Models for pooled and panel data

Political Science 307

February 8, 2021

Nowadays, the availability of the data often makes it possible to combine cross sectional and longitudinal data. The two types of data with variation across units and time are *pooled data* and *panel data*.

A cumulation of independent cross sectional data with the same measurements at multiple points in time is a *pooled* data set. A typical structure of pooled data is

Unit	Period	V1	V2	V3	V4	V5	V6
1	1	13.3	1	1.9	331	1208.9	0
2	1	13.7	1	3.4	277	2473.0	0
3	1	15.6	0	2.9	404	1552.1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
528	1	12.8	1	4.0	371	927.3	0
529	2	18.7	0	2.1	536	781.4	0
530	2	16.1	1	2.0	583	1031.8	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

In a pooled data set, the observations have time period in common with other units in the same sample and differ in time period from units in different samples. In pooled data, then, the time period is among the variables of interest. The data matrix has $N_1 + N_2 + \cdots + N_T$ rows, where N_t is the sample size for the tth of T time periods in the pool. Examples of pooled data are

- The Current Population Survey (CPS) November Supplement (Series P-20), which questions independent samples of the voting age population every other November about participation in the federal election just past.
- The American National Election Studies (ANES) Time Series Cumulative Data File, which pools the responses to all of the ANES surveys since 1952.
- The ANES Continuous Monitoring Study, combining data from independent samples (of size 76 on average) drawn in each of 46 weeks in 1984.

A collection of data from the same sample of units each observed over multiple time periods is a *panel* data set. A typical structure of panel data is

Unit	Period	V1	V2	V3	V4	V5	V6
1	1	13.3	1	1.9	331	1208.9	0
1	2	13.7	1	3.4	277	2473.0	0
1	3	15.6	0	2.9	404	1552.1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	11	12.8	1	4.0	371	927.3	0
2	1	18.7	0	2.1	536	781.4	0
2	2	16.1	1	2.0	583	1031.8	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Each observation is a unit-period, designated by a double index, i.e., y_{it} and x_{it}. With N units and T time periods, the data matrix has (maximally) NT rows, one for each unit-period observation. It might be considered a set of T cross sections of size N or a set of N time series of length T. Some examples are

- The American National Election Studies (ANES) panel studies, the first of which was 1956, 1958, and 1960: the investigators reinterviewed respondents to the 1956 election study during the 1958 midterm election and again during the 1960 presidential election.
- Rosenstone's data for his election forecasting study, assembled for each state for each presidential election year from 1948 to 1972 (plus 1976 and 1980, for model validation).
- The data used by Alvarez, Cheibub, Przeworski, and Limongi for their study of democracy and development: annual data for 135 countries from 1950 or independence through 1990.

Panel data are often described as "long and narrow" (large T, small N) or "short and wide" (large N, small T). Survey panel studies are usually short and wide, with large numbers of respondents and few time periods (or "waves"). Long and narrow panel data are sometimes called *time-series cross-section* (TSCS) data.

Panel data sets in which every unit is observed in every period are called *balanced.* Panels with some units represented in some periods but not others are called *unbalanced.* Whether balanced or unbalanced, the panel is *fixed* if it contains the same units over the duration of the study. More extended survey studies "replenish" the panel to offset "panel mortality." Other studies, like the Current Population Survey, enroll new panel respondents in each wave and discharge old ones after a set number of waves.

1 Least squares and pooled data

Ordinary least squares applied to pooled data poses few specific complications. The samples are independent across time so serial autocorrelation is not a possibility. Heteroscedasticity associated with time is a prospect, as it would be across any difference in a cross section. The considerable benefit of pooling cross sections is the increase in sample size.

The analytical opportunities in pooled data can be addressed by specification and hypothesis testing. The mean of the distribution of an outcome variable, for example, might be expected to vary over time. Voters' support for the Republican presidential candidate, for instance, might vary from election to election due to unobservable (or not independently observable) factors like the relative "charisma" of the Republican and Democratic nominees, which affect all voters in a given election. Variation in means across time periods can be modeled with time-period dummy variables $p_{ti} = 1$ if i is observed in period t,

$$y_i = b_1 + b_2 p_{2i} + \cdots + b_T p_{Ti} + \mathbf{b}'_x \mathbf{x}_i + e_i,$$

and assessed across individual time periods with a t test or across all time periods jointly with an F test. For reasons that will become clear later, time-period dummy variables in models for pooled data are sometimes referred to as time-period *fixed effects*.

Time-period dummy variables can also be used to model secular trends that are not wholly explained by variation in the characteristics of the sample. Public opinion on gay rights, for example, has lately changed much faster than the social and political characteristics that usually explain it, perhaps suggesting a "bandwagon" effect. Secular trends might be especially well suited for estimation using dummy variables coded cumulatively:

$$\begin{aligned} t2t &= 1 \qquad \text{for } t = 2, \ldots, T \\ t3t &= 1 \qquad \text{for } t = 3, \ldots, T \\ &\vdots \qquad\qquad \vdots \\ tt &= 1 \qquad \text{for } t = T. \end{aligned}$$

The coefficient for each dummy variable represents the difference between the mean for period t units and the mean for period $t-1$ units so the t-test for each coefficient assesses the change from the previous period. (In contrast, the coefficients for dummy variables that represent single years assess the change from the baseline period.) F tests can be used to test for the existence of a secular trend or for reversals of secular trends.

Time-period interactions can also be used to test for differences in the slope coefficients across time. Defining $p_{it} = 1$ if i is observed in period t and x_K as the variable whose coefficient we believe varies from time period to time period, the model is

$$y_i = b_1 + \sum_{t=2}^{T} b_t p_{it} + b_K x_{iK} + b_{K2} p_{i2} x_{iK} + \cdots + b_{KT} p_{iT} x_{iK} + \mathbf{b}'_{(K)} \mathbf{x}_{i(K)} + e_i.$$

The effect of religious confession on voting, for example, might be different in 1960 than it was in election years before or after. The Chow test of structural difference can be used to assess differences in slope (and intercept) coefficients over time.

A popular recent use of pooled data has been the estimation of effects of discrete events on outcomes. Suppose we wish to estimate the impact that an African American candidate, Barack Obama, had on the preferences of racial groups. Using data from the 2008 ANES, we could estimate an equation for voters' ratings of Obama on a feeling thermometer:

$$ObamaRating_i = b_0 + d_{08} Black_i + \mathbf{b}_x' \mathbf{x}_i + e_i,$$

where $Black$ takes the value 1 if the respondent is African American and the value 0 if the respondent is white. (For purposes of illustration, we can exclude other racial categories for the time being.) The coefficient d_{08} represents the difference between the average rating of Obama by blacks and the average rating by whites,

$$d_{08} = \text{Mean}(Black) - \text{Mean}(White),$$

net of other influences.

But of course, Obama had characteristics other than his racial identification that might have made him differentially appealing to blacks and whites: he was a Democrat, a liberal, a northerner, an opponent of the war in Iraq, and so on. To what degree, then, is d_{08} an accurate assessment of the impact of an African American candidate on the preferences of racial groups? Using data from the 2004 ANES, we could estimate the same equation for voters' ratings of John Kerry, also a Democrat, also a liberal, also a northerner, also an opponent (belatedly) of the war in Iraq, and so forth:

$$KerryRating_i = b_0 + d_{04} Black_i + \mathbf{b}_x' \mathbf{x}_i + e_i.$$

As in the first model, d_{04} is the difference between the average rating of the Democratic nominee by blacks and the average rating by whites.

Now suppose that we compare the difference in differences. The difference between the two coefficients,

$$d_{08} - d_{04} = (\bar{r}_{08,B} - \bar{r}_{08,W}) - (\bar{r}_{04,B} - \bar{r}_{04,W}),$$

indicates the differential impact of an African American candidate on the racial difference in voting for the Democratic nominee.

The difference of dummy-variable coefficients across cross sections has come to be called the *difference-in-differences estimator*. It is widely used to assess the causal impact of public policies and other interventions and events. By pooling the two cross sections, we can test whether the difference in differences is distinguishable from 0, that is, we can test

$$\text{H}_0\colon\ d_{08} = d_{04}.$$

Define $y08_i = 1$ if i participated in the 2008 survey year. Then the unrestricted model is

$$DemocratRating_i = b_{0,08}y08_i + b_{0,04}(1-y08_i) + d_{08}y08_iBlack_i + d_{04}(1-y08_i)Black_i + \mathbf{b}'_x\mathbf{x}_i + e_i$$

and the restricted model is

$$DemocratRating_i = b_{0,08}y08_i + b_{0,04}(1-y08_i) + dBlack_i + \mathbf{b}'_x\mathbf{x}_i + e_i.$$

The comparison of the unrestricted and restricted sums of squares in an F test calibrates the difference it made to have an African American candidate in 2008, above and beyond the greater affinity that blacks have for Democratic candidates relative to whites. If the null hypothesis is rejected, further testing might establish whether Obama's candidacy differentially attracted blacks, repelled whites, or something of both.

2 Least squares and panel data

The opportunity (and challenge) in panel data is that sets of observations pertain to the same unit followed over time. Individuals, institutions, and nations can be compared not only each to the other but also to themselves at earlier or later points in time.

