CHAPTER 10

Regression with Panel Data

Multiple regression is a powerful tool for controlling for the effect of variables on which we have data. If data are not available for some of the variables, however, they cannot be included in the regression and the OLS estimators of the regression coefficients could have omitted variable bias.

This chapter describes a method for controlling for some types of omitted variables without actually observing them. This method requires a specific type of data, called panel data, in which each observational unit, or entity, is observed at two or more time periods. By studying *changes* in the dependent variable over time, it is possible to eliminate the effect of omitted variables that differ across entities but are constant over time.

The empirical application in this chapter concerns drunk driving: What are the effects of alcohol taxes and drunk driving laws on traffic fatalities? We address this question using data on traffic fatalities, alcohol taxes, drunk driving laws, and related variables for the 48 contiguous U.S. states for each of the seven years from 1982 to 1988. This panel data set lets us control for unobserved variables that differ from one state to the next, such as prevailing cultural attitudes toward drinking and driving, but do not change over time. It also allows us to control for variables that vary through time, like improvements in the safety of new cars, but do not vary across states.

Section 10.1 describes the structure of panel data and introduces the drunk driving data set. Fixed effects regression, the main tool for regression analysis of panel data, is an extension of multiple regression that exploits panel data to control for variables that differ across entities but are constant over time. Fixed effects regression is introduced in Sections 10.2 and 10.3, first for the case of only two time periods and then for multiple time periods. In Section 10.4, these methods are extended to incorporate so-called time fixed effects, which control for unobserved variables that are constant across entities but change over time. Section 10.5 discusses the panel data regression assumptions and standard errors for panel data regression. In Section 10.6, we use these methods to study the effect of alcohol taxes and drunk driving laws on traffic deaths.

KEY CONCEPT 10.1

Notation for Panel Data

Panel data consist of observations on the same n entities at two or more time periods T, as is illustrated in Table 1.3. If the data set contains observations on the variables X and Y, then the data are denoted

$$(X_{it}, Y_{it}), i = 1, \ldots, n \text{ and } t = 1, \ldots, T, \tag{10.1}$$

where the first subscript, i, refers to the entity being observed and the second subscript, t, refers to the date at which it is observed.

10.1 Panel Data

Recall from Section 1.3 that **panel data** (also called longitudinal data) refers to data for n different entities observed at T different time periods. The state traffic fatality data studied in this chapter are panel data. Those data are for $n = 48$ entities (states), where each entity is observed in $T = 7$ time periods (each of the years 1982, . . . , 1988), for a total of $7 \times 48 = 336$ observations.

When describing cross-sectional data it was useful to use a subscript to denote the entity; for example, Y_i referred to the variable Y for the i^{th} entity. When describing panel data, we need some additional notation to keep track of both the entity and the time period. We do so by using two subscripts rather than one: The first, i, refers to the entity, and the second, t, refers to the time period of the observation. Thus Y_{it} denotes the variable Y observed for the i^{th} of n entities in the t^{th} of T periods. This notation is summarized in Key Concept 10.1.

Some additional terminology associated with panel data describes whether some observations are missing. A **balanced panel** has all its observations; that is, the variables are observed for each entity and each time period. A panel that has some missing data for at least one time period for at least one entity is called an **unbalanced panel.** The traffic fatality data set has data for all 48 contiguous U.S. states for all seven years, so it is balanced. If, however, some data were missing (for example, if we did not have data on fatalities for some states in 1983), then the data set would be unbalanced. The methods presented in this chapter are described for a balanced panel; however, all these methods can be used with an unbalanced panel, although precisely how to do so in practice depends on the regression software being used.

Example: Traffic Deaths and Alcohol Taxes

There are approximately 40,000 highway traffic fatalities each year in the United States. Approximately one-fourth of fatal crashes involve a driver who was drinking, and this fraction rises during peak drinking periods. One study (Levitt and Porter, 2001) estimates that as many as 25% of drivers on the road between 1 A.M. and 3 A.M. have been drinking and that a driver who is legally drunk is at least 13 times as likely to cause a fatal crash as a driver who has not been drinking.

In this chapter, we study how effective various government policies designed to discourage drunk driving actually are in reducing traffic deaths. The panel data set contains variables related to traffic fatalities and alcohol, including the number of traffic fatalities in each state in each year, the type of drunk driving laws in each state in each year, and the tax on beer in each state. The measure of traffic deaths we use is the fatality rate, which is the number of annual traffic deaths per 10,000 people in the population in the state. The measure of alcohol taxes we use is the "real" tax on a case of beer, which is the beer tax, put into 1988 dollars by adjusting for inflation.[1] The data are described in more detail in Appendix 10.1.

Figure 10.1a is a scatterplot of the data for 1982 on two of these variables, the fatality rate and the real tax on a case of beer. A point in this scatterplot represents the fatality rate in 1982 and the real beer tax in 1982 for a given state. The OLS regression line obtained by regressing the fatality rate on the real beer tax is also plotted in the figure; the estimated regression line is

$$\widehat{FatalityRate} = \underset{(0.15)}{2.01} + \underset{(0.13)}{0.15}BeerTax \quad \text{(1982 data).} \tag{10.2}$$

The coefficient on the real beer tax is positive, but not statistically significant at the 10% level.

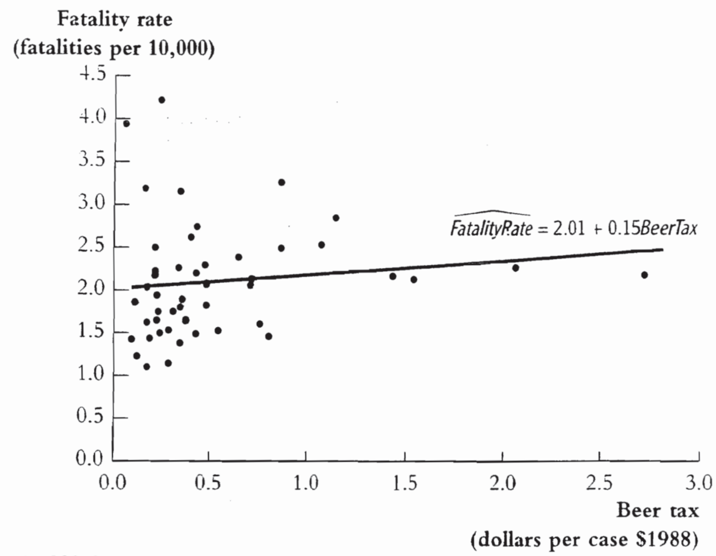
Because we have data for more than one year, we can reexamine this relationship for another year. This is done in Figure 10.1b, which is the same scatterplot as before except that it uses the data for 1988. The OLS regression line through these data is

$$\widehat{FatalityRate} = \underset{(0.11)}{1.86} + \underset{(0.13)}{0.44}BeerTax \quad \text{(1988 data).} \tag{10.3}$$

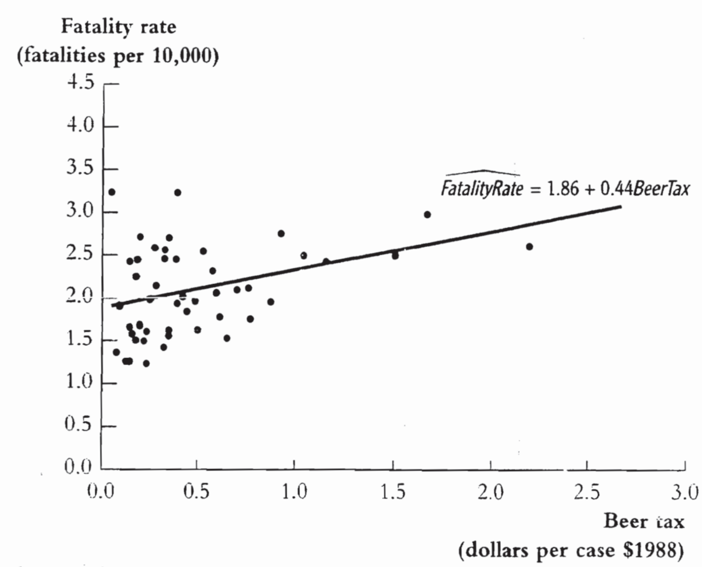
[1] To make the taxes comparable over time, they are put into "1988 dollars" using the Consumer Price Index (CPI). For example, because of inflation a tax of $1 in 1982 corresponds to a tax of $1.23 in 1988 dollars.

