BUSS975 Causal Inference in Financial Research

Ji-Woong Chung
chung_jiwoong@korea.ac.kr
Korea University Business School

Chapter 6

Panel Data

6.1 Introduction: Why Panel Data is Helpful

Omitted variable bias is a fundamental challenge in empirical economics. When important explanatory variables are not included in a regression model, the estimated coefficients on included variables can be biased and inconsistent. This problem is especially severe if the omitted factors are correlated with the regressors of interest. Unfortunately, many such omitted variables are inherently unobservable (e.g. innate ability, firm culture, risk preferences), making it impossible to include them directly as controls. In this context, panel data (also called longitudinal data) can offer a powerful solution by allowing researchers to control for certain unobserved characteristics.

To motivate the usefulness of panel data, consider an example from corporate finance. Suppose we are interested in the relationship between a firm's leverage (debt-to-assets ratio) and its profitability (net income-to-assets). We might start with a simple pooled regression model for firm i, operating in industry j, in year t:

$$\text{leverage}_{i,j,t} = \beta_0 + \beta_1 \text{profit}_{i,j,t} + u_{i,j,t}\,,$$

where $u_{i,j,t}$ is the regression error term. The coefficient β_1 would capture the marginal effect of profitability on leverage under the assumption that, conditional on profit, the error term is uncorrelated with profit. However, this assumption is dubious: it is easy to think of many omitted variables that might affect a firm's leverage and also be correlated with its profit.

What are some likely omitted factors in this leverage example? One possibility is managerial talent or risk aversion: more talented or risk-tolerant managers might both achieve higher profits and choose a different leverage policy. Another potential omitted variable is industry conditions: favorable industry demand shocks could raise profits and simultaneously enable firms to borrow more (or reduce their need to borrow). The cost of capital faced by the firm could influence leverage (firms with access to cheaper credit borrow more) and might be related to profitability. Other examples include investment opportunities (firms with better growth opportunities might have higher profits and also choose particular leverage levels) and market sentiment or credit market conditions that vary over time and affect both profits and leverage. All these factors are typically unobserved by the econometrician or hard to measure, yet they could bias our estimate of β_1. For instance, if more profitable firms operate in booming industries (unobserved positive industry shock) and simultaneously carry higher leverage, a simple OLS regression of leverage on profit would attribute to profit an effect that partly reflects industry conditions.

The omitted variables problem becomes even more pronounced when using data across heterogeneous groups such as different regions or countries. For example, if we pool firms from various countries, there may be unobserved country-level differences like the strength of institutions, enforcement of property rights, financial market development, or investor sentiment. A firm's location might proxy for these factors. If these country-specific factors influence both firm profitability and capital structure, then comparing firms across countries without accounting for such differences would be misleading.

One common approach to mitigating omitted variable bias is to include proxy variables—observable variables that are thought to be correlated with the unobservable factor. In our leverage example, one might try to proxy for managerial ability using characteristics like the manager's education or for investor sentiment using stock market indicators. However, for a proxy variable to yield consistent estimates of other coefficients, a strong assumption is required: the proxy must be so good that, after controlling for it, the remaining unobserved component is uncorrelated with the other regressors. In other words, the proxy effectively absorbs all the correlation between the unobserved factor and the included explanatory variables. In practice, finding such perfect proxies is difficult, and using imperfect proxies can still leave substantial bias.

Panel data to the rescue: If we have data on the same observational units (e.g. the same firms) over multiple time periods, we can exploit the structure of the data to deal with a particular type of omitted variable. Specifically, panel data allows us to control for unobserved time-invariant factors. Many omitted variables, such as innate ability (in individual wage regressions) or corporate culture (in firm outcomes), are plausibly constant over time for a given individual or firm. Panel data methods enable us to difference or subtract out these time-invariant factors, eliminating their influence on the estimated effect of interest. The key insight is that while we cannot observe or measure f_i (say, the ability of individual i or the management quality of firm i), if f_i remains fixed over time, then changes in the outcome for unit i over time cannot be due to changes in f_i. By focusing on within-unit variation (how y_{it} changes as x_{it} changes for the same i), we can control for any characteristic of unit i that does not change over time.

In summary, panel data provides an effective way to control for omitted variables that are constant within an entity but vary across entities. This helps us get closer to causal inference by removing one important source of bias. In the rest of this chapter, we will develop the fixed effects model, which formalizes this idea, and discuss its statistical properties, benefits, and limitations. We will also compare it to alternative panel data methods such as random effects and first differencing, and address special cases like dynamic models with lagged dependent variables.

6.2 The Fixed Effects Model

6.2.1 Panel Data Basics and Notation

A dataset is called **panel data** (or longitudinal data) if it contains multiple observations on each individual or entity in the sample. We will use i to index the individual (or cross-sectional unit) and t to index time. The total number of individuals is N and the number of time periods (observations per individual) is T. In a **balanced panel**, each individual is observed in all T periods, so the total number of observations is $N \times T$. (If different individuals have different numbers of observations, the panel is unbalanced. Most of the methods we discuss apply to unbalanced panels as well, though for simplicity we often assume a balanced panel in theoretical exposition.)

Panel data examples in economics and finance include:

- Firm-level data: e.g. 5,000 firms observed annually for 20 years ($N = 5000$, $T = 20$).
- Household or individual data: e.g. 1,000 households tracked over 10 years ($N = 1000$, $T = 10$).
- Repeated observations of countries or regions: e.g. 50 states in the US observed over multiple decades.

In panel datasets, we can think of each individual i having a vector of observations $(y_{i1}, y_{i2}, \ldots, y_{iT})$ for the dependent variable and similarly $(x_{i1}, \ldots, x_{iT})$ for each explanatory variable. The power of panel data comes from the possibility of controlling for **individual-specific effects**. In a regression setting, a general model for panel data can be written as:

$$y_{it} = \alpha + \beta x_{it} + f_i + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where

- y_{it} is the outcome for individual i at time t,
- x_{it} is a vector of one or more observed explanatory variables (we can treat β as a vector if there are multiple regressors),
- α is a constant term (common intercept),
- f_i represents the unobserved individual-specific effect (the **fixed effect** for individual i),
- u_{it} is the idiosyncratic error term that varies over both i and t.

The term f_i captures all unobservable influences on y_{it} that are peculiar to individual i and do not change over time. We often call f_i **unobserved heterogeneity** or an **unobserved fixed effect**. For example, in a wage regression y_{it} might be the wage of person i in year t, x_{it} could include education and experience, and f_i could represent that person's innate ability or family background, which is not observed but affects wages and remains constant for that person.

It is important to clarify the assumptions we make about the error components f_i and u_{it}. A basic set of assumptions for the fixed effects model is:

1. $E(u_{it}) = 0$ for all i, t. (The idiosyncratic errors have mean zero, given the regressors and individual effect.)

2. $\text{Var}(u_{it}) = \sigma_u^2$ and $\text{Cov}(u_{it}, u_{js}) = 0$ for $i \neq j$ or $t \neq s$. (Different individuals' errors are uncorrelated, and a given individual's errors are uncorrelated over time. This is a baseline assumption of no autocorrelation in the idiosyncratic errors and no cross-sectional error dependence, which can be relaxed later. We also often assume homoskedasticity of u_{it} here for simplicity.)

3. f_i is constant for each i (by construction) and may be correlated with the regressors x_{it}. However, f_i itself does not vary over t, and crucially we assume f_i has zero mean or is absorbed into the intercept α (so we don't introduce bias in the intercept).

4. The regressors are uncorrelated with the idiosyncratic errors across all time periods: for all i, t, and s,

$$\text{Cov}(x_{it}, u_{is}) = 0\,,$$

   or equivalently $E(u_{is} \mid x_{i1}, x_{i2}, \ldots, x_{iT}, f_i) = 0$. This assumption is known as **strict exogeneity** of the regressors with respect to the idiosyncratic error. It is stronger than the usual contemporaneous exogeneity (which would only require $E(u_{it} \mid x_{it}) = 0$) because it rules out any correlation between x at time t and the error at any other time s. In words, strict exogeneity means that after controlling for the observed x and the fixed effect f_i, there are no feedback effects or anticipation effects: the explanatory variables in any period are effectively predetermined and unaffected by past shocks u, and they do not predict future shocks either.

Assumption (4) of strict exogeneity is critical for the fixed effects estimator (and the first-difference estimator) to be consistent. If x_{it} is correlated with $u_{i,t+1}$ or any future (or past) error term, then the within-group transformations we use in fixed effects estimation will not eliminate all sources of endogeneity. For instance, if a firm experiences an unusually low leverage $u_{i,t}$ shock and then responds by adjusting its future $x_{i,t+1}$ (profitability or some other regressor), this would violate strict exogeneity. We will discuss later what can be done when strict exogeneity fails (e.g. using instrumental variables).

6.2.2 Omitted Variable Bias in Panel Data

Before introducing the fixed effects solution, let us examine the bias that arises if we ignore the unobserved f_i. If we estimate a naive pooled OLS regression of y_{it} on x_{it}, omitting f_i, the model is:

$$y_{it} = \alpha + \beta x_{it} + v_{it},$$

where the combined error term is $v_{it} = f_i + u_{it}$. Here v_{it} contains the unobserved heterogeneity f_i which was left out of the regression, as well as the idiosyncratic error u_{it}. By construction, f_i is constant over t for each i, so v_{it} is serially correlated (since $v_{i,t}$ and $v_{i,s}$ share the common component f_i for any t, s).

If $\mathrm{Cov}(x_{it}, f_i) \neq 0$, which is the usual case we worry about, then x_{it} will be correlated with the composite error v_{it}. This violates the OLS assumption of exogeneity of regressors and leads to **omitted variable bias (OVB)** in the OLS estimate of β.

We can derive the direction and magnitude of this bias in a simple scenario for intuition. Consider a simplified case with a single regressor (scalar x) and suppose we ignore f_i. The probability limit of the OLS estimator can be expressed as:

$$plim\ \hat{\beta}_{\mathrm{OLS}} = \beta + \frac{\mathrm{Cov}(x_{it}, f_i)}{\mathrm{Var}(x_{it})}\,\delta\,,$$

where δ is the coefficient on f_i in the true model (in our formulation above, the true model was $y_{it} = \alpha + \beta x_{it} + \delta f_i + u_{it}$). This formula shows that the OLS estimator is biased by an amount equal to the coefficient on the omitted variable (δ) times the regression coefficient of f_i on x_{it} (i.e. $\mathrm{Cov}(x, f)/\mathrm{Var}(x)$).

In practice we don't know δ or the correlation of x with f, but we often can sign this bias. If f_i is positively correlated with x_{it} and $\delta > 0$ (i.e. the omitted factor has a positive effect on y), then $\hat{\beta}_{\text{OLS}}$ will be upward-biased (too large). If $\delta > 0$ but the correlation is negative, the bias is downward (too small, potentially even flipping the sign of the estimate).

