Regression with Panel Data [1]

Jasmine(Yu) Hao

Faculty of Business and Economics
Hong Kong University

August 19, 2021

[1] This section is based on Stock and Watson (2020), Chapter 10.

Motivation

- 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**.
- Panel regression is a method for controlling for some types of omitted variables without actually observing them.
- 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.

Example

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* from 1982 to 1988(*7 years*).
  - Unobserved variables:
    - prevailing cultural attitudes toward drinking and driving(*vary by entity*).
    - improvements in the safety of new cars(*vary through time*).

Panel Data I

- 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
  $(X_{it}, Y_{it}), i = 1, \ldots, n,$ and $t = 1, \ldots, T, (10.1)$
  - i, refers to the **entity** being observed
  - t, refers to the **date** at which it is observed.
- 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 $1982, \ldots, 1988$), for a total of $7 * 48 = 336$ observations.
  - Compare cross-sectional data, i denote the entity.
  - Panel data, we need some additional notation to keep track of both the entity and the time period.
  - Use 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 ith of n entities in the t-th of T periods.

Balanced v.s. unbalanced panel I

- 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**.
- E.g. The traffic fatality data set has data for all 48 contiguous U.S. states for all seven years, so it is **balanced**.
- If we did not have data on fatalities for some states in 1983), then the data set would be **unbalanced**.
- The methods presented focus on **balanced panel**;
- 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.

Background I

- 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
  - 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.

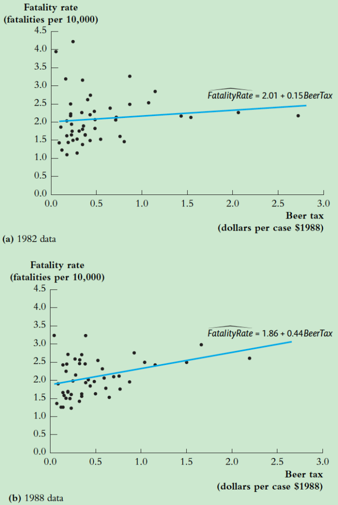
From Model to Results

- 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(**fatality rate**, which is the number of annual traffic deaths per 10,000 people in the population in the state)
  - the type of drunk driving laws in each state in each year,
  - the tax on beer in each state("real" tax on a case of beer, which is the beer tax, put into 1988 dollars by adjusting for inflation.).

OLS regression I

FIGURE 10.1 The Traffic Fatality Rate and the Tax on Beer

Figure 10.1a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Figure 10.1b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



OLS regression II

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 estimated regression line is

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

- The coefficient on the real beer tax is positive but not statistically significant at the 10% level.
- We reexamine this relationship for another year(1988):

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

- The coefficient real beer tax is statistically significant at the 1% level (the t-statistic is 3.43).

OLS regression III

- 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.
- 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.

OLS regression IV

- 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.

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

- When data for each state are obtained for $T = 2$ time periods, it is possible to compare values of the dependent variable when $t = 2$ to values when $t = 1$.
  - Example? Cultural factors.
- 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 **time-invariant** variable that determines the fatality rate in the i-th state.
  - local cultural attitude toward drinking and driving, which changes slowly and thus could be considered to be constant between 1982 and 1988.

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

The population linear regression relating Z_i and the real beer tax to the fatality rate is

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_i$$

- where u_{it} is the error term, $i = 1, \ldots, n$, and $t = 1, \ldots, T$. Because
- Z_i does not change over time, the influence of Z_i can be eliminated by analyzing the change in the fatality rate between the two periods.

Mathematical Derivation: Bofore and After I

$$FatalityRate_{i,1982} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{i,1982}$$
$$FatalityRate_{i,1988} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{i,1988}.$$

- Interpretation: cultural attitudes toward drinking and driving affect the level of drunk driving and thus the traffic fatality rate in a state.
  - If 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.
  - changes in the tax on beer and changes in the error term (which captures changes in other factors that determine traffic deaths).

Difference Regression I

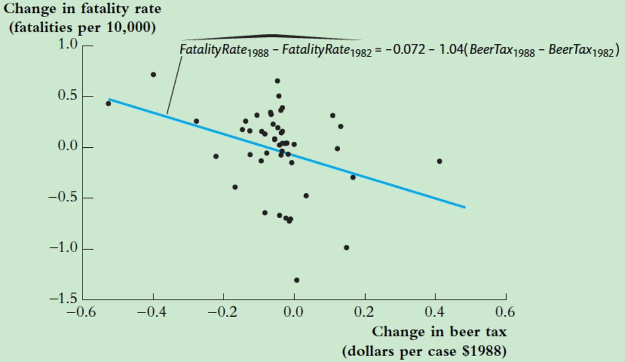
$$
\begin{aligned}
&FatalityRate_{i,1988} - FatalityRate_{i,1982} \\
&= \beta_1(BeerTax_{i,1988} - BeerTax_{i,1982}) + u_{i,1988} - u_{i,1982}.
\end{aligned}
$$

- The specification 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.