2.1 Pooled regression models of panel data

Ordinary least squares applied to panel data estimates

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \epsilon_{it}, \tag{1}$$

a pooled regression, where the sampled values $i = 1, \ldots, N$ and $t = 1, \ldots, T$ include NT total observations. Assuming that the regressors and the residuals are conditionally independent,

$$E(\epsilon_{it}|\mathbf{x}_{it}) = \mathbf{0},$$

and that the product moment matrix of the regressors $\sum_{it}\mathbf{x}_{it}\mathbf{x}'_{it} = \mathbf{X}'\mathbf{X}$ has full rank, ordinary least squares estimates are unbiased.

If, in addition, the residuals are i.i.d. across units and time, the ordinary least squares estimator is efficient. White's estimator applied to the pooled data gives consistent estimates of the least squares coefficient variance matrix under heteroscedasticity. Very often, however, we will expect that the residual variance is heterogeneous across units (which, as we will see, will induce correlation across time within units). We can get at the idea by writing (1) for the T observations of unit i as

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \upsilon_i,$$

so that we can treat $\text{Var}(\upsilon_i)$ as comprising two parts,

$$\text{Var}(\upsilon_i|\mathbf{x}_i) = \sigma_\epsilon^2 \mathbf{I}_T + \mathbf{\Sigma}_i,$$

the common variance of the i.i.d. residuals ϵ_{it} and variances and covariances specific to each unit i. In the spirit of the White estimator, the asymptotic coefficient variance matrix is

$$\text{Est.Asy.Var}(\mathbf{b}) = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1},$$

where $\mathbf{X}_i$ is the $T \times K$ matrix of regressors for the T observations for unit i and $\mathbf{u}_i$ is the $T \times 1$ vector of least squares residuals (from (1)) for unit i. The standard errors so derived are often called *panel robust standard errors* or *clustered standard errors*.

2.1.1 Between groups and within groups estimators

Panel data vary in the cross section across units and in time series across periods. If we designate the grand mean of $\mathbf{x}_{it}$ over i and t by $\bar{\bar{\mathbf{x}}}$ and the grand mean of y_{it} by $\bar{\bar{y}}$, then the pooled estimator of $\boldsymbol{\beta}$ in Equation (1) is

$$\mathbf{b}_P = (\mathbf{S}_{xx}^P)^{-1}\mathbf{S}_{xy}^P,$$

where

$$\mathbf{S}_{xx}^P = \sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\bar{\mathbf{x}}})(\mathbf{x}_{it} - \bar{\bar{\mathbf{x}}})'$$

is the matrix of sums of squares and cross products of the mean-deviated independent variables and

$$\mathbf{S}_{xy}^P = \sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\bar{\mathbf{x}}})(\bar{y}_i - \bar{\bar{y}})$$

is the vector of cross products of the mean-deviated dependent variable and the mean-deviated independent variables.

Part of the variation in y_{it} and $\mathbf{x}_{it}$ is variation across units, indexed by i, and part is variation across time, indexed by t. The contribution of the cross-sectional variation to the pooled estimates might be summarized by averaging the time series observations for each unit to obtain the cross-sectional regression

$$\begin{aligned} \tfrac{1}{T}(y_{i1} + y_{i2} + \cdots + y_{iT}) &= \boldsymbol{\beta}'\big(\tfrac{1}{T}(\mathbf{x}_{i1} + \mathbf{x}_{i2} + \cdots + \mathbf{x}_{iT})\big) + \tfrac{1}{T}(\epsilon_{i1} + \epsilon_{i2} + \cdots + \epsilon_{iT}) \\ \bar{y}_i &= \boldsymbol{\beta}'\bar{\mathbf{x}}_i + \bar{\epsilon}_i. \end{aligned} \tag{2}$$

The ordinary least squares estimator of (2) is

$$\mathbf{b}_B = (\mathbf{S}_{xx}^B)^{-1}\mathbf{S}_{xy}^B$$

where

$$\mathbf{S}_{xx}^{B} = \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})' = \sum_{i=1}^{N} T(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})'$$

is the matrix of sums of squares and cross products of the deviations of the group means of the independent variables from their grand means and

$$\mathbf{S}_{xy}^{B} = \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{y}_i - \bar{\bar{y}}) = \sum_{i=1}^{N} T(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{y}_i - \bar{\bar{y}}),$$

is the vector of cross products of the deviations of the group mean of the dependent variable from its grand mean. The estimator $\mathbf{b}_B$ is called the *between groups estimator* of $\boldsymbol{\beta}$.

Finally, by subtracting (2) from (1) we can express the pooled regression model in terms of deviations from the group means,

$$y_{it} - \bar{y}_i \;=\; \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \epsilon_{it} - \bar{\epsilon}_i. \tag{3}$$

The ordinary least squares estimator of (3) is

$$\mathbf{b}_W = (\mathbf{S}_{xx}^{W})^{-1}\mathbf{S}_{xy}^{W} \tag{4}$$

where

$$\mathbf{S}_{xx}^{W} = \sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'$$

is the matrix of sums of squares and cross products of the deviations of the independent variables from their group means and

$$\mathbf{S}_{xy}^{W} = \sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i)$$

is the vector of cross products of the deviations of the dependent variable from its group mean. The estimator $\mathbf{b}_W$ is called the *within groups estimator* of $\boldsymbol{\beta}$. An average of N time-series regressions of length T, it represents the contribution of the longitudinal variation to the pooled estimates.

The pooled regression (1), the between regression (2), and the within regression (3) are related by

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \epsilon_{it} = \boldsymbol{\beta}'\bar{\mathbf{x}}_i + \bar{\epsilon}_i + (\boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \epsilon_{it} - \bar{\epsilon}_i) = \bar{y}_i + (y_{it} - \bar{y}_i),$$

that is, the pooled regression is a combination of a cross-sectional regression on N unit averaged over T observations and an average over N time-series regressions with T observations. Consequently, the matrices of moments are also so related:

$$\mathbf{S}_{xx}^{P} = \mathbf{S}_{xx}^{B} + \mathbf{S}_{xx}^{W} \;\text{ and }\; \mathbf{S}_{xy}^{P} = \mathbf{S}_{xy}^{B} + \mathbf{S}_{xy}^{W}.$$

We can therefore rewrite the pooled estimator as

$$\mathbf{b}_P = (\mathbf{S}^P_{xx})^{-1}\mathbf{S}^P_{xy} = (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}(\mathbf{S}^B_{xy} + \mathbf{S}^W_{xy}).$$

Furthermore, we can substitute $\mathbf{S}^B_{xy} = \mathbf{S}^B_{xx}\mathbf{b}_B$ and $\mathbf{S}^W_{xy} = \mathbf{S}^W_{xx}\mathbf{b}_W$ to find

$$\begin{aligned}
\mathbf{b}_P &= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}(\mathbf{S}^B_{xx}\mathbf{b}_B + \mathbf{S}^W_{xx}\mathbf{b}_W) \\
&= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}(\mathbf{S}^W_{xx}\mathbf{b}_W + \mathbf{S}^B_{xx}\mathbf{b}_B) \\
&= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_W + (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^B_{xx}\mathbf{b}_B \\
&= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_W + (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}(\mathbf{S}^W_{xx} + \mathbf{S}^B_{xx} - \mathbf{S}^W_{xx})\mathbf{b}_B \\
&= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_W + (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}(\mathbf{S}^W_{xx} + \mathbf{S}^B_{xx})\mathbf{b}_B - (\mathbf{S}^W_{xx} + \mathbf{S}^B_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_B \\
&= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_W + \mathbf{b}_B - (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_B \\
&= (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}\mathbf{b}_W + (\mathbf{I} - (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx})\mathbf{b}_B.
\end{aligned}$$

Defining $\mathbf{\Lambda} = (\mathbf{S}^B_{xx} + \mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx}$,

$$\mathbf{b}_P = \mathbf{\Lambda}\mathbf{b}_W + (\mathbf{I} - \mathbf{\Lambda})\mathbf{b}_B.$$

Thus, the pooled regression estimator from (1) is a matrix-weighted average of the within groups estimator and the between groups estimator. The weights $\mathbf{\Lambda}$ are in the form of a matrix "fraction," the ratio of the within variation to the total variation of the regressors. If the time-series variation outweighs the cross-sectional variation, the pooled estimator puts more weight on the within estimates, and if the cross-section variation outweighs the time-series variation, it puts more weight on the between estimates.

We will see the within and between groups estimators again.

2.2 Models with unobserved effects

Ordinary least squares applied to (1) ignores that T of the observations pertain to each of the N units i. It treats panel data as pooled data, as the designation "pooled regression" implies. Suppose, however, that each unit has characteristics that influence the outcome variable that are specific to itself and constant across time, that is, the true model is

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \epsilon_{it}. \tag{5}$$

The variable α_i is called an *unobserved effect*, *unobserved component*, *unobserved heterogeneity*, or *latent variable*. If y_{it} is test scores, α_i might represent a student's intellectual ability; if y_{it} is military expenditure, α_i might represent an intrinsic value a country's leaders place on force capability.