FIGURE 10.1 The Traffic Fatality Rate and the Tax on Beer

Panel (a) is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel (b) shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



(a) 1982 data



(b) 1988 data

In contrast to the regression using the 1982 data, the coefficient on the real beer tax is statistically significant at the 1% level (the t-statistic is 3.43). Curiously, the estimated coefficients for the 1982 and the 1988 data are *positive*: Taken literally, higher real beer taxes are associated with *more*, not fewer, traffic fatalities.

Should we conclude that an increase in the tax on beer leads to more traffic deaths? Not necessarily, because these regressions could have substantial omitted variable bias. Many factors affect the fatality rate, including the quality of the automobiles driven in the state, whether the state highways are in good repair, whether most driving is rural or urban, the density of cars on the road, and whether it is socially acceptable to drink and drive. Any of these factors may be correlated with alcohol taxes, and if they are, they will lead to omitted variable bias. One approach to these potential sources of omitted variable bias would be to collect data on all these variables and add them to the annual cross-sectional regressions in Equations (10.2) and (10.3). Unfortunately, some of these variables, such as the cultural acceptance of drinking and driving, might be very hard or even impossible to measure.

If these factors remain constant over time in a given state, however, then another route is available. Because we have panel data, we can in effect hold these factors constant even though we cannot measure them. To do so, we use OLS regression with fixed effects.

10.2 Panel Data with Two Time Periods: "Before and After" Comparisons

When data for each state are obtained for $T = 2$ time periods, it is possible to compare values of the dependent variable in the second period to values in the first period. By focusing on *changes* in the dependent variable, this "before and after" or "differences" comparison in effect holds constant the unobserved factors that differ from one state to the next but do not change over time within the state.

Let Z_i be a variable that determines the fatality rate in the i^{th} state, but does not change over time (so the t subscript is omitted). For example, Z_i might be the local cultural attitude toward drinking and driving, which changes slowly and thus could be considered to be constant between 1982 and 1988. Accordingly, the population linear regression relating Z_i and the real beer tax to the fatality rate is

$$\widetilde{FatalityRate}_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{it}, \tag{10.4}$$

where u_{it} is the error term and $i = 1, \ldots, n$ *and* $t = 1, \ldots, T$.

Because Z_i does not change over time, in the regression model in Equation (10.4) it will not produce any *change* in the fatality rate between 1982 and 1988. Thus, in this regression model, the influence of Z_i can be eliminated by analyzing the change in the fatality rate between the two periods. To see this mathematically, consider Equation (10.4) for each of the two years 1982 and 1988:

$$FatalityRate_{i1982} = \beta_0 + \beta_1 BeerTax_{i1982} + \beta_2 Z_i + u_{i1982}, \quad (10.5)$$

$$FatalityRate_{i1988} = \beta_0 + \beta_1 BeerTax_{i1988} + \beta_2 Z_i + u_{i1988}. \quad (10.6)$$

Subtracting Equation (10.5) from Equation (10.6) eliminates the effect of Z_i:

$$\begin{aligned} &FatalityRate_{i1988} - FatalityRate_{i1982} \\ &\quad = \beta_1(BeerTax_{i1988} - BeerTax_{i1982}) + u_{i1988} - u_{i1982}. \end{aligned} \quad (10.7)$$

This specification has an intuitive interpretation. Cultural attitudes toward drinking and driving affect the level of drunk driving and thus the traffic fatality rate in a state. If, however, they did not change between 1982 and 1988, then they did not produce any *change* in fatalities in the state. Rather, any changes in traffic fatalities over time must have arisen from other sources. In Equation (10.7), these other sources are changes in the tax on beer and changes in the error term (which captures changes in other factors that determine traffic deaths).

Specifying the regression in changes in Equation (10.7) eliminates the effect of the unobserved variables Z_i that are constant over time. In other words, analyzing changes in Y and X has the effect of controlling for variables that are constant over time, thereby eliminating this source of omitted variable bias.

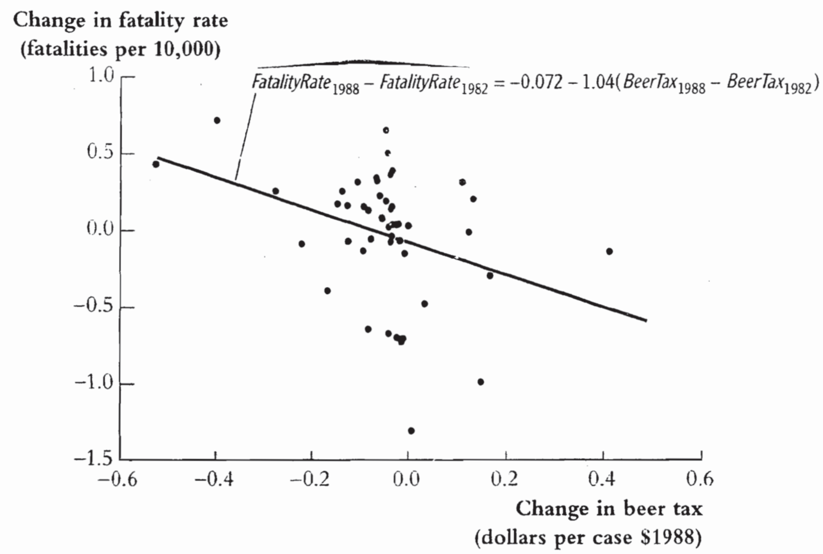
Figure 10.2 presents a scatterplot of the *change* in the fatality rate between 1982 and 1988 against the *change* in the real beer tax between 1982 and 1988 for the 48 states in our data set. A point in Figure 10.2 represents the change in the fatality rate and the change in the real beer tax between 1982 and 1988 for a given state. The OLS regression line, estimated using these data and plotted in the figure, is

$$\begin{aligned} &\widehat{FatalityRate_{1988} - FatalityRate_{1982}} \\ &\quad = \underset{(0.065)}{-0.072} - \underset{(0.36)}{1.04}(BeerTax_{1988} - BeerTax_{1982}). \end{aligned} \quad (10.8)$$

Including an intercept in Equation (10.8) allows for the possibility that the mean change in the fatality rate, in the absence of a change in the real beer tax, is nonzero. For example, the negative intercept (−0.072) could reflect improvements in auto safety from 1982 to 1988 that reduced the average fatality rate.

FIGURE 10.2 Changes in Fatality Rates and Beer Taxes, 1982–1988

This is a scatterplot of the *change* in the traffic fatality rate and the *change* in real beer taxes between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.



In contrast to the cross-sectional regression results, the estimated effect of a change in the real beer tax is negative, as predicted by economic theory. The hypothesis that the population slope coefficient is zero is rejected at the 5% significance level. According to this estimated coefficient, an increase in the real beer tax by $1 per case reduces the traffic fatality rate by 1.04 deaths per 10,000 people. This estimated effect is very large: The average fatality rate is approximately 2 in these data (that is, 2 fatalities per year per 10,000 members of the population), so the estimate suggests that traffic fatalities can be cut in half merely by increasing the real tax on beer by $1 per case.

By examining changes in the fatality rate over time, the regression in Equation (10.8) controls for fixed factors such as cultural attitudes toward drinking and driving. But there are many factors that influence traffic safety, and if they change over time and are correlated with the real beer tax, then their omission will produce omitted variable bias. In Section 10.5, we undertake a more careful analysis that controls for several such factors, so for now it is best to refrain from drawing any substantive conclusions about the effect of real beer taxes on traffic fatalities.

This "before and after" analysis works when the data are observed in two different years. Our data set, however, contains observations for seven different years, and it seems foolish to discard those potentially useful additional data. But the "before and after" method does not apply directly when $T > 2$. To analyze all the observations in our panel data set, we use the method of fixed effects regression.

10.3 Fixed Effects Regression

Fixed effects regression is a method for controlling for omitted variables in panel data when the omitted variables vary across entities (states) but do not change over time. Unlike the "before and after" comparisons of Section 10.2, fixed effects regression can be used when there are two or more time observations for each entity.

The fixed effects regression model has n different intercepts, one for each entity. These intercepts can be represented by a set of binary (or indicator) variables. These binary variables absorb the influences of all omitted variables that differ from one entity to the next but are constant over time.

The Fixed Effects Regression Model

Consider the regression model in Equation (10.4) with the dependent variable (*FatalityRate*) and observed regressor (*BeerTax*) denoted as Y_{it} and X_{it}, respectively:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}, \tag{10.9}$$

where Z_i is an unobserved variable that varies from one state to the next but does not change over time (for example, Z_i represents cultural attitudes toward drinking and driving). We want to estimate β_1, the effect on Y of X holding constant the unobserved state characteristics Z.

