

Chapter 17

Panel Data

17.1 Introduction

Economists traditionally use the term **panel data** to refer to data structures consisting of observations on individuals for multiple time periods. Other fields such as statistics typically call this structure **longitudinal data**. The observed “individuals” can be, for example, people, households, workers, firms, schools, production plants, industries, regions, states, or countries. The distinguishing feature relative to cross-sectional data sets is the presence of multiple observations for each individual. More broadly, panel data methods can be applied to any context with cluster-type dependence.

There are several distinct advantages of panel data relative to cross-section data. One is the possibility of controlling for unobserved time-invariant endogeneity without the use of instrumental variables. A second is the possibility of allowing for broader forms of heterogeneity. A third is modeling dynamic relationships and effects.

There are two broad categories of panel data sets in economic applications: micro panels and macro panels. Micro panels are typically surveys or administrative records on individuals and are characterized by a large number of individuals (often in the 1000’s or higher) and a relatively small number of time periods (often 2 to 20 years). Macro panels are typically national or regional macroeconomic variables and are characterized by a moderate number of individuals (e.g. 7-20) and a moderate number of time periods (20-60 years).

Panel data was once relatively esoteric in applied economic practice. Now, it is a dominant feature of applied research.

A typical maintained assumption for micro panels (which we follow in this chapter) is that the individuals are mutually independent while the observations for a given individual are correlated across time periods. This means that the observations follow a clustered dependence structure. Because of this, current econometric practice is to use cluster-robust covariance matrix estimators when possible. Similar assumptions are often used for macro panels though the assumption of independence across individuals (e.g. countries) is much less compelling.

The application of panel data methods in econometrics started with the pioneering work of Mundlak (1961) and Balestra and Nerlove (1966).

Several excellent monographs and textbooks have been written on panel econometrics, including Arellano (2003), Hsiao (2003), Wooldridge (2010), and Baltagi (2013). This chapter will summarize some of the main themes but for a more in-depth treatment see these references.

One challenge arising in panel data applications is that the computational methods can require meticulous attention to detail. It is therefore advised to use established packages for routine applications. For most panel data applications in economics Stata is the standard package.

17.2 Time Indexing and Unbalanced Panels

It is typical to index observations by both the individual i and the time period t , thus Y_{it} denotes a variable for individual i in period t . We index individuals as $i = 1, \dots, N$ and time periods as $t = 1, \dots, T$. Thus N is the number of individuals in the panel and T is the number of time series periods.

Panel data sets can involve data at any time series frequency though the typical application involves annual data. The observations in a data set will be indexed by calendar time which for the case of annual observations is the year. For notational convenience it is customary to denote the time periods as $t = 1, \dots, T$, so that $t = 1$ is the first time period observed and T is the final time period.

When observations are available on all individuals for the same time periods we say that the panel is **balanced**. In this case there are an equal number T of observations for each individual and the total number of observations is $n = NT$.

When different time periods are available for the individuals in the sample we say that the panel is **unbalanced**. This is the most common type of panel data set. It does not pose a problem for applications but does make the notation cumbersome and also complicates computer programming.

To illustrate, consider the data set Invest1993 on the textbook webpage. This is a sample of 1962 U.S. firms extracted from Compustat, assembled by Bronwyn Hall, and used in the empirical work in Hall and Hall (1993). In Table 17.1 we display a set of variables from the data set for the first 13 observations. The first variable is the firm code number. The second variable is the year of the observation. These two variables are essential for any panel data analysis. In Table 17.1 you can see that the first firm (#32) is observed for the years 1970 through 1977. The second firm (#209) is observed for 1987 through 1991. You can see that the years vary considerably across the firms so this is an unbalanced panel.

For unbalanced panels the time index $t = 1, \dots, T$ denotes the full set of time periods. For example, in the data set Invest1993 there are observations for the years 1960 through 1991, so the total number of time periods is $T = 32$. Each individual is observed for a subset of T_i periods. The set of time periods for individual i is denoted as S_i so that individual-specific sums (over time periods) are written as $\sum_{t \in S_i}$.

The observed time periods for a given individual are typically contiguous (for example, in Table 17.1, firm #32 is observed for each year from 1970 through 1977) but in some cases are non-contiguous (if, for example, 1973 was missing for firm #32). The total number of observations in the sample is $n = \sum_{i=1}^N T_i$.

Table 17.1: Observations from Investment Data Set

Firm Code Number	Year	I_{it}	\bar{I}_i	\dot{I}_{it}	Q_{it}	\bar{Q}_i	\dot{Q}_{it}	\hat{e}_{it}
32	1970	0.122	0.155	-0.033	1.17	0.62	0.55	.
32	1971	0.092	0.155	-0.063	0.79	0.62	0.17	-0.005
32	1972	0.094	0.155	-0.061	0.91	0.62	0.29	-0.005
32	1973	0.116	0.155	-0.039	0.29	0.62	-0.33	0.014
32	1974	0.099	0.155	-0.057	0.30	0.62	-0.32	-0.002
32	1975	0.187	0.155	0.032	0.56	0.62	-0.06	0.086
32	1976	0.349	0.155	0.194	0.38	0.62	-0.24	0.248
32	1977	0.182	0.155	0.027	0.57	0.62	-0.05	0.081
209	1987	0.095	0.071	0.024	9.06	21.57	-12.51	.
209	1988	0.044	0.071	-0.027	16.90	21.57	-4.67	-0.244
209	1989	0.069	0.071	-0.002	25.14	21.57	3.57	-0.257
209	1990	0.113	0.071	0.042	25.60	21.57	4.03	-0.226
209	1991	0.034	0.071	-0.037	31.14	21.57	9.57	-0.283

17.3 Notation

This chapter focuses on panel data regression models whose observations are pairs (Y_{it}, X_{it}) where Y_{it} is the dependent variable and X_{it} is a k -vector of regressors. These are the observations on individual i for time period t .

It will be useful to cluster the observations at the level of the individual. We borrow the notation from Section 4.21 to write \mathbf{Y}_i as the $T_i \times 1$ stacked observations on Y_{it} for $t \in S_i$, stacked in chronological order. Similarly, we write \mathbf{X}_i as the $T_i \times k$ matrix of stacked X'_{it} for $t \in S_i$, stacked in chronological order.

We will also sometimes use matrix notation for the full sample. To do so, let $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_N)'$ denote the $n \times 1$ vector of stacked \mathbf{Y}_i , and set $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_N)'$ similarly.

17.4 Pooled Regression

The simplest model in panel regression is pooled regression

$$\begin{aligned} Y_{it} &= X'_{it}\beta + e_{it} \\ \mathbb{E}[X_{it}e_{it}] &= 0. \end{aligned} \tag{17.1}$$

where β is a $k \times 1$ coefficient vector and e_{it} is an error. The model can be written at the level of the individual as

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\beta + \mathbf{e}_i \\ \mathbb{E}[\mathbf{X}'_i\mathbf{e}_i] &= 0 \end{aligned}$$

where \mathbf{e}_i is $T_i \times 1$. The equation for the full sample is $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ where \mathbf{e} is $n \times 1$.

The standard estimator of β in the pooled regression model is least squares, which can be written as

$$\begin{aligned} \hat{\beta}_{\text{pool}} &= \left(\sum_{i=1}^N \sum_{t \in S_i} X_{it} X'_{it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t \in S_i} X_{it} Y_{it} \right) \\ &= \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{Y}_i \right) \\ &= (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{Y}). \end{aligned}$$

In the context of panel data $\hat{\beta}_{\text{pool}}$ is called the **pooled regression estimator**. The vector of residuals for the i^{th} individual is $\hat{\mathbf{e}}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\text{pool}}$.