Example (continued): In the leverage-profit regression example, one omitted factor was **managerial risk aversion**. Suppose more risk-averse managers tend to both maintain lower leverage ratios (negative δ if we treat f_i = managers' risk aversion level) and also achieve perhaps somewhat lower profits (imagine risk-averse managers forgo some risky profitable projects, so $\text{Cov}(\text{profit}, f_i) > 0$ if f_i represents risk aversion since higher f_i means more risk aversion correlated with lower profit). In this case, x (profit) is negatively correlated with f (risk aversion), and $\delta < 0$ (risk aversion lowers leverage). The product $\frac{\text{Cov}(x,f)}{\text{Var}(x)}\delta$ would be positive (because $\text{Cov}(x, f) < 0$ and $\delta < 0$ gives a positive product). Thus $\hat{\beta}_{\text{OLS}}$ would be β plus a positive bias term, meaning OLS overestimates the true effect of profit on leverage. Intuitively, some of the negative effect of risk aversion on leverage is being falsely attributed to profit, because risk-averse (high f_i) managers produce lower profits and also choose lower leverage.

This analysis underscores why pooling data without accounting for unobserved heterogeneity can lead us astray. Many variables of interest in economics are correlated with traits that vary across individuals (or firms, countries, etc.) but are stable over time. If those traits also influence the outcome, a simple cross-sectional or pooled analysis will be biased. Panel data offers a way out of this conundrum by allowing us to difference out or control for the f_i term.

6.2.3 The Within Transformation and Fixed Effects Estimation

The key to eliminating the bias due to f_i is to remove f_i from the regression equation. Since f_i is constant for each individual i, one very useful transformation is to **demean** the data for each individual. Define the time average for individual i of each variable:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \quad \bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}\,.$$

(Note: if the panel is unbalanced, we would average over the periods available for each i; assume balanced panel for simplicity.) Also, since f_i is constant over t, $\bar{f}_i = f_i$ for all i. Now consider the average of the regression model over t for a given i:

$$\bar{y}_i = \alpha + \beta\bar{x}_i + \delta f_i + \bar{u}_i\,.$$

Next, subtract this time-averaged equation from the original equation $y_{it} = \alpha + \beta x_{it} + \delta f_i + u_{it}$ for each t. The result is:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + \delta(f_i - f_i) + (u_{it} - \bar{u}_i)\,.$$

This simplifies to:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)\,. \tag{6.1}$$

All mention of f_i has disappeared from the equation, as has the intercept α (because subtracting the mean removes any constant term). The transformation we just applied is called the **within transformation** or **demeaning within groups**. It subtracts the individual-specific mean from each variable, yielding variables expressed as deviations from the individual's average.

Equation (6.1) is the core of the **fixed effects (FE) estimator**. We can now estimate β consistently by applying OLS to this transformed equation:

$$\tilde{y}_{it} = \beta\tilde{x}_{it} + \tilde{u}_{it}\,,$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_i$ and $\tilde{x}_{it} = x_{it} - \bar{x}_i$ are the demeaned variables, and $\tilde{u}_{it} = u_{it} - \bar{u}_i$ is the transformed error term. The unobserved effect f_i was the source of omitted variable bias; since it is time-invariant, demeaning the data has exactly removed it.

Intuitively, we are now using **within-individual variation** over time to estimate β. Each individual's mean has been subtracted out, so any purely cross-sectional difference between individuals (which could be due to different f_i values or other time-invariant factors) no longer influences the estimate. The coefficient β in this transformed regression is identified by the question:

for a given individual, when x is higher than usual, is y higher than usual? By comparing each individual to themselves at different times, we control for all stable characteristics of that individual.

For the FE estimator to be consistent, we require that the transformed regressor $\tilde{x}_{it}$ is uncorrelated with the transformed error $\tilde{u}_{it}$. Under the earlier assumption of strict exogeneity, this will hold. To see this, note that

$$\tilde{u}_{it} = u_{it} - \bar{u}_i = u_{it} - \frac{1}{T}\sum_{s=1}^{T} u_{is} \,.$$

Given strict exogeneity, $E(x_{it}u_{is}) = 0$ for all s, and thus $E(x_{it}\bar{u}_i) = \frac{1}{T}\sum_s E(x_{it}u_{is}) = 0$. Also $E(x_{it}u_{it}) = 0$. Therefore $E(x_{it}(u_{it} - \bar{u}_i)) = 0$. This implies $E(\tilde{x}_{it}\tilde{u}_{it}) = 0$, i.e. $\tilde{x}$ is uncorrelated with $\tilde{u}$. In words, if the original x had no correlation with past, present, or future u (strict exogeneity), then deviations of x from its mean are uncorrelated with deviations of u from its mean, making the within-estimator unbiased and consistent.

Performing OLS on the within-transformed data yields the **within estimator $\hat{\beta}_{\text{FE}}$**. This is also commonly referred to as the **fixed effects estimator**. We will often use the term "fixed effects model" to refer to the model that includes f_i for each individual, and "fixed effects estimation" to refer to the within transformation approach that effectively estimates that model. Note that when we run this regression, we do not include an intercept, because the intercept would be absorbed by the demeaning (every individual's mean was subtracted, so the demeaned data automatically have zero mean for both y and x).

A small but important detail: by performing the within transformation, we have used up one degree of freedom per individual (the individual's mean, essentially equivalent to estimating one f_i for each individual). Thus, the total degrees of freedom for the regression will be $NT - K - N$, where K is the number of regressors (elements in x_{it} not including the intercept). In other words, we lose N degrees of freedom compared to a pooled OLS regression without fixed effects. If N is large, this is usually not a big concern, but it matters for standard error calculation and statistical significance; we will let statistical software handle this automatically.

To summarize, the fixed effects estimator differences out time-invariant omitted variables by using within-individual variation. Under the maintained assumptions (notably strict exogeneity), $\hat{\beta}_{\text{FE}}$ is a consistent estimator of the true effect β.

6.2.4 The Dummy Variable (LSDV) Approach to Fixed Effects

The within transformation is algebraically elegant, but there is another equivalent way to arrive at the fixed effects estimator that can provide intuition: include a dummy variable for each individual. This approach is sometimes called the **Least Squares Dummy Variable (LSDV)** model. The idea is simple: instead of subtracting the means, we explicitly model the individual-specific intercepts. That is, we write:

$$y_{it} = \alpha + \beta x_{it} + \gamma_2 D_{2,i} + \gamma_3 D_{3,i} + \cdots + \gamma_N D_{N,i} + u_{it} \,,$$

where $D_{j,i}$ is a dummy (indicator) variable that equals 1 if $i = j$ and 0 otherwise. In this formulation, we have an intercept α (which will serve as the base level for individual 1, say) and a dummy for each of the other $N-1$ individuals. The coefficient γ_i on the dummy D_i essentially estimates that individual's fixed effect f_i. For example, if individual i consistently has outcomes higher than predicted by $\alpha + \beta x_{it}$, then γ_i will pick up that excess as a positive number. Collectively, these dummies absorb all between-individual variation.

Because we introduced a full set of individual dummies along with a constant, we have perfect multicollinearity (the dummies sum to one and equal the intercept). In practice, one way to handle this is to omit the overall intercept α and include N dummies (one for each individual). Alternatively, one can include the intercept and omit one of the D_i dummies (e.g. drop $D_{1,i}$ for individual 1). Either way, the model has N intercept terms (one for each individual, either explicitly or implicitly). The coefficient on any particular dummy γ_i represents $\alpha + f_i$ for that individual (if we included a common intercept α and dropped one dummy, then γ_i for $i > 1$ would represent $f_i - f_1$, the difference between individual i's fixed effect and the first individual's fixed effect which is absorbed into α).

The LSDV estimator is obtained by running OLS on this dummy-variable model. This will yield identical estimates of β (and identical standard errors for β) as the within estimator described earlier. The reason is given by the Frisch-Waugh-Lovell theorem (partial regression): if you run a regression of y on x and a full set of dummies, the coefficient on x is the same as if you first removed (projected out) the influence of the dummies from both y and x and then regressed the residualized y on the residualized x. But "removing the influence of the dummies" means demeaning within each individual, since regressing a variable on the set of dummies for individuals effectively yields the individual-specific mean as the fitted value. Thus y residualized on the dummies is $y_{it} - \hat{\alpha} - \hat{f}_i = y_{it} - \bar{y}_i$ (since the OLS fit within each group is just the group average), and similarly x residualized on dummies is $x_{it} - \bar{x}_i$. Regressing these residuals is exactly the within estimator. So $\hat{\beta}_{\text{LSDV}} = \hat{\beta}_{\text{FE}}$. In short, including a dummy for each individual produces the same within-group differences result as explicitly demeaning the data.

The LSDV approach is conceptually straightforward — you are literally controlling for unobserved heterogeneity by introducing an intercept for each cross-sectional unit. However, it is often not practical to estimate a model with N dummy variables when N is large (imagine thousands of firms, each needing a dummy). It also clutters the output, and the individual dummy coefficients are usually not of direct interest (since f_i itself is not usually our focus). Therefore, in practice researchers use the within-transformation (via specialized routines in statistical software) to estimate fixed effects models without ever explicitly creating all those dummies.

A practical note: if you do use the dummy variable approach, recall that one dummy (or the intercept) must be omitted to avoid the dummy trap. The interpretation of the intercept in that case is just the average y when all x's are zero for the omitted group. Usually this intercept is not meaningful, and software might output it in a possibly confusing way. For example, Stata's `xtreg, fe` command automatically handles the demeaning internally and then reports an intercept which is actually the average of all the individual fixed effects (the mean of $\hat{f}_i$). This intercept is not generally of substantive interest, and one should not interpret it as one would a usual regression constant.

Even though LSDV and the within estimator yield the same $\hat{\beta}$, you will notice differences in reported fit measures like R^2. If you include N dummy variables, the model will typically explain much more of the total variance in y (possibly the R^2 will be very high) because those dummies soak up all between-individual variation. In contrast, the fixed-effects (within) regression reports an R^2_{within}, which is computed based only on variation within individuals. Many software packages output three R^2 values for panel regressions: within R^2, between R^2, and overall R^2. The within R^2 measures the proportion of variance in $\tilde{y}_{it} = y_{it} - \bar{y}_i$ explained by $\tilde{x}_{it}$; it is the relevant measure of goodness-of-fit for the fixed effects model. The between R^2 would be from a regression on individual means (how much of the variation in $\bar{y}_i$ is explained by $\bar{x}_i$), and the overall is like an R^2 from pooled OLS on the original y_{it}, x_{it}. In fixed effects analysis, we tend to focus on within R^2. The within R^2 is often smaller than the overall R^2 because much of the variation in y (between individuals) is removed before estimation. That is not a problem; it simply reflects that once we control for individual heterogeneity, the remaining variation to be explained by x is less.

6.2.5 Interpreting Fixed Effects Estimates

The coefficient estimates from a fixed effects model have a specific interpretation: they measure the effect of a change in the regressor on the outcome **holding constant all time-invariant characteristics of the individual**. Since each individual serves as their own control, any time-constant attribute (observed or unobserved) is accounted for. For example, if we estimate a wage equation with individual fixed effects, the coefficient on education would be identified by individuals whose education level changes over time (e.g. they obtain an extra qualification) and how that change affects their wage, controlling for their personal fixed ability, etc. If education does not change for most people in the sample (say we consider only prime-age workers over a short horizon), the fixed effect estimate of the return to education may rely on very limited variation (only those who, for instance, complete a degree during the panel). In extreme cases, if a regressor does not vary at all for a given individual, that individual's data does not contribute to identifying that regressor's effect.