If we apply ordinary least squares to (5) without taking into account the unobserved effects, then we estimate

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \upsilon_{it},$$

where $\upsilon_{it} = \alpha_i + \epsilon_{it}$ is a composite residual. If α_i is correlated with $\mathbf{x}_{it}$, then υ_{it} is also correlated with the regressors and ordinary least squares is biased and inconsistent. (The bias results from an omitted variable, α_i.) Alternatively, if α_i is uncorrelated with $\mathbf{x}_{it}$, then the composite residuals are not independently distributed and ordinary least squares is inefficient.

- If the unobserved effect α_i is assumed to correlate with the regressors, we should estimate the *fixed effects model*. The effect α_i is "fixed" in the same way that the regressors are fixed in repeated sampling: α_i is nonstochastic.

- Alternatively, if the unobserved effect α_i is assumed to be uncorrelated with the regressors, we should estimate the *random effects model*. The effect α_i is "random" in the sense that it is the outcome of a stochastic process (perhaps purely stochastic, perhaps partly structural).

In either case, the panel structure of the data allows the estimation of the structural parameters of the model taking into account the influence of the unobserved heterogeneity across the units. Multiple observations for each unit give purchase on an effect common to the time-period observations of each unit i.

2.3 Fixed effects estimation

In the fixed effects model, the unobserved effect α_i is a nonstochastic constant that varies across units but is fixed across time. For the unobserved effects model

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \epsilon_{it}, \tag{5}$$

α_i is the intercept for each of the T longitudinal observations for unit i. Thus, (5) is a regression of y_{it} on $N + K$ regressors: the K variables $\mathbf{x}_{it}$ plus N unit-specific constants, either N dummy variables d_i or a constant and $N - 1$ dummy variables. We allow that the observed effect might be correlated with the explanatory variables,

$$E(\alpha_i|\mathbf{x}_i) \neq 0,$$

but we assume that the regressors are strictly exogenous conditional on α_i,

$$E(\epsilon_{it}|\mathbf{x}_i, \alpha_i) = 0, \ t = 1, \ldots, T.$$

Then, so long as the $NT \times (K + N)$ regressor matrix has full column rank, ordinary least squares estimates of (5) are unbiased and consistent. The regression of y_{it} on time-varying independent variables $\mathbf{x}_{it}$ and unit-specific constants is often called the *least squares dummy variable* estimator (and the estimated intercepts are often called, in a loose application of terminology, the "fixed effects").

So far we have said little about the regressors $\mathbf{x}_{it}$, other than that they all vary across time. Suppose, though, that have independent variables in the fixed effects model that vary across units but not over time, that is, one or more of our variables $x_{it} = x_i \; \forall \; t = 1, \ldots, T$. They would either be representations of immutable characteristics, like sex for individuals or region for nations, or representations of attributes that do not vary for any unit over the observed time periods, like educational attainment in a sample of older adults or signature requirements for qualification for the ballot in a sample of states followed over a time of stability in election law. It will be immediately obvious that the effect of a time-invariant variable cannot be distinguished from the unobserved effect α_i. The time-invariant variable and the unit-specific constant are perfectly collinear. The effects of *all* time-invariant differences across units are summarized by the unit-specific intercepts.

Because α_i summarizes the effects of all differences across units, the estimates of the unit-specific intercepts themselves often hold little interest. The inclusion of unit-specific dummy variables in the regressor matrix, moreover, requires the inversion of a $(K+N) \times (K+N)$ matrix in order to estimate the coefficients. We can very easily transform the equations, however, to eliminate the unobserved effect α_i.

If we average equation (5) over $t = 1, \ldots, T$, we get the cross sectional equation

$$\bar{y}_i = \boldsymbol{\beta}'\bar{\mathbf{x}}_i + \alpha_i + \bar{\epsilon}_i, \tag{6}$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \;\; \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}, \text{ and } \bar{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it}.$$

Subtracting the averaged equation (6) from (5),

$$\begin{aligned} y_{it} - \bar{y}_i &= \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \epsilon_{it} - \bar{\epsilon}_i, \\ \breve{y}_{it} &= \boldsymbol{\beta}'\breve{\mathbf{x}}_{it} + \breve{\epsilon}_{it}, \end{aligned} \tag{7}$$

where $\breve{y}_{it} = y_{it} - \bar{y}_i$, $\breve{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, and $\breve{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_i$. The transformation, which mean-deviates the variables within units and eliminates the unobserved effect α_i, is called the *fixed effects transformation*.

The fixed effects transformation is a means to an end, the consistent estimation of $\boldsymbol{\beta}$, the structural coefficients in (5) and (7) both. Given the assumption of strict exogeneity,

$$E(\epsilon_{it}|\mathbf{x}_i, \alpha_i) = 0, \;\; t = 1, \ldots, T,$$

the de-meaned residual is also uncorrelated with the regressors,

$$E(\breve{\epsilon}_{it}) = E(\epsilon_{it}|\mathbf{x}_i) - E(\bar{\epsilon}_{it}) = 0,$$

which means that

$$E(\breve{\epsilon}_{it}|\breve{\mathbf{x}}_{i1}, \ldots, \breve{\mathbf{x}}_{iT}) = 0$$

as each vector $\breve{\mathbf{x}}_{it}$ is a function of $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$. Under the strict exogeneity assumption, that is, ordinary least squares applied to the time-mean-deviated data is unbiased and consistent. The estimator is called the *fixed effects estimator*.

To derive the fixed effects estimator, write (5) for all time periods as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \alpha_i \mathbf{1} + \boldsymbol{\epsilon}_i, \tag{8}$$

where $\mathbf{y}_i$, $\mathbf{1}$, and $\boldsymbol{\epsilon}_i$ are $T \times 1$ and $\mathbf{X}_i$ is $T \times K$. Define a $T \times T$ centering matrix

$$\mathbf{C}_T = \mathbf{I}_T - \bar{\mathbf{J}}_T = \mathbf{I}_T - \tfrac{1}{T}\mathbf{1}\mathbf{1}',$$

which is symmetric and idempotent and has rank $T-1$. Premultiplication of (8) by $\mathbf{C}$ gives

$$\begin{aligned}
\mathbf{C}_T \mathbf{y}_i &= \mathbf{C}_T \mathbf{X}_i \boldsymbol{\beta} + \alpha_i \mathbf{C}_T \mathbf{1} + \mathbf{C}_T \boldsymbol{\epsilon}_i \\
\mathbf{y}_i - \bar{\mathbf{y}}_i &= \left(\mathbf{X}_i - \bar{\mathbf{X}}_i\right) \boldsymbol{\beta} + \alpha_i \left(\mathbf{1} - \tfrac{1}{T}\mathbf{1}\mathbf{1}'\mathbf{1}\right) + \boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}}_i \\
\breve{\mathbf{y}}_i &= \breve{\mathbf{X}}_i \boldsymbol{\beta} + \breve{\boldsymbol{\epsilon}}_i,
\end{aligned} \tag{9}$$

which is (7) written for all time periods.

Now assume that the columns of $\breve{\mathbf{X}}_i$ are linearly independent: $\operatorname{rank}(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i) = K$. Note that the assumption rules out regressors that do not vary over time for any i; in such case a column of $\breve{\mathbf{X}}_i = \mathbf{X}_i - \bar{\mathbf{X}}_i$ will equal $\mathbf{0}$ for all cross sections i. In large samples, the equivalent assumption is $\operatorname{rank}(E(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i)) = K$.

With strict exogeneity and linear independence for the regressors $\breve{\mathbf{X}}_i$, the fixed effects estimator $\mathbf{b}_{FE}$ is the regression of $\breve{\mathbf{y}}_i$ on $\breve{\mathbf{X}}_i$:

$$\mathbf{b}_{FE} = \left(\sum_{i=1}^{N} \breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \breve{\mathbf{X}}_i'\breve{\mathbf{y}}_i\right) = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{C}_T\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{C}_T\mathbf{y}_i\right). \tag{10}$$

Comparing (7) to (3) and (10) to (4), we see that the fixed effects estimator is also the *within groups estimator*. It uses only the time variation within each cross section in the estimation, that is, the estimator is the sum of N time-series regressions with T observations each. (With unobserved effects, the pooled estimator is ordinary least squares applied to (5) and the *between groups estimator* is OLS applied to (6), the data for each unit averaged over time. The between groups estimator is not consistent under the assumptions in this section.) The fixed effects estimator is unbiased and consistent under the assumptions of strict exogeneity and linear independence of the regressors. Given the equivalence of (5) and (7), the fixed effects estimator (10) is also frequently called the *least squares dummy variable* (LSDV) estimator.

A last detail is the unobserved effects, α_i. The fixed effects estimator is based on a model, Equation (7), that eliminates them. The least squares normal equations for the dummy variable model (5), though, show that

$$a_i = \bar{y}_i - \mathbf{b}'_{FE}\bar{\mathbf{x}}_i = \frac{1}{T}\left(\sum_{t=1}^{T} y_{it} - \mathbf{b}'\sum_{t=1}^{T}\mathbf{x}_{it}\right) = \frac{1}{T}\sum_{t=1}^{T} u_{it},$$

the average of the T residuals generated from the structural components (excluding the fixed effects) from the least squares dummy variable regression for unit i, is an unbiased estimator of α_i (see Hsaio 2014: 36). Given that $\upsilon_{it} = \alpha_i + \epsilon_{it}$,

$$E(\upsilon_{it}) = \alpha_i + E(\epsilon_{it}) = \alpha_i.$$

As the expression indicates, a_i can be calculated from the data using $\mathbf{b}_{FE}$. Thus, it is possible to estimate the unit-specific unobserved effects and the degree of heterogeneity across the units.