The pooled regression model is ideally suited for the context where the errors e_{it} satisfy **strict mean independence**:

$$\mathbb{E}[e_{it} | \mathbf{X}_i] = 0. \tag{17.2}$$

This occurs when the errors e_{it} are mean independent of all regressors X_{ij} for all time periods $j = 1, \dots, T$. Strict mean independence is stronger than pairwise mean independence $\mathbb{E}[e_{it} | X_{it}] = 0$ as well as projection (17.1). Strict mean independence requires that neither lagged nor future values of X_{it} help to forecast e_{it} . It excludes lagged dependent variables (such as Y_{it-1}) from X_{it} (otherwise e_{it} would be predictable given X_{it+1}). It also requires that X_{it} is exogenous in the sense discussed in Chapter 12.

We now describe some statistical properties of $\hat{\beta}_{\text{pool}}$ under (17.2). First, notice that by linearity and the cluster-level notation we can write the estimator as

$$\hat{\beta}_{\text{pool}} = \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}'_i (\mathbf{X}_i \beta + \mathbf{e}_i) \right) = \beta + \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{e}_i \right).$$

Using (17.2)

$$\mathbb{E}[\hat{\beta}_{\text{pool}} | \mathbf{X}] = \beta + \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbb{E}[\mathbf{e}_i | \mathbf{X}_i] \right) = \beta$$

so $\hat{\beta}_{\text{pool}}$ is unbiased for β .

Under the additional assumption that the error e_{it} is serially uncorrelated and homoskedastic the covariance estimator takes a classical form and the classical homoskedastic variance estimator can be used. If the error e_{it} is heteroskedastic but serially uncorrelated then a heteroskedasticity-robust covariance matrix estimator can be used.

In general, however, we expect the errors e_{it} to be correlated across time t for a given individual. This does not necessarily violate (17.2) but invalidates classical covariance matrix estimation. The conventional solution is to use a cluster-robust covariance matrix estimator which allows arbitrary within-cluster dependence. Cluster-robust covariance matrix estimators for pooled regression equal

$$\hat{\mathbf{V}}_{\text{pool}} = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

As in (4.55) this can be multiplied by a degree-of-freedom adjustment. The adjustment used by the Stata `regress` command is

$$\hat{\mathbf{V}}_{\text{pool}} = \left(\frac{n-1}{n-k} \right) \left(\frac{N}{N-1} \right) (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

The pooled regression estimator with cluster-robust standard errors can be obtained using the Stata command `regress cluster(id)` where `id` indicates the individual.

When strict mean independence (17.2) fails the pooled least squares estimator $\hat{\beta}_{\text{pool}}$ is not necessarily consistent for β . Since strict mean independence is a strong and undesirable restriction it is typically preferred to adopt one of the alternative estimators described in the following sections.

To illustrate the pooled regression estimator consider the data set `Invest1993` described earlier. We consider a simple investment model

$$I_{it} = \beta_1 Q_{it-1} + \beta_2 D_{it-1} + \beta_3 CF_{it-1} + \beta_4 T_i + e_{it} \quad (17.3)$$

where I is investment/assets, Q is market value/assets, D is long term debt/assets, CF is cash flow/assets, and T is a dummy variable indicating if the corporation's stock is traded on the NYSE or AMEX. The regression also includes 19 dummy variables indicating an industry code. The Q theory of investment suggests that $\beta_1 > 0$ while $\beta_2 = \beta_3 = 0$. Theories of liquidity constraints suggest that $\beta_2 < 0$ and $\beta_3 > 0$. We will be using this example throughout this chapter. The values of I and Q for the first 13 observations are also displayed in Table 17.1.

In Table 17.2 we present the pooled regression estimates of (17.3) in the first column with cluster-robust standard errors.

17.5 One-Way Error Component Model

One approach to panel data regression is to model the correlation structure of the regression error e_{it} . The most common choice is an error-components structure. The simplest takes the form

$$e_{it} = u_i + \varepsilon_{it} \quad (17.4)$$

Table 17.2: Estimates of Investment Equation

	Pooled	Random Effects	Fixed Effects	Two-Way	Hausman-Taylor
Q_{it-1}	0.0024 (0.0010)	0.0019 (0.0009)	0.0017 (0.0008)	0.0016 (0.0008)	0.0017 (0.0008)
D_{it-1}	0.0096 (0.0041)	-0.0092 (0.0039)	-0.0139 (0.0049)	-0.0140 (0.0051)	0.0132 (0.0050)
CF_{it-1}	0.0261 (0.0111)	0.0412 (0.0125)	0.0491 (0.0132)	0.0476 (0.0129)	0.0408 (0.0119)
T_i	-0.0167 (0.0024)	-0.0181 (0.0028)			-0.0348 (0.0048)
Industry Dummies	Yes	Yes	No	No	Yes
Time Effects	No	No	No	Yes	Yes

Cluster-robust standard errors in parenthesis.

where u_i is an individual-specific effect and ε_{it} are idiosyncratic (i.i.d.) errors. This is known as a **one-way error component model**.

In vector notation we can write $\mathbf{e}_i = \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i$ where $\mathbf{1}_i$ is a $T_i \times 1$ vector of 1's.

The one-way error component regression model is

$$Y_{it} = X'_{it}\beta + u_i + \varepsilon_{it}$$

written at the level of the observation, or $\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i$ written at the level of the individual.

To illustrate why an error-component structure such as (17.4) might be appropriate, examine Table 17.1. In the final column we have included the pooled regression residuals \hat{e}_{it} for these observations. (There is no residual for the first year for each firm due to the lack of lagged regressors for this observation.) What is quite striking is that the residuals for the second firm (#209) are all negative, clustering around -0.25 . While informal, this suggests that it may be appropriate to model these errors using (17.4), expecting that firm #209 has a large negative value for its individual effect u .

17.6 Random Effects

The random effects model assumes that the errors u_i and ε_{it} in (17.4) are conditionally mean zero, uncorrelated, and homoskedastic.

Assumption 17.1 Random Effects. Model (17.4) holds with

$$\mathbb{E}[\varepsilon_{it} | \mathbf{X}_i] = 0 \quad (17.5)$$

$$\mathbb{E}[\varepsilon_{it}^2 | \mathbf{X}_i] = \sigma_\varepsilon^2 \quad (17.6)$$

$$\mathbb{E}[\varepsilon_{it}\varepsilon_{js} | \mathbf{X}_i] = 0 \quad (17.7)$$

$$\mathbb{E}[u_i | \mathbf{X}_i] = 0 \quad (17.8)$$

$$\mathbb{E}[u_i^2 | \mathbf{X}_i] = \sigma_u^2 \quad (17.9)$$

$$\mathbb{E}[u_i\varepsilon_{it} | \mathbf{X}_i] = 0 \quad (17.10)$$

where (17.7) holds for all $s \neq t$.

Assumption 17.1 is known as a **random effects** specification. It implies that the vector of errors \mathbf{e}_i for individual i has the covariance structure

$$\begin{aligned}\mathbb{E}[\mathbf{e}_i | \mathbf{X}_i] &= \mathbf{0} \\ \mathbb{E}[\mathbf{e}_i \mathbf{e}_i' | \mathbf{X}_i] &= \mathbf{1}_i \mathbf{1}_i' \sigma_u^2 + \mathbf{I}_i \sigma_\varepsilon^2 \\ &= \begin{pmatrix} \sigma_u^2 + \sigma_\varepsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_\varepsilon^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_\varepsilon^2 \end{pmatrix} \\ &= \sigma_\varepsilon^2 \Omega_i,\end{aligned}$$

say, where \mathbf{I}_i is an identity matrix of dimension T_i . The matrix Ω_i depends on i since its dimension depends on the number of observed time periods T_i .

Assumptions 17.1.1 and 17.1.4 state that the idiosyncratic error ε_{it} and individual-specific error u_i are strictly mean independent so the combined error e_{it} is strictly mean independent as well.