It is crucial to remember that fixed effects remove *all* between-individual variation. Thus, the FE estimator cannot identify the impact of any variable that has no within-individual variation. If a variable z_i is constant over time for each individual (e.g. gender in a short panel, or a firm's founding year, or a country's legal origin), then we cannot estimate a coefficient on z_i in a model that includes individual fixed effects. Such a variable is perfectly collinear with the individual dummies (or absorbed in the demeaning). The fixed effect model has in a sense "swept out" those effects entirely. This is often fine (we generally include fixed effects precisely because we suspect those constant factors cause bias and we do not necessarily care to estimate their coefficients). But if one of your key variables of interest is time-invariant, a fixed effects approach will not be able to estimate its effect. In that case, you must use alternative strategies (see discussion of random effects or Hausman-Taylor estimators below, which rely on additional assumptions to identify such effects).

Finally, when using fixed effects, one must consider statistical inference. The within-estimator essentially assumes u_{it} are i.i.d. or at least uncorrelated over time (once f_i is removed). In many panel applications, the idiosyncratic errors u_{it} might still have serial correlation or heteroskedasticity within an individual's time series (for example, a firm might have persistent shocks over time). The fixed effects transformation does not automatically fix serial correlation or heteroskedasticity in u_{it}. Therefore, it is common to use **clustered standard errors** at the individual level when reporting fixed effects regression results. Clustering by i allows for an arbitrary variance-covariance structure for u_{it} across t within the same i (while still assuming independence across i). This adjustment produces consistent standard error estimates even if u_{it} is autocorrelated or heteroskedastic within each individual's observations. In practical terms, one might report robust standard errors clustered at the panel unit level for valid inference.

6.3 Statistical Properties of the Fixed Effects Estimator (Optional)

In this section, we consider the statistical properties (consistency, asymptotic distribution) of the fixed effects estimator. We focus on the large N, fixed T asymptotic scenario, which is the typical panel data context: we have many individuals but only a few time periods per individual. (There is an alternative large T asymptotic scenario for panel data, but it is less common in microeconometrics applications except in macroeconomic panels or when yearly data spans many decades.)

Consistency: Under the assumptions stated earlier (notably strict exogeneity of x with respect to u and that f_i is allowed to correlate with x arbitrarily), the fixed effects estimator $\hat{\beta}_{\text{FE}}$ is consistent as $N \to \infty$ (with T fixed). Intuitively, as the number of individuals grows, we get more and more independent pieces of within-person variation to pin down β. The key requirement is that each individual's error term u_{it} is well-behaved (e.g. zero mean, no serial correlation or at least not too strong, finite variance) and that x_{it} has enough variation within each person and is exogenous as described. Then $\hat{\beta}_{\text{FE}}$ converges in probability to the true β. To sketch a proof of consistency, one can examine the formula for $\hat{\beta}_{\text{FE}}$. The OLS estimator on the transformed model can be written in closed form as:

$$\hat{\beta}_{\text{FE}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}^2}\,,$$

where $\tilde{x}_{it} = x_{it} - \bar{x}_i$ and $\tilde{y}_{it} = y_{it} - \bar{y}_i$. Substitute $y_{it} = \beta x_{it} + f_i + u_{it}$ (assuming α absorbed into f_i for simplicity) into $\tilde{y}_{it}$: note that $\bar{y}_i = \beta\bar{x}_i + f_i + \bar{u}_i$, so $\tilde{y}_{it} = \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) = \beta\tilde{x}_{it} + \tilde{u}_{it}$. Then we have:

$$\hat{\beta}_{\text{FE}} = \frac{\sum_{i,t}\tilde{x}_{it}(\beta\tilde{x}_{it} + \tilde{u}_{it})}{\sum_{i,t}\tilde{x}_{it}^2} = \beta + \frac{\sum_{i,t}\tilde{x}_{it}\tilde{u}_{it}}{\sum_{i,t}\tilde{x}_{it}^2}\,.$$

By assumption $E(\tilde{x}_{it}\tilde{u}_{it}) = 0$ (no correlation between transformed regressor and error). We can consider the numerator $\frac{1}{N}\sum_{i,t}\tilde{x}_{it}\tilde{u}_{it}$ and the denominator $\frac{1}{N}\sum_{i,t}\tilde{x}_{it}^2$. As $N \to \infty$, by the Law of Large Numbers,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{u}_{it} \to_p E(\tilde{x}_{it}\tilde{u}_{it}) = 0\,,$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}^2 \to_p E(\tilde{x}_{it}^2)\,,$$

which is some positive number (assuming there is within-variation in x). Thus, in the probability limit, $\hat{\beta}_{\text{FE}} \to_p \beta + \frac{0}{E(\tilde{x}^2)} = \beta$. This establishes consistency (again under the maintained assumptions, including no serious multicollinearity issues etc.). The important point is that we need $N \to \infty$; if the number of individuals is small, fixed effects may not give reliable estimates (in fact β will not converge to a point as you increase T alone while N fixed unless $T \to \infty$ too, but that is another asymptotic regime).

Asymptotic Normality: Not only is $\hat{\beta}_{\text{FE}}$ consistent, it is also asymptotically normal (again for large N, fixed T). The asymptotic distribution can be derived using a central limit theorem argument. Continuing from the expression above:

$$\sqrt{N}(\hat{\beta}_{\text{FE}} - \beta) = \left(\frac{1}{N} \sum_{i,t} \tilde{x}_{it}^2 \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i,t} \tilde{x}_{it} \tilde{u}_{it} \,.$$

The first term (inverse of average $\tilde{x}^2$) converges in probability to $[E(\tilde{x}^2)]^{-1}$, a constant matrix (or scalar in the single-regressor case). The second term $\frac{1}{\sqrt{N}} \sum_{i,t} \tilde{x}_{it}\tilde{u}_{it}$ is a sum of N independent random vectors (when conditioned on the xs, each individual's sum $\sum_t \tilde{x}_{it}\tilde{u}_{it}$ is i.i.d. across i). By the Lyapunov or Lindeberg Central Limit Theorem, this sum is asymptotically normal with mean zero and variance equal to its cross-sectional variance. That is,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{u}_{it} \to_d \mathcal{N}(0, \Omega) \,,$$

where $\Omega = \text{Var}\left(\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right)$. The matrix Ω simplifies to $E\left[(\sum_t \tilde{x}_{it}\tilde{u}_{it})(\sum_t \tilde{x}_{it}\tilde{u}_{it})\right]$. Provided no degeneracies, Ω will be positive-definite. Therefore, by Slutsky's theorem (continuous mapping applied to the first term which converges in probability), we have:

$$\sqrt{N}(\hat{\beta}_{\text{FE}} - \beta) \to_d \mathcal{N}(0, \Sigma_x^{-1} \Omega \Sigma_x^{-1}) \,,$$

where $\Sigma_x = E\left(\sum_t \tilde{x}_{it}\tilde{x}_{it}\right)$ (the average within-group sum of squares of x). In a simpler notation, we often write the asymptotic variance as $(X'_{\text{within}}X_{\text{within}})^{-1}(\Omega_{u_{it}})(X'_{\text{within}}X_{\text{within}})^{-1}$ where $\Omega_{u_{it}}$ captures the within-group error correlation structure.

In practice, to estimate the variance of $\hat{\beta}_{\text{FE}}$ consistently, one can use a robust (sandwich) estimator. One common robust estimator of $\text{Var}(\hat{\beta})$ that is consistent as $N \to \infty$ is:

$$\widehat{\text{Var}}(\hat{\beta}_{\text{FE}}) = \left(\sum_i \sum_t \tilde{x}_{it}\tilde{x}_{it} \right)^{-1} \left(\sum_i \left(\sum_t \tilde{x}_{it}\hat{u}_{it} \right) \left(\sum_t \tilde{x}_{it}\hat{u}_{it} \right)' \right) \left(\sum_i \sum_t \tilde{x}_{it}\tilde{x}_{it} \right)^{-1},$$

where $\hat{u}_{it}$ are the residuals from the fixed effects regression. This variance formula corresponds to clustering at the individual level (it allows an arbitrary covariance matrix for u_{it} within each i). Many software packages will report these "cluster-robust" standard errors if specified, which is usually recommended in panel applications. Under classical assumptions of no serial correlation and homoskedastic u_{it}, a simpler formula could be used, but the robust one is safer.

Thus, $\hat{\beta}_{FE}$ is $\sqrt{N}$-consistent and asymptotically normal, facilitating inference in large samples.

6.4 Advantages of Fixed Effects Estimation

The fixed effects model has several major advantages in applied research:

- Control for time-invariant unobservables: As we have emphasized, the FE estimator allows arbitrary correlation between the observed regressors x_{it} and any unobserved individual-specific effect f_i. This is a very general way to handle omitted variable bias stemming from omitted factors that do not change within an individual. We don't need to specify or measure f_i; it is enough that it is fixed over time and we difference it out. This flexibility is a stark contrast to methods like random effects (or simple OLS) which require that the unobserved f_i be uncorrelated with the x's.

- Intuitive interpretation: The fixed effect coefficient β can be interpreted as the effect of x on y based on variation *within* the same individual (or entity). Many researchers informally describe fixed effects as "each person serves as their own control." This resonates with the idea of a controlled experiment where each individual is compared to themselves under different conditions (here, different values of x). For policy evaluation or causal inference, this within-unit comparison often bolsters credibility, as it eliminates many sources of spurious association that come from comparing different individuals.

- Flexible use of multiple fixed effects: The concept of fixed effects can be extended beyond just one dimension (the individual). We can include additional sets of dummy variables to control for other forms of unobserved heterogeneity. For example:

- We might include time fixed effects (year dummies) δ_t to capture any common shocks or trends affecting all individuals in a given time period (e.g. macroeconomic conditions, technological progress).
- We could include group fixed effects such as industry fixed effects or region fixed effects to control for unobservable differences across industries or regions.
- In a more complicated panel such as employees within firms, we could include firm fixed effects and person fixed effects simultaneously if workers switch firms (this would be a two-way fixed effects model).
- We can even include interaction fixed effects like industry-by-year fixed effects to capture shocks that are common to all firms in an industry in a particular year.

The inclusion of these additional fixed effects is straightforward in estimation (just add the corresponding dummy variables or use methods to absorb them). The interpretation remains similar: e.g. if you include year fixed effects, you are now looking at deviations from the overall year mean, so effectively you compare individuals to others in the same year, focusing on idiosyncratic deviations net of the year shock.

Each fixed effect absorbs a certain pattern of variation: - Individual FE absorb any level differences across individuals. - Year FE absorb any aggregate time-series movements common to all individuals. - Industry-year FE would absorb any shock that is specific to that industry in that year, and so on.

Because of this flexibility, fixed effects models can dramatically reduce bias. For instance, if you fear that your outcome and regressor are both affected by general economic booms and busts, including year dummies will control for that. If you worry that some industries have systematically different levels of productivity (affecting both x and y), including industry dummies will control for that. Essentially, fixed effects can handle any omitted variable that can be described as a group effect for some known group or category.

