**12.9** A researcher is interested in the effect of military service on human capital. He collects data from a random sample of 4000 workers aged 40 and runs the OLS regression $Y_i = \beta_0 + \beta_1 X_i + u_i$, where $Y_i$ is the worker's annual earnings and $X_i$ is a binary variable that is equal to 1 if the person served in the military and is equal to 0 otherwise.

**a.** Explain why the OLS estimates are likely to be unreliable. (*Hint:* Which variables are omitted from the regression? Are they correlated with military service?)

**b.** During the Vietnam War there was a draft, where priority for the draft was determined by a national lottery. (Birthdates were randomly selected and ordered 1 through 365. Those with birthdates ordered first were drafted before those with birthdates ordered second, and so forth.) Explain how the lottery might be used as an instrument to estimate the effect of military service on earnings. (For more about this issue, see Joshua D. Angrist, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records," *American Economic Review*, June 1990: 313–336.)

**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 that $Z_i$ and $W_i$ are uncorrelated. Is the IV estimator consistent?

**b.** Suppose that $Z_i$ 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.pearsonhighered.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.

**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.pearsonhighered.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?

[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.

[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.

**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.

**E12.3** (This requires Appendix 12.5) On the textbook Web site **www.pearsonhighered.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 $\beta_1$.

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

**c.** Compute a 95% confidence interval for $\beta_1$ 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 contiguous 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-