The random effects model is equivalent to an **equi-correlation** model. That is, suppose that the error e_{it} satisfies

$$\begin{aligned}\mathbb{E}[e_{it} | \mathbf{X}_i] &= 0 \\ \mathbb{E}[e_{it}^2 | \mathbf{X}_i] &= \sigma^2\end{aligned}$$

and

$$\mathbb{E}[e_{is}e_{it} | \mathbf{X}_i] = \rho\sigma^2$$

for $s \neq t$. These conditions imply that e_{it} can be written as (17.4) with the components satisfying Assumption 17.1 with $\sigma_u^2 = \rho\sigma^2$ and $\sigma_\varepsilon^2 = (1-\rho)\sigma^2$. Thus random effects and equi-correlation are identical.

The random effects regression model is

$$Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$$

or $\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i$ where the errors satisfy Assumption 17.1.

Given the error structure the natural estimator for β is GLS. Suppose σ_u^2 and σ_ε^2 are known. The GLS estimator of β is

$$\hat{\beta}_{\text{glS}} = \left(\sum_{i=1}^N \mathbf{X}_i' \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \Omega_i^{-1} \mathbf{Y}_i \right).$$

A feasible GLS estimator replaces the unknown σ_u^2 and σ_ε^2 with estimators. See Section 17.15.

We now describe some statistical properties of the estimator under Assumption 17.1. By linearity

$$\hat{\beta}_{\text{glS}} - \beta = \left(\sum_{i=1}^N \mathbf{X}_i' \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \Omega_i^{-1} \mathbf{e}_i \right).$$

Thus

$$\mathbb{E}[\hat{\beta}_{\text{glS}} - \beta | \mathbf{X}] = \left(\sum_{i=1}^N \mathbf{X}_i' \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \Omega_i^{-1} \mathbb{E}[\mathbf{e}_i | \mathbf{X}_i] \right) = \mathbf{0}.$$

Thus $\hat{\beta}_{\text{glS}}$ is conditionally unbiased for β . The conditional variance of $\hat{\beta}_{\text{glS}}$ is

$$\mathbf{V}_{\text{glS}} = \left(\sum_{i=1}^n \mathbf{X}_i' \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \sigma_\varepsilon^2. \quad (17.11)$$

Now let's compare $\hat{\beta}_{\text{gls}}$ with the pooled estimator $\hat{\beta}_{\text{pool}}$. Under Assumption 17.1 the latter is also conditionally unbiased for β and has conditional variance

$$\mathbf{V}_{\text{pool}} = \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}'_i \Omega_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)^{-1}. \quad (17.12)$$

Using the algebra of the Gauss-Markov Theorem we deduce that

$$\mathbf{V}_{\text{gls}} \leq \mathbf{V}_{\text{pool}} \quad (17.13)$$

and thus the random effects estimator $\hat{\beta}_{\text{gls}}$ is more efficient than the pooled estimator $\hat{\beta}_{\text{pool}}$ under Assumption 17.1. (See Exercise 17.1.) The two variance matrices are identical when there is no individual-specific effect (when $\sigma_u^2 = 0$) for then $\mathbf{V}_{\text{gls}} = \mathbf{V}_{\text{pool}} = (\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2$.

Under the assumption that the random effects model is a useful approximation but not literally true then we may consider a cluster-robust covariance matrix estimator such as

$$\hat{\mathbf{V}}_{\text{gls}} = \left(\sum_{i=1}^N \mathbf{X}'_i \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}'_i \Omega_i^{-1} \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \Omega_i^{-1} \mathbf{X}_i \right) \left(\sum_{i=1}^N \mathbf{X}'_i \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \quad (17.14)$$

where $\hat{\mathbf{e}}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\text{gls}}$. This may be re-scaled by a degree of freedom adjustment if desired.

The random effects estimator $\hat{\beta}_{\text{gls}}$ can be obtained using the Stata command `xtreg`. The default covariance matrix estimator is (17.11). For the cluster-robust covariance matrix estimator (17.14) use the command `xtreg vce(robust)`. (The `xtset` command must be used first to declare the group identifier. For example, `cusip` is the group identifier in Table 17.1.)

To illustrate, in the second column of Table 17.2 we present the random effect regression estimates of the investment model (17.3) with cluster-robust standard errors (17.14). The point estimates are reasonably different from the pooled regression estimator. The coefficient on debt switches from positive to negative (the latter consistent with theories of liquidity constraints) and the coefficient on cash flow increases significantly in magnitude. These changes appear to be greater in magnitude than would be expected if Assumption 17.1 were correct. In the next section we consider a less restrictive specification.

17.7 Fixed Effect Model

Consider the one-way error component regression model

$$Y_{it} = \mathbf{X}'_{it} \beta + u_i + \varepsilon_{it} \quad (17.15)$$

or

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i. \quad (17.16)$$

In many applications it is useful to interpret the individual-specific effect u_i as a time-invariant unobserved missing variable. For example, in a wage regression u_i may be the unobserved ability of individual i . In the investment model (17.3) u_i may be a firm-specific productivity factor.

When u_i is interpreted as an omitted variable it is natural to expect it to be correlated with the regressors \mathbf{X}_{it} . This is especially the case when \mathbf{X}_{it} includes choice variables.

To illustrate, consider the entries in Table 17.1. The final column displays the pooled regression residuals \hat{e}_{it} for the first 13 observations which we interpret as estimates of the error $e_{it} = u_i + \varepsilon_{it}$. As described before, what is particularly striking about the residuals is that they are all strongly negative for firm #209, clustering around -0.25 . We can interpret this as an estimate of u_i for this firm. Examining the values

of the regressor Q for the two firms we can see that firm #209 has very large values (in all time periods) for Q . (The average value \bar{Q}_i for the two firms appears in the seventh column.) Thus it appears (though we are only looking at two observations) that u_i and Q_{it} are correlated. It is not reasonable to infer too much from these limited observations, but the relevance is that such correlation violates strict mean independence.

In the econometrics literature if the stochastic structure of u_i is treated as unknown and possibly correlated with X_{it} then u_i is called a **fixed effect**.

Correlation between u_i and X_{it} will cause both pooled and random effect estimators to be biased. This is due to the classic problems of omitted variables bias and endogeneity. To see this in a generated example view Figure 17.1. This shows a scatter plot of three observations (Y_{it}, X_{it}) from three firms. The true model is $Y_{it} = 9 - X_{it} + u_i$. (The true slope coefficient is -1 .) The variables u_i and X_{it} are highly correlated so the fitted pooled regression line through the nine observations has a slope close to $+1$. (The random effects estimator is identical.) The apparent positive relationship between Y and X is driven entirely by the positive correlation between X and u . Conditional on u , however, the slope is -1 . Thus regression techniques which do not control for u_i will produce biased and inconsistent estimators.

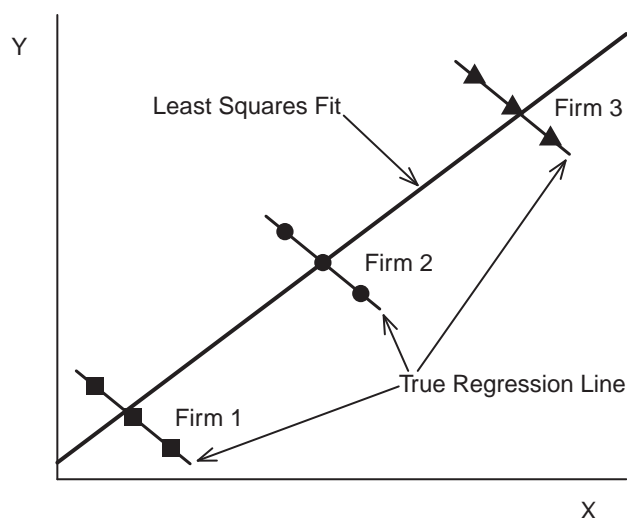


Figure 17.1: Scatter Plot and Pooled Regression Line

The presence of the unstructured individual effect u_i means that it is not possible to identify β under a simple projection assumption such as $\mathbb{E}[X_{it}\varepsilon_t] = 0$. It turns out that a sufficient condition for identification is the following.

Definition 17.1 The regressor X_{it} is **strictly exogenous** for the error ε_{it} if

$$\mathbb{E}[X_{is}\varepsilon_{it}] = 0 \quad (17.17)$$

for all $s = 1, \dots, T$.