Broad applicability to grouped data: While textbooks often present fixed effects in the context of individuals observed over time, the approach is applicable in any setting where observations can be grouped and there is potential omitted heterogeneity at the group level. For example, if you have data on students grouped by class (with no time dimension), you could include class fixed effects to control for class-level attributes. Or if you have repeated observations of the same city under different policies, city fixed effects control for all time-invariant city characteristics. In panel data terminology, the "individual" index i need not be an actual person; it could be a firm, a country, a school, etc., and t could index time or any ordered sequence of observations for that unit. The main requirement for consistency is that the number of groups N is large, as N effectively plays the role of sample size in the asymptotic analysis.

As a side note, when we include a large number of fixed effects (say industry, year, region, etc. simultaneously), we have to be careful about interpretation and multicollinearity. If we include fixed effects for many dimensions, we cannot separately identify the effect of variables that are themselves functions of those fixed effects (e.g., if we include state fixed effects and year fixed effects, we cannot also include a state-specific time trend unless we omit something or use a different method, because a state-specific trend can look like an interaction of state and year fixed effects).

In summary, the fixed effects model is very general and imposes minimal structure on the data: basically that a certain form of heterogeneity is constant within a group. This is often far more plausible than assuming the heterogeneity is absent or uncorrelated with the regressors. This generality is a major reason fixed effects regressions are ubiquitous in empirical economics, especially in fields like labor economics, development, and corporate finance, where unobserved ability, culture, or institutional differences abound.

6.5 Limitations of Fixed Effects Models

Despite their many advantages, fixed effects models also come with several important limitations and potential costs. Researchers must be mindful of these when choosing the fixed effects approach and interpreting results.

6.5.1 Inability to Estimate Time-Invariant Regressors

The most obvious limitation of the fixed effects transformation is that it wipes out all time-invariant information. If an explanatory variable does not change over time for a given individual, we cannot identify its effect separately from the individual fixed effect. The variable is perfectly collinear with the fixed effect. For example, suppose we are studying CEO compensation with a panel of CEO-year observations (where a CEO might switch firms, so we track CEOs over time possibly in different firms). We include CEO fixed effects to control for unobserved CEO talent. Now one regressor of interest might be a dummy $female_i$ indicating whether the CEO is female. This variable is constant for each individual CEO (their gender does not change over time). If we include CEO fixed effects in the model:

$$ln(\text{total pay})_{ijt} = \alpha + \beta_1 ln(\text{firm size}_{ijt}) + \beta_2 \text{female}_i + \delta_t + f_i + \lambda_j + u_{ijt},$$

where i indexes the CEO, j the firm, and t year, and we have included CEO fixed effect f_i, year fixed effects δ_t, and firm fixed effects λ_j. In this setup, the female dummy is perfectly collinear with the CEO fixed effect (each CEO i has a fixed effect f_i and either female$=0$ or 1 which is just a different mean for that CEO). The regression software will thus automatically drop the female dummy or the CEO dummy to resolve collinearity. We will not get an estimate for β_2 at all from a fixed-effects regression. The effect of CEO gender on pay cannot be disentangled from the generic CEO-specific effect.

This is a general point: any characteristic that is fixed for an individual (or group) cannot be studied within that individual using fixed effects. In some cases, this is acceptable because those effects are nuisance parameters we didn't care to estimate (e.g. you might not care to estimate the effect of being a particular person, you just want to control for it). But in other cases, it is a drawback: if the effect of a policy or trait that doesn't vary over time is of interest, fixed effects won't recover it. For instance, if you wanted to estimate the effect of a country being landlocked on its trade volume, a country fixed effects model would drop the landlocked dummy (since a country is either always landlocked or not).

Caution: Statistical software like Stata's `xtreg, fe` is smart enough to automatically omit time-invariant variables. However, if a researcher manually creates dummy variables or fixed effect categories and includes them in OLS, one must be careful. Some software will handle multicollinearity by arbitrarily dropping one of the collinear variables. If one naively includes a full set of dummies and a time-invariant regressor, the program might drop one of the dummy variables (perhaps the first individual's dummy) instead of the time-invariant regressor. In that case, the coefficient reported for the time-invariant regressor is meaningless — it's not actually identified; it effectively absorbed what was the fixed effect of the omitted individual. This can lead to incorrect interpretation if one does not realize the variable was collinear. The safe practice is to let the fixed effects estimator or a careful manual procedure handle such cases, and to understand that any variable with no within-group variation cannot be identified.

Is there any solution? If we truly need to estimate the effect of a time-invariant regressor, one approach is to relax the assumption that f_i can be arbitrarily correlated with x_{it}. The random effects model (discussed below) allows identification of time-invariant variables by assuming f_i is uncorrelated with all regressors. However, this assumption is often hard to justify. Another approach is the Hausman-Taylor method (Hausman and Taylor 1981[1]), which is an instrumental variables technique: one uses the fact that some regressors might be uncorrelated with f_i to serve as instruments for the time-invariant regressors. The details of this method are beyond our scope here, but it essentially blends fixed and random effects assumptions to allow estimation of coefficients on time-invariant variables. Alternatively, if an external instrument is available for the time-invariant variable, one could employ IV methods in a panel context. In summary, pure fixed effects alone cannot handle time-invariant regressor estimation — additional assumptions or external variation are needed.

Hausman-Taylor method (Optional)

If we truly need to estimate the effect of a time-invariant regressor, one approach is to relax the assumption that f_i can be arbitrarily correlated with x_{it}. The random effects (RE) model allows identification of time-invariant

[1] Hausman, J.A. and Taylor, W.E. (1981). "Panel data and unobservable individual effects." *Econometrica*, 49(6), 1377-1398.

variables by assuming f_i is uncorrelated with all regressors. However, this assumption is often hard to justify. Another approach is the Hausman–Taylor (HT) method (Hausman and Taylor 1981[2]), which is an instrumental-variables technique that blends fixed- and random-effects logic: some regressors are allowed to be correlated with f_i, and other regressors—assumed uncorrelated with f_i—supply internal instruments that identify the coefficients on the problematic variables, including time-invariant ones.

Setup and classification. Write the standard one-way error-components model

$$y_{it} \;=\; x_{it}'\beta \;+\; z_i'\gamma \;+\; f_i \;+\; \varepsilon_{it}, \quad i = 1, \ldots, N,\; t = 1, \ldots, T_i,$$

where x_{it} are time-varying regressors and z_i are time-invariant regressors. Partition the regressors into four groups:

$$x_{it} = \big(X_{1it},\, X_{2it}\big), \qquad z_i = \big(Z_{1i},\, Z_{2i}\big),$$

where X_1 and Z_1 are *exogenous w.r.t.* f_i (uncorrelated with f_i), and X_2 and Z_2 may be *endogenous w.r.t.* f_i (correlated with f_i). All regressors are assumed uncorrelated with ε_{it}, and the usual variance-components structure holds, $\varepsilon_{it} \sim (0, \sigma_\varepsilon^2)$, $f_i \sim (0, \sigma_f^2)$, independent of each other.

HT estimation in practice (step-by-step).

1. **Within (FE) step for time-varying coefficients.** Demean all time-varying variables within i so that f_i drops out:

   $$\tilde{y}_{it} \;=\; \tilde{X}_{1it}\beta_1 \;+\; \tilde{X}_{2it}\beta_2 \;+\; \tilde{\varepsilon}_{it}.$$

   Estimate β_1, β_2 by FE/OLS using only within variation. These are consistent even if X_2 is correlated with f_i because f_i is removed by demeaning.

2. **Estimate variance components and form the quasi-RE transform.** Using residuals from the FE step, obtain method-of-moments (or QMLE) estimates of σ_ε^2 and σ_f^2. Compute the usual RE weight

   $$\hat{\theta}_i \;=\; 1 - \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_f^2}},$$

[2] Hausman, J.A. and Taylor, W.E. (1981). "Panel data and unobservable individual effects." *Econometrica*, 49(6), 1377–1398.

and apply the quasi-RE ("partial between") transformation to any variable w_{it}:

$$w_{it}^* \;=\; w_{it} - \hat{\theta}_i \bar{w}_i,$$

where $\bar{w}_i$ is the time average for unit i. This transformation reintroduces between-unit information (needed to identify time-invariant effects) while accounting for the estimated variance components.

3. **Construct instruments.** Use *internal* instruments based on exogenous regressors:

   - The *within* variation of X_1, $\tilde{X}_{1it}$, and the *between* variation (unit means) $\bar{X}_{1i}$ are valid instruments because X_1 is uncorrelated with f_i.
   - The time-invariant exogenous variables Z_{1i} are also valid instruments.

   These instruments are used for the potentially endogenous X_2 (in the transformed equation) and, crucially, for Z_2, whose identification relies on between variation. A necessary rank condition is that the number of exogenous instruments, $\{\tilde{X}_{1it}, \bar{X}_{1i}, Z_{1i}\}$, is at least as large as the number of endogenous regressors, $\{X_{2it}, Z_{2i}\}$.

4. **Final 2SLS on the quasi-RE–transformed model.** Run IV/2SLS of y_{it}^* on $\left(X_{1it}^*, X_{2it}^*, Z_{1i}^*, Z_{2i}^*\right)$ using instruments $\{\tilde{X}_{1it}, \bar{X}_{1i}, Z_{1i}\}$. This yields the HT estimates $(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_1, \hat{\gamma}_2)$, including consistent estimates for the time-invariant coefficients γ under the maintained exogeneity classification. Inference should use standard errors robust to arbitrary serial correlation and heteroskedasticity within i (e.g., cluster by i).

Remarks. (i) When all regressors are exogenous w.r.t. f_i, HT reduces to RE/GLS. (ii) If the HT exogeneity classification is misspecified (e.g., some X_1 or Z_1 are actually correlated with f_i), the IV exclusion restrictions fail. In practice, researchers often compare FE and (quasi-)RE estimates via a Hausman test and use subject-matter knowledge to justify the HT partition. (iii) If a credible *external* instrument exists for a time-invariant regressor, panel IV with fixed effects is an alternative. In summary, pure fixed effects cannot identify coefficients on time-invariant regressors, but HT recovers them by leveraging exogenous within and between variation as instruments while allowing correlation between f_i and a subset of regressors.

6.5.2 Amplification of Measurement Error Bias

Another often under-appreciated cost of fixed effects (and first-difference) estimators is that they can exacerbate biases due to measurement error in the regressors. Recall that classical measurement error in an explanatory variable tends to attenuate OLS estimates towards zero (bias them towards zero). The severity of attenuation depends on the signal-to-noise ratio — roughly, β is multiplied by $Var(x^*)/[Var(x^*)+Var(e)]$, where x^* is the true variable and e the measurement error.

The fixed effects transformation can worsen the signal-to-noise ratio if the true variation in x is mostly between individuals rather than within individuals. Think of decomposing the variation in x into two components: meaningful variation (possibly including between-individual differences and persistent trends) and noise variation (transitory fluctuations or measurement error). When we apply the within transformation, we remove the between-individual variance (which may have been largely the "good" variation if individuals have very different levels of x). What remains is only the within-individual variation. If the noise in x is i.i.d. over time, the variance of the noise in the within dimension might actually increase relative to the variance of the signal. In extreme cases, if x varies little over time (highly persistent or almost fixed) but each observation is measured with a similar amount of error, then taking differences or deviations could leave mostly noise. Consequently, the attenuation bias on β can become more severe in a fixed effects regression than it would have been in a cross-sectional regression.