Because Z_i varies from one state to the next but is constant over time, the population regression model in Equation (10.9) can be interpreted as having n intercepts, one for each state. Specifically, let $\alpha_i = \beta_0 + \beta_2 Z_i$. Then Equation (10.9) becomes

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}. \tag{10.10}$$

Equation (10.10) is the **fixed effects regression model**, in which $\alpha_1, \ldots, \alpha_n$ are treated as unknown intercepts to be estimated, one for each state. The interpretation of α_i as a state-specific intercept in Equation (10.10) comes from considering the population regression line for the i^{th} state; this population regression line is $\alpha_i + \beta_1 X_{it}$. The slope coefficient of the population regression line, β_1, is the same for all states, but the intercept of the population regression line varies from one state to the next.

Because the intercept α_i in Equation (10.10) can be thought of as the "effect" of being in entity i (in the current application, entities are states), the terms $\alpha_1, \ldots, \alpha_n$ are known as **entity fixed effects**. The variation in the entity fixed effects comes from omitted variables that, like Z_i in Equation (10.9), vary across entities but not over time.

The state-specific intercepts in the fixed effects regression model also can be expressed using binary variables to denote the individual states. Section 8.3 considered the case in which the observations belong to one of two groups and the population regression line has the same slope for both groups but different intercepts (see Figure 8.8a). That population regression line was expressed mathematically using a single binary variable indicating one of the groups (case #1 in Key Concept 8.4). If we had only two states in our data set, that binary variable regression model would apply here. Because we have more than two states, however, we need additional binary variables to capture all the state-specific intercepts in Equation (10.10).

To develop the fixed effects regression model using binary variables, let $D1_i$ be a binary variable that equals 1 when $i = 1$ and equals 0 otherwise, let $D2_i$ equal 1 when $i = 2$ and equal 0 otherwise, and so on. We cannot include all n binary variables plus a common intercept, for if we do the regressors will be perfectly multicollinear (this is the "dummy variable trap" of Section 6.7), so we arbitrarily omit the binary variable $D1_i$ for the first group. Accordingly, the fixed effects regression model in Equation (10.10) can be written equivalently as

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it}, \tag{10.11}$$

where $\beta_0, \beta_1, \gamma_2, \ldots, \gamma_n$ are unknown coefficients to be estimated. To derive the relationship between the coefficients in Equation (10.11) and the intercepts in Equation (10.10), compare the population regression lines for each state in the two equations. In Equation (10.11), the population regression equation for the first state is $\beta_0 + \beta_1 X_{it}$, so $\alpha_1 = \beta_0$. For the second and remaining states, it is $\beta_0 + \beta_1 X_{it} + \gamma_i$, so $\alpha_i = \beta_0 + \gamma_i$ for $i \geq 2$.

Thus there are two equivalent ways to write the fixed effects regression model, Equations (10.10) and (10.11). In Equation (10.10), it is written in terms of n state-specific intercepts. In Equation (10.11), the fixed effects regression model has a common intercept and $n - 1$ binary regressors. In both formulations, the slope coefficient on X is the same from one state to the next. The state-specific intercepts in Equation (10.10) and the binary regressors in Equation (10.11) have the same source: the unobserved variable Z_i that varies across states but not over time.

Extension to multiple X's. If there are other observed determinants of Y that are correlated with X and that change over time, then these should also be included in the regression to avoid omitted variable bias. Doing so results in the fixed effects regression model with multiple regressors, summarized in Key Concept 10.2.

KEY CONCEPT 10.2

The Fixed Effects Regression Model

The fixed effects regression model is

$$Y_{it} = \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it} + \alpha_i + u_{it}, \tag{10.12}$$

where $i = 1, \ldots, n$; $t = 1, \ldots, T$; $X_{1,it}$ 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 $\alpha_1, \ldots, \alpha_n$ are entity-specific intercepts.

Equivalently, the fixed effects regression model can be written in terms of a common intercept, the X's, and $n - 1$ binary variables representing all but one entity:

$$\begin{aligned} Y_{it} = {} & \beta_0 + \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it} + \gamma_2 D2_i \\ & + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it}, \end{aligned} \tag{10.13}$$

where $D2_i = 1$ if $i = 2$ and $D2_i = 0$ otherwise, and so forth.

Estimation and Inference

In principle the binary variable specification of the fixed effects regression model [Equation (10.13)] can be estimated by OLS. This regression, however, has $k + n$ regressors (the k X's, the $n - 1$ binary variables, and the intercept), so in practice this OLS regression is tedious or, in some software packages, impossible to implement if the number of entities is large. Econometric software therefore has special routines for OLS estimation of fixed effects regression models. These special routines are equivalent to using OLS on the full binary variable regression, but are faster because they employ some mathematical simplifications that arise in the algebra of fixed effects regression.

The "entity-demeaned" OLS algorithm. Regression software typically computes the OLS fixed effects estimator in two steps. In the first step, the entity-specific average is subtracted from each variable. In the second step, the regression is estimated using "entity-demeaned" variables. Specifically, consider the case of a single regressor in the version of the fixed effects model in Equation (10.10) and take the average of both sides of Equation (10.10); then $\overline{Y}_i = \beta_1 \overline{X}_i + \alpha_i + \overline{u}_i$, where $\overline{Y}_i = (1/T)\sum_{t=1}^{T} Y_{it}$, and $\overline{X}_i$ and $\overline{u}_i$ are defined similarly. Thus Equation (10.10) implies that $Y_{it} - \overline{Y}_i = \beta_1(X_{it} - \overline{X}_i) + (u_{it} - \overline{u}_i)$. Let $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i$, $\widetilde{X}_{it} = X_{it} - \overline{X}_i$ and $\widetilde{u}_{it} = u_{it} - \overline{u}_i$; accordingly,

$$\widetilde{Y}_{it} = \beta_1 \widetilde{X}_{it} + \widetilde{u}_{it}. \tag{10.14}$$

Thus β_1 can be estimated by the OLS regression of the "entity-demeaned" variables $\widetilde{Y}_{it}$ on $\widetilde{X}_{it}$. In fact, this estimator is identical to the OLS estimator of β_1 obtained by estimation of the fixed effects model in Equation (10.11) using $n - 1$ binary variables (Exercise 18.6).

The "before and after" (differences) regression versus the binary variables specification. Although Equation (10.11) with its binary variables looks quite different from the "before and after" regression model in Equation (10.7), in the special case that $T = 2$ the OLS estimator of β_1 from the binary variable specification and that from the "before and after" specification are identical if the intercept is excluded from the "before and after" specifications. Thus, when $T = 2$, there are three ways to estimate β_1 by OLS: the "before and after" specification in Equation (10.7) (without an intercept), the binary variable specification in Equation (10.11), and the "entity-demeaned" specification in Equation (10.14). These three methods are equivalent; that is, they produce identical OLS estimates of β_1 (Exercise 10.11).

The sampling distribution, standard errors, and statistical inference. In multiple regression with cross-sectional data, if the four least squares assumptions in Key Concept 6.4 hold, then the sampling distribution of the OLS estimator is normal in large samples. The variance of this sampling distribution can be estimated from the data, and the square root of this estimator of the variance—that is, the standard error—can be used to test hypotheses using a t-statistic and to construct confidence intervals.

Similarly, in multiple regression with panel data, if a set of assumptions—called the fixed effects regression assumptions—hold, then the sampling distribution of the fixed effects OLS estimator is normal in large samples, the variance of that distribution can be estimated from the data, the square root of that estimator is the standard error, and the standard error can be used to construct t-statistics and confidence intervals. Given the standard error, statistical inference—testing hypotheses (including joint hypotheses using F-statistics) and constructing confidence intervals—proceeds in exactly the same way as in multiple regression with cross-sectional data.

The fixed effects regression assumptions and standard errors for fixed effects regression are discussed further in Section 10.5.

Application to Traffic Deaths

The OLS estimate of the fixed effects regression line relating the real beer tax to the fatality rate, based on all 7 years of data (336 observations), is

$$\underset{}{\widehat{FatalityRate}} = \underset{(0.29)}{-0.66}BeerTax + StateFixedEffects, \tag{10.15}$$

where, as is conventional, the estimated state fixed intercepts are not listed to save space and because they are not of primary interest in this application.