Strict exogeneity is a strong projection condition, meaning that if X_{is} for any $s \neq t$ is added to (17.15) it will have a zero coefficient. Strict exogeneity is a projection analog of strict mean independence

$$\mathbb{E}[\varepsilon_{it} | \mathbf{X}_i] = 0. \quad (17.18)$$

(17.18) implies (17.17) but not conversely. While (17.17) is sufficient for identification and asymptotic theory we will also use the stronger condition (17.18) for finite sample analysis.

While (17.17) and (17.18) are strong assumptions they are much weaker than (17.2) or Assumption 17.1, which require that the individual effect u_i is also strictly mean independent. In contrast, (17.17) and (17.18) make no assumptions about u_i .

Strict exogeneity (17.17) is typically inappropriate in dynamic models. In Section 17.41 we discuss estimation under the weaker assumption of predetermined regressors.

17.8 Within Transformation

In the previous section we showed that if u_i and X_{it} are correlated then pooled and random-effects estimators will be biased and inconsistent. If we leave the relationship between u_i and X_{it} fully unstructured then the only way to consistently estimate the coefficient β is by an estimator which is invariant to u_i . This can be achieved by transformations which eliminate u_i .

One such transformation is the **within transformation**. In this section we describe this transformation in detail.

Define the mean of a variable for a given individual as

$$\bar{Y}_i = \frac{1}{T_i} \sum_{t \in S_i} Y_{it}.$$

We call this the **individual-specific mean** since it is the mean of a given individual. Contrarywise, some authors call this the **time-average** or **time-mean** since it is the average over the time periods.

Subtracting the individual-specific mean from the variable we obtain the deviations

$$\dot{Y}_{it} = Y_{it} - \bar{Y}_i.$$

This is known as the **within transformation**. We also refer to \dot{Y}_{it} as the **demeaned values** or **deviations from individual means**. Some authors refer to \dot{Y}_{it} as **deviations from time means**. What is important is that the demeaning has occurred at the individual level.

Some algebra may also be useful. We can write the individual-specific mean as $\bar{Y}_i = (\mathbf{1}'_i \mathbf{1}_i)^{-1} \mathbf{1}'_i \mathbf{Y}_i$. Stacking the observations for individual i we can write the within transformation using the notation

$$\begin{aligned} \dot{\mathbf{Y}}_i &= \mathbf{Y}_i - \mathbf{1}_i \bar{Y}_i \\ &= \mathbf{Y}_i - \mathbf{1}_i (\mathbf{1}'_i \mathbf{1}_i)^{-1} \mathbf{1}'_i \mathbf{Y}_i \\ &= \mathbf{M}_i \mathbf{Y}_i \end{aligned}$$

where $\mathbf{M}_i = \mathbf{I}_i - \mathbf{1}_i (\mathbf{1}_i' \mathbf{1}_i)^{-1} \mathbf{1}_i'$ is the individual-specific demeaning operator. Notice that \mathbf{M}_i is an idempotent matrix.

Similarly for the regressors we define the individual-specific means and demeaned values:

$$\begin{aligned}\bar{X}_i &= \frac{1}{T_i} \sum_{t \in S_i} X_{it} \\ \dot{X}_{it} &= X_{it} - \bar{X}_i \\ \dot{\mathbf{X}}_i &= \mathbf{M}_i \mathbf{X}_i.\end{aligned}$$

We illustrate demeaning in Table 17.1. In the fourth and seventh columns we display the firm-specific means \bar{I}_i and \bar{Q}_i and in the fifth and eighth columns the demeaned values \dot{I}_{it} and \dot{Q}_{it} .

We can also define the full-sample within operator. Define $\mathbf{D} = \text{diag}\{\mathbf{1}_{T_1}, \dots, \mathbf{1}_{T_N}\}$ and $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$. Note that $\mathbf{M}_D = \text{diag}\{\mathbf{M}_1, \dots, \mathbf{M}_N\}$. Thus

$$\mathbf{M}_D \mathbf{Y} = \dot{\mathbf{Y}} = \begin{pmatrix} \dot{Y}_1 \\ \vdots \\ \dot{Y}_N \end{pmatrix}, \quad \mathbf{M}_D \mathbf{X} = \dot{\mathbf{X}} = \begin{pmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \end{pmatrix}. \quad (17.19)$$

Now apply these operations to equation (17.15). Taking individual-specific averages we obtain

$$\bar{Y}_i = \bar{X}_i' \beta + u_i + \bar{\varepsilon}_i \quad (17.20)$$

where $\bar{\varepsilon}_i = \frac{1}{T_i} \sum_{t \in S_i} \varepsilon_{it}$. Subtracting from (17.15) we obtain

$$\dot{Y}_{it} = \dot{X}_{it}' \beta + \dot{\varepsilon}_{it} \quad (17.21)$$

where $\dot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$. The individual effect u_i has been eliminated!

We can alternatively write this in vector notation. Applying the demeaning operator \mathbf{M}_i to (17.16) we obtain

$$\dot{\mathbf{Y}}_i = \dot{\mathbf{X}}_i \beta + \dot{\boldsymbol{\varepsilon}}_i. \quad (17.22)$$

The individual-effect u_i is eliminated because $\mathbf{M}_i \mathbf{1}_i = 0$. Equation (17.22) is a vector version of (17.21).

The equation (17.21) is a linear equation in the transformed (demeaned) variables. As desired the individual effect u_i has been eliminated. Consequently estimators constructed from (17.21) (or equivalently (17.22)) will be invariant to the values of u_i . This means that the endogeneity bias described in the previous section will be eliminated.

Another consequence, however, is that all time-invariant regressors are also eliminated. That is, if the original model (17.15) had included any regressors $X_{it} = X_i$ which are constant over time for each individual then for these regressors the demeaned values are identically 0. What this means is that if equation (17.21) is used to estimate β it will be impossible to estimate (or identify) a coefficient on any regressor which is time invariant. This is not a consequence of the estimation method but rather a consequence of the model assumptions. In other words, if the individual effect u_i has no known structure then it is impossible to disentangle the effect of any time-invariant regressor X_i . The two have observationally equivalent effects and cannot be separately identified.

The within transformation can greatly reduce the variance of the regressors. This can be seen in Table 17.1 where you can see that the variation between the elements of the transformed variables \dot{I}_{it} and \dot{Q}_{it} is less than that of the untransformed variables, as much of the variation is captured by the firm-specific means.

It is not typically needed to directly program the within transformation, but if it is desired the following Stata commands easily do so.

Stata Commands for Within Transformation

```
*   x is the original variable
*   id is the group identifier
*   xdot is the within-transformed variable
egen xmean = mean(x), by(id)
gen xdot = x - xmean
```

17.9 Fixed Effects Estimator

Consider least squares applied to the demeaned equation (17.21) or equivalently (17.22). This is

$$\begin{aligned}\hat{\beta}_{fe} &= \left(\sum_{i=1}^N \sum_{t \in S_i} \dot{X}_{it} \dot{X}_{it}' \right)^{-1} \left(\sum_{i=1}^N \sum_{t \in S_i} \dot{X}_{it} \dot{Y}_{it} \right) \\ &= \left(\sum_{i=1}^N \dot{X}_i' \dot{X}_i \right)^{-1} \left(\sum_{i=1}^N \dot{X}_i' \dot{Y}_i \right) \\ &= \left(\sum_{i=1}^N X_i' M_i X_i \right)^{-1} \left(\sum_{i=1}^N X_i' M_i Y_i \right).\end{aligned}$$

This is known as the **fixed-effects** or **within** estimator of β . It is called the fixed-effects estimator because it is appropriate for the fixed effects model (17.15). It is called the within estimator because it is based on the variation of the data within each individual.

The above definition implicitly assumes that the matrix $\sum_{i=1}^N \dot{X}_i' \dot{X}_i$ is full rank. This requires that all components of X_{it} have time variation for at least some individuals in the sample.