To illustrate, suppose $x_{it} = x_{it}^* + e_{it}$, where x_{it}^* is the true regressor and e_{it} is classical (mean-zero) measurement error. In a simple cross-sectional OLS, the probability limit of the slope is

$$plim\hat{\beta}_{\text{OLS}} \;=\; \beta\,\frac{Var(x^*)}{Var(x^*)+Var(e)}\,.$$

Now consider the fixed effects (within) estimator. The "signal" part of the demeaned regressor is $\tilde{x}_{it}^* \equiv x_{it}^* - \bar{x}_i^*$ and the "noise" part is $\tilde{e}_{it} \equiv e_{it} - \bar{e}_i$, so the demeaned regressor is $\tilde{x}_{it} = \tilde{x}_{it}^* + \tilde{e}_{it}$. The variance of $\tilde{x}_{it}^*$ is generally smaller than $Var(x_{it}^*)$ because demeaning removes across-individual (and some low-frequency) variation. The variance of $\tilde{e}_{it}$ also changes relative to $Var(e_{it})$: if

e_{it} is i.i.d. over time with variance σ_e^2, then in a balanced panel

$$Var(\tilde{e}_{it}) \;=\; Var(e_{it} - \bar{e}_i) \;=\; \sigma_e^2\left(1 - \frac{1}{T}\right).$$

Meanwhile, if x_{it}^* is highly persistent with autocorrelation $\rho \approx 1$, then $Var(\tilde{x}_{it}^*)$ can be much smaller than $Var(x_{it}^*)$ (e.g., if x_{it}^* is nearly constant within i, $\tilde{x}_{it}^*$ is tiny). Consequently, the reliability ratio in the within regression becomes

$$plim\hat{\beta}_{\text{FE}} \;=\; \beta\,\frac{Var(\tilde{x}^*)}{Var(\tilde{x}^*) + Var(\tilde{e})},$$

which is typically smaller than the cross-sectional ratio above, implying stronger attenuation toward zero. In short, measurement-error bias is often more severe with fixed effects.[3]

Practical implication: If you see a coefficient estimate on a key regressor shrink toward zero (or become statistically insignificant) after adding fixed effects, it might be because the fixed effects removed a lot of the variation in the regressor, leaving mostly noise. One should be cautious in interpreting a "zero result" in such cases. It could be a true zero effect, but it could also be that any effect is drowned out by measurement error after using only within-group variation.

[3] Another way to show this using FD: Remember, with ME in independent variable, we have

$$plim(\hat{\beta}) = \beta\frac{\text{var}(x^*)}{\text{var}(x^*) + \text{var}(e)}$$

In first-difference estimator:

$$plim(\hat{\beta}) = \beta\frac{\text{var}(\Delta x^*)}{\text{var}(\Delta x^*) + \text{var}(\Delta e)}$$

, where $\text{var}(\Delta x^*) = \text{var}(x_t^*) - 2\text{cov}(x_t^*, x_{t-1}^*) + \text{var}(x_{t-1}^*)$. If x_t is stationary, $\text{var}(\Delta x^*) = 2\sigma_x^2 - 2\text{cov}(x_t^*, x_{t-1}^*) = 2\sigma_x^2(1-\rho)$. Define r to be the autocorrelation coefficient in u_t so we can write

$$plim(\hat{\beta}) = \beta\frac{2\sigma_x^2(1-\rho)}{2\sigma_x^2(1-\rho) + 2\sigma_u^2(1-r)} = \beta\frac{1}{1 + \frac{\sigma_u^2(1-r)}{\sigma_x^2(1-\rho)}}$$

When $r = \rho = 0$, traditional attenuation bias. When $r = 0$ only (ME is serially uncorrelated, but the signal is correlated), worrisome. When $r = 1$ (i.e., ME is fixed), FE eliminates ME.

Another related issue arises even when variables are perfectly measured: sometimes the outcome y responds to long-term changes in x but not to short-term fluctuations. If we use fixed effects (or first differences), we are essentially relying on possibly short-term deviations for identification. If those short-term movements in x do not have a substantive effect on y (because only sustained, long-run changes in x matter for y), then the FE estimator might find no effect, even though a long-run cross-sectional comparison would show an effect. In other words, y_{it} might react to the persistent component of x_{it} but not to the transitory component, whereas FE throws away the persistent component (if it is constant or slow-moving) and uses mainly transitory variation. This phenomenon has been noted by McKinnish2008[4] , among others. The lesson is that one must consider the time-scale of the effect of interest. If we suspect the effect of x on y operates over long horizons, a short panel with FE might not capture it well. Techniques like looking at long differences or distributed lags can sometimes be more informative in such cases.

Example 6.1. A study by Paravisini, Rappoport, Schnabl, and Wolfenzon (2014)[5] investigated how shocks to bank credit supply affect firm output (exports). A major challenge was that credit shocks often coincide with demand shocks (in recessions, banks cut lending and firms face lower demand). To isolate credit supply effects, the authors exploited a unique dataset of Peruvian firms that export products to different countries and have relationships with different banks. They included firm fixed effects (controlling for any time-invariant firm trait), bank fixed effects (controlling for banks that consistently lend more or less), and even product-destination fixed effects (controlling for, say, a particular product exported to a particular country, to net out demand shifts in that product market). This is a very exhaustive set of FE, ensuring that identification comes purely from within-firm changes in credit from different banks and how that affects exports of a given product to a given destination. The result was a relatively small estimated effect of credit on output. However, one concern is measurement error: firm-level credit might be measured with noise (loans not fully observed, timing issues

[4] McKinnish, T. (2008). "Panel data models and transitory fluctuations in the explanatory variable." *Journal of Econometrics*, 144(1), 39-52.

[5] Paravisini, D., Rappoport, V., Schnabl, P., & Wolfenzon, D. (2014). "Dissecting the effect of credit supply on trade: Evidence from matched credit-export data." *Quarterly Journal of Economics*, 129(2), 861-921.

in measurement, etc.). By using so many fixed effects, the variation in credit used to estimate the effect is quite limited (e.g., firm deviations from their usual credit, controlling also for average credit from each bank, etc.), potentially amplifying attenuation bias. The authors, aware of this, are careful not to over-interpret the small coefficient as evidence that credit supply matters little. Instead, they acknowledge that due to the extensive fixed effects (necessary for identification), the estimate could understate the true importance of finance (because the remaining variation in credit could be quite noisy). This example highlights how combining heavy FE strategies with imperfect data can lead to coefficients that are biased towards zero.

What can be done about measurement error? Measurement error is a tough nut to crack in any context. With panel data, some specialized methods have been proposed. For instance, Griliches and Hausman (1986)[6] discuss how one might use the difference between within and between estimates to assess bias. The intuition is that if measurement error is classical, the within estimator (using short-term changes) might be more attenuated than the between (cross-sectional) estimator. Under some assumptions, one can correct for measurement error by comparing the two. Other works, like Biorn (2000)[7] and Erickson and Whited (2000, 2012)[8] propose methods to adjust for measurement error in panel settings (often requiring instrumental variables or strong assumptions about the structure of the error). Another approach is to collect better data or proxies that can instrument for the mismeasured variable (for instance, using multiple measures of the same concept). In summary, there is no simple fix within the OLS/FE framework aside from being aware of the issue. If attenuation bias is suspected, one might interpret small coefficients as a lower bound of the possible effect, or try sensitivity analyses.

[6] Griliches, Z. & Hausman, J.A. (1986). "Errors in variables in panel data." *Journal of Econometrics*, 31(1), 93-118.

[7] Biorn, E. (2000). "Panel data with measurement errors: Instrumental variables versus some GMM estimators." *Econometric Reviews*, 19(3), 219-250.

[8] Erickson, T. & Whited, T.M. (2000). "Measurement error and the relationship between investment and q." *Journal of Political Economy*, 108(5), 1027-1057; Erickson, T. & Whited, T.M. (2012). "Treating measurement error in Tobin's *q*." *Review of Financial Studies*, 25(4), 1286-1329.

6.5.3 High-Dimensional Fixed Effects and Computational Issues

Including fixed effects for very large numbers of groups can strain computational resources and statistical power. In a typical regression, including N individual dummies is manageable if N is, say, a few thousand. But what if we want to include multiple high-dimensional fixed effects, e.g. firm FE ($N \approx 10000$) and year FE ($T \approx 30$)? Year FE are fine (only 30 dummies), but firm FE are 10,000 dummies. That many parameters can be estimated, but now imagine adding industry-by-year FE (if 100 industries over 30 years, that is another 3000 dummies), and maybe region FE, etc. The number of dummy variables can blow up, making the regression computationally heavy and possibly imprecise (each dummy uses up a degree of freedom).

When only one type of fixed effect is present, the within transformation deals with it elegantly (as we did for individual FE, we avoid ever estimating those N intercepts explicitly). However, when we have multiple types of fixed effects, we cannot simultaneously demean in two dimensions at once by a simple subtraction. For instance, consider we want firm fixed effects and year fixed effects in:

$$y_{it} = \alpha + \beta x_{it} + f_i + \delta_t + u_{it} \, ,$$

with i indexing firm and t indexing year. The within transformation can remove one dimension at a time. If we transform by demeaning within each firm, we eliminate f_i. But year fixed effects δ_t remain (since they vary over time, not constant within i). We can then include year dummies in the regression to handle δ_t. That is straightforward since T is usually not huge. Alternatively, we could first demean within years (subtract year means, which removes δ_t) and then include firm dummies. Either way, one dimension of FE can be partialled out perfectly; the other has to be explicitly included or partialled out in a second step.

With modern software, this is not such a hurdle. Programs exist to absorb two or more sets of high-dimensional fixed effects without explicitly creating all dummy variables (e.g. methods using iterative demeaning or the Frisch-Waugh approach multiple times). In R, the `lfe` or `fixest` packages and in Stata, commands like `areg` (absorbing one FE) or user-written `reghdfe` (for multiple FEs) handle these scenarios efficiently. They use clever algorithms to avoid the memory blow-up of creating huge dummy matrices.

However, historically and conceptually, as you add more and more fixed effects, you must be cautious. You are using up degrees of freedom and possibly encountering collinearity issues (for example, including firm and industry FE together is fine, but including firm and industry and industry-by-year FE is redundant if firm FE are there, etc.). One must ensure each set of fixed effects is needed and correctly specified.

Example 6.2. (computational difficulty) If we tried to include firm FE and industry×year FE in a single regression with a large panel of firms, the number of dummy variables could be enormous (say 10,000 firms + (100 industries $\times$ 30 years = 3,000) = 13,000 dummies). This is on the edge of feasibility for OLS with large N. If instead we had 100,000 firms, it becomes even more daunting. The design matrix would be huge. In such cases, specialized algorithms or tricks (like absorbing one effect at a time) are essential. The specific details of these computational techniques go beyond our scope, but it is good to be aware that naive inclusion of many FE can slow down estimation or even crash your software if you run out of memory.

Typically, if you encounter a model that requires multiple high-dimension FEs, you might consider:

- Is each set of FE absolutely necessary?
- Can any be replaced with a more parsimonious control or a random effects structure?
- Use appropriate software that can handle multi-way fixed effects efficiently (as mentioned, rather than creating 0/1 dummies manually).

One way to think about multi-way fixed effects is to consider them as a generalization of the within transformation. For example, to incorporate both firm and year fixed effects, one can transform the data by subtracting the firm mean (removing firm FE) and also subtracting the year mean, and then add back the overall mean (to avoid subtracting the constant twice). This is a demeaning in two dimensions (also known as the iterative Helmert transformation for two-way effects). The resulting data will have both firm and year means zero. Estimating β on that transformed data yields the two-way FE estimate. Doing this manually is error-prone, but conceptually it's what software does.

In summary, multiple fixed effects are doable but require care. The cost is primarily computational and possibly interpretational (lots of effects and potential multicollinearity if you're not careful). There is no finite-sample bias introduced by having many FE (aside from using up degrees of freedom), but see the next point about incidental parameters in non-linear models which addresses a different issue.

6.5.4 Incidental Parameter Problem and Non-Linear Models

Thus far, we have focused on linear panel models (where y_{it} enters linearly and we estimate via OLS). In linear models, the fixed effects approach is straightforward due to linearity: we can eliminate f_i by demeaning without having to estimate f_i explicitly. In non-linear models (such as binary outcome models like logit/probit, count models like Poisson, etc.), the fixed effects cannot be removed by a simple transformation in general. Instead, they would appear as additional parameters to estimate (one for each individual). When N is large and T is fixed, we face what is known as the incidental parameter problem: we have N additional parameters (the f_i's) which grow with the sample size. Maximum Likelihood Estimation (MLE) in such settings can lead to inconsistency of the parameters of interest.

In a fixed effects logit model, for example, if we include a dummy for each individual in a logistic regression, we technically have N person-specific intercepts. As $N \to \infty$ (with T fixed), the number of parameters grows without bound. Neyman and Scott (1948) first pointed out that when the number of parameters increases with the sample size, the usual consistency properties of MLE can fail for the parameters of interest. Intuitively, the MLE tries to fit N intercepts exactly for each individual. When T is small, those intercept estimates $\hat{f}_i$ can soak up a lot of variation, potentially even overfitting the idiosyncratic noise. The β coefficients (which are common across individuals) can get biased as a result.

In linear models, we avoided this by using the within transformation. We never explicitly estimated N separate f_i parameters; we differenced them out. As a result, the incidental parameters (f_i) did not need to be estimated, and $\hat{\beta}$ is consistent. In non-linear models, such clean differencing is not always possible: - Logit model: There's a special case where fixed effects logit can be addressed by a conditional likelihood approach. For binary outcomes, one can condition on the total number of successes for each individual (a sufficient statistic for f_i in a logit model). This leads to the conditional logit estimator (also called logit with fixed effects), which actually gives consistent estimates for β without needing to estimate f_i. However, this conditional likelihood method is specific to logit and a few other models (like Poisson). - Probit model: There is no analogous conditioning trick for probit. If you include individual dummies in a probit and N is large, the coefficient estimates will be biased (even as N grows) unless $T \to \infty$. Essentially, fixed effects probit demands that f_i be treated as random or else you have the incidental parameter issue. The bias does not vanish with more individuals, only with more time periods. - Poisson model: It turns out a fixed effects Poisson (for count data) can be estimated consistently via a conditional likelihood as well (conditioning on the total count for each individual). So Poisson is a nice case where fixed effects are tractable without bias. - Other non-linear models: Tobit, Cox proportional hazards, multinomial logit, etc., all have difficulties with including numerous fixed effects without additional assumptions or bias corrections.

Therefore, for many non-linear panel models, one should exercise caution with naive fixed effects inclusion. Either use models specifically designed for fixed effects (like conditional logit or Poisson FE), or consider alternative estimation methods (like generalized estimating equations, or treat the effects as random but perhaps use a correlated random effects approach like Mundlak's adjustment, where one includes the individual means of time-varying covariates to account for correlation with f_i).

Another byproduct of the incidental parameter issue is that the actual estimates of the fixed effects themselves ($\hat{f}_i$ or the dummies coefficients) are generally inconsistent when T is small. Even in the linear case, while $\hat{\beta}$ is consistent, each $\hat{f}_i$ is only based on T observations and does not improve as N grows; it's an unbiased estimator of f_i but with variance that doesn't vanish (it remains on the order of $\sigma_u/\sqrt{T}$). So one should not over-interpret individual fixed effect estimates, especially if T is low. For example, you might estimate student fixed effects in test scores; a student with just a couple of test observations might have a high fixed effect estimate, but that could be largely noise (since it includes the average of their idiosyncratic errors). If you tried to rank individuals by fixed effect, the ranking could be noisy for those with few time observations.

Researchers sometimes want to interpret the distribution of $\hat{f}_i$—for instance, Bertrand and Schoar (2003)[9] estimated CEO fixed effects on corporate policies and then analyzed those fixed effects, labeling them as "managerial styles." They found that the CEO dummies were jointly significant (rejecting that all $\gamma_i = 0$ in LSDV, meaning CEOs do differ systematically). However, interpreting these fixed effect estimates as true measures of a CEO's skill or style must be done with care. If one took those $\hat{f}_i$ to a second-stage regression, formally one should correct for the estimation error in $\hat{f}_i$ (since they are noisy estimates of the true effect). Simply regressing something on $\hat{f}_i$ can lead to attenuation or other issues.

Moreover, tests involving fixed effects (like the F-test for all $f_i = 0$) assume certain conditions for validity. The standard F-test for joint significance of fixed effects assumes homoskedastic, iid errors. In practice, with panels, errors can be heteroskedastic or autocorrelated, so a robust test or bootstrap might be needed. There is also a concern pointed out by Fee, Hadlock, and Pierce (2013)[10]: they found that if they randomly assign managers to firms (breaking the true link between manager and performance), they still found significant "manager effects" using a fixed effects approach. This suggests that under certain circumstances, one might find spurious significance of fixed effects due to either multiple testing or the underlying distribution of the errors violating assumptions. Thus, while fixed effects are extremely useful, one should not automatically interpret the presence of significant fixed effects as deep evidence of structural differences without further scrutiny.

To summarize the advice on this issue: - In linear models, use fixed effects freely to remove bias, but don't focus on the estimated fixed effect values themselves as if they were precise or structural. - In non-linear models, be cautious: either avoid too many fixed effects or use methods specifically designed for them (like conditional likelihood where available, or accept bias if T is modest). - If you need to include fixed effects in a non-linear model and T is not large, understand that point estimates of β may be biased (often called "Nickell bias" in the context of dynamic linear models, and a similar

[9] Bertrand, M. & Schoar, A. (2003). "Managing with style: The effect of managers on firm policies." *Quarterly Journal of Economics*, 118(4), 1169-1208.

[10] Fee, C.E., Hadlock, C.J., & Pierce, J.R. (2013). "Managers with and without style: Evidence using exogenous variation." *Review of Financial Studies*, 26(3), 567-601.

phenomenon occurs in non-linear ones). - There are some bias correction methods proposed in the econometric literature for the incidental parameter problem (e.g. for probit, or for short T dynamic panels), but they can be complex.

6.6 The Random Effects Model

The fixed effects model makes no assumption about whether f_i is correlated with the regressors. It treats f_i as a parameter to be estimated (or eliminated). The random effects (RE) model takes a different approach: it treats f_i as a random draw from a population distribution and crucially assumes that f_i is *uncorrelated* with the regressors x_{it}. In other words, the random effects model assumes:

$$Cov(f_i, x_{it}) = 0 \quad \text{for all } t\,.$$

Under this assumption, the omitted variable bias problem disappears: if f_i is uncorrelated with the x's, then even though f_i is in the error term of the pooled model, that error component is uncorrelated with x, satisfying the OLS exogeneity condition. In that case, a pooled OLS regression would actually yield a consistent estimate of β.

However, pooled OLS would not be fully efficient because the composite error $v_{it} = f_i + u_{it}$ has a specific structure: it is correlated over time for the same i (since $\text{Cov}(v_{i,t}, v_{i,s}) = Var(f_i)$ for $t \neq s$). OLS ignoring this correlation would still be consistent (if the assumption holds) but the standard errors would be wrong and OLS wouldn't be the best linear unbiased estimator. Instead, one can do Generalized Least Squares (GLS) which accounts for the intraclass correlation of errors to get more efficient estimates. Random effects estimation typically refers to GLS (or Feasible GLS if variances are unknown and need to be estimated) under the assumption of zero correlation between f_i and x_{it}.

The mechanics of random effects GLS often involve transforming the data by a partial demeaning. Specifically, the GLS estimator for random effects can be seen as a weighted combination of the between-group estimator and the within-group (fixed effects) estimator. It effectively does:

$$y_{it} = \theta \bar{y}_i = \beta(x_{it} - \theta \bar{x}_i) + \text{transformed error},$$

where θ is a function of the variance components (σ_f^2 and σ_u^2). If σ_f^2 is large relative to σ_u^2, θ is close to 1 and the model puts more weight on within variation (approaching fixed effects); if σ_f^2 is small, θ is close to 0, putting more weight on between variation (approaching pooled OLS). When $Cov(f_i, x_{it}) = 0$, this GLS is the Best Linear Unbiased Estimator (BLUE) and also consistent. If the correlation assumption is false, RE GLS is inconsistent.

Is the random effects assumption realistic? In many economic contexts, it is hard to justify that the unobserved f_i (which could be ability, preferences, institutional quality, etc.) is uncorrelated with the regressors of interest. For example, imagine y_{it} is earnings and x_{it} includes education. The unobserved f_i might be "ability." It's unlikely that ability is uncorrelated with education; more able individuals tend to get more education. Thus, the random effects assumption would fail and RE estimates of the return to education would be biased. Similarly, in our earlier firm example, f_i might be the firm's general risk aversion or culture, which plausibly correlates with financial decisions like leverage, violating the RE assumption.

In practice, because the zero-correlation assumption is often implausible, random effects models are not trusted as much. A typical approach is to run a Hausman test: this test compares the estimates from a consistent estimator (fixed effects) to those from an efficient estimator under H_0 (random effects). Under the null hypothesis that $Cov(f_i, x_{it}) = 0$, both FE and RE are consistent, but FE is inefficient. Under the alternative that $Cov(f_i, x_{it}) \neq 0$, FE is consistent but RE is inconsistent. If the Hausman test (which essentially checks if the estimates differ significantly) rejects, it means the data are not consistent with the RE assumption, and one should prefer the FE estimator. If it fails to reject, one might use RE for efficiency (and plus RE allows estimation of time-invariant covariates).

However, as Angrist and Pischke (2009)[11] note, the efficiency gain of RE over FE is often small, especially if T is modest or the explanatory power of f_i is not overwhelming. Moreover, if there is any doubt about the assumption, most researchers opt for the safer course (FE). Also, if the assumption truly holds, pooled OLS itself would be consistent (just with wrong standard errors). The main point of RE GLS is theoretical efficiency improvement.

[11] Angrist, J.D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics.* Princeton University Press. See p. 223.

When might random effects be useful? - If you have reason to believe the unobserved effect is not correlated with regressors (perhaps the regressors are randomly assigned or quasi-random, which is rare in observational data). - If you need to estimate coefficients on time-invariant regressors which FE cannot do. For example, if you're studying the effect of a time-invariant policy or geographic feature, and you are willing to assume that omitted fixed effect is uncorrelated (or you include enough other controls to mitigate correlation), you might use RE. - If N is not huge but T is fairly large, RE might save some degrees of freedom. But usually, large T also means FE bias from any correlation might diminish anyway, so FE could still be fine.

Summary on RE vs FE: The fixed effects model addresses the primary concern of many studies: omitted variable bias from time-invariant unobservables. The random effects model yields similar results only under the strong assumption of no omitted variable bias (in that sense). Since this assumption is typically suspect, random effects is often not trusted. Many practitioners simply use FE as the default for panel data. RE is sometimes taught and used for specific cases, but one should always perform a Hausman test or other diagnostics to check its validity. Even if valid, the benefits (slightly smaller standard errors, ability to estimate coefficients on time-invariant variables) have to be weighed against the risk that the assumption might not hold perfectly.

In the context of corporate finance, Angrist and Pischke's viewpoint is apt: the conditions for RE to be correct are strong, and even if they hold, FE was consistent anyway and likely not much less efficient, whereas if they don't hold, RE is seriously misleading. Therefore, RE is seldom the preferred strategy in practice.

6.7 The First-Difference Estimator

First-differencing is an alternative method to remove fixed effects. Instead of subtracting the time mean, we subtract the previous observation of the same individual. The idea is: between two adjacent time periods, the individual effect f_i does not change, so it will cancel out when we take a difference. For simplicity, consider $T = 2$ first. If we have two time periods, the first-difference (FD) estimator is nearly identical to the fixed effects (within) estimator; in fact, with $T = 2$, demeaning each individual's data is the same as subtracting the period 1 value from the period 2 value (up to a sign). So for $T = 2$, FE and FD produce exactly the same $\hat{\beta}$.

For $T > 2$, FE and FD are not identical transformations, but they both eliminate f_i. Let's derive the first-differenced model for general T. The model is again:

$$y_{it} = \alpha + \beta x_{it} + f_i + u_{it} \,.$$

The first-difference between period t and $t-1$ is:

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + (f_i - f_i) + (u_{it} - u_{i,t-1}) \,.$$

The f_i cancels out, as desired. We typically assume the model holds for $t = 2, \ldots, T$ differences (we lose the first period as it has no previous value to difference with). Define $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\Delta x_{it} = x_{it} - x_{i,t-1}$. The differenced equation is:

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta u_{it}, \quad t = 2, ..., T$$

We can now run OLS on this first-differenced equation. This yields the FD estimator $\hat{\beta}_{FD}$.

If the strict exogeneity assumption holds (i.e. $E(x_{it}\Delta u_{is}) = 0$ for all t, s), then the FD estimator is consistent for β. Strict exogeneity in this context implies $x_{i,t}$ is uncorrelated with $u_{i,t}$ (contemporaneous) and also with $u_{i,t-1}$ (one-period lagged) because $\Delta u_{i,t} = u_{i,t} - u_{i,t-1}$ includes $u_{i,t-1}$. If x at time t is correlated with the prior period's error $u_{i,t-1}$, then $x_{i,t}$ will be correlated with $\Delta u_{i,t}$ and the FD estimator will be biased. This is a subtle but important difference: FE requires $x_{i,t}$ uncorrelated with $u_{i,s}$ for all s (same requirement), and FD effectively requires a similar condition. So both need strict exogeneity for consistency. If x reacts to past shocks, both FE and FD are in trouble (though the nature of the bias differs, as we discuss soon).

The first-difference approach, like FE, wipes out any time-invariant f_i. It's also possible to do "differences" in a panel that is not strictly time-based (for example, if observations are in some inherent order, one could subtract the previous observation's value; but this is rarely used unless there is a logical ordering).

FE vs. FD for $T > 2$: Efficiency considerations. If strict exogeneity holds and errors are i.i.d. across t, both FE and FD are consistent, but they are not numerically identical unless $T = 2$. Which is more efficient depends on the serial correlation in u_{it}:

- **No serial correlation in u_{it}.** If u_{it} is serially uncorrelated (and homoskedastic), FE is more efficient than FD. Differencing creates $\Delta u_{it} = u_{it} - u_{i,t-1}$, which induces MA(1) correlation:

  $$Cov(\Delta u_{i,t}, \Delta u_{i,t-1}) = -Var(u_{i,t-1}) \quad \text{when } Cov(u_{i,t}, u_{i,t-1}) = 0.$$

  Thus FD inherits serial correlation even when the level errors are white noise, reducing efficiency relative to FE (unless one uses GLS that accounts for the MA(1) structure).

- **Positive serial correlation in u_{it}.** If u_{it} follows an AR(1), $u_{i,t} = \phi u_{i,t-1} + \varepsilon_{it}$ with $\phi > 0$, the within (FE) transformation does not eliminate this persistence; the transformed errors $\tilde{u}_{it}$ remain serially correlated. FD, however, reduces persistence: differencing an AR(1) yields an MA(1) with smaller autocorrelation. In the extreme random-walk case ($\phi = 1$),

  $$u_{i,t} = u_{i,t-1} + \varepsilon_{it} \;\Rightarrow\; \Delta u_{it} = \varepsilon_{it} \text{ (i.i.d.)},$$

  so FD removes all serial correlation in the errors and can be more efficient than FE in levels (unless FE is estimated by appropriate GLS).

- **Inference.** With substantial serial correlation (of either sign), FE and FD remain consistent under strict exogeneity, but standard errors must allow for within-i dependence (e.g., cluster-robust by i). The relative finite-sample efficiency of FE vs. FD depends on the exact error process and whether one exploits it via GLS.

Another consideration: - FE is slightly more complicated with heteroskedastic or non-normal errors, because the inclusion of many dummy variables can potentially soak up some variation and might rely on large N normal approximations. FD is essentially a transformation that also can be used in GMM contexts more readily if needed. - Both FE and FD are equally sensitive to measurement error in x (both will suffer attenuation, though as discussed FD might suffer more if x is persistent). - If the strict exogeneity assumption is violated in the sense of $x_{i,t}$ being correlated with future errors but not past ones (that is, x is predetermined but not strictly exogenous), an interesting difference emerges: *if x is correlated with $u_{i,t+1}$, the FE estimator is actually inconsistent (because the demeaning uses future errors in the mean), but the FD estimator might still be inconsistent as well (since $x_{i,t}$ will correlate with $\Delta u_{i,t+1}$ which includes $u_{i,t}$)? Let's clarify:* - Actually, if x is predetermined (meaning $Cov(x_{i,t}, u_{i,s}) = 0$ for $s \leq t$ but possibly not for $s > t$), then FE estimator of β is still consistent as $T \to \infty$ (the bias from correlation with future errors dies out at rate $1/T$ because each observation's weight in the mean shrinks with T). For fixed T, FE is biased but the bias is $O(1/T)$. - FD in that case might not enjoy a similar reduction with larger T, since each difference still involves a contemporaneous correlation if it exists. Thus, if x reacts to past shocks (so not strict exogeneity), often FE is preferred because the bias might be smaller (especially if T is moderately large) whereas FD might have a bias that doesn't diminish unless you instrument.

Given these nuances, a common suggestion in empirical practice is to try both FE and FD estimators. If they give very similar results, that increases confidence in the findings (since each has different small-sample properties). If they differ significantly, that flags a potential issue: either serial correlation, or violation of strict exogeneity, or measurement error might be affecting them differently. As mentioned earlier, Griliches and Hausman (1986) proposed leveraging the difference between FE and FD estimates to diagnose measurement error: if measurement error is a big problem, the FE estimate might be attenuated more than the FD estimate (or vice versa depending on persistence), so comparing them can provide clues. In some cases, one can even compute an estimate of the true coefficient by combining the two biased estimates (this is an advanced technique requiring assumptions on error structure).

In general, FE is more popular than FD for panel data because it's easier to implement with standard software (most packages have a built-in FE estimator but not all have an easy FD routine, though it's not hard to manually difference). Also, FE naturally extends to more than 2 time periods and multiple fixed effects, whereas FD is most natural for one dimension at a time. But conceptually, both are accomplishing the same goal: removing f_i.

It's worth noting that if one uses FD, adding period fixed effects (year dummies) is equivalent to removing any secular trends before differencing. Often one might include a time dummy in the FD regression to capture mean growth from $t-1$ to t that would otherwise appear as a nonzero intercept in the FD equation.

To illustrate FE vs FD concretely: Suppose we have an individual wage panel with 5 years of data. A fixed effects estimator would effectively use deviations from each person's 5-year average wage and average x (say education doesn't change, but experience does year by year). A first-difference estimator would use at most 4 differences (year2-year1, year3-year2, etc.) per person. If the wage shocks are i.i.d., FE is using all 5 data points optimally; FD is throwing away one data point and having correlated errors, so FE is better. If wage shocks follow a random walk, FD uses essentially the innovation in wage which is clean, and FE might erroneously treat the accumulated shocks as noise. Typically, wage shocks have some persistence but not a pure random walk, so either method could be fine.

In summary: FE and FD are two sides of the same coin (both difference out fixed effects). They yield the same estimate when $T = 2$. For $T > 2$, both are consistent (if exogeneity holds) but differ in efficiency depending on error autocorrelation. If you suspect strong autocorrelation or want to avoid assumptions of no serial correlation, FD with appropriate standard errors is robust. If you trust no serial correlation, FE might be more efficient.

6.8 Dynamic Panel Models (Lagged Dependent Variables)

A common extension to our model is to allow for dynamics: include lagged dependent variables as regressors. For example, a firm's current leverage might depend on last year's leverage (perhaps due to adjustment costs or target behavior). Or an individual's current consumption might depend on past consumption. When we include a lagged y, the fixed effects approach faces a new challenge. Consider the model:

$$y_{it} = \alpha + \rho y_{i,t-1} + \beta x_{it} + f_i + u_{it} \, .$$

with $|\rho| < 1$ to ensure stability. This is a dynamic panel model with fixed effects. The presence of $y_{i,t-1}$ (the lagged dependent variable) violates the strict exogeneity assumption, because $y_{i,t-1}$ obviously depends on $u_{i,t-1}$ (in fact $y_{i,t-1}$ contains $u_{i,t-1}$). Even if u_{it} itself is serially uncorrelated, $y_{i,t-1}$ will generally be correlated with f_i and with $u_{i,t-1}$. This creates problems for both OLS and FE estimators:

Pooled OLS (ignoring fixed effects). If we estimate the dynamic model

$$y_{it} = \alpha + \rho\, y_{i,t-1} + \beta x_{it} + f_i + u_{it}$$

by OLS without accounting for f_i, the composite error is $v_{it} = f_i + u_{it}$. Because $y_{i,t-1}$ contains f_i (since $y_{i,t-1} = \alpha + \rho y_{i,t-2} + \beta x_{i,t-1} + f_i + u_{i,t-1}$), we have $Cov(y_{i,t-1}, v_{it}) \neq 0$. Hence the lag regressor is endogenous and OLS is biased and inconsistent.