Like the "differences" specification in Equation (10.8), the estimated coefficient in the fixed effects regression in Equation (10.15) is negative, so, as predicted by economic theory, higher real beer taxes are associated with fewer traffic deaths, which is the opposite of what we found in the initial cross-sectional regressions of Equations (10.2) and (10.3). The two regressions are not identical because the "differences" regression in Equation (10.8) uses only the data for 1982 and 1988 (specifically, the difference between those two years), whereas the fixed effects regression in Equation (10.15) uses the data for all 7 years. Because of the additional observations, the standard error is smaller in Equation (10.15) than in Equation (10.8).

Including state fixed effects in the fatality rate regression lets us avoid omitted variables bias arising from omitted factors, such as cultural attitudes toward drinking and driving, that vary across states but are constant over time within a state. Still, a skeptic might suspect that other factors could lead to omitted variables bias. For example, over this period cars were getting safer and occupants were increasingly wearing seat belts; if the real tax on beer rose on average during the mid-1980s, then *BeerTax* could be picking up the effect of overall automobile safety improvements. If, however, safety improvements evolved over time but were the same for all states, then we can eliminate their influence by including time fixed effects.

10.4 Regression with Time Fixed Effects

Just as fixed effects for each entity can control for variables that are constant over time but differ across entities, so can time fixed effects control for variables that are constant across entities but evolve over time.

Because safety improvements in new cars are introduced nationally, they serve to reduce traffic fatalities in all states. So, it is plausible to think of automobile safety as an omitted variable that changes over time but has the same value for all states. The population regression in Equation (10.9) can be modified to make explicit the effect of automobile safety, which we will denote S_t:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + \beta_3 S_t + u_{it}, \tag{10.16}$$

where S_t is unobserved and where the single t subscript emphasizes that safety changes over time but is constant across states. Because $\beta_3 S_t$ represents variables that determine Y_{it}, if S_t is correlated with X_{it}, then omitting S_t from the regression leads to omitted variable bias.

Time Effects Only

For the moment, suppose that the variables Z_i are not present so that the term $\beta_2 Z_i$ can be dropped from Equation (10.16), although the term $\beta_3 S_t$ remains. Our objective is to estimate β_1, controlling for S_t.

Although S_t is unobserved, its influence can be eliminated because it varies over time but not across states, just as it is possible to eliminate the effect of Z_i, which varies across states but not over time. In the entity fixed effects model, the presence of Z_i leads to the fixed effects regression model in Equation (10.10), in which each state has its own intercept (or fixed effect). Similarly, because S_t varies over time but not over states, the presence of S_t leads to a regression model in which each time period has its own intercept.

The **time fixed effects regression model** with a single X regressor is

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}. \tag{10.17}$$

This model has a different intercept, λ_t, for each time period. The intercept λ_t in Equation (10.17) can be thought of as the "effect" on Y of year t (or, more generally, time period t), so the terms $\lambda_1, \ldots, \lambda_T$ are known as **time fixed effects**. The variation in the time fixed effects comes from omitted variables that, like S_t in Equation (10.16), vary over time but not across entities.

Just as the entity fixed effects regression model can be represented using $n - 1$ binary indicators, so, too, can the time fixed effects regression model be represented using $T - 1$ binary indicators:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \cdots + \delta_T BT_t + u_{it}, \tag{10.18}$$

where $\delta_2, \ldots, \delta_T$ are unknown coefficients and where $B2_t = 1$ if $t = 2$ and $B2_t = 0$ otherwise, and so forth. As in the fixed effects regression model in Equation

(10.11), in this version of the time effects model the intercept is included, and the first binary variable ($B1_t$) is omitted to prevent perfect multicollinearity.

When there are additional observed "X" regressors, then these regressors appear in Equations (10.17) and (10.18) as well.

In the traffic fatalities regression, the time fixed effects specification allows us to eliminate bias arising from omitted variables like nationally introduced safety standards that change over time but are the same across states in a given year.

Both Entity and Time Fixed Effects

If some omitted variables are constant over time but vary across states (such as cultural norms) while others are constant across states but vary over time (such as national safety standards), then it is appropriate to include *both* entity (state) *and* time effects.

The combined **entity and time fixed effects regression model** is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}, \tag{10.19}$$

where α_i is the entity fixed effect and λ_t is the time fixed effect. This model can equivalently be represented using $n-1$ entity binary indicators and $T-1$ time binary indicators, along with an intercept:

$$\begin{aligned} Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \cdots + \gamma_n Dn_i \\ + \delta_2 B2_t + \cdots + \delta_T BT_t + u_{it}, \end{aligned} \tag{10.20}$$

where $\beta_0, \beta_1, \gamma_2, \ldots, \gamma_n$, and $\delta_2, \ldots, \delta_T$ are unknown coefficients.

When there are additional observed "X" regressors, then these appear in Equations (10.19) and (10.20) as well.

The combined state and time fixed effects regression model eliminates omitted variables bias arising both from unobserved variables that are constant over time and from unobserved variables that are constant across states.

Estimation. The time fixed effects model and the entity and time fixed effects model are both variants of the multiple regression model. Thus their coefficients can be estimated by OLS by including the additional time binary variables. Alternatively, in a balanced panel the coefficients on the X's can be computed by first deviating Y and the X's from their entity *and* time-period means and then by estimating the multiple regression equation of deviated Y on the deviated X's.

This algorithm, which is commonly implemented in regression software, eliminates the need to construct the full set of binary indicators that appear in Equation (10.20). An equivalent approach is to deviate Y, the X's, and the time indicators from their entity (but not time) means and to estimate $k + T$ coefficients by multiple regression of the deviated Y on the deviated X's and the deviated time indicators. Finally, if $T = 2$, the entity and time fixed effects regression can be estimated using the "before and after" approach of Section 10.2, including the intercept in the regression. Thus the "before and after" regression reported in Equation (10.8), in which the change in *FatalityRate* from 1982 to 1988 is regressed on the change in *BeerTax* from 1982 to 1988 including an intercept, provides the same estimate of the slope coefficient as the OLS regression of *FatalityRate* on *BeerTax*, including entity and time fixed effects, estimated using data for the two years 1982 and 1988.

Application to traffic deaths. Adding time effects to the state fixed effects regression results in the OLS estimate of the regression line:

$$\widehat{FatalityRate} = \underset{(0.36)}{-0.64}BeerTax + StateFixedEffects + TimeFixedEffects. \tag{10.21}$$

This specification includes the beer tax, 47 state binary variables (state fixed effects), 6 single-year binary variables (time fixed effects), and an intercept, so this regression actually has $1 + 47 + 6 + 1 = 55$ right-hand variables! The coefficients on the time and state binary variables and the intercept are not reported because they are not of primary interest.

Including time effects has little impact on the coefficient on the real beer tax [compare Equations (10.15) and (10.21)]. Although this coefficient is less precisely estimated when time effects are included, it is still significant at the 10%, but not 5%, significance level ($t = -0.64/0.36 = -1.78$).

This estimated relationship between the real beer tax and traffic fatalities is immune to omitted variable bias from variables that are constant either over time or across states. However, many important determinants of traffic deaths do not fall into this category, so this specification could still be subject to omitted variable bias. Section 10.6 therefore undertakes a more complete empirical examination of the effect of the beer tax and of laws aimed directly at eliminating drunk driving, controlling for a variety of factors. Before turning to that study, we first discuss the assumptions underlying panel data regression and the construction of standard errors for fixed effects estimators.

10.5 The Fixed Effects Regression Assumptions and Standard Errors for Fixed Effects Regression

In panel data, the regression error can be correlated over time within an entity. Like heteroskedasticity, this correlation does not introduce bias into the fixed effects estimator, but it affects the variance of the fixed effects estimator and therefore it affects how one computes standard errors. The standard errors for fixed effects regressions reported in this chapter are so-called clustered standard errors, which are robust both to heteroskedasticity and to correlation over time within an entity. When there are many entities (when n is large), hypothesis tests and confidence intervals can be computed using the usual large-sample normal and F critical values.

This section describes clustered standard errors. We begin with the fixed effects regression assumptions, which extend the least squares regression assumptions to panel data; under these assumptions, the fixed effects estimator is asymptotically normally distributed when n is large. To keep the notation as simple as possible, this section focuses on the entity fixed effects regression model of Section 10.3, in which there are no time effects.

The Fixed Effects Regression Assumptions

The four fixed effects regression assumptions are summarized in Key Concept 10.3. These assumptions extend the four least squares assumptions, stated for cross-sectional data in Key Concept 6.4, to panel data.