The fixed effects residuals are

$$\begin{aligned}\hat{\varepsilon}_{it} &= \dot{Y}_{it} - \dot{X}_{it}' \hat{\beta}_{fe} \\ \hat{\varepsilon}_i &= \dot{Y}_i - \dot{X}_i' \hat{\beta}_{fe}.\end{aligned}\tag{17.23}$$

Let us describe some of the statistical properties of the estimator under strict mean independence (17.18). By linearity and the fact $M_i \mathbf{1}_i = 0$, we can write

$$\hat{\beta}_{fe} - \beta = \left(\sum_{i=1}^N X_i' M_i X_i \right)^{-1} \left(\sum_{i=1}^N X_i' M_i \varepsilon_i \right).$$

Then (17.18) implies

$$\mathbb{E} [\hat{\beta}_{fe} - \beta \mid \mathbf{X}] = \left(\sum_{i=1}^N X_i' M_i X_i \right)^{-1} \left(\sum_{i=1}^N X_i' M_i \mathbb{E} [\varepsilon_i \mid X_i] \right) = 0.$$

Thus $\hat{\beta}_{fe}$ is unbiased for β under (17.18).

Let $\Sigma_i = \mathbb{E} [\varepsilon_i \varepsilon_i' \mid X_i]$ denote the $T_i \times T_i$ conditional covariance matrix of the idiosyncratic errors. The variance of $\hat{\beta}_{fe}$ is

$$V_{fe} = \text{var} [\hat{\beta}_{fe} \mid \mathbf{X}] = \left(\sum_{i=1}^N \dot{X}_i' \dot{X}_i \right)^{-1} \left(\sum_{i=1}^N \dot{X}_i' \Sigma_i \dot{X}_i \right) \left(\sum_{i=1}^N \dot{X}_i' \dot{X}_i \right)^{-1}.\tag{17.24}$$

This expression simplifies when the idiosyncratic errors are homoskedastic and serially uncorrelated:

$$\mathbb{E}[\varepsilon_{it}^2 | \mathbf{X}_i] = \sigma_\varepsilon^2 \quad (17.25)$$

$$\mathbb{E}[\varepsilon_{ij}\varepsilon_{it} | \mathbf{X}_i] = 0 \quad (17.26)$$

for all $j \neq t$. In this case, $\Sigma_i = \mathbf{I}_i \sigma_\varepsilon^2$ and (17.24) simplifies to

$$\mathbf{V}_{\text{fe}}^0 = \sigma_\varepsilon^2 \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \right)^{-1}. \quad (17.27)$$

It is instructive to compare the variances of the fixed-effects and pooled estimators under (17.25)-(17.26) and the assumption that there is no individual-specific effect $u_i = 0$. In this case we see that

$$\mathbf{V}_{\text{fe}}^0 = \sigma_\varepsilon^2 \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \right)^{-1} \geq \sigma_\varepsilon^2 \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} = \mathbf{V}_{\text{pool}}. \quad (17.28)$$

The inequality holds since the demeaned variables $\dot{\mathbf{X}}_i$ have reduced variation relative to the original observations \mathbf{X}_i . (See Exercise 17.28.) This shows the cost of using fixed effects relative to pooled estimation. The estimation variance increases due to reduced variation in the regressors. This reduction in efficiency is a necessary by-product of the robustness of the estimator to the individual effects u_i .

17.10 Differenced Estimator

The within transformation is not the only transformation which eliminates the individual-specific effect. Another important transformation which does the same is first-differencing.

The **first-differencing** transformation is $\Delta Y_{it} = Y_{it} - Y_{it-1}$. This can be applied to all but the first observation (which is essentially lost). At the level of the individual this can be written as $\Delta \mathbf{Y}_i = \mathbf{D}_i \mathbf{Y}_i$ where \mathbf{D}_i is the $(T_i - 1) \times T_i$ matrix differencing operator

$$\mathbf{D}_i = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Applying the transformation Δ to (17.15) or (17.16) we obtain $\Delta Y_{it} = \Delta X_{it}' \beta + \Delta \varepsilon_{it}$ or

$$\Delta \mathbf{Y}_i = \Delta \mathbf{X}_i \beta + \Delta \boldsymbol{\varepsilon}_i. \quad (17.29)$$

We can see that the individual effect u_i has been eliminated.

Least squares applied to the differenced equation (17.29) is

$$\begin{aligned} \hat{\beta}_\Delta &= \left(\sum_{i=1}^N \sum_{t \geq 2} \Delta X_{it} \Delta X_{it}' \right)^{-1} \left(\sum_{i=1}^N \sum_{t \geq 2} \Delta X_{it} \Delta Y_{it} \right) \\ &= \left(\sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{Y}_i \right) \\ &= \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{D}_i' \mathbf{D}_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{D}_i' \mathbf{D}_i \mathbf{Y}_i \right). \end{aligned} \quad (17.30)$$

(17.30) is called the **differenced estimator**. For $T = 2$, $\hat{\beta}_\Delta = \hat{\beta}_{fe}$ equals the fixed effects estimator. See Exercise 17.6. They differ, however, for $T > 2$.

When the errors ε_{it} are serially uncorrelated and homoskedastic then the error $\Delta \varepsilon_i = \mathbf{D}_i \varepsilon_i$ in (17.29) has covariance matrix $\mathbf{H}\sigma_\varepsilon^2$ where

$$\mathbf{H} = \mathbf{D}_i \mathbf{D}_i' = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}. \quad (17.31)$$

We can reduce estimation variance by using GLS. When the errors ε_{it} are i.i.d. (serially uncorrelated and homoskedastic), this is

$$\begin{aligned} \tilde{\beta}_\Delta &= \left(\sum_{i=1}^N \Delta \mathbf{X}_i' \mathbf{H}^{-1} \Delta \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \Delta \mathbf{X}_i' \mathbf{H}^{-1} \Delta \mathbf{Y}_i \right) \\ &= \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{D}_i' (\mathbf{D}_i \mathbf{D}_i')^{-1} \mathbf{D}_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{D}_i' (\mathbf{D}_i \mathbf{D}_i')^{-1} \mathbf{D}_i \mathbf{Y}_i \right) \\ &= \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_i \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_i \mathbf{Y}_i \right) \end{aligned}$$

where $\mathbf{M}_i = \mathbf{D}_i' (\mathbf{D}_i \mathbf{D}_i')^{-1} \mathbf{D}_i$. Recall, the matrix \mathbf{D}_i is $(T_i - 1) \times T_i$ with rank $T_i - 1$ and is orthogonal to the vector of ones $\mathbf{1}_i$. This means \mathbf{M}_i projects orthogonally to $\mathbf{1}_i$ and thus equals the within transformation matrix. Hence $\tilde{\beta}_\Delta = \hat{\beta}_{fe}$, the fixed effects estimator!

What we have shown is that under i.i.d. errors, GLS applied to the first-differenced equation precisely equals the fixed effects estimator. Since the Gauss-Markov theorem shows that GLS has lower variance than least squares, this means that the fixed effects estimator is more efficient than first differencing under the assumption that ε_{it} is i.i.d.

This argument extends to any other transformation which eliminates the fixed effect. GLS applied after such a transformation is equal to the fixed effects estimator and is more efficient than least squares applied after the same transformation under i.i.d. errors. This shows that the fixed effects estimator is Gauss-Markov efficient in the class of estimators which eliminate the fixed effect, under these assumptions.

17.11 Dummy Variables Regression

An alternative way to estimate the fixed effects model is by least squares of Y_{it} on X_{it} and a full set of dummy variables, one for each individual in the sample. It turns out that this is algebraically equivalent to the within estimator.

To see this start with the error-component model without a regressor:

$$Y_{it} = u_i + \varepsilon_{it}. \quad (17.32)$$

Consider least squares estimation of the vector of fixed effects $u = (u_1, \dots, u_N)'$. Since each fixed effect u_i is an individual-specific mean and the least squares estimate of the intercept is the sample mean it follows that the least squares estimate of u_i is $\hat{u}_i = \bar{Y}_i$. The least squares residual is then $\hat{\varepsilon}_{it} = Y_{it} - \bar{Y}_i = \dot{Y}_{it}$, the within transformation.