Fixed effects (within) estimation. Demeaning to remove f_i yields

$$y_{it} - \bar{y}_i \;=\; \rho\,(y_{i,t-1} - \bar{y}_i) \;+\; \beta\,(x_{it} - \bar{x}_i) \;+\; (u_{it} - \bar{u}_i).$$

However, $(y_{i,t-1} - \bar{y}_i)$ is correlated with $(u_{it} - \bar{u}_i)$ because $\bar{u}_i = \frac{1}{T}\sum_{s=1}^{T} u_{is}$ includes $u_{i,t-1}$, which also helps determine $y_{i,t-1}$. Formally, for finite T,

$$Cov\big(y_{i,t-1} - \bar{y}_i,\; u_{it} - \bar{u}_i\big) \neq 0,$$

so the within estimator of ρ is biased (typically downward when $\rho > 0$). This is the *Nickell bias* (Nickell, 1981): with $N \to \infty$ and fixed T, $\hat{\rho}_{\text{FE}}$ is inconsistent, and the bias is $O(1/T)$. As $T \to \infty$, the bias vanishes and both FE and OLS (with f_i observed) become consistent.

The bias of $\hat{\rho}_{FE}$ in a dynamic panel is known as the Nickell bias (after Nickell, 1981). For moderate T (say $T = 5$ or 10), this bias can be non-negligible. For example, with ρ around 0.5 and $T = 5$, the bias might be on the order of -0.1 (estimated ρ too low by 0.1 or so). As T increases, the bias shrinks; e.g. at $T = 30$, it's much smaller, and in the limit $T \to \infty$, it vanishes.

First Differences: If we difference the dynamic model: $y_{it} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + \beta(x_{it} - x_{i,t-1}) + (u_{it} - u_{i,t-1})$, we also encounter an issue: $y_{i,t-1}$ in the right-hand side difference is correlated with $u_{i,t-1}$ in the error term (because $y_{i,t-1}$ includes $u_{i,t-1}$ from the original equation). So $\Delta y_{i,t-1}$ is correlated with Δu_{it} (since Δu_{it} contains $-u_{i,t-1}$). Thus OLS on the differenced equation is also biased for ρ. In fact, the FD estimator has a similar magnitude of bias in finite samples (also $O(1/T)$, but it might differ slightly in sign or magnitude depending on ρ).

Given that neither FE nor FD (nor pooled OLS) produce consistent estimates in a dynamic panel with fixed N, how do we estimate such models? The solution lies in using instrumental variables or GMM techniques that exploit further lags as instruments. This approach was pioneered by Anderson and Hsiao (1982) who suggested instrumenting $\Delta y_{i,t-1}$ with $y_{i,t-2}$ (which is correlated with $y_{i,t-1}$ but not with $u_{i,t} - u_{i,t-1}$, under assumptions of no serial correlation in u). More efficiently, Arellano and Bond (1991) and subsequent works developed a general GMM estimator that uses all available lagged values of y as instruments for the differenced (or level) equations. These are known as dynamic panel GMM estimators (difference GMM and system GMM, etc.). Those are beyond the scope of this chapter, but essentially they tackle the endogeneity of the lagged dependent variable by instrumenting it with deeper lags (assuming no autocorrelation in u beyond some order and assuming initial conditions are such that those lags are valid).

The key point for our discussion is: **do not use basic fixed effects or OLS when you have a lagged dependent variable in a short panel**. They will be biased. If T is reasonably large, one might tolerate the bias, but often researchers prefer to address it directly via instrumental variable methods.

Sometimes, researchers face a model where they are not sure if a lagged dependent variable should be included or not. One trick sometimes mentioned is "bracketing" the true effect of a variable by comparing the static FE model and the dynamic model. Suppose our variable of interest is x_{it} and we want its effect on y_{it}. If the true model is static (no lagged y) but we erroneously include $y_{i,t-1}$, the coefficient on x_{it} in the dynamic model might be biased (since $y_{i,t-1}$ will soak up some effect that belongs to x or might itself be endogenous). If the true model is dynamic but we estimate a static FE model (omitting the lag), then the coefficient on x_{it} could be biased because it picks up some dynamic effects.

Typically, if β (the effect of x) is positive, one can argue: - If the true model is dynamic but we use static FE (omit $y_{i,t-1}$), then some of the effect of past y on current y might be attributed to x, potentially leading to an *overestimate* of β. Intuition: suppose x has a positive effect, and y is highly persistent ($\rho > 0$). Then x_{it} might be correlated with future $y_{i,t+1}$ as well through persistence. In a static model, you don't control for that, so you might see a bigger immediate effect to compensate. - If the true model is static but we mistakenly include $y_{i,t-1}$, that lag term might pick up some of the contemporaneous effect of x. The β on x would then appear smaller (underestimated) because the model is giving some explanatory power to $y_{i,t-1}$ that actually should be attributed to x_{it} (which influenced $y_{i,t-1}$ if x is autocorrelated or something, or simply the lag is acting as a proxy for omitted factors which x also captured). - So the static FE estimate of β might be an upper bound, and the dynamic (with lag) estimate might be a lower bound, on the true effect (if those are the only two issues).

This idea of bracketing suggests that if you run a static FE model and a dynamic panel model (with appropriate correction for the lag, say using an IV or at least seeing the bias direction), the true β might lie between the two estimates. If you find that in your data the static FE $\hat{\beta}$ is, say, 0.8 and the dynamic model $\hat{\beta}$ is 0.4 (just as an example), you might suspect the true effect is somewhere in between, maybe around 0.6. If you find instead that the static gave 0.5 and the dynamic gave 0.7 (reversing the order), then something is inconsistent with the typical pattern—maybe the dynamic estimation isn't reliable or other omitted variables are playing a role, because we expected static to overshoot if anything.

However, one has to be careful: both these estimates (especially naive dynamic using OLS/FE) are biased in unknown ways, so one shouldn't literally take them as bounds without more justification. The bracketing argument is more heuristic. Ideally, one would directly address the dynamic panel bias via a proper estimator (like Arellano-Bond GMM). The bounded idea is sometimes mentioned to guess the magnitude of bias.

In summary for the dynamic case: - Including lagged dependent variables in panel regressions introduces special bias issues if using FE/OLS. - Use appropriate instruments or estimators for dynamic panels (this is a whole topic by itself, often taught under panel data econometrics or advanced econometrics). - Recognize that a coefficient difference between a model with and without a lagged y might suggest some dynamic structure. If theory suggests dynamics, it's better to use a proper dynamic panel method. - If one incorrectly ignores dynamics, coefficients on other variables might be biased (often upwards, because the omitted lag can cause serial correlation in errors). - If one incorrectly includes a lag when not needed, one might unnecessarily soak up explanatory power (leading to wider standard errors or multicollinearity issues, but consistency isn't at stake if T large, though if done by FE with small T, you're introducing Nickell bias and hurting β too).

To close this section: dynamic panels are common (e.g. growth regressions, adjustment models). The good news is that when T is moderately large (e.g. annual data over 30 years), simply doing FE with a lag might be okay; the Nickell bias is roughly $(1 + \rho)/(T - 1)$ for large N scenarios, so with $T = 30$ and $\rho = 0.5$, bias ≈ 0.05 which might be negligible. But with $T = 5$, it's not negligible. So always check the context and possibly use specialized estimators like difference GMM or system GMM if needed.

6.9 Summary and Conclusions

Panel data methods such as fixed effects and first differences are powerful tools to control for unobserved heterogeneity. They allow us to remove the influence of omitted variables that are constant within an entity, thereby mitigating a major source of bias in estimating causal relationships.

Key takeaways: - **Fixed Effects (FE) estimation** differences out time-invariant characteristics. By using only within-unit variation, FE provides consistent estimates even when each unit has its own intercept f_i correlated with the regressors. This greatly broadens the situations where OLS can be used without bias, making minimal assumptions about unobserved differences across units. - The FE estimator has an intuitive interpretation (effect of x on y for a given entity's changes) and is very general (you can have individual FE, time FE, and other group FE to control for various heterogeneities). - Random Effects (RE) estimation is an alternative that assumes no correlation between f_i and x_{it}. While it can be more efficient and can estimate coefficients on time-invariant variables, this assumption is often questionable. In practice, if RE assumptions hold, pooled OLS is consistent and one might just use that; if they fail, RE is inconsistent. Thus, the safe approach favored in most studies is FE, unless one has strong justification to use RE. The Hausman test is a classic way to compare FE and RE; typically it leads to rejection of RE in observational data. - **First Differences (FD) estimation** achieves the same asymptotic goals as FE (eliminate f_i)

by looking at changes between periods. FD and FE are numerically identical when $T = 2$; for larger T, they differ in efficiency depending on error structure, but both are consistent under similar assumptions. Trying both can be a useful robustness check. - **Limitations of FE:** - It cannot identify effects of variables that do not change within a unit (time-invariant regressors). To learn about those, other methods or assumptions are needed (e.g., RE model or instrumental variables). - It can exacerbate attenuation bias if key regressors are measured with error, because the within-variation may have a lower signal-to-noise ratio. - It uses a lot of degrees of freedom when N is large, and multiple fixed effects can be computationally heavy (though solvable with modern algorithms). In extreme cases, including too many fixed effects can lead to multicollinearity or overfitting issues as well. - In non-linear models or models with short time series, the inclusion of many fixed effects introduces the incidental parameter problem, potentially biasing the estimates. Special methods or caution are required in those contexts. - **Dynamic panels:** A lagged dependent variable with fixed effects causes standard FE/OLS estimators to be biased. One should use specialized techniques (IV/GMM) or be aware of the approximate nature if using a short panel.

General advice: Use fixed effects when you suspect important omitted variables are fixed within units. This often greatly improves the credibility of your estimates by controlling for unobserved heterogeneity. Always remember, though, that fixed effects only control for time-constant heterogeneity; time-varying omitted variables can still bias results. Thus, you may need additional control variables or instruments for those. Also, fixed effects "use up" variation, so check that your key regressor does vary sufficiently within units; if not, FE won't be informative for that effect.

If you have panel data, it's often wise to start with a pooled OLS, then see how results change with fixed effects. A big change suggests that unobserved heterogeneity was indeed biasing the OLS results (or at least the between-individual differences were important). If OLS and FE are similar (and Hausman test says no difference), then maybe unobserved f_i wasn't a big problem (or x is primarily within-variation anyway). In that case, you might even prefer the simpler model or RE if you want to include time-invariant factors.

Avoid blindly using fixed effects in situations they aren't suitable (short panels with lagged y, non-linear models with many dummies, etc.) without further techniques.

In conclusion, panel data greatly expands the toolkit of empirical economics, allowing more credible causal inference in the presence of unobserved heterogeneity. The fixed effects model is a workhorse method, and understanding its proper use and limitations is essential for graduate-level applied research.