Before proceeding further, it may be well to present the model in terms of the data matrices. In the data, $\mathbf{y}$ is a "stacked" vector of N $T \times 1$ vectors $\mathbf{y}_i$ and $\mathbf{X}$ is a "stacked" matrix of N $T \times K$ matrices $\mathbf{X}_i$, with the unobserved effect α_i and the residuals $\boldsymbol{\epsilon}$ defined likewise: (8) is

$$\begin{aligned}
\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} &= \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \alpha_1\mathbf{1} \\ \alpha_2\mathbf{1} \\ \vdots \\ \alpha_N\mathbf{1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix} \\
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} \otimes \mathbf{1}_T + \boldsymbol{\epsilon},
\end{aligned}$$

where $\boldsymbol{\alpha}$ is a $N \times 1$ vector of unobserved effects, $\mathbf{1}_T$ is a $T \times 1$ vector of 1s, and $\otimes$ is the *Kronecker product*

$$\mathbf{A}_{M\times N} \otimes \mathbf{B}_{K\times J} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MN}\mathbf{B} \end{bmatrix}_{MK\times NJ}.$$

The centering matrix that time-mean-deviates the stacked data is

$$\mathbf{I}_N \otimes \mathbf{C}_T,$$

where $\mathbf{C}_T$ is defined as before, so that $\mathbf{I}_N \otimes \mathbf{C}_T$ is a $NT \times NT$ *block diagonal* matrix with $\mathbf{C}_T$ on the N main diagonal blocks and $\mathbf{0}_{T\times T}$ blocks off the diagonal. The fixed effects estimator

(10), then, is

$$
\begin{aligned}
\mathbf{b}_{FE} &= (\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{C}_T)\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{C}_T)\mathbf{y} \\
&= \left([\mathbf{X}_1' \ \ \mathbf{X}_2' \ \ \cdots \ \ \mathbf{X}_N'] \begin{bmatrix} \mathbf{C}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_T \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \right)^{-1} \\
&\qquad \times \left([\mathbf{X}_1' \ \ \mathbf{X}_2' \ \ \cdots \ \ \mathbf{X}_N'] \begin{bmatrix} \mathbf{C}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_T \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} \right) \\
&= (\mathbf{X}_1'\mathbf{C}_T\mathbf{X}_1 + \mathbf{X}_2'\mathbf{C}_T\mathbf{X}_2 + \cdots + \mathbf{X}_N'\mathbf{C}_T\mathbf{X}_N)^{-1} (\mathbf{X}_1'\mathbf{C}_T\mathbf{y}_1 + \mathbf{X}_2'\mathbf{C}_T\mathbf{y}_2 + \cdots + \mathbf{X}_N'\mathbf{C}_T\mathbf{y}_N) \\
&= \left(\breve{\mathbf{X}}_1'\breve{\mathbf{X}}_1 + \breve{\mathbf{X}}_2'\breve{\mathbf{X}}_2 + \cdots + \breve{\mathbf{X}}_N'\breve{\mathbf{X}}_N\right)^{-1} \left(\breve{\mathbf{X}}_1'\breve{\mathbf{y}}_1 + \breve{\mathbf{X}}_2'\breve{\mathbf{y}}_2 + \cdots + \breve{\mathbf{X}}_N'\breve{\mathbf{y}}_N\right).
\end{aligned}
$$

2.3.1 Estimation of the fixed effects coefficient variance matrix

In order to derive the coefficient variance matrix, we assume that the residuals in (4) are independent and identically distributed (i.i.d.):

$$E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma_\epsilon^2 \mathbf{I}_T.$$

We can rewrite the coefficient estimator (10) as

$$\mathbf{b}_{FE} = \boldsymbol{\beta} + \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{C}_T\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{C}_T\boldsymbol{\epsilon}_i\right)$$

and so using the i.i.d. assumption we find the variance of $\mathbf{b}_{FE}$ to be

$$\mathrm{Var}(\mathbf{b}_{FE}) = \sigma_\epsilon^2 \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{C}_T\mathbf{X}_i\right)^{-1} = \sigma_\epsilon^2 \left(\sum_{i=1}^{N} \breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i\right)^{-1}.$$

The standard errors of the coefficients, of course, are the square roots of the diagonal elements of $\mathrm{Var}(\mathbf{b}_{FE})$.

The residual variance σ_ϵ^2 is the variance of the residuals in the dummy variable model (8) for the untransformed data, not the fixed effects model (9) estimated on the time-mean-deviated data. The residual variance estimator for (8) is

$$s_e^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - a_i - \mathbf{b}_{FE}'\mathbf{x}_{it})^2}{NT - N - K} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2}{N(T-1) - K},$$

where the numerator is the sum of the squared residuals from the dummy variable regression and where the degrees of freedom in the denominator equals the total number of observations, NT, minus the total number of parameter estimates, $N+K$: N unobserved effects a_i and K coefficients $\mathbf{b}_{FE}$. Recalling that $a_i = \bar{y}_i - \mathbf{b}'_{FE}\bar{\mathbf{x}}_i$, the itth residual is

$$\begin{aligned} e_{it} &= y_{it} - a_i - \mathbf{b}'_{FE}\mathbf{x}_{it} \\ &= y_{it} - (\bar{y}_i - \mathbf{b}'_{FE}\bar{\mathbf{x}}_i) - \mathbf{b}'_{FE}\mathbf{x}_{it} \\ &= (y_{it} - \bar{y}_i) - \mathbf{b}'_{FE}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \\ &= \breve{y}_{it} - \mathbf{b}'_{FE}\breve{\mathbf{x}}_{it} \\ &= \breve{e}_{it}, \end{aligned}$$

the residuals from the fixed effects regression (9), estimated on the time-mean-deviated data. Thus, the residual sum of squares in the numerator of

$$s_e^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2}{N(T-1)-K} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \breve{e}_{it}^2}{N(T-1)-K}$$

is identical to the residual sum of squares estimated by the fixed effects regression (9). Ordinary least squares applied to (9), however, will compute the degrees of freedom in the denominator as $NT-K$ rather than $N(T-1)-K$ and so it will report an estimate of the residual variance that is incorrect and smaller than s_e^2. It is straightforward to apply a correction to the residual variance estimator and the coefficient variance estimator. Computer programs with routines for fixed effects estimation will always make the correct calculation.

2.3.2 Inference

The fixed effects estimator of the structural coefficients, $\mathbf{b}_{FE}$, is unbiased and it is also consistent for large N, large T, or both. (In short and wide panels, the large sample properties are usually justified by a large cross section of units.) Unsurprisingly, the unobserved effects estimator a_i is biased and it is consistent only for large T. In smaller samples, inference can be supported with a normality assumption for the residuals. For tests of specific coefficients,

$$t = \frac{b_k - \beta_k}{SE(b_k)} \sim t(N(T-1)-K),$$

and for tests of linear restrictions,

$$F = \frac{(\mathrm{RSS}_r - \mathrm{RSS}_u)/q}{\mathrm{RSS}_u/(N(T-1)-K)} \sim F(q, N(T-1)-K),$$

where RSS_r is the residual sum of squares from the restricted regression, RSS_r is the residual sum of squares from the unrestricted model, and q is the number of linear restrictions. The unobserved effects α_i for specific units are rarely of interest (although they might be examined in certain circumstances to see which are particularly large). The existence of heterogeneity across the panel units, however, often is of interest. The test of $\mathrm{H}_0\colon\ \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$ compares the residual sum of squares from the fixed effects model (9) (which is the same as the residual sum of squares from (8)) to the residual sum of squares from the pooled regression

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \epsilon_{it},$$

where α is a regression constant. The test statistic is

$$F = \frac{(\mathrm{RSS}_P - \mathrm{RSS}_{FE})/(N-1)}{\mathrm{RSS}_{FE}/(N(T-1)-K)} \sim F(N-1, N(T-1)-K),$$

where RSS_P is the residual sum of squares from the pooled regression, RSS_{FE} is the residual sum of squares from the fixed effects regression, and $N-1$ indicates the restriction of N unobserved effects to zero and the estimation of a pooled constant. It can be considered a test of the necessity of treating the data as a panel rather than as a pool.

In the pooled model, the residual combines the unobserved effects and the stochastic errors so that $\upsilon_{it} = \alpha_i + \epsilon_{it}$. The variance of the pooled residual is $\sigma_\upsilon^2 = \sigma_\alpha^2 + \sigma_\epsilon^2$. We have a consistent estimator of σ_ϵ^2,

$$s_e^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2}{N(T-1)-K} = \frac{\sum_{i=1}^N \sum_{t=1}^T \breve{e}_{it}^2}{N(T-1)-K},$$

and we can estimate σ_υ^2 consistently either by estimating the residuals from a pooled regression or estimating them using the (unbiased and consistent) fixed effects coefficients:

$$s_u^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{b}'_{FE}\mathbf{x}_{it})}{NT-K}.$$

With $\upsilon_{it} = \alpha_i + \epsilon_{it}$, we can decompose the residual sum of squares for the fixed effects regression,

$$\begin{aligned} \mathrm{RSS}_u &= \mathrm{RSS}_e + \mathrm{RSS}_a \\ \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{b}'_{FE}\mathbf{x}_{it})^2 &= \sum_{i=1}^N \sum_{t=1}^T \breve{e}_{it}^2 + \sum_{i=1}^N T a_i^2, \end{aligned}$$

and therefore obtain an estimate of the variance of the unobserved effects,

$$s_a^2 = s_u^2 - s_e^2.$$

We might then apportion the total sum of squares between the estimated sum of squares (explained by the "within" model), the sum of squares for the unobserved effects (explained by unobserved heterogeneity "between" units), and the residual sum of squares (unexplained by either the model or the unobserved effects).

2.3.3 Unbalanced panels

So far, we have assumed that the panel is balanced, that we have T over-time observations for each of the N units. Very frequently, however, panel data are unbalanced, owing to

- Missing data. For some time periods, the data were not gathered or not reported.
- Panel mortality. In panel surveys, participants drop out or cannot be recontacted.
- Panel eligibility. In a panel of nations or organizations, observations only begin when a country or organization forms, or observations end when a country or organization ceases to be independent.

As Greene shows, the complications of unbalanced panels in the fixed effects model pertain largely to the computation of means and variances, and they are minor. See Greene (2000: 566–67) for further details.

2.3.4 Robust estimation

Fixed effects estimation takes into account unobserved heterogeneity that might induce heteroscedasticity or serial autocorrelation in pooled regression residuals. Even having modeled unobserved heterogeneity, though, residual variances might differ across units and residuals might be correlated across time within units. Following White, Arellano proposes a robust fixed effects coefficient variance estimator

$$\text{Est.Asy.Var}(\mathbf{b}_{FE}) = \left(\sum_{i=1}^{N} \breve{\mathbf{X}}_i' \breve{\mathbf{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \breve{\mathbf{X}}_i' \mathbf{e}_i \mathbf{e}_i' \breve{\mathbf{X}}_i\right) \left(\sum_{i=1}^{N} \breve{\mathbf{X}}_i' \breve{\mathbf{X}}_i\right)^{-1},$$

where $\breve{\mathbf{X}}_i = \mathbf{C}_T \mathbf{X}_i$ is the time-demeaned $T \times K$ matrix of regressors for the ith unit and $\mathbf{e}_i$ is the $T \times 1$ vector of either fixed effects or pooled residuals. It is consistent for heteroscedasticity and for serial autocorrelation so long as N is large relative to T.

2.3.5 Panel corrected standard errors

The nature of panel data raises another possibility, that the residuals may be correlated across units, in the manner of spatial autocorrelation. Because of the longitudinal variation in panel data sets, the contemporaneous correlation in the errors across units may be estimated and used to adjust the ordinary least squares coefficient variance matrix. Beck and Katz (1995) propose a "panel corrected" OLS variance matrix

$$\text{Est.Asy.Var}(\mathbf{b}_{FE}) = (\breve{\mathbf{X}}'\breve{\mathbf{X}})^{-1}(\breve{\mathbf{X}}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Sigma}}_i)\breve{\mathbf{X}})(\breve{\mathbf{X}}'\breve{\mathbf{X}}_i)^{-1},$$

where the (i, j)th element of $\hat{\boldsymbol{\Sigma}}_i$ is

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^{T} e_{it} e_{jt}.$$

The procedure corrects for heteroscedasticity and spatial autocorrelation across units. Serial autocorrelation must be addressed before the correction is applied.

2.3.6 Difference in differences estimation

A frequent use of fixed effects models is the estimation of intervention effects. In an unobserved effects model with two time periods,

$$y_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \delta_s s_{it} + \delta_p p_{it} + \delta_{sp} s_{it} p_{it} + \alpha_i + \epsilon_{it}, \ t = 0, 1$$

suppose we have a variable to represent an intervention at time $t = 1$,

$$p_{it} = \begin{cases} 0 & \text{if } t = 0 \\ 1 & \text{if } t = 1, \end{cases}$$

and a variable that distinguishes a "treatment" group exposed to the intervention and a "control" group not so exposed,

$$s_{it} = \begin{cases} 1 & \text{if } i \text{ is in the "treatment" group} \\ 0 & \text{if } i \text{ is in the "control" group.} \end{cases}$$

The intervention p_{it} might be an actual experimental intervention, with the placement in treatment and control groups, s_{it}, determined by random assignment. Green and Gerber, for example, selected registered voters randomly to receive ($s_{it} = 1$) or not receive ($s_{it} = 0$) a message encouraging them to turn out to vote ($p_{it} = 1$). Very often, s_{it} is the outcome of a "natural experiment," an arguably exogenous event or circumstance that exposed one population to the intervention exclusively (or to a much greater extent than the other population). Often cited as an example is Card's study of the effect of immigration on the wages of low-skilled workers, comparing workers' earnings in Miami ($s_{it} = 1$) and Los Angeles ($s_{it} = 0$) each before ($p_{it} = 0$) and after ($p_{it} = 1$) the sudden influx of 125,000 immigrants from Cuba during the Mariel boatlift of 1980. (The similarity of natural experiments to controlled and randomized experiments is often disputed.) Maintaining proper respect for the limits placed on inference by nonexperimental data, s_{it} might also be a distinction between units that we expect to be consequential in the response to a stimulus that varies from one time period to the other, as in our earlier example of a differential effect of Obama's candidacy ($p_{it} = 1$) on blacks ($s_{it} = 1$) and whites ($s_{it} = 0$).

The special attraction of panel data in estimating treatment effects is that change may be measured within the same units, holding constant the characteristics of the treatment and control groups. Given the coding, the control group at $t = 0$ is the excluded category with conditional mean $E(y_{it}|s_{it} = 0, p_{it} = 0) = \bar{y}_{00} = 0$, counting the constant among the regressors $\mathbf{x}_{it}$. At $t = 1$, after the intervention, the conditional mean for the control group is $E(y_{it}|s_{it} = 0, p_{it} = 1) = \bar{y}_{01} = \delta_p$. For the treatment group, on the other hand, the conditional mean before the intervention is $E(y_{it}|s_{it} = 1, p_{it} = 0) = \bar{y}_{10} = \delta_s$, and at $t = 1$, after the intervention, the conditional mean is $E(y_{it}|s_{it} = 1, p_{it} = 1) = \bar{y}_{11} = \delta_s + \delta_p + \delta_{sp}$, because

$$s_{it}p_{it} = \begin{cases} 0 & \text{if } i \text{ is in the control group and } t = 0 \; or \; t - 1 \\ 0 & \text{if } i \text{ is in the treatment group and } t = 0 \\ 1 & \text{if } i \text{ is in the treatment group and } t = 1. \end{cases}$$

Thus, the difference in differences estimate of the effect of the intervention on the treatment group is the change in $\bar{y}$ in the treatment group relative to the change in $\bar{y}$ in the control group:

$$(\bar{y}_{11} - \bar{y}_{10}) - (\bar{y}_{01} - \bar{y}_{00}) = \big((\delta_s + \delta_p + \delta_{sp}) - \delta_s\big) - \big(\delta_p - 0\big) = \delta_{sp}.$$

The fixed effects transformation, applied to y_{it} and $\mathbf{x}_{it}$ and to s_{it} and p_{it} as well,

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \delta_p p_{it} + \delta_{sp} s_{it} p_{it} + \epsilon_{it} - \bar{\epsilon}_i,$$

eliminates the unobserved effects α_i and the unit-specific variable s_{it}. Note that $p_{it} - \bar{p}_i = 0.5$ for $t = 1$ and $p_{it} - \bar{p}_i = -0.5$ for $t = 0$; analysts usually specify the intervention with p_{it} instead, with no difference for the estimates. The difference in differences estimate is still

$$(\delta_p + \delta_{sp}) - \delta_p = \delta_{sp}.$$

Analyses of intervention effects often, however, use a different transformation, the *first-difference transformation*,

$$\begin{aligned} y_{i1} - y_{i0} &= \boldsymbol{\beta}'(\mathbf{x}_{i1} - \mathbf{x}_{i0}) + \delta_p + \delta_{sp} s_{it} + \epsilon_{i1} - \epsilon_{i2} \\ \Delta y_{it} &= \boldsymbol{\beta}'\Delta\mathbf{x}_{it} + \delta_p + \delta_{sp} s_{it} + \Delta\epsilon_{it}. \end{aligned}$$

By subtracting the data for $t = 0$ from the data for $t = 1$, the first-difference transformation reduces the number of cases from $2N$ to N. Like the fixed-effects transformation, it eliminates the unobserved effects α_i and the unit-specific variable s_{it}. The differenced treatment variable $p_{i1} - p_{i0} = 1$ for all i, and the treatment-intervention interaction $s_{it}p_{it} = s_{it}$. The difference in difference estimate is still

$$(\delta_p + \delta_{sp}) - \delta_p = \delta_{sp}.$$

For panel data with $T = 2$, the fixed effects model and the first differenced model are mathematically equivalent.

2.4 Random effects estimation

In the random effects model, the unobserved effect α_i is a stochastic constant that varies across units but is fixed across time. If

$$E(\epsilon_{it}|\mathbf{x}_i, \alpha_i) = 0,\ t = 1, \ldots, T,$$

and if, in addition,

$$E(\alpha_i|\mathbf{x}_i) = 0,$$

meaning that the unobserved effects are *uncorrelated* with the regressors $\mathbf{x}_i$, then the least squares pooled regression

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \epsilon_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \upsilon_{it} \tag{5}$$

yields consistent but inefficient estimates of the structural coefficients $\boldsymbol{\beta}$.

The residual in (5),

$$\upsilon_{it} = \alpha_i + \epsilon_{it},$$

is a composite of two stochastic variables, ϵ_{it}, which varies across units and time, and α_i, which varies only across units. The residual variance is

$$\mathrm{Var}(\upsilon_{it}) = E(\alpha_i + \epsilon_{it})^2 = E(\alpha_i^2 + 2\alpha_i\epsilon_{it} + \epsilon_{it}^2) = \sigma_\alpha^2 + \sigma_\epsilon^2,$$

given the first assumption above and the usual i.i.d. assumption,

$$E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i') = \sigma_\epsilon^2\mathbf{I}_T.$$

The division of the residuals and their variance into parts means that (5) is an *error components model* (or *variance components model*).

Because the stochastic variable ϵ_{it} varies over time but the stochastic variable α_i does not,

$$E(\upsilon_{it}\upsilon_{is}) = E(\alpha_i + \epsilon_{it})(\alpha_i + \epsilon_{is}) = E(\alpha_i^2 + \alpha_i\epsilon_{it} + \alpha_i\epsilon_{is} + \epsilon_{it}\epsilon_{is}) = E(\alpha_i^2) = \sigma_\alpha^2,\ \forall\ t \neq s,$$

that is, ϵ_{it} is serially uncorrelated but α_i, a serial constant, is not. Thus, the residual variance matrix for (5) is

$$\boldsymbol{\Omega}_T = E(\boldsymbol{\upsilon}\boldsymbol{\upsilon}') = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\epsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{bmatrix}. \tag{11}$$

The residual variance matrix $\mathbf{\Omega}_T$ has the *random effects structure*: it depends on only two parameters, σ^2_α and σ^2_ϵ, regardless of T, and the covariance between the composite residuals υ_{it} and υ_{is} does not depend upon $t - s$. In fact, the correlation between υ_{it} and υ_{is},

$$\rho_{\upsilon_{it}\upsilon_{is}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\epsilon} \ \forall\ t \neq s.$$

The squared correlation is also the proportion of the total residual variance that stems from the unobserved effects α_i; in that sense it is a measure of the importance of the unobserved heterogeneity.

Given the structure of $\mathbf{\Omega}_T$, the ordinary least squares residuals are not i.i.d. and OLS is inefficient. As before, however, we can derive efficient estimates of $\boldsymbol{\beta}$ by applying generalized least squares. For (5), the GLS estimator is

$$b_{RE} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{\Omega}_T^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{\Omega}_T^{-1} \mathbf{y}_i\right).$$

It is called the *random effects estimator*.

Its main element $\mathbf{\Omega}_T$, the residual variance matrix in (11), may be written

$$\mathbf{\Omega}_T = \sigma^2_\epsilon \mathbf{I}_T + \sigma^2_\alpha \mathbf{1}_T \mathbf{1}_T' = \sigma^2_\epsilon \mathbf{I}_T + T\sigma^2_\alpha \bar{\mathbf{J}}_T,$$

where $\bar{\mathbf{J}}_T = (1/T)\mathbf{1}\mathbf{1}'$. Recalling the centering matrix $\mathbf{C}_T = \mathbf{I}_T - \bar{\mathbf{J}}_T$,

$$\mathbf{\Omega}_T = \sigma^2_\epsilon \mathbf{I}_T + T\sigma^2_\alpha \bar{\mathbf{J}}_T = \sigma^2_\epsilon(\mathbf{C}_T + \bar{\mathbf{J}}_T) + T\sigma^2_\alpha \bar{\mathbf{J}}_T = \sigma^2_\epsilon \mathbf{C}_T + (\sigma^2_\epsilon + T\sigma^2_\alpha)\bar{\mathbf{J}}_T = \sigma^2_\epsilon\left(\mathbf{C}_T + \frac{\sigma^2_\epsilon + T\sigma^2_\alpha}{\sigma^2_\epsilon}\bar{\mathbf{J}}_T\right).$$

Both $\mathbf{C}_T$ and $\bar{\mathbf{J}}_T$ are idempotent and $\mathbf{C}_T\bar{\mathbf{J}}_T = \bar{\mathbf{J}}_T - \bar{\mathbf{J}}_T\bar{\mathbf{J}}_T = \mathbf{0}$. Therefore,

$$\mathbf{\Omega}_T^{-1} = \frac{1}{\sigma^2_\epsilon}\left(\mathbf{C}_T + \frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + T\sigma^2_\alpha}\bar{\mathbf{J}}_T\right),$$

as may be confirmed by multiplication with $\mathbf{\Omega}_T$.

As with other GLS estimators, we can derive a transformation of the variables to implement GLS using OLS on the transformed data. Assume, for now, that σ^2_ϵ and σ^2_α are known. Substituting back for $\mathbf{C}_T$,

$$\mathbf{\Omega}_T^{-1} = \frac{1}{\sigma^2_\epsilon}\left(\mathbf{I}_T - \bar{\mathbf{J}}_T + \frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + T\sigma^2_\alpha}\bar{\mathbf{J}}_T\right) = \frac{1}{\sigma^2_\epsilon}\left(\mathbf{I}_T - \left(1 - \frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + T\sigma^2_\alpha}\right)\bar{\mathbf{J}}_T\right),$$

which is a symmetric, positive definite matrix. Letting

$$\theta = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma^2_\epsilon + T\sigma^2_\alpha}},$$

the square root of $\mathbf{\Omega}_T^{-1}$ is

$$\mathbf{\Omega}_T^{-1/2} = \frac{1}{\sigma_\epsilon}\big(\mathbf{I}_T - \theta\bar{\mathbf{J}}_T\big).$$

(For confirmation, square $\mathbf{\Omega}_T^{-1/2}$, noting that $\sigma_\epsilon^2/(\sigma_\epsilon^2 + T\sigma_\alpha^2) = (1-\theta)^2$.) Premultiplying (5),

$$\mathbf{\Omega}_T^{-1/2}\mathbf{y}_i = \frac{1}{\sigma_\epsilon}(\mathbf{y}_i - \theta\bar{\mathbf{y}}_i) = \check{\mathbf{y}}_i,$$

or

$$\check{y}_{it} = (y_{it} - \theta\bar{y}_i)/\sigma_\epsilon,$$

and

$$\mathbf{\Omega}_T^{-1/2}\mathbf{X}_i = \frac{1}{\sigma_\epsilon}(\mathbf{X}_i - \theta\bar{\mathbf{X}}_i) = \check{\mathbf{X}}_i,$$

or

$$\check{x}_{kit} = (x_{kit} - \theta\bar{x}_{ki})/\sigma_\epsilon,$$

and finally,

$$\mathbf{\Omega}_T^{-1/2}\boldsymbol{\upsilon}_i = \frac{1}{\sigma_\epsilon}(\boldsymbol{\upsilon}_i - \theta\bar{\boldsymbol{\upsilon}}_i) = \check{\boldsymbol{\upsilon}}_i,$$

or

$$\check{\upsilon}_{it} = (\upsilon_{it} - \theta\bar{\upsilon}_i)/\sigma_\epsilon.$$

The *partial deviations* or *quasi-demeaned variables* give the transformed model

$$\begin{aligned} \mathbf{\Omega}_T^{-1/2}\mathbf{y}_i &= \mathbf{\Omega}_T^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{\Omega}_T^{-1/2}\boldsymbol{\upsilon}_i \\ \check{\mathbf{y}}_i &= \check{\mathbf{X}}_i\boldsymbol{\beta} + \check{\boldsymbol{\upsilon}}_i, \end{aligned}$$

which may be estimated by ordinary least squares. The GLS estimator, then, is

$$\mathbf{b}_{RE} = \left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{\Omega}_T^{-1}\mathbf{X}\right)^{-1}\left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{\Omega}_T^{-1}\mathbf{y}_i\right) = \left(\sum_{i=1}^N \check{\mathbf{X}}_i'\check{\mathbf{X}}_i\right)^{-1}\left(\sum_{i=1}^N \check{\mathbf{X}}_i'\check{\mathbf{y}}_i\right). \tag{12}$$

Note the similarity to the GLS model for serially autocorrelated residuals. Also note the similarity to the fixed effects estimator (10): $\mathbf{\Omega}_T^{-1} = \mathbf{C}_T$ and the random effects transformations are the same as the fixed effects transformations when $\theta = 1$.

Like the ordinary least squares estimator of (1), the generalized least squares estimator of (5) is a matrix-weighted average of the within groups estimator

$$\mathbf{b}_W = (\mathbf{S}_{\check{x}\check{x}}^W)^{-1}\mathbf{S}_{\check{x}\check{y}}^W = \left(\sum_{i=1}^N\sum_{t=1}^T(\check{\mathbf{x}}_{it} - \bar{\check{\mathbf{x}}}_i)(\check{\mathbf{x}}_{it} - \bar{\check{\mathbf{x}}}_i)'\right)^{-1}\sum_{i=1}^N\sum_{t=1}^T(\check{\mathbf{x}}_{it} - \bar{\check{\mathbf{x}}}_i)(\check{y}_{it} - \bar{\check{y}}_i)$$

and the between groups estimator

$$\mathbf{b}_B = (\mathbf{S}^B_{\check{x}\check{x}})^{-1}\mathbf{S}^B_{\check{x}\check{y}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\check{\mathbf{x}}}_i - \bar{\bar{\check{\mathbf{x}}}})(\bar{\check{\mathbf{x}}}_i - \bar{\bar{\check{\mathbf{x}}}})'\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\check{\mathbf{x}}}_i - \bar{\bar{\check{\mathbf{x}}}})(\bar{\check{y}}_i - \bar{\bar{\check{y}}}),$$

whose sums of squares and cross products are calculated on the partial-deviated data, $\check{y}_{it}$ and $\check{\mathbf{x}}_{it}$. The regressor product moment matrix for the within groups estimator is $\mathbf{S}^W_{\check{x}\check{x}} = \mathbf{S}^W_{xx}$, the same as the regressor moments for the within groups estimator using un-deviated data. The moment matrix for the between groups estimator, however, is $\mathbf{S}^B_{\check{x}\check{x}} = (1-\theta)^2\mathbf{S}^B_{xx}$.

Accordingly, the random effects estimator is

$$\mathbf{b}_{RE} = \mathbf{\Lambda}\mathbf{b}_W + (\mathbf{I} - \mathbf{\Lambda})\mathbf{b}_B,$$

also a matrix-weighted average of the within-groups and between-groups estimates. The weighting matrix is

$$\mathbf{\Lambda} = (\mathbf{S}^W_{xx} + (1-\theta)^2\mathbf{S}^B_{xx})^{-1}\mathbf{S}^W_{xx},$$

with

$$\theta = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma^2_\epsilon + T\sigma^2_\alpha}} = 1 - \left(\frac{1}{1 + T(\sigma^2_\alpha/\sigma^2_\epsilon)}\right)^{1/2}.$$

If the variance of the unobserved effects, σ^2_α, is small relative to the total residual variance, then $\sigma^2_\alpha/\sigma^2_\epsilon \to 0$, $\theta \to 0$,

$$\mathbf{\Lambda} = (\mathbf{S}^W_{xx} + (1-\theta)^2\mathbf{S}^B_{xx})^{-1}\mathbf{S}^W_{xx} \to (\mathbf{S}^W_{xx} + \mathbf{S}^B_{xx})^{-1}\mathbf{S}^W_{xx},$$

and

$$\mathbf{b}_{RE} \to \mathbf{b}_P,$$

that is, the generalized least squares estimator of the unobserved effects model (5) is closer and closer to the ordinary least squares estimator for pooled data (1). This should be no surprise, because OLS is an efficient estimator when the unobserved effects equal zero. If (5), including unobserved effects, is the correct model, the expression also shows, then pooled least squares regression gives greater weight to the between groups estimator relative to the within groups estimator in comparison with the random effects estimator.

Conversely, if the variance of the unobserved effects dominates the residual variance, or if the number of longitudinal observations T is very large, then $\sigma^2_\alpha/\sigma^2_\epsilon \to \infty$ or $T \to \infty$, $\theta \to 1$,

$$\mathbf{\Lambda} = (\mathbf{S}^W_{xx} + (1-\theta)^2\mathbf{S}^B_{xx})^{-1}\mathbf{S}^W_{xx} \to (\mathbf{S}^W_{xx})^{-1}\mathbf{S}^W_{xx} = \mathbf{I},$$

and

$$\mathbf{b}_{RE} \to \mathbf{b}_W.$$

Recalling that $\mathbf{b}_W = \mathbf{b}_{FE}$, the fixed effects estimator, we see that the difference between the random effects estimator and the fixed effects estimator narrows to zero when the unobserved effects are very large. The random effects and fixed effects estimators are also very close when the sample size is very large. Because $E(\epsilon_{it}) = 0$, with a sufficiently large sample the unobserved effects may be estimated by the group-mean residuals to a very high degree of accuracy.

2.4.1 Estimation of σ_α^2 and σ_υ^2

The random effects estimator (12) is feasible only if we can estimate the components of $\mathbf{\Omega}_T^{-1}$, σ_ϵ^2 and σ_α^2. As long as we have consistent estimates of υ_{it} and ϵ_{it} the task is not difficult. The problem, though, is not a shortage of estimates of the variance components but a surfeit. Depending upon the choice of estimator, the same model of the same data can produce very different results. Moreover, depending on the data, each method can also produce variance estimates that are negative! As Greene (276) warns, "the practioner is strongly advised to consult the [statistical] program documentation for resolution."

That said, the main alternatives are fairly similar. As we have seen, the composite residual $\upsilon_{it} = \alpha_i + \epsilon_{it}$ has variance $\sigma_\upsilon^2 = \sigma_\alpha^2 + \sigma_\epsilon^2$. As an estimator of $\mathrm{Var}(\upsilon_{it})$, Greene (375) proposes

$$s_u^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \mathbf{b}_P'\mathbf{x}_{it})^2}{NT - K - 1},$$

where $\mathbf{b}_P$ is the pooled OLS coefficient vector estimated from (5). (Wooldridge (260) gives the same estimator but with $NT - K$ degrees of freedom.) As an estimator of $\mathrm{Var}(\epsilon_{it})$, Greene proposes

$$s_e^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \mathbf{b}_{FE}'\mathbf{x}_{it})^2}{N(T-1) - K},$$

where $\mathbf{b}_{FE}$ is the fixed effects (or within groups) estimator (10). (Hsaio (44) gives the same estimator except using the time-mean-deviated variables $y_{it} - \bar{y}_i$ and $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$. As we have seen, the residuals are the same mathematically.) Greene then estimates $\mathrm{Var}(\alpha_i)$ using

$$s_a^2 = s_u^2 - s_e^2.$$

Hsaio (44) gives a direct estimator for $\mathrm{Var}(\alpha_i)$ based on the between groups estimator. Wooldridge (261) takes an unusual approach to the estimation of $\mathrm{Var}(\alpha_i)$, proposing the average of the $T(T-1)/2$ unique residual covariances $E(\upsilon_{it}\upsilon_{is}) = \sigma_\alpha^2$, using the residuals from a pooled OLS regression as the estimates of υ_{it}.

2.4.2 Inference

With satisfactory variance components estimates in hand, we can estimate the feasible GLS coefficients and, more importantly, the random effects coefficient variances,

$$\mathrm{Var}(\mathbf{b}_{RE}) = \left(\sum_{i=1}^{N} \mathbf{X}_i' \hat{\mathbf{\Omega}}_T^{-1} \mathbf{X}_i\right)^{-1} = s_e^2 \left(\sum_{i=1}^{N} \mathbf{X}_i' (\mathbf{C}_T + \frac{s_e^2}{s_e^2 + T s_a^2} \bar{\mathbf{J}}_T) \mathbf{X}_i\right)^{-1}.$$

The coefficient standard errors, as always, are the square roots of the diagonal elements of $\mathrm{Var}(\mathbf{b}_{RE})$.

We can also test for the presence of the unobserved effects, α_i. Breusch and Pagan offer a Lagrange multiplier test of $\mathrm{H}_0\colon\ \sigma_\alpha^2 = 0$ (or, equivalently, $\mathrm{H}_0\colon\ \rho_{\upsilon_{it}\upsilon_{is}} = 0$). The test statistic is

$$\mathrm{LM} = \frac{NT}{2(T-1)} \left(\frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T} e_{it}\right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2} - 1 \right)^2 \overset{a}{\sim} \chi^2(1).$$

The residuals e_{it} are from (5), a pooled OLS regression. In the numerator of the ratio is (conceptually) the average of the N squared unobserved effects, estimated as the average residual for each unit i, scaled to the average squared residual in the denominator. The limiting distribution of the LM test statistic is χ^2 with 1 degree of freedom. It may be considered a test of the appropriateness of estimating the coefficients with a pooled least squares regression: if $\sigma_\alpha^2 = 0$, then the model is (1), not (5).

2.4.3 Unbalanced panels

Unbalanced panels create more formidable difficulties for random effects models. The residual variance matrix $E(\boldsymbol{\upsilon}\boldsymbol{\upsilon}') = \mathbf{\Omega}_T$ is $T \times T$ for all units i in balanced panels, but of course $\mathbf{\Omega}_i$ is $T_i \times T_i$ in unbalanced panels. The parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_i$ is also specific to each unit, making the combined residual variance matrix heteroscedastic.

Unbalanced panels also complicate the estimation of the residual variance σ_υ^2 and its components σ_α^2 and σ_ϵ^2, σ_υ^2 in particular. See Greene (2000: 577–78) for details.