If you would prefer an algebraic argument, let d_i be a vector of N dummy variables where the i^{th} element indicates the i^{th} individual. Thus the i^{th} element of d_i is 1 and the remaining elements are zero. Notice that $u_i = d_i' u$ and (17.32) equals $Y_{it} = d_i' u + \varepsilon_{it}$. This is a regression with the regressors d_i and coefficients u . We can also write this in vector notation at the level of the individual as $Y_i = \mathbf{1}_i d_i' u + \varepsilon_i$ or using full matrix notation as $Y = D u + \varepsilon$ where $D = \text{diag}\{\mathbf{1}_{T_1}, \dots, \mathbf{1}_{T_N}\}$.

The least squares estimate of u is

$$\begin{aligned}\hat{u} &= (D' D)^{-1} (D' Y) \\ &= \text{diag}(\mathbf{1}_i' \mathbf{1}_i)^{-1} \{\mathbf{1}_i' Y_i\}_{i=1, \dots, n} \\ &= \left\{ (\mathbf{1}_i' \mathbf{1}_i)^{-1} \mathbf{1}_i' Y_i \right\}_{i=1, \dots, n} \\ &= \{\bar{Y}_i\}_{i=1, \dots, n}.\end{aligned}$$

The least squares residuals are

$$\hat{\varepsilon} = (I_n - D(D' D)^{-1} D') Y = \dot{Y}$$

as shown in (17.19). Thus the least squares residuals from the simple error-component model are the within transformed variables.

Now consider the error-component model with regressors, which can be written as

$$Y_{it} = X_{it}' \beta + d_i' u + \varepsilon_{it} \quad (17.33)$$

since $u_i = d_i' u$ as discussed above. In matrix notation

$$Y = X \beta + D u + \varepsilon. \quad (17.34)$$

We consider estimation of (β, u) by least squares and write the estimates as $Y = X \hat{\beta} + D \hat{u} + \hat{\varepsilon}$. We call this the **dummy variable estimator** of the fixed effects model.

By the Frisch-Waugh-Lovell Theorem (Theorem 3.5) the dummy variable estimator $\hat{\beta}$ and residuals $\hat{\varepsilon}$ may be obtained by the least squares regression of the residuals from the regression of Y on D on the residuals from the regression of X on D . We learned above that the residuals from the regression on D are the within transformations. Thus the dummy variable estimator $\hat{\beta}$ and residuals $\hat{\varepsilon}$ may be obtained from least squares regression of the within transformed \dot{Y} on the within transformed \dot{X} . This is exactly the fixed effects estimator $\hat{\beta}_{fe}$. Thus the dummy variable and fixed effects estimators of β are identical.

This is sufficiently important that we state this result as a theorem.

Theorem 17.1 The fixed effects estimator of β algebraically equals the dummy variable estimator of β . The two estimators have the same residuals.

This may be the most important practical application of the Frisch-Waugh-Lovell Theorem. It shows that we can estimate the coefficients either by applying the within transformation or by inclusion of dummy variables (one for each individual in the sample). This is important because in some cases one approach is more convenient than the other and it is important to know that the two methods are algebraically equivalent.

When N is large it is advisable to use the within transformation rather than the dummy variable approach. This is because the latter requires considerably more computer memory. To see this consider

the matrix \mathbf{D} in (17.34) in the balanced case. It has TN^2 elements which must be created and stored in memory. When N is large this can be excessive. For example, if $T = 10$ and $N = 10,000$, the matrix \mathbf{D} has one billion elements! Whether or not a package can technically handle a matrix of this dimension depends on several particulars (system RAM, operating system, package version), but even if it can execute the calculation the computation time is slow. Hence for fixed effects estimation with large N it is recommended to use the within transformation rather than dummy variable regression.

The dummy variable formulation may add insight about how the fixed effects estimator achieves invariance to the fixed effects. Given the regression equation (17.34) we can write the least squares estimator of β using the residual regression formula:

$$\begin{aligned}\hat{\beta}_{fe} &= (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_D\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_D(\mathbf{X}\beta + \mathbf{D}u + \epsilon)) \\ &= \beta + (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_D\epsilon)\end{aligned}\quad (17.35)$$

since $\mathbf{M}_D\mathbf{D} = 0$. The expression (17.35) is free of the vector u and thus $\hat{\beta}_{fe}$ is invariant to u . This is another demonstration that the fixed effects estimator is invariant to the actual values of the fixed effects, and thus its statistical properties do not rely on assumptions about u_i .

17.12 Fixed Effects Covariance Matrix Estimation

First consider estimation of the classical covariance matrix \mathbf{V}_{fe}^0 as defined in (17.27). This is

$$\hat{\mathbf{V}}_{fe}^0 = \hat{\sigma}_\epsilon^2 (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \quad (17.36)$$

with

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n - N - k} \sum_{i=1}^n \sum_{t \in S_i} \hat{\epsilon}_{it}^2 = \frac{1}{n - N - k} \sum_{i=1}^n \hat{\epsilon}_i' \hat{\epsilon}_i. \quad (17.37)$$

The $N + k$ degree of freedom adjustment is motivated by the dummy variable representation. You can verify that $\hat{\sigma}_\epsilon^2$ is unbiased for σ_ϵ^2 under assumptions (17.18), (17.25) and (17.26). See Exercise 17.8.

Notice that the assumptions (17.18), (17.25), and (17.26) are identical to (17.5)-(17.7) of Assumption 17.1. The assumptions (17.8)-(17.10) are not needed. Thus the fixed effect model weakens the random effects model by eliminating the assumptions on u_i but retaining those on ϵ_{it} .

The classical covariance matrix estimator (17.36) for the fixed effects estimator is valid when the errors ϵ_{it} are homoskedastic and serially uncorrelated but is invalid otherwise. A covariance matrix estimator which allows ϵ_{it} to be heteroskedastic and serially correlated across t is the cluster-robust covariance matrix estimator, clustered by individual

$$\hat{\mathbf{V}}_{fe}^{\text{cluster}} = (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \hat{\epsilon}_i \hat{\epsilon}_i' \dot{\mathbf{X}}_i \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \quad (17.38)$$

where $\hat{\epsilon}_i$ as the fixed effects residuals as defined in (17.23). (17.38) was first proposed by Arellano (1987). As in (4.55) $\hat{\mathbf{V}}_{fe}^{\text{cluster}}$ can be multiplied by a degree-of-freedom adjustment. The adjustment recommended by the theory of C. Hansen (2007) is

$$\hat{\mathbf{V}}_{fe}^{\text{cluster}} = \left(\frac{N}{N-1} \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \hat{\epsilon}_i \hat{\epsilon}_i' \dot{\mathbf{X}}_i \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \quad (17.39)$$

and that corresponding to (4.55) is

$$\widehat{\mathbf{V}}_{\text{fe}}^{\text{cluster}} = \left(\frac{n-1}{n-N-k} \right) \left(\frac{N}{N-1} \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{X}}'_i \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}'_i \dot{\mathbf{X}}_i \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1}. \quad (17.40)$$

These estimators are convenient because they are simple to apply and allow for unbalanced panels.

In typical micropanel applications N is very large and k is modest. Thus the adjustment in (17.39) is minor while that in (17.40) is approximately $\bar{T}/(\bar{T}-1)$ where $\bar{T} = n/N$ is the average number of time periods per individual. When \bar{T} is small this can be a very large adjustment. Hence the choice between (17.38), (17.39), and (17.40) can be substantial.