Difference Regression II

FIGURE 10.2 Changes in Fatality Rates and Beer Taxes from 1982 to 1988

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



- 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.

Difference Regression III

- 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 estimated OLS regression line is

$$\widehat{FatalityRate_{i,1988} - FatalityRate_{i,1982}} = \underset{(0.065)}{-0.072} - \underset{(0.36)}{1.04}(BeerTax_{i,1988} - BeerTax_{i,1982}).$$

- Intercept: 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 between 1982 and 1988 that reduced the average fatality rate.
- In contrast to the cross-sectional regression results, the estimated effect of a change in the real beer tax is **negative**.

Difference Regression IV

- The hypothesis that the population slope coefficient is 0 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,
  - the estimate suggests that traffic fatalities can be cut in half by increasing the real tax on beer by \$1 per case.
- By examining changes in the fatality rate over time, the regression in controls for fixed factors such as cultural attitudes toward drinking and driving.

Issues I

- Time-varying factors that influence traffic safety, correlated with the real beer tax, then their omission will produce omitted variable bias.
- for now cannot draw substantive conclusions about the effect of real beer taxes on traffic fatalities.
- This "before and after" or "differences" analysis works when the data are observed in two different years.
  - Not directly applicable when $T = 7$.
- 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.

Issues II

- represented by a set of binary (or indicator) 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 I

- 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:

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

- 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 can be interpreted as having n intercepts, one for each state. Specifically, let $\alpha_i = \beta_0 + \beta_2 Z_i$:

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

The Fixed Effects Regression Model II

- The regression 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 comes from considering the population regression line for the ith state; this population regression line is $a_i + \beta_1 X_{it}$.
- 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 can be thought of as the "effect" of being in entity i (in the current application, entities are states), the terms $\alpha_1, \ldots, \alpha_n$, an are known as entity fixed effects.
- The variation in the entity fixed effects comes from omitted variables that, like Z_i, that 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.

The Fixed Effects Regression Model III

- That population regression line was expressed mathematically using a single binary variable indicating one of the groups.
  - 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.
- To develop the fixed effects regression model using binary variables, let D_{1i} be a binary variable that equals 1 when $i = 1$ and equals 0 otherwise, let D_{2i} equal 1 when $i = 2$ and equal 0 otherwise, and so on.
- Note: we cannot include all n binary variables plus a common intercept, the regressors will be perfectly multicollinear.
  - we arbitrarily omit the binary variable D_{1i} for the first entity.

Fixed Effect Regression I

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + +\gamma_n D_{ni} + u_{it}, (10.11)$$

- where $\beta_0, \beta_1, \gamma_2, \ldots, \gamma_n$ are unknown coefficients to be estimated.
- 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 $\beta_0 + \beta_1 X_{it}$, so $\alpha_i = \beta_0 + \gamma_i$ for $i = 2, \ldots, n$.
  - Thus there are two equivalent ways to write the fixed effects regression model.
    - write in terms of n statespecific intercepts.
    - 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.

Extension to multiple X's I

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.

- The binary variable specification of the fixed effects regression model can be estimated by OLS.
- This regression 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 but 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.

Extension to multiple X's II

1. the entity-specific average is subtracted from each variable.
2. the regression is estimated using "entity-demeaned" variables.

Extension to multiple X's III

Consider the case of a single regressor fixed effects model, and take the average of both sides; then $Y_i = \beta_1 X_i + a_i + u_i$, where $\bar{Y}_i = \sum_{t=1}^{T} Y_{it}$, and X_i and u_i are defined similarly.
Then $Y_{it} - \bar{Y}_i = \beta_1(X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i)$.
Let $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, $\tilde{X}_{it} = X_{it} - \bar{X}_i$ and $\tilde{u}_{it} = u_{it} - \bar{u}_i$ accordingly, $\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$.(10.14)
Thus β_1 can be estimated by the OLS regression of the "entity-demeaned" variables

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$$

This estimator is identical to the OLS estimator of β_1 obtained by estimation of the fixed effects model using $n - 1$ binary variables **(Exercise 19.6)**.

The “before and after” (differences) regression versus the binary variables specification I

- When $T = 2$ 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” specification.
- Thus, when $T = 2$, there are three ways to estimate b1 by OLS: the “before and after” specification in Equation (10.7) (without an intercept), the binary variable specification in Equation (textbook 10.11), and the entitydemeaned specification in Equation (10.14). These three methods are equivalent; that is, they produce identical OLS estimates of β_1 **(Exercise 10.11)**.

Application to Traffic Deaths I

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

$$\mathit{FatalityRate} = \underset{(0.29)}{-0.66}\mathit{BeerTax} + \mathit{state\ fixed\ effects}$$

- Conventionally, the estimated state fixed intercepts are not listed.
- the estimated coefficient in the fixed effects regression is negative
  - higher real beer taxes are associated with fewer traffic deaths,
  - the regression is different from the difference regression.
  - The difference regression uses only the data for 1982 and 1988 (specifically, the difference between those two years), whereas the fixed effects regression uses the data for all 7 years.
  - Because of the additional observations, the standard error is smaller in the fixed effect regression.

Application to Traffic Deaths II

- 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.
- Other (time-varying) 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.

Regression with Time Fixed Effects I

- Time fixed effects can control for variables that are constant across entities but evolve over time.
  - Safety improvements in new cars are introduced nationally, they serve to reduce traffic fatalities in all states(omitted variable that changes over time.)
- The population regression with explicit the effect of automobile safety(S_t):

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

- where S_t is unobserved and where the single t subscript emphasizes that safety changes over time but is constant across states.
- $\beta_3 S_t$ 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 I

- Suppose that the variables Z_i are not present, so that the term $\beta_2 Z_i$ can be dropped, 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
- Recall: in the entity fixed effects model, the presence of Z_i leads to the fixed effects regression model: state-specific intercept.
- 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

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

- This model has a different intercept, λ_t, for each time period: the "effect" on Y of time period t.
- The terms $\lambda_1, \ldots, \lambda_T$ are known as **time fixed effects**.

Time Effects Only II

- The variation in the time fixed effects comes from omitted variables that, like S_t in, vary over time but not across entities.

The time fixed effects regression model be represented using $T - 1$ binary indicators

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

- where $\delta_2, \ldots, \delta_T$ are unknown coefficients and where $B2_t = 1$ if $t = 2$ and $B2_t = 0$ otherwise, etc.
- The first binary variable $B1_t$ is omitted to prevent perfect multicollinearity.

Both Entity and Time Fixed Effects I

- 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

$$Y_{it} = \beta_1 X_{it} + a_i + \lambda_t + u_{it},$$

- where a_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:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \ldots + \gamma_n D_n i + \delta_2 B2_t + \ldots + \delta_T BT_t + u_{it}.$$

Application to traffic deaths I

Adding time effects to the state fixed effects regression results in the OLS estimate of the regression line:

$$\widehat{FatalityRate} = -\underset{(0.36)}{0.64} BeerTax + State\ Fixed\ Effects + Time\ Fixed\ Effects. \quad (10.21)$$

- 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.

Application to traffic deaths II

- Although this coefficient is less precisely estimated when time effects are included, it is still significant at the 10%, but not the 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.

The sampling distribution, standard errors, and statistical inference I

- In multiple regression with cross-sectional data, if the four least squares assumptions hold, then the sampling distribution of the *OLS* estimator is normal in large samples.
- Similarly, in multiple regression with panel data, if the fixed effects regression assumptions holds,
  - 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,
  - 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.

Assumptions

Model:

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

1. u_{it} has conditional mean 0: $E[u_{it}|X_{i1}, X_{i2}, \ldots, X_{iT}] = 0$.
2. $(X_{i1}, X_{i2}, \ldots, X_{iT}, u_{i1}, \ldots, u_{iT})$ are i.i.d draws from a joint distribution.
3. Large outliers are unlikely: (X_{it}, u_{it}) have finite fourth moment.
4. There is no perfect collinearity.

Assumptions I

These assumptions extend the four least squares assumptions for causal inference, stated for cross-sectional data to panel data.

- The first assumption is that the error term has conditional mean 0 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 and implies that there is no omitted variable bias.
- The requirement that the conditional mean of uit 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 are i.i.d. across entities for $i = 1, \ldots, n$.

Assumptions II

- ◇ The second assumption for fixed effects regression holds if entities are selected by simple random sampling from the population.
- ▷ The third and fourth assumptions for fixed effects regression are analogous to the third and fourth least squares assumptions for cross-sectional data.