2.4.4 Robust estimation

Random effects estimation by generalized least squares takes into account the serial autocorrelation induced by stochastic unobserved effects. The model might still be subject to residual variances that differ across units or to serial correlation within units and across time. A robust estimator of the coefficient standard errors under heteroscedasticity or under serial autocorrelation in the random effects model is

$$\mathrm{Est.Asy.Var}(\mathbf{b}_{RE}) = \left(\sum_{i=1}^{N} \check{\mathbf{X}}_i' \check{\mathbf{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \check{\mathbf{X}}_i' \check{\mathbf{e}}_i \check{\mathbf{e}}_i' \check{\mathbf{X}}_i\right) \left(\sum_{i=1}^{N} \check{\mathbf{X}}_i' \check{\mathbf{X}}_i\right)^{-1},$$

where $\check{\mathbf{X}}_i = \sigma_\epsilon^{-1}(\mathbf{X}_i - \theta\bar{\mathbf{X}}_i)$ is the $T \times K$ partial-differenced regressor matrix for the ith unit and $\check{\mathbf{e}}_i$ is the $T \times 1$ vector of random effects residuals for the ith unit.

2.5 Comparing fixed effects and random effects

We have now developed a fixed effects approach and a random effects approach to the problem of unobserved effects in panel data. But which is better? The answer, of course, depends on the substance of the problem. Conceptually, we can compare the estimators using four criteria.

- Correlation between the regressors and the unobserved effects.

  The random effects model assumes that the unobserved effects α_i are uncorrelated with the regressors $\mathbf{x}_{it}$. The fixed effects model does not. If the unobserved effects are in fact correlated with the regressors, the fixed effects estimator is unbiased and consistent but the random effects estimator is not. While the assumption about the correlation of the unobserved effects with the independent variables is a consequential difference, though, it is not an automatic point in favor of fixed effects. If $\mathbf{x}_{it}$ is a vector of treatments in a randomized experiment, for example – i.e., each observation is one of multiple "runs" of the experiment for each subject – then α_i will be uncorrelated in expectation as a matter of course. The assumption may also be defensible with observational data.

- Estimation of observed effects that vary across units but not within units over time

  In the fixed effects model, the unobserved heterogeneity α_i is a structural variable and perfectly correlated with all other time-invariant differences across units, whether measured or unmeasured. Separate effects of unit-level characteristics cannot be estimated. In the random effects model, the unobserved heterogeneity is a component of the residual and uncorrelated with all of the measured differences within and between units. The effects of time-invariant variables can be estimated.

- Exploitation of variation over time and across units in estimation

  As we have seen, the fixed effects estimator is the within groups estimator. In effect, it derives its coefficient estimates as an average of N separate time series regressions. By treating the cross-unit effects as nonstochastic but unobserved, that is, the fixed effects estimator "throws away" information about known sources of variation across units. The random effects estimator, on the other hand, is a weighted combination of a within groups estimator and a between groups estimator. The ability in random effects analysis to incorporate regressors that vary in cross section but not in time series also presumably mitigates the limitation imposed by the random effects assumption that the unobserved effects are uncorrelated with the regressors.

- Inferences to samples and populations

  The last point is subtle. The fixed effects model estimates the coefficient vector assuming that the unobserved effects α_i and the observed regressors $\mathbf{x}_{it}$ may be correlated. The random effects model estimates $\boldsymbol{\beta}$ assuming that the unobserved effects are assigned to each unit stochastically. If the unobserved effects are nonstochastic and correlated with the regressors, however, we have no knowledge of it beyond the information in the sample. Because the effects are unobserved but nonstochastic, and because the coefficient estimates depend upon the correlation between the unobserved effects and the observed independent variables, we cannot really make inferences about the population from the sample. If the unobserved effects α_i are stochastic, on the other hand, we know that they are assigned to each unit in the same way that the residuals ϵ_{it} are assigned, as a random draw from a distribution whose characteristics we can estimate from the sample. We can make inferences about the population from the sample.

2.5.1 Specification tests

Hausman has devised a test between the fixed effects and the random effects specification of the unobserved effects model. The fixed effects estimator is consistent whether the unobserved effects are stochastic or nonstochastic but it is inefficient if the unobserved heterogeneity is stochastic. The random effects estimator, on the other hand, is consistent and efficient if the unobserved effects are stochastic but inconsistent if they are nonstochastic. Hausman's approach tests the null hypothesis $\mathrm{H}_0\colon\ \mathbf{b}_{FE} - \mathbf{b}_{RE} = \mathbf{0}$, that is, it assesses the correspondence between the fixed effects estimates and the random effects. Because both estimators are consistent when α_i is stochastic, the null hypothesis implies that the assumptions of the random effects model are correct and the efficient random effects specification is to be preferred. Because the fixed effects estimator is consistent but the random effects estimator is not when α_i is nonstochastic, the alternative hypothesis implies that the assumptions of the fixed effects model are correct and the consistent fixed effects specification is to be preferred.

The Hausman test statistic is

$$W = (\mathbf{b}_{FE} - \mathbf{b}_{RE})'(\mathrm{Var}(\mathbf{b}_{FE}) - \mathrm{Var}(\mathbf{b}_{RE}))^{-1}(\mathbf{b}_{FE} - \mathbf{b}_{RE}) \overset{a}{\sim} \chi^2(K-1),$$

where the random effects coefficient vector and coefficient variance matrix both delete the constant. The variance of the difference of the coefficient vectors

$$\mathrm{Var}(\mathbf{b}_{FE} - \mathbf{b}_{RE}) = \mathrm{Var}(\mathbf{b}_{FE}) + \mathrm{Var}(\mathbf{b}_{RE}) - 2\mathrm{Cov}(\mathbf{b}_{FE}, \mathbf{b}_{RE}),$$

but Hausman shows that the covariance of an efficient estimator and the difference between an efficient and an inefficient estimator is zero, so that

$$\mathrm{Cov}((\mathbf{b}_{FE} - \mathbf{b}_{RE}), \mathbf{b}_{RE}) = \mathbf{0} \Longrightarrow \mathrm{Cov}(\mathbf{b}_{FE}, \mathbf{b}_{RE}) - \mathrm{Var}(\mathbf{b}_{RE}) = \mathbf{0}.$$

Substituting $\text{Cov}(\mathbf{b}_{FE}, \mathbf{b}_{RE}) = \text{Var}(\mathbf{b}_{RE})$ into the variance,

$$\text{Var}(\mathbf{b}_{FE} - \mathbf{b}_{RE}) = \text{Var}(\mathbf{b}_{FE}) - \text{Var}(\mathbf{b}_{RE}),$$

hence the form of the variance matrix in the test statistic. Its limiting distribution is χ^2 with $(K-1)$ degrees of freedom (the number of elements in the coefficient vectors).

In some cases, the difference in the variance matrices is not a positive definite matrix and the test fails. See Greene (380) and references therein for a discussion of alternatives.

2.5.2 Correlated random effects

The random effects approach assumes that the unobserved effects α_i are uncorrelated with $\mathbf{x}_{it}$. The fixed effects approach subtracts the over-time averages from $\mathbf{x}_i$ to remove the unobserved effects α_i. Suppose instead that we allow the unobserved effects, which are constant over time, to be correlated with the within-unit mean level of $\mathbf{x}_{it}$, $\bar{\mathbf{x}}_i$. Indeed, let

$$\alpha_i = \boldsymbol{\gamma}'\bar{\mathbf{x}}_i + \eta_i,$$

where $E(\mathbf{x}_{it}\eta_i) = \mathbf{0}$. Substituting in (5) for α_i, the unobserved effects model becomes

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\bar{\mathbf{x}}_i + \eta_i + \epsilon_{it}. \tag{13}$$

Transferring the earlier assumptions about α_i to η_i, namely

$$E(\epsilon_{it}|\mathbf{x}_i, \eta_i) = 0, \; t = 1, \ldots, T,$$

and taking note of the new composite residual

$$\upsilon_{it} = \eta_i + \epsilon_{it},$$

(14) clearly calls for random effects estimation. Now, however, the specification allows for correlation between the unobserved effects $\alpha_i = \boldsymbol{\gamma}'\bar{\mathbf{x}}_i + \eta_i$ and the regressors $\mathbf{x}_{it}$. The approach was first suggested by Mundlak. Wooldridge (2010) calls it *correlated random effects.*

Designate the random effects coefficient estimators in (14) by $\mathbf{b}_{CRE}$ and $\mathbf{g}_{CRE}$. Wooldridge (2010) shows that

$$\mathbf{b}_{CRE} = \mathbf{b}_{FE},$$

which is (10), the fixed effects estimator of (5). The equivalence gives an attractive interpretation of fixed effects: it estimates the partial effect of $\mathbf{x}_{it}$ on y_{it} controlling for $\bar{\mathbf{x}}_{it}$. For example, suppose our analysis seeks to explain levels of social welfare spending across nations over time as a function of covariates including national income. Controlling for average national income allows for systematic differences between nations with higher and lower levels of "permanent income." Controlling for average income explicitly using the correlated random effects approach also helps us to see why short-term fluctuations in national income might have a negative, not positive, effect on social welfare spending. The between group and within group differences might have contrasting implications for the outcome.

The correlated random effects approach also provides an alternative to the Hausman test. The random effects model sets $\boldsymbol{\gamma} = \mathbf{0}$, in which case $\alpha_i = \eta_i$. The fixed effects model estimates $\boldsymbol{\gamma}$. The distinction suggests an F test of H_0: $\boldsymbol{\gamma} = \mathbf{0}$. If we can reject H_0 at a sufficient level of confidence we can reject the random effects assumptions in favor of the fixed effects assumptions.