead to more seat belt use? What about secondary enforcement?

g. In 2000, New Jersey changed from secondary enforcement to primary enforcement. Estimate the number of lives saved per year by making this change.

APPENDIX 10.1 The State Traffic Fatality Data Set

The data are for the "lower 48" U.S. states (excluding Alaska and Hawaii), annually for 1982 through 1988. The traffic fatality rate is the number of traffic deaths in a given state in a given year, per 10,000 people living in that state in that year. Traffic fatality data were obtained from the U.S. Department of Transportation Fatal Accident Reporting System. The beer tax is the tax on a case of beer, which is a measure of state alcohol taxes more generally. The drinking age variables in Table 10.1 are binary variables indicating whether the legal drinking age is 18, 19, or 20. The two binary punishment variables in Table 10.1 describe the state's minimum sentencing requirements for an initial drunk driving conviction: "Mandatory jail?" equals 1 if the state requires jail time and equals 0 otherwise, and "Mandatory community service?" equals 1 if the state requires community service and equals 0 otherwise. Data on the total vehicle miles traveled annually by state were obtained from the Department of Transportation. Personal income was obtained from the U.S. Bureau of Economic Analysis, and the unemployment rate was obtained from the U.S. Bureau of Labor Statistics.

These data were graciously provided to us by Professor Christopher J. Ruhm of the Department of Economics at the University of North Carolina.

b. Construct a 95% confidence interval for your answer to (a).

c. Think of an important omitted variable that might bias the answer in (a). What is it and how would it bias the results?

11.10 (Requires Section 11.3 and calculus) Suppose that a random variable Y has the following probability distribution: $\Pr(Y = 1) = p$, $\Pr(Y = 2) = q$, and $\Pr(Y = 3) = 1 - p - q$. A random sample of size n is drawn from this distribution and the random variables are denoted $Y_1, Y_2, \ldots, Y_n$.

a. Derive the likelihood function for the parameters p and q.

b. Derive formulas for the MLE of p and q.

11.11 (Requires Appendix 11.3) Which model would you use for:

a. A study explaining the number of minutes that a person spends talking on a cellular phone during the month?

b. A study explaining grades (A–F) in a large Principles of Economics class?

c. A study of consumers' choices for Coke, Pepsi, or generic cola?

d. A study of the number of cellular phones owned by a family?

Empirical Exercises

E11.1 It has been conjectured that workplace smoking bans induce smokers to quit by reducing their opportunities to smoke. In this assignment you will estimate the effect of workplace smoking bans on smoking using data on a sample of 10,000 U.S. indoor workers from 1991–1993, available on the textbook Web site **www.aw-bc.com/stock_watson** in the file **Smoking**. The data set contains information on whether individuals were or were not subject to a workplace smoking ban, whether the individuals smoked, and other individual characteristics.[7] A detailed description is given in **Smoking_Description**, available on the Web site.

a. Estimate the probability of smoking for (i) all workers, (ii) workers affected by workplace smoking bans, and (iii) workers not affected by workplace smoking bans.

[7]These data were provided by Professor William Evans of the University of Maryland and were used in his paper with Matthew Farrelly and Edward Montgomery, "Do Workplace Smoking Bans Reduce Smoking?" *American Economic Review* 1999; 89(4): 728–747.

b. What is the difference in the probability of smoking between workers affected by a workplace smoking ban and workers not affected by a workplace smoking ban? Use a linear probability model to determine whether this difference is statistically significant.

c. Estimate a linear probability model with *smoker* as the dependent variable and the following regressors: *smkban*, *female*, *age*, age^2, *hsdrop*, *hsgrad*, *colsome*, *colgrad*, *black*, and *hispanic*. Compare the estimated effect of a smoking ban from this regression with your answer from (b). Suggest a reason, based on the substance of this regression, explaining the change in the estimated effect of a smoking ban between (b) and (c).

d. Test the hypothesis that the coefficient on *smkban* is zero in the population version of the regression in (c) against the alternative that it is nonzero, at the 5% significance level.

e. Test the hypothesis that the probability of smoking does not depend on the level of education in the regression in (c). Does the probability of smoking increase or decrease with the level of education?

f. Based on the regression in (c), is there a nonlinear relationship between *age* and the probability of smoking? Plot the relationship between the probability of smoking and *age* for $18 \leq age \leq 65$ for a white, non-Hispanic male college graduate with no workplace smoking ban.

E11.2 This exercise uses the same data as Empirical Exercise 11.1.

a. Estimate a probit model using the same regressors as in Empirical Exercise 11.1(c).

b. Test the hypothesis that the coefficient on *smkban* is zero in the population version of this probit regression against the alternative that it is nonzero, at the 5% significance level. Compare your *t*-statistic and your conclusion with those of Empirical Exercise 11.1(d) based on the linear probability model.

c. Test the hypothesis that the probability of smoking does not depend on the level of education in this probit model. Compare your results with those in question Empirical Exercise 11.1(e) using the linear probability model.

d. Mr. A is white, non-Hispanic, 20 years old, and a high school dropout. Using the probit regression from (a), and assuming that Mr. A is not subject to a workplace smoking ban, calculate the probability that

Mr. A smokes. Carry out the calculation again assuming that he is subject to a workplace smoking ban. What is the effect of the smoking ban on the probability of smoking?

e. Repeat (d) for Ms. B, a female, black, 40-year-old, college graduate.

f. Repeat (d) and (e) using the linear probability model from Empirical Exercise 11.1(c).

g. Based on the answers to (d)–(f), do the probit and linear probability model results differ? If they do, which results make more sense? Are the estimated effects large in a real-world sense?

h. Are there important remaining threats to internal validity?

E11.3 In this exercise you will study health insurance, health status, and employment using a random sample of more than 8000 workers in the United States. The data are available on the textbook Web site www.aw-bc.com/stock_watson in the file **Insurance**.[8] A detailed description is given in **Insurance_Description**, available on the Web site.

a. Are the self-employed less likely to have health insurance than wage earners? If so, is the difference large in a real-world sense? Is the difference statistically significant?

b. The self-employed might systematically differ from wage earners in their age, education, and so forth. After you control for these other factors, are the self-employed less likely to have health insurance?

c. How does health insurance status vary with age? Are older workers more likely to have health insurance? Less likely?

d. Is the effect of self-employment on insurance status different for older workers than it is for younger workers?

e. It has been argued that the self-employed are less likely to be insured, but despite this, they are just as healthy as wage-earners. Is this right? Does the argument hold up for young workers? For older workers? Are there potential two-way causality problems that might undermine the internal validity of this kind of statistical analysis?

[8]These data were provided by Professor Harvey Rosen of Princeton University and were used in his paper with Craig Perry, "The Self-Employed Are Less Likely Than Wage-Earners to Have Health Insurance. So What?" in Douglas Holtz-Eakin and Harvey S. Rosen, eds., *Entrepreneurship and Public Policy*, MIT Press, 2004.

12.10 Consider the instrumental variable regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$, where Z_i is an instrument. Suppose that data on W_i are not available and the model is estimated omitting W_i from the regression.

a. Suppose Z_i and W_i are uncorrelated. Is the IV estimator consistent?

b. Suppose Z_i and W_i are correlated. Is the IV estimator consistent?

Empirical Exercises

E12.1 During the 1880s, a cartel known as the Joint Executive Committee (JEC) controlled the rail transport of grain from the Midwest to eastern cities in the United States. The cartel preceded the Sherman Antitrust Act of 1890, and it legally operated to increase the price of grain above what would have been the competitive price. From time to time, cheating by members of the cartel brought about a temporary collapse of the collusive price-setting agreement. In this exercise, you will use variations in supply associated with the cartel's collapses to estimate the elasticity of demand for rail transport of grain. On the textbook Web site **www.aw-bc.com/stock_watson**, you will find a data file **JEC** that contains weekly observations on the rail shipping price and other factors from 1880 to 1886.[4] A detailed description of the data is contained in **JEC_Description** available on the Web site.

Suppose that the demand curve for rail transport of grain is specified as $\ln(Q_i) = \beta_0 + \beta_1 \ln(P_i) + \beta_2 Ice_i + \sum_{j=1}^{12} \beta_{2+j} Seas_{j,i} + u_i$, where Q_i is the total tonnage of grain shipped in week i, P_i is the price of shipping a ton of grain by rail, Ice_i is a binary variable that is equal to 1 if the Great Lakes are not navigable because of ice, and $Seas_j$ is a binary variable that captures seasonal variation in demand. *Ice* is included because grain could also be transported by ship when the Great Lakes were navigable.

a. Estimate the demand equation by OLS. What is the estimated value of the demand elasticity and its standard error?

b. Explain why the interaction of supply and demand could make the OLS estimator of the elasticity biased.

c. Consider using the variable *cartel* as instrumental variable for $\ln(P)$. Use economic reasoning to argue whether *cartel* plausibly satisfies the two conditions for a valid instrument.

[4]These data were provided by Professor Robert Porter of Northwestern University and were used in his paper "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886," *The Bell Journal of Economics* 1983; 14(2): 301–314.

d. Estimate the first-stage regression. Is *cartel* a weak instrument?

e. Estimate the demand equation by instrumental variable regression. What is the estimated demand elasticity and its standard error?

f. Does the evidence suggest that the cartel was charging the profit-maximizing monopoly price? Explain. (*Hint:* What should a monopolist do if the price elasticity is less than 1?)

E12.2 How does fertility affect labor supply? That is, how much does a woman's labor supply fall when she has an additional child? In this exercise you will estimate this effect using data for married women from the 1980 U.S. Census.[5] The data are available on the textbook Web site **www.aw-bc.com/stock_watson** in the file **Fertility** and described in the file **Fertility_Description**. The data set contains information on married women aged 21–35 with two or more children.

a. Regress *weeksworked* on the indicator variable *morekids* using OLS. On average, do women with more than two children work less than women with two children? How much less?

b. Explain why the OLS regression estimated in (a) is inappropriate for estimating the causal effect of fertility (*morekids*) on labor supply (*weeksworked*).

c. The data set contains the variable *samesex*, which is equal to 1 if the first two children are of the same sex (boy–boy or girl–girl) and equal to 0 otherwise. Are couples whose first two children are of the same sex more likely to have a third child? Is the effect large? Is it statistically significant?

d. Explain why *samesex* is a valid instrument for the instrumental variable regression of *weeksworked* on *morekids*.

e. Is *samesex* a weak instrument?

f. Estimate the regression of *weeksworked* on *morekids* using *samesex* as an instrument. How large is the fertility effect on labor supply?

g. Do the results change when you include the variables *agem1*, *black*, *hispan*, and *othrace* in the labor supply regression (treating these variable as exogenous)? Explain why or why not.

[5]These data were provided by Professor William Evans of the University of Maryland and were used in his paper with Joshua Angrist, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review* 1998, 88(3): 450–477.

E12.3 (This requires Appendix 12.5) On the textbook Web site **www.aw-bc.com/stock_watson** you will find the data set **WeakInstrument** that contains 200 observations on (Y_i, X_i, Z_i) for the instrumental regression $Y_i = \beta_0 + \beta_1 X_i + u_i$.

a. Construct $\hat{\beta}_1^{TSLS}$, its standard error, and the usual 95% confidence interval for β_1.

b. Compute the F-statstic for the regression of X_i on Z_i. Is there evidence of a "weak instrument" problem?

c. Compute a 95% confidence interval for β_1 using the Anderson-Rubin procedure. (To implement the procedure, assume that $-5 \leq \beta_1 \leq 5$.)

d. Comment on the differences in the confidence intervals in (a) and (c). Which is more reliable?

APPENDIX 12.1 The Cigarette Consumption Panel Data Set

The data set consists of annual data for the 48 continental U.S. states from 1985 to 1995. Quantity consumed is measured by annual per capita cigarette sales in packs per fiscal year, as derived from state tax collection data. The price is the real (that is, inflation-adjusted) average retail cigarette price per pack during the fiscal year, including taxes. Income is real per capita income. The general sales tax is the average tax, in cents per pack, due to the broad-based state sales tax applied to all consumption goods. The cigarette-specific tax is the tax applied to cigarettes only. All prices, income, and taxes used in the regressions in this chapter are deflated by the Consumer Price Index and thus are in constant (real) dollars. We are grateful to Professor Jonathan Gruber of MIT for providing us with these data.

APPENDIX 12.2 Derivation of the Formula for the TSLS Estimator in Equation (12.4)

The first stage of TSLS is to regress X_i on the instrument Z_i by OLS, and to compute the OLS predicted value $\hat{X}_i$, and the second stage is to regress Y_i on $\hat{X}_i$ by OLS. Accordingly,