The first assumption is that the error term has conditional mean zero, given all T values of X for that entity. This assumption plays the same role as the first least squares assumption for cross-sectional data in Key Concept 6.4 and implies that there is no omitted variable bias. The requirement that the conditional mean of u_{it} not depend on *any* of the values of X for that entity—past, present, or future—adds an important subtlety beyond the first least squares assumption for cross-sectional data. This assumption is violated if current u_{it} is correlated with past, present, or future values of X.

The second assumption is that 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 $i = 1, \ldots, n$. Like the second least squares assumption in Key Concept 6.4, the second assumption for fixed effects regression holds if entities are selected by simple random sampling from the population.

The Fixed Effects Regression Assumptions

KEY CONCEPT 10.3

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

where

1. u_{it} has conditional mean zero: $E(u_{it}|X_{i1}, X_{i2}, \ldots, X_{iT}, \alpha_i) = 0$.
2. $(X_{i1}, X_{i2}, \ldots, X_{iT}, u_{i1}, u_{i2}, \ldots, u_{iT}), i = 1, \ldots, n$ are i.i.d. draws from their joint distribution.
3. Large outliers are unlikely: (X_{it}, u_{it}) have nonzero finite fourth moments.
4. There is no perfect multicollinearity.

For multiple regressors, X_{it} should be replaced by the full list $X_{1,it}, X_{2,it}, \ldots, X_{k,it}$.

The third and fourth assumptions for fixed effects regression are analogous to the third and fourth least squares assumptions for cross-sectional data in Key Concept 6.4.

Under the least squares assumptions for panel data in Key Concept 10.3, the fixed effects estimator is consistent and is normally distributed when n is large. The details are discussed in Appendix 10.2.

An important difference between the panel data assumptions in Key Concept 10.3 and the assumptions for cross-sectional data in Key Concept 6.4 is Assumption 2. The cross-sectional counterpart of Assumption 2 holds that each observation is independent, which arises under simple random sampling. In contrast, Assumption 2 for panel data holds that the variables are independent across entities, but makes no such restriction within an entity. For example, Assumption 2 allows X_{it} to be correlated over time within an entity.

If X_{it} is correlated with X_{is} for different values of s and t—that is, if X_{it} is correlated over time for a given entity—then X_{it} is said to be **autocorrelated** (correlated with itself, at different dates) or **serially correlated**. Autocorrelation is a pervasive feature of time series data: What happens one year tends to be correlated with what happens the next year. In the traffic fatality example, X_{it}, the beer tax in state i in year t, is autocorrelated: Most of the time, the legislature does not change the beer tax, so if it is high one year relative to its mean value for state i, it will tend to be high the next year, too. Similarly, it is possible to think of reasons why u_{it} would be autocorrelated. Recall that u_{it} consists of time-varying factors that are determinants of Y_{it} but are not included as regressors, and some of these omitted factors might be autocorrelated. For example, a downturn in the local economy might produce layoffs and diminish commuting traffic, thus reducing traffic fatalities for 2 or more years. Similarly, a major road improvement project might reduce traffic accidents not only in the year of completion but also in future years. Such omitted factors which persist over multiple years, produce autocorrelated regression errors. Not all omitted factors will produce autocorrelation in u_{it}; for example, severe winter driving conditions plausibly affect fatalities, but if winter weather conditions for a given state are independently distributed from one year to the next, then this component of the error term would be serially uncorrelated. In general, though, as long as some omitted factors are autocorrelated, then u_{it} will be autocorrelated.

Standard Errors for Fixed Effects Regression

If the regression errors are autocorrelated, then the usual heteroskedasticity-robust standard error formula for cross-section regression [Equations (5.3) and (5.4)] is not valid. One way to see this is to draw an analogy to heteroskedasticity. In a regression with cross-sectional data, if the errors are heteroskedastic, then (as discussed in Section 5.4) the homoskedasticity-only standard errors are not valid because they were derived under the false assumption of homoskedasticity. Similarly, if the errors are autocorrelated, then the usual standard errors will not be valid because they were derived under the false assumption of no serial correlation.

Standard errors that are valid if u_{it} is potentially heteroskedastic and potentially correlated over time within an entity are referred to as **heteroskedasticity- and autocorrelation-consistent (HAC) standard errors**. The standard errors used in this chapter are one type of HAC standard errors, **clustered standard errors**. The term *clustered* arises because these standard errors allow the regression errors to have an arbitrary correlation within a cluster, or grouping, but assume that the regression errors are uncorrelated across clusters. In the context of panel data, each cluster consists of an entity. Thus clustered standard errors allow for heteroskedasticity and for arbitrary autocorrelation within an entity, but treat the errors as uncorrelated across entities. That is, clustered standard errors allow for heteroskedasticity and autocorrelation in a way that is consistent with the second fixed effects regression assumption in Key Concept 10.3.

Like heteroskedasticity-robust standard errors in regression with cross-sectional data, clustered standard errors are valid whether or not there is heteroskedasticity, autocorrelation, or both. If the number of entities n is large, inference using clustered standard errors can proceed using the usual large-sample normal critical values for t-statistics and $F_{q,\infty}$ critical values for F-statistics testing q restrictions.

In practice, there can be a large difference between clustered standard errors and standard errors that do not allow for autocorrelation of u_{it}. For example, the usual (cross-section data) heteroskedasticity-robust standard error for the *BeerTax* coefficient in Equation (10.21) is 0.25, substantially smaller than the clustered standard error, 0.36, and the respective t-statistics testing $\beta_1 = 0$ are -2.51 and -1.78. The reason we report the clustered standard error is that it allows for serial correlation of u_{it} within an entity, whereas the usual heteroskedasticity-robust standard error does not. The formula for clustered standard errors is given in Appendix 10.2.

10.6 Drunk Driving Laws and Traffic Deaths

Alcohol taxes are only one way to discourage drinking and driving. States differ in their punishments for drunk driving, and a state that cracks down on drunk driving could do so by toughening driving laws as well as raising taxes. If so, omitting these laws could produce omitted variable bias in the OLS estimator of the effect of real beer taxes on traffic fatalities, even in regressions with state and time fixed effects. In addition, because vehicle use depends in part on whether drivers have jobs and because tax changes can reflect economic conditions (a state budget deficit can lead to tax hikes), omitting state economic conditions also could result in omitted variable bias. In this section, we therefore extend the preceding analysis of traffic fatalities to include other driving laws and economic conditions.

The results are summarized in Table 10.1. The format of the table is the same as that of the tables of regression results in Chapters 7 through 9: Each column reports a different regression, and each row reports a coefficient estimate and standard error, F-statistic and p-value, or other information about the regression.

Column (1) in Table 10.1 presents results for the OLS regression of the fatality rate on the real beer tax without state and time fixed effects. As in the cross-sectional regressions for 1982 and 1988 [Equations (10.2) and (10.3)], the coefficient on the real beer tax is *positive* (0.36): According to this estimate, increasing beer taxes *increases* traffic fatalities! However, the regression in column (2) [reported previously as Equation (10.15)], which includes state fixed effects, suggests that the positive coefficient in regression (1) is the result of omitted variable bias (the coefficient on the real beer tax is -0.66). The regression $\overline{R}^2$ jumps from 0.091 to 0.889 when fixed effects are included; evidently, the state fixed effects account for a large amount of the variation in the data.

Little changes when time effects are added, as reported in column (3) [reported previously as Equation (10.21)], except that the beer tax coefficient is now estimated less precisely. The results in columns (1) through (3) are consistent with the omitted fixed factors—historical and cultural factors, general road conditions, population density, attitudes toward drinking and driving, and so forth—being important determinants of the variation in traffic fatalities across states.

TABLE 10.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

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

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36**	−0.66*	−0.64⁺	−0.45	−0.69*	−0.46	−0.93**
	(0.05)	(0.29)	(0.36)	(0.30)	(0.35)	(0.31)	(0.34)
Drinking age 18				0.028	−0.010		0.037
				(0.070)	(0.083)		(0.102)
Drinking age 19				−0.018	−0.076		−0.065
				(0.050)	(0.068)		(0.099)
Drinking age 20				0.032	−0.100⁺		−0.113
				(0.051)	(0.056)		(0.125)
Drinking age						−0.002	
						(0.021)	
Mandatory jail or community service?				0.038	0.085	0.039	0.089
				(0.103)	(0.112)	(0.103)	(0.164)
Average vehicle miles per driver				0.008	0.017	0.009	0.124
				(0.007)	(0.011)	(0.007)	(0.049)
Unemployment rate				−0.063**		−0.063**	−0.091**
				(0.013)		(0.013)	(0.021)
Real income per capita (logarithm)				1.82**		1.79**	1.00
				(0.64)		(0.64)	(0.68)
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			4.22	10.12	3.48	10.28	37.49
			(0.002)	(< 0.001)	(0.006)	(< 0.001)	(< 0.001)
Drinking age coefficients = 0				0.35	1.41		0.42
				(0.786)	(0.253)		(0.738)
Unemployment rate, income per capita = 0				29.62		31.96	25.20
				(< 0.001)		(< 0.001)	(< 0.001)
$\bar{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

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. The data set is described in Appendix 10.1. 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.