To understand if the degree of freedom adjustment in (17.40) is appropriate, consider the simplified setting where the residuals are constructed with the true β but estimated fixed effects u_i . This is a useful approximation since the number of estimated slope coefficients β is small relative to the sample size n . Then $\widehat{\boldsymbol{\varepsilon}}_i = \boldsymbol{\varepsilon}_i = \mathbf{M}_i \boldsymbol{\varepsilon}_i$ so $\dot{\mathbf{X}}'_i \widehat{\boldsymbol{\varepsilon}}_i = \dot{\mathbf{X}}'_i \boldsymbol{\varepsilon}_i$ and (17.38) equals

$$\widehat{\mathbf{V}}_{\text{fe}}^{\text{cluster}} = (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{X}}'_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i \dot{\mathbf{X}}_i \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1}$$

which is the idealized estimator with the true errors rather than the residuals. Since $\mathbb{E}[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i | \mathbf{X}_i] = \Sigma_i$ it follows that $\mathbb{E}[\widehat{\mathbf{V}}_{\text{fe}}^{\text{cluster}} | \mathbf{X}] = \mathbf{V}_{\text{fe}}$ and $\widehat{\mathbf{V}}_{\text{fe}}^{\text{cluster}}$ is unbiased for \mathbf{V}_{fe} ! Thus no degree of freedom adjustment is required. This is despite the fact that N fixed effects have been estimated. While this analysis concerns the idealized case where the residuals have been constructed with the true coefficients β so does not translate into a direct recommendation for the feasible estimator, it still suggests that the strong *ad hoc* adjustment in (17.40) is unwarranted.

This (crude) analysis suggests that for the cluster robust covariance estimator for fixed effects regression the adjustment recommended by C. Hansen (17.39) is the most appropriate. It is typically well approximated by the unadjusted estimator (17.38). Based on current theory there is no justification for the *ad hoc* adjustment (17.40). The main argument for the latter is that it produces the largest standard errors and is thus the most conservative choice.

In current practice the estimators (17.38) and (17.40) are the most commonly used covariance matrix estimators for fixed effects estimation.

In Sections 17.22 and 17.23 we discuss covariance matrix estimation under heteroskedasticity but no serial correlation.

To illustrate, in Table 17.2 we present the fixed effect regression estimates of the investment model (17.3) in the third column with cluster-robust standard errors. The trading indicator T_i and the industry dummies cannot be included as they are time-invariant. The point estimates are similar to the random effects estimates, though the coefficients on debt and cash flow increase in magnitude.

17.13 Fixed Effects Estimation in Stata

There are several methods to obtain the fixed effects estimator $\widehat{\beta}_{\text{fe}}$ in Stata.

The first method is dummy variable regression. This can be obtained by the Stata `regress` command, for example `reg y x i.id, cluster(id)` where `id` is the group (individual) identifier. In most cases, as discussed in Section 17.11, this is not recommended due to the excessive computer memory requirements and slow computation. If this command is done it may be useful to suppress display of the full list of coefficient estimates. To do so, type quietly `reg y x i.id, cluster(id)` followed by `estimates table, keep(x _cons) be se`. The second command will report the coefficient(s) on `x` only, not those on the index variable `id`. (Other statistics can be reported as well.)

The second method is to manually create the within transformed variables as described in Section 17.8, and then use `regress`.

The third method is `xtreg fe` which is specifically written for panel data. This estimates the slope coefficients using the partialling-out approach. The default covariance matrix estimator is classical as defined in (17.36). The cluster-robust covariance matrix (17.38) can be obtained using the options `vce(robust)` or `r`.

The fourth method is `areg absorb(id)`. This command is an alternative implementation of partialling-out regression. The default covariance matrix estimator is the classical (17.36). The cluster-robust covariance matrix estimator (17.40) can be obtained using the `cluster(id)` option. The heteroskedasticity-robust covariance matrix is obtained when `r` or `vce(robust)` is specified but this is not recommended unless T_i is large as will be discussed in Section 17.22.

An important difference between the Stata `xtreg` and `areg` commands is that they implement different cluster-robust covariance matrix estimators: (17.38) in the case of `xtreg` and (17.40) in the case of `areg`. As discussed in the previous section the adjustment used by `areg` is *ad hoc* and not well-justified but produces the largest and hence most conservative standard errors.

Another difference between the commands is how they report the equation R^2 . This difference can be huge and stems from the fact that they are estimating distinct population counter-parts. Full dummy variable regression and the `areg` command calculate R^2 the same way: the squared correlation between Y_{it} and the fitted regression with all predictors including the individual dummy variables. The `xtreg fe` command reports three values for R^2 : within, between, and overall. The “within” R^2 is identical to what is obtained from a second stage regression using the within transformed variables. (The second method described above.) The “overall” R^2 is the squared correlation between Y_{it} and the fitted regression excluding the individual effects.

Which R^2 should be reported? The answer depends on the baseline model before regressors are added. If we view the baseline as an individual-specific mean, then the **within** calculation is appropriate. If the baseline is a single mean for all observations then the full regression (`areg`) calculation is appropriate. The latter (`areg`) calculation is typically much higher than the within calculation, as the fixed effects typically “explain” a large portion of the variance. In any event as there is not a single definition of R^2 it is important to be explicit about the method if it is reported.

In current econometric practice both `xtreg` and `areg` are used, though `areg` appears to be the more popular choice. Since the latter typically produces a much higher value of R^2 , reported R^2 values should be viewed skeptically unless their calculation method is documented by the author.

17.14 Between Estimator

The **between estimator** is calculated from the individual-mean equation (17.20)

$$\bar{Y}_i = \bar{X}_i' \beta + u_i + \bar{\varepsilon}_i. \quad (17.41)$$

Estimation can be done at the level of individuals or at the level of observations. Least squares applied to (17.41) at the level of the N individuals is

$$\hat{\beta}_{be} = \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i' \right)^{-1} \left(\sum_{i=1}^N \bar{X}_i \bar{Y}_i \right).$$

Least squares applied to (17.41) at the level of observations is

$$\tilde{\beta}_{be} = \left(\sum_{i=1}^N \sum_{t \in S_i} \bar{X}_i \bar{X}_i' \right)^{-1} \left(\sum_{i=1}^N \sum_{t \in S_i} \bar{X}_i \bar{Y}_i \right) = \left(\sum_{i=1}^N T_i \bar{X}_i \bar{X}_i' \right)^{-1} \left(\sum_{i=1}^N T_i \bar{X}_i \bar{Y}_i \right).$$

In balanced panels $\tilde{\beta}_{be} = \hat{\beta}_{be}$ but they differ on unbalanced panels. $\tilde{\beta}_{be}$ equals weighted least squares applied at the level of individuals with weight T_i .

Under the random effects assumptions (Assumption 17.1) $\hat{\beta}_{be}$ is unbiased for β and has variance

$$\mathbf{V}_{be} = \text{var}[\hat{\beta}_{be} | \mathbf{X}] = \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i' \right)^{-1} \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i' \sigma_i^2 \right) \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i' \right)^{-1}$$

where

$$\sigma_i^2 = \text{var}[u_i + \bar{\varepsilon}_i] = \sigma_u^2 + \frac{\sigma_\varepsilon^2}{T_i}$$

is the variance of the error in (17.41). When the panel is balanced the variance formula simplifies to

$$\mathbf{V}_{be} = \text{var}[\hat{\beta}_{be} | \mathbf{X}] = \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i' \right)^{-1} \left(\sigma_u^2 + \frac{\sigma_\varepsilon^2}{T} \right).$$

Under the random effects assumption the between estimator $\hat{\beta}_{be}$ is unbiased for β but is less efficient than the random effects estimator $\hat{\beta}_{gls}$. Consequently there seems little direct use for the between estimator in linear panel data applications.