Auto-correlation I

- 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 uit 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.

Auto-correlation II

- 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 I

- If the regression errors are autocorrelated, then the usual **heteroskedasticity-robust** standard error formula for cross-section regression 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 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-robust (HAR) standard errors. The standard errors used in this chapter are one type of HAR standard errors, clustered standard errors.

Standard Errors for Fixed Effects Regression II

- 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.
- 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, critical values for F-statistics testing q restrictions.

Standard Errors for Fixed Effects Regression III

- In practice, there can be a large difference between clustered standard errors and standard errors that do not allow for autocorrelation of uit. For example, the usual (cross-sectional data) heteroskedasticity-robust standard error for the Beer-Tax coefficient 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 uit within an entity, whereas the usual heteroskedasticity-robust standard error does not.

Drunk Driving Laws and Traffic Deaths I

- 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.

Drunk Driving Laws and Traffic Deaths II

- 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, a 95% confidence interval for the coefficients on the policy variables of interest, a F-statistic and p-value, or other information about the regression.

TABLE 11.2 Mortgage Denial Regressions Using the Boston HMDA Data

Dependent variable: *deny* = 1 if mortgage application is denied, = 0 if accepted; 2380 observations.

Regression Model / Regressor	*LPM* (1)	*Logit* (2)	*Probit* (3)	*Probit* (4)	*Probit* (5)	*Probit* (6)
black	0.084** (0.023)	0.688** (0.182)	0.389** (0.098)	0.371** (0.099)	0.363** (0.100)	0.246 (0.448)
P/I ratio	0.449** (0.114)	4.76** (1.33)	2.44** (0.61)	2.46** (0.60)	2.62** (0.61)	2.57** (0.66)
housing expense-to-income ratio	−0.048 (0.110)	−0.11 (1.29)	−0.18 (0.68)	−0.30 (0.68)	−0.50 (0.70)	−0.54 (0.74)
medium loan-to-value ratio (0.80 ≤ *loan-value ratio* ≤ 0.95)	0.031* (0.013)	0.46** (0.16)	0.21** (0.08)	0.22** (0.08)	0.22** (0.08)	0.22** (0.08)
high loan-to-value ratio (*loan-value ratio* > 0.95)	0.189** (0.050)	1.49** (0.32)	0.79** (0.18)	0.79** (0.18)	0.84** (0.18)	0.79** (0.18)
consumer credit score	0.031** (0.005)	0.29** (0.04)	0.15** (0.02)	0.16** (0.02)	0.34** (0.11)	0.16** (0.02)
mortgage credit score	0.021 (0.011)	0.28* (0.14)	0.15* (0.07)	0.11 (0.08)	0.16 (0.10)	0.11 (0.08)
public bad credit record	0.197** (0.035)	1.23** (0.20)	0.70** (0.12)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)
denied mortgage insurance	0.702** (0.045)	4.55** (0.57)	2.56** (0.30)	2.59** (0.29)	2.59** (0.30)	2.59** (0.29)
self-employed	0.060** (0.021)	0.67** (0.21)	0.36** (0.11)	0.35** (0.11)	0.34** (0.11)	0.35** (0.11)
single				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
high school diploma				−0.61** (0.23)	−0.60* (0.24)	−0.62** (0.23)
unemployment rate				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
condominium					−0.05 (0.09)	
black × *P/I ratio*						−0.58 (1.47)
black × *housing expense-to-income ratio*						1.23 (1.69)
additional credit rating indicator variables	no	no	no	no	yes	no
constant	−0.183** (0.028)	−5.71** (0.48)	−3.04** (0.23)	−2.57** (0.34)	−2.90** (0.39)	−2.54** (0.35)

Comments I

- 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 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.

Comments II

- 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.
- 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.

Comments III

- 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 * 0.50 $=$ 0.23 deaths per 10,000. This estimated effect is large: Because the average fatality rate is 2 deaths 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 $-0.45 * 0.50 \pm 1.96 * 0.30 * 0.50 = (-0.52, 0.08)$. This wide 95% confidence interval includes 0, 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 0 cannot be rejected at the 10% significance level: The F-statistic testing the joint hypothesis that the three coefficients are 0 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 0 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 1 percentage point is estimated to reduce traffic fatalities by 0.063 deaths 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 deaths per 10,000. 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 there is 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 regressionin 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 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.

References I

Stock, J. H. and Watson, M. W. (2020). *Introduction to econometrics*, volume 4. Pearson New York.