The next four regressions in Table 10.1 include additional potential determinants of fatality rates along with state and time effects. The base specification, reported in column (4), includes variables related to drunk driving laws plus variables that control for the amount of driving and overall state economic conditions. The first legal variables are the minimum legal drinking age, represented by three binary variables for a minimum legal drinking age of 18, 19, and 20 (so the omitted group is a minimum legal drinking age of 21 or older). The other legal variable is the punishment associated with the first conviction for driving under the influence of alcohol, either mandatory jail time or mandatory community service (the omitted group is less severe punishment). The three measures of driving and economic conditions are average vehicle miles per driver, the unemployment rate, and the logarithm of real (1988 dollars) personal income per capita (using the logarithm of income permits the coefficient to be interpreted in terms of percentage changes of income; see Section 8.2). The final regression in Table 10.1 follows the "before and after" approach of Section 10.2 and uses only data from 1982 and 1988; thus regression (7) extends the regression in Equation (10.8) to include the additional regressors.

The regression in column (4) has four interesting results.

1. Including the additional variables reduces the estimated effect of the beer tax from -0.64 in column (3) to -0.45 in column (4). One way to evaluate the magnitude of this coefficient is to imagine a state with an average real beer tax doubling its tax; because the average real beer tax in these data is approximately \$0.50 per case (in 1988 dollars), this entails increasing the tax by \$0.50 per case. The estimated effect of a \$0.50 increase in the beer tax is to decrease the expected fatality rate by $0.45 \times 0.50 = 0.23$ death per 10,000. This estimated effect is large: Because the average fatality rate is 2 per 10,000, a reduction of 0.23 corresponds to reducing traffic deaths by nearly one-eighth. This said, the estimate is quite imprecise: Because the standard error on this coefficient is 0.30, the 95% confidence interval for this effect is $-0.45 \times 0.50 \pm 1.96 \times 0.30 \times 0.50 = (-0.52, 0.07)$. This wide 95% confidence interval includes zero, so the hypothesis that the beer tax has no effect cannot be rejected at the 5% significance level.

2. The minimum legal drinking age is precisely estimated to have a small effect on traffic fatalities. According to the regression in column (4), the 95% confidence interval for the increase in the fatality rate in a state with a minimum legal drinking age of 18, relative to age 21, is $(-0.11, 0.17)$. The joint hypothesis that the coefficients on the minimum legal drinking age variables are zero cannot be rejected at the 10% significance level: The F-statistic testing the joint hypothesis that the three coefficients are zero is 0.35, with a p-value of 0.786.

3. The coefficient on the first offense punishment variable is also estimated to be small and is not significantly different from zero at the 10% significance level.
4. The economic variables have considerable explanatory power for traffic fatalities. High unemployment rates are associated with fewer fatalities: An increase in the unemployment rate by one percentage point is estimated to reduce traffic fatalities by 0.063 death per 10,000. Similarly, high values of real per capita income are associated with high fatalities: The coefficient is 1.82, so a 1% increase in real per capita income is associated with an increase in traffic fatalities of 0.0182 death per 10,000 (see Case I in Key Concept 8.2 for interpretation of this coefficient). According to these estimates, good economic conditions are associated with higher fatalities, perhaps because of increased traffic density when the unemployment rate is low or greater alcohol consumption when income is high. The two economic variables are jointly significant at the 0.1% significance level (the F-statistic is 29.62).

Columns (5) through (7) of Table 10.1 report regressions that check the sensitivity of these conclusions to changes in the base specification. The regression in column (5) drops the variables that control for economic conditions. The result is an increase in the estimated effect of the real beer tax, which becomes significant at the 5% level, but no appreciable change in the other coefficients. The sensitivity of the estimated beer tax coefficient to including the economic variables, combined with the statistical significance of the coefficients on those variables in column (4), indicates that the economic variables should remain in the base specification. The regression in column (6) shows that the results in column (4) are not sensitive to changing the functional form when the three drinking age indicator variables are replaced by the drinking age itself. When the coefficients are estimated using the changes of the variables from 1982 to 1988 [column (7)], as in Section 10.2, the findings from column (4) are largely unchanged except that the coefficient on the beer tax is larger and is significant at the 1% level.

The strength of this analysis is that including state and time fixed effects mitigates the threat of omitted variable bias arising from unobserved variables that either do not change over time (like cultural attitudes toward drinking and driving) or do not vary across states (like safety innovations). As always, however, it is important to think about possible threats to validity. One potential source of omitted variable bias is that the measure of alcohol taxes used here, the real tax on beer, could move with other alcohol taxes, which suggests interpreting the results as pertaining more broadly than just to beer. A subtler possibility is that hikes in the real beer tax could be associated with public education campaigns. If so, changes in the real beer tax could pick up the effect of a broader campaign to reduce drunk driving.

Taken together, these results present a provocative picture of measures to control drunk driving and traffic fatalities. According to these estimates, neither stiff punishments nor increases in the minimum legal drinking age have important effects on fatalities. In contrast, there is some evidence that increasing alcohol taxes, as measured by the real tax on beer, does reduce traffic deaths, presumably through reduced alcohol consumption. The imprecision of the estimated beer tax coefficient means, however, that we should be cautious about drawing policy conclusions from this analysis and that additional research is warranted.[2]

10.7 Conclusion

This chapter showed how multiple observations over time on the same entity can be used to control for unobserved omitted variables that differ across entities but are constant over time. The key insight is that if the unobserved variable does not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics. If cultural attitudes toward drinking and driving do not change appreciably over 7 years within a state, then explanations for changes in the traffic fatality rate over those 7 years must lie elsewhere.

To exploit this insight, you need data in which the same entity is observed at two or more time periods; that is, you need panel data. With panel data, the multiple regression model of Part II can be extended to include a full set of entity binary variables; this is the fixed effects regression model, which can be estimated by OLS. A twist on the fixed effects regression model is to include time fixed effects, which control for unobserved variables that change over time but are constant across entities. Both entity and time fixed effects can be included in the regression to control for variables that vary across entities but are constant over time and for variables that vary over time but are constant across entities.

Despite these virtues, entity and time fixed effects regression cannot control for omitted variables that vary *both* across entities *and* over time. And, obviously, panel data methods require panel data, which often are not available. Thus there remains a need for a method that can eliminate the influence of unobserved omitted variables when panel data methods cannot do the job. A powerful and general method for doing so is instrumental variables regression, the topic of Chapter 12.

[2]For further analysis of these data, see Ruhm (1996). A recent meta-analysis of 112 studies of the effect of alcohol prices and taxes on consumption found elasticities of −0.46 for beer, −0.69 for wine, and −0.80 for spirits, and concluded that alcohol taxes have large effects on reducing consumption, relative to other programs [Wagenaar, Salois, and Komro (2009)]. To learn more about drunk driving and alcohol, and about the economics of alcohol more generally, also see Cook and Moore (2000), Chaloupka, Grossman, and Saffer (2002), Young and Bielinska-Kwapisz (2006), and Dang (2008).

Summary

1. Panel data consist of observations on multiple (n) entities—states, firms, people, and so forth—where each entity is observed at two or more time periods (T).
2. Regression with entity fixed effects controls for unobserved variables that differ from one entity to the next but remain constant over time.
3. When there are two time periods, fixed effect regression can be estimated by a "before and after" regression of the change in Y from the first period to the second on the corresponding change in X.
4. Entity fixed effects regression can be estimated by including binary variables for $n - 1$ entities plus the observable independent variables (the X's) and an intercept.
5. Time fixed effects control for unobserved variables that are the same across entities but vary over time.
6. A regression with time and entity fixed effects can be estimated by including binary variables for $n - 1$ entities and binary variables for $T - 1$ time periods plus the X's and an intercept.
7. In panel data, variables are typically autocorrelated, that is, correlated over time within an entity. Standard errors need to allow both for this autocorrelation and for potential heteroskedasticity, and one way to do so is to use clustered standard errors.

Key Terms

panel data (348)
balanced panel (348)
unbalanced panel (348)
fixed effects regression model (354)
entity fixed effects (354)
time fixed effects regression model (359)
time fixed effects (359)
entity and time fixed effects regression model (360)
autocorrelated (363)
serially correlated (363)
heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (364)
clustered standard errors (364)

Review the Concepts

10.1 Why is it necessary to use two subscripts, i and t, to describe panel data? What does i refer to? What does t refer to?

10.2 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. Give some examples of unobserved person-specific variables that are correlated with both education and earnings. Can you think of examples of time-specific variables that might be correlated with education and earnings? How would you control for these person-specific and time-specific effects in a panel data regression?

10.3 Can the regression that you suggested in response to Question 10.2 be used to estimate the effect of gender on an individual's earnings? Can that regression be used to estimate the effect of the national unemployment rate on an individual's earnings? Explain.

10.4 In the context of the regression you suggested for Question 10.2, explain why the regression error for a given individual might be serially correlated.

Exercises

10.1 This exercise refers to the drunk driving panel data regression summarized in Table 10.1.

a. New Jersey has a population of 8.1 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. Construct a 95% confidence interval for your answer.

b. 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. Construct a 95% confidence interval for your answer.

c. 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 the number of traffic fatalities in the next year. Construct a 90% confidence interval for your answer.

d. Should time effects be included in the regression? Why or why not?

e. 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? (Be specific about the specification of the regression and the statistical test you would use.)

10.2 Consider the binary variable version of the fixed effects model in Equation (10.11), except with an additional regressor, $D1_i$; that is, let

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_1 D1_i + \gamma_2 D2_i + \cdots + \gamma_n Dn_i + u_{it}.$$

a. Suppose that $n = 3$. 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, where $X_{0,it} = 1$ for all i, t.

b. Show the result in (a) for general n.

c. What will happen if you try to estimate the coefficients of the regression by OLS?

10.3 Section 9.2 gave a list of five potential threats to the internal validity of a regression study. Apply this list to the empirical analysis in Section 10.6 and thereby draw conclusions about its internal validity.

10.4 Using the regression in Equation (10.11), what is the slope and intercept for

a. Entity 1 in time period 1?

b. Entity 1 in time period 3?

c. Entity 3 in time period 1?

d. Entity 3 in time period 3?

10.5 Consider the model with a single regressor $Y_{it} = \beta_1 X_{1,it} + \alpha_i + \lambda_t + u_{it}$. This model also can be written as

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \delta_2 B2_t + \cdots + \delta_T BT_t + \gamma_2 D2_i + \cdots + \gamma_n Dn_i + u_{it},$$

where $B2_t = 1$ if $t = 2$ and 0 otherwise, $D2_i = 1$ if $i = 2$ and 0 otherwise, and so forth. How are the coefficients $(\beta_0, \delta_2, \ldots, \delta_T, \gamma_2, \ldots, \gamma_n)$ related to the coefficients $(\alpha_1, \ldots, \alpha_n, \lambda_1, \ldots, \lambda_T)$?

10.6 Do the fixed effects regression assumptions in Key Concept 10.3 imply that $\text{cov}(\widetilde{v}_{it}, \widetilde{v}_{is}) = 0$ for $t \neq s$ in Equation (10.28)? Explain.

10.7 A researcher believes that traffic fatalities increase when roads are icy 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:

a. The researcher collects data on the average snowfall for each state and adds this regressor ($AverageSnow_i$) to the regressions given in Table 10.1.

b. The researcher collects data on the snowfall in each state for each year in the sample ($Snow_{it}$) and adds this regressor to the regressions.

10.8 Consider observations (Y_{it}, X_{it}) from the linear panel data model

$$Y_{it} = X_{it}\beta_1 + \alpha_i + \lambda_i t + u_{it},$$

where $t = 1, \ldots, T$; $i = 1, \ldots, N$; and $\alpha_i + \lambda_i t$ is an unobserved individual-specific time trend. How would you estimate β_1?

10.9 **a.** In the fixed effects regression model, are the fixed entity effects, α_i, consistently estimated as $n \longrightarrow \infty$ with T fixed? (*Hint:* Analyze the model with no X's: $Y_{it} = \alpha_i + u_{it}$.)

b. If n is large (say, $n = 2000$) but T is small (say, $T = 4$), do you think that the estimated values of α_i are approximately normally distributed? Why or why not? (*Hint:* Analyze the model $Y_{it} = \alpha_i + u_{it}$.)

10.10 In a study of the effect on earnings of education using panel data on annual earnings for a large number of workers, a researcher regresses earnings in a given year on age, education, union status, and the worker's earnings in the 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.)

10.11 Let $\hat{\beta}_1^{DM}$ denote the entity-demeaned estimator given in Equation (10.22), and let $\hat{\beta}_1^{BA}$ denote the "before and after" estimator without an intercept, so that $\hat{\beta}_1^{BA} = [\sum_{i=1}^{n}(X_{i2} - X_{i1})(Y_{i2} - Y_{i1})]/[\sum_{i=1}^{n}(X_{i2} - X_{i1})^2]$. Show that, if $T = 2$, $\hat{\beta}_1^{DM} = \hat{\beta}_1^{BA}$. [*Hint:* Use the definition of $\widetilde{X}_{it}$ before Equation (10.22) to show that $\widetilde{X}_{i1} = -\frac{1}{2}(X_{i2} - X_{i1})$ and $\widetilde{X}_{i2} = \frac{1}{2}(X_{i2} - X_{i1})$.]

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.pearsonhighered.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 through 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?

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

---

[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.

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.pearsonhighered.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 through 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 seat belt, but must have another 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.

---

[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.

APPENDIX 10.1 The State Traffic Fatality Data Set

The data are for the contiguous 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 (the tax on a case of beer) was obtained from Beer Institute's *Brewers Almanac*. The drinking age variables in Table 10.1 are binary variables indicating whether the legal drinking age is 18, 19, or 20. The binary punishment variable in Table 10.1 describes the state's minimum sentencing requirements for an initial drunk driving conviction: This variable equals 1 if the state requires jail time or community service and equals 0 otherwise (a lesser punishment). 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 by Professor Christopher J. Ruhm of the Department of Economics at the University of North Carolina.

APPENDIX 10.2 Standard Errors for Fixed Effects Regression

This appendix provides formulas for standard errors for fixed effects regression with a single regressor. These formulas are extended to multiple regressors in Exercise 18.15.

The Asymptotic Distribution of the Fixed Effects Estimator with Large n

The fixed effects estimator. The fixed effects estimator of β_1 is the OLS estimator obtained using the entity-demeaned regression of Equation (10.14) in which $\widetilde{Y}_{it}$ is regressed on $\widetilde{X}_{it}$, where $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i$, $\widetilde{X}_{it} = X_{it} - \overline{X}_i$, $\overline{Y}_i = T^{-1}\sum_{t=1}^{T} Y_{it}$, and $\overline{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$. The formula for the OLS estimator is obtained by replacing $X_i - \overline{X}$ by $\widetilde{X}_{it}$ and $Y_i - \overline{Y}$ by $\widetilde{Y}_{it}$ in

Equation (4.7) and by replacing the single summation in Equation (4.7) by two summations, one over entities ($i = 1, \ldots, n$) and one over time periods ($t = 1, \ldots, T$),[5] so

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}\widetilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^2}. \tag{10.22}$$

The derivation of the sampling distribution of $\hat{\beta}_1$ parallels the derivation in Appendix 4.3 of the sampling distribution of the OLS estimator with cross-sectional data. First, substitute $\widetilde{Y}_{it} = \beta_1\widetilde{X}_{it} + \widetilde{u}_{it}$ [Equation (10.14)] into the numerator of Equation (10.22) to obtain the panel data counterpart of Equation (4.30):

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}\widetilde{u}_{it}}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^2}. \tag{10.23}$$

Next, rearrange this expression and multiply both sides by $\sqrt{nT}$ to obtain

$$\sqrt{nT}(\hat{\beta}_1 - \beta_1) = \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\eta_i}{\hat{Q}_{\widetilde{X}}}, \text{ where } \eta_i = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\widetilde{X}_{it}\widetilde{u}_{it} \text{ and } \hat{Q}_{\widetilde{X}} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^2. \tag{10.24}$$

The scaling factor in Equation (10.24), nT, is the total number of observations.

Distribution and standard errors when n is large. In most panel data applications n is much larger than T, which motivates approximating sampling distributions by letting $n \rightarrow \infty$ while keeping T fixed. Under the fixed effects regression assumptions of Key Concept 10.3, $\hat{Q}_{\widetilde{X}} \xrightarrow{p} Q_{\widetilde{X}} = ET^{-1}\sum_{t=1}^{T}\widetilde{X}_{it}^2$ as $n \rightarrow \infty$. Also, η_i is i.i.d. over $i = 1, \ldots, n$ (by Assumption 2) with mean zero (by Assumption 1) and variance σ_η^2 (which is finite by

[5]The double summation is the extension to double subscripts of a single summation:

$$\begin{aligned}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it} &= \sum_{i=1}^{n}\left(\sum_{t=1}^{T}X_{it}\right)\\ &= \sum_{i=1}^{n}(X_{i1} + X_{i2} + \cdots + X_{iT})\\ &= (X_{11} + X_{12} + \cdots + X_{1T}) + (X_{21} + X_{22} + \cdots + X_{2T}) + \cdots + (X_{n1} + X_{n2} + \cdots + X_{nT}).\end{aligned}$$

Assumption 3), so by the central limit theorem $\sqrt{1/n}\sum_{i=1}^{n}\eta_i \xrightarrow{d} N(0, \sigma_\eta^2)$. It follows from Equation (10.24) that

$$\sqrt{nT}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N\left(0, \frac{\sigma_\eta^2}{Q_{\tilde{X}}^2}\right). \tag{10.25}$$

From Equation (10.25), the variance of the large-sample distribution of $\hat{\beta}_1$ is

$$\operatorname{var}(\hat{\beta}_1) = \frac{1}{nT}\frac{\sigma_\eta^2}{Q_{\tilde{X}}^2}. \tag{10.26}$$

The clustered standard error formula replaces the population moments in Equation (10.26) by their sample counterparts:

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{nT}\frac{s_\eta^2}{\hat{Q}_{\tilde{X}}^2}},$$
$$\text{where } s_\eta^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\eta}_i - \bar{\hat{\eta}})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\hat{\eta}_i^2 \text{ (clustered standard error)} \tag{10.27}$$

where $\hat{\eta}_i = \sqrt{1/T}\sum_{t=1}^{T}\tilde{X}_{it}\hat{u}_{it}$ is the sample counterpart of η_i [$\hat{\eta}_i$ is η_i in Equation (10.24), with $\tilde{u}_{it}$ replaced by the fixed effect regression residual $\hat{u}_{it}$], and $\bar{\hat{\eta}} = (1/n)\sum_{i=1}^{n}\hat{\eta}_i$. The final equality in Equation (10.27) arises because $\bar{\hat{\eta}} = 0$, which in turn follows from the residuals and regressors being uncorrelated [Equation (4.34)]. Note that s_η^2 is just the sample variance of $\hat{\eta}_i$ [see Equation (3.7)].

The estimator s_η^2 is a consistent estimator of σ_η^2 as $n \to \infty$ even if there is heteroskedasticity or autocorrelation (Exercise 18.15); thus the clustered standard error in Equation (10.27) is heteroskedasticity- and autocorrelation-consistent. Because the clustered standard error is consistent, the t-statistic testing $\beta_1 = \beta_{1,0}$ has a standard normal distribution under the null hypothesis as $n \to \infty$.

All the foregoing results apply if there are multiple regressors. In addition, if n is large, then the F-statistic testing q restrictions (computed using the clustered variance formula) has its usual asymptotic $F_{q,\infty}$ distribution.

Why isn't the usual heteroskedasticity-robust estimator of Chapter 5 valid for panel data? There are two reasons. The most important reason is that the heteroskedasticity-robust estimator of Chapter 5 does not allow for serial correlation within a cluster. Recall that, for two random variables U and V, $\text{var}(U + V) = \text{var}(U) + \text{var}(V) + 2\text{cov}(U, V)$. The variance η_i in Equation (10.24) therefore can be written as the sum of variances plus covariances. Let $\tilde{v}_{it} = \widetilde{X}_{it}\tilde{u}_{it}$; then

$$\begin{aligned}\text{var}(\eta_i) &= \text{var}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}\right) = \frac{1}{T}\text{var}(\tilde{v}_{i1} + \tilde{v}_{i2} + \cdots + \tilde{v}_{iT}) \\ &= \frac{1}{T}[\text{var}(\tilde{v}_{i1}) + \text{var}(\tilde{v}_{i2}) + \cdots + \text{var}(\tilde{v}_{iT}) \\ &\quad + 2\text{cov}(\tilde{v}_{i1}, \tilde{v}_{i2}) + \cdots + 2\text{cov}(\tilde{v}_{iT-1}, \tilde{v}_{iT})].\end{aligned} \tag{10.28}$$

The heteroskedasticity-robust variance formula of Chapter 5 misses all the covariances in the final part of Equation (10.28), so if there is serial correlation, the usual heteroskedasticity-robust variance estimator is inconsistent.

The second reason is that if T is small, the estimation of the fixed effects introduces bias into the Chapter 5 heteroskedasticity-robust variance estimator. This problem does not arise in cross-sectional regression.

The one case in which the usual heteroskedasticity-robust standard errors can be used with panel data is with fixed effects regression with $T = 2$ observations. In this case, fixed effects regression is equivalent to the "before and after" differences regression in Section 10.2, and heteroskedasticity-robust and clustered standard errors are equivalent.

For empirical examples showing the importance of using clustered standard errors in economic panel data, see Bertrand, Duflo, and Mullainathan (2004).

Standard Errors When u_{it} Is Correlated Across Entities. In some cases, u_{it} might be correlated across entities. For example, in a study of earnings, suppose that the sampling scheme selects families by simple random sampling, then tracks all siblings within a family. Because the omitted factors that enter the error term could have common elements for siblings, it is not reasonable to assume that the errors are independent for siblings (even though they are independent across families).

In the siblings example, families are natural clusters, or groupings, of observations, where u_{it} is correlated within the cluster but not across clusters. The derivation leading to Equation (10.27) can be modified to allow for clusters across entities (for example, families) or across both entities and time, as long as there are many clusters.

Distribution and Standard Errors When n Is Small

If n is small and T is large, then it remains possible to use clustered standard errors; however, t-statistics need to be compared with critical values from the t_{n-1} tables, and the F-statistic testing q restrictions needs to be compared to the $F_{q,n-q}$ critical value multiplied by

$(n-1)/(n-q)$. These distributions are valid under the assumptions in Key Concept 10.3 plus some additional assumptions on the joint distribution of X_{it} and u_{it} over time within an entity. Although the validity of the t-distribution in cross-sectional regression requires normality and homoskedasticity of the regression errors (Section 5.6), neither requirement is needed to justify using the t-distribution with clustered standard errors in panel data when T is large.

To see why the cluster t-statistic has a t_{n-1} distribution when n is small and T is large, even if u_{it} is neither normally distributed nor homoskedastic, first note that if T is large, then under additional assumptions η_i in Equation (10.24) will obey a central limit theorem, so $\eta_i \xrightarrow{d} N(0, \sigma_\eta^2)$. (The additional assumptions required for this result are substantial and technical, and we defer further discussion of them to our treatment of time series data in Chapter 14.) Thus, if T is large, then $\sqrt{nT}(\hat{\beta}_1 - \beta_1)$ in Equation (10.24) is a scaled average of the n normal random variables η_i. Moreover, the clustered formula s_η^2 in Equation (10.27) is the usual formula for the sample variance, and if it could be computed using η_i, then $(n-1)s_\eta^2/\sigma_\eta^2$ would have a χ_{n-1}^2 distribution, so the t-statistic would have a t_{n-1} distribution [see Section (3.6)]. Using the residuals to compute $\hat{\eta}_i$ and s_η^2 does not change this conclusion. In the case of multiple regressors, analogous reasoning leads to the conclusion that the F-statistic testing q restrictions, computed using the cluster variance estimator, is distributed as $(\frac{n-1}{n-q})F_{q,n-q}$. [For example, the 5% critical value for this F-statistic when $n = 10$ and $q = 4$ is $(\frac{10-1}{10-4}) \times 4.53 = 6.80$, where 4.53 is the 5% critical value from the $F_{4,6}$ distribution given in Appendix Table 5B.] Note that, as n increases, the t_{n-1} and $(\frac{n-1}{n-q})F_{q,n-q}$ distributions approach the usual standard normal and $F_{q,\infty}$ distributions.[6]

If both n and T are small, then in general $\hat{\beta}_1$ will not be normally distributed and clustered standard errors will not provide reliable inference.

[6]Not all software implements clustered standard errors using the t_{n-1} and $(\frac{n-1}{n-q})F_{q,n-q}$ distributions that apply if n is small, so you should check how your software implements and treats clustered standard errors.