Instead, its primary application is to construct an estimate of σ_u^2 . First, consider estimation of

$$\sigma_b^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2 = \sigma_u^2 + \frac{1}{N} \sum_{i=1}^N \frac{\sigma_\varepsilon^2}{T_i} = \sigma_u^2 + \frac{\sigma_\varepsilon^2}{\bar{T}}$$

where $\bar{T} = N / \sum_{i=1}^N T_i^{-1}$ is the harmonic mean of T_i . (In the case of a balanced panel $\bar{T} = T$.) A natural estimator of σ_b^2 is

$$\hat{\sigma}_b^2 = \frac{1}{N-k} \sum_{i=1}^N \hat{e}_{bi}^2. \quad (17.42)$$

where $\hat{e}_{bi} = \bar{Y}_i - \bar{X}_i' \hat{\beta}_{be}$ are the between residuals. (Either $\hat{\beta}_{be}$ or $\tilde{\beta}_{be}$ can be used.)

From the relation $\sigma_b^2 = \sigma_u^2 + \sigma_\varepsilon^2 / \bar{T}$ and (17.42) we can deduce an estimator for σ_u^2 . We have already described an estimator $\hat{\sigma}_\varepsilon^2$ for σ_ε^2 in (17.37) for the fixed effects model. Since the fixed effects model holds under weaker conditions than the random effects model, $\hat{\sigma}_\varepsilon^2$ is valid for the latter as well. This suggests the following estimator for σ_u^2

$$\hat{\sigma}_u^2 = \hat{\sigma}_b^2 - \frac{\hat{\sigma}_\varepsilon^2}{\bar{T}}. \quad (17.43)$$

To summarize, the fixed effect estimator is used for $\hat{\sigma}_\varepsilon^2$, the between estimator for $\hat{\sigma}_b^2$, and $\hat{\sigma}_u^2$ is constructed from the two.

It is possible for (17.43) to be negative. It is typical to use the constrained estimator

$$\hat{\sigma}_u^2 = \max \left[0, \hat{\sigma}_b^2 - \frac{\hat{\sigma}_\varepsilon^2}{\bar{T}} \right]. \quad (17.44)$$

(17.44) is the most common estimator for σ_u^2 in the random effects model.

The between estimator $\hat{\beta}_{be}$ can be obtained using the Stata command `xtreg be`. The estimator $\tilde{\beta}_{be}$ can be obtained by `xtreg be wls`.

17.15 Feasible GLS

The random effects estimator can be written as

$$\hat{\beta}_{\text{re}} = \left(\sum_{i=1}^N \mathbf{X}'_i \Omega_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}'_i \Omega_i^{-1} \mathbf{Y}_i \right) = \left(\sum_{i=1}^N \tilde{\mathbf{X}}'_i \tilde{\mathbf{X}}_i \right)^{-1} \left(\sum_{i=1}^N \tilde{\mathbf{X}}'_i \tilde{\mathbf{Y}}_i \right) \quad (17.45)$$

where $\tilde{\mathbf{X}}_i = \Omega_i^{-1/2} \mathbf{X}_i$ and $\tilde{\mathbf{Y}}_i = \Omega_i^{-1/2} \mathbf{Y}_i$. It is instructive to study these transformations.

Define $\mathbf{P}_i = \mathbf{1}_i (\mathbf{1}'_i \mathbf{1}_i)^{-1} \mathbf{1}'_i$ so that $\mathbf{M}_i = \mathbf{I}_i - \mathbf{P}_i$. Thus while \mathbf{M}_i is the within operator, \mathbf{P}_i can be called the individual-mean operator since $\mathbf{P}_i \mathbf{Y}_i = \mathbf{1}_i \bar{Y}_i$. We can write

$$\Omega_i = \mathbf{I}_i + \mathbf{1}_i \mathbf{1}'_i \sigma_u^2 / \sigma_\varepsilon^2 = \mathbf{I}_i + \frac{T_i \sigma_u^2}{\sigma_\varepsilon^2} \mathbf{P}_i = \mathbf{M}_i + \rho_i^{-2} \mathbf{P}_i$$

where

$$\rho_i = \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i \sigma_u^2}}. \quad (17.46)$$

Since the matrices \mathbf{M}_i and \mathbf{P}_i are idempotent and orthogonal we find that $\Omega_i^{-1} = \mathbf{M}_i + \rho_i^2 \mathbf{P}_i$ and

$$\Omega_i^{-1/2} = \mathbf{M}_i + \rho_i \mathbf{P}_i = \mathbf{I}_i - (1 - \rho_i) \mathbf{P}_i. \quad (17.47)$$

Therefore the transformation used by the GLS estimator is

$$\tilde{\mathbf{Y}}_i = (\mathbf{I}_i - (1 - \rho_i) \mathbf{P}_i) \mathbf{Y}_i = \mathbf{Y}_i - (1 - \rho_i) \mathbf{1}_i \bar{Y}_i$$

which is a partial within transformation.

The transformation as written depends on ρ_i which is unknown. It can be replaced by the estimator

$$\hat{\rho}_i = \frac{\hat{\sigma}_\varepsilon}{\sqrt{\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_u^2}} \quad (17.48)$$

where the estimators $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_u^2$ are given in (17.37) and (17.44). We obtain the feasible transformations

$$\tilde{\mathbf{Y}}_i = \mathbf{Y}_i - (1 - \hat{\rho}_i) \mathbf{1}_i \bar{Y}_i \quad (17.49)$$

and

$$\tilde{\mathbf{X}}_i = \mathbf{X}_i - (1 - \hat{\rho}_i) \mathbf{1}_i \bar{X}_i. \quad (17.50)$$

The feasible random effects estimator is (17.45) using (17.49) and (17.50).

In the previous section we noted that it is possible for $\hat{\sigma}_u^2 = 0$. In this case $\hat{\rho}_i = 1$ and $\hat{\beta}_{\text{re}} = \hat{\beta}_{\text{pool}}$.

What this shows is the following. The random effects estimator (17.45) is least squares applied to the transformed variables $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$ defined in (17.50) and (17.49). When $\hat{\rho}_i = 0$ these are the within transformations, so $\tilde{\mathbf{X}}_i = \hat{\mathbf{X}}_i$, $\tilde{\mathbf{Y}}_i = \hat{\mathbf{Y}}_i$, and $\hat{\beta}_{\text{re}} = \hat{\beta}_{\text{fe}}$ is the fixed effects estimator. When $\hat{\rho}_i = 1$ the data are untransformed $\tilde{\mathbf{X}}_i = \mathbf{X}_i$, $\tilde{\mathbf{Y}}_i = \mathbf{Y}_i$, and $\hat{\beta}_{\text{re}} = \hat{\beta}_{\text{pool}}$ is the pooled estimator. In general, $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$ can be viewed as partial within transformations.

Recalling the definition $\hat{\rho}_i = \hat{\sigma}_\varepsilon / \sqrt{\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_u^2}$ we see that when the idiosyncratic error variance $\hat{\sigma}_\varepsilon^2$ is large relative to $T_i \hat{\sigma}_u^2$ then $\hat{\rho}_i \approx 1$ and $\hat{\beta}_{\text{re}} \approx \hat{\beta}_{\text{pool}}$. Thus when the variance estimates suggest that the individual effect is relatively small the random effect estimator simplifies to the pooled estimator. On the other hand when the individual effect error variance $\hat{\sigma}_u^2$ is large relative to $\hat{\sigma}_\varepsilon^2$ then $\hat{\rho}_i \approx 0$ and $\hat{\beta}_{\text{re}} \approx \hat{\beta}_{\text{fe}}$. Thus when the variance estimates suggest that the individual effect is relatively large the random effect estimator is close to the fixed effects estimator.

17.16 Intercept in Fixed Effects Regression

The fixed effect estimator does not apply to any regressor which is time-invariant for all individuals. This includes an intercept. Yet some authors and packages (e.g. Amemiya (1971) and `xtreg` in Stata) report an intercept. To see how to construct an estimator of an intercept take the components regression equation adding an explicit intercept

$$Y_{it} = \alpha + X'_{it}\beta + u_i + \varepsilon_{it}.$$

We have already discussed estimation of β by $\hat{\beta}_{fe}$. Replacing β in this equation with $\hat{\beta}_{fe}$ and then estimating α by least squares, we obtain

$$\hat{\alpha}_{fe} = \bar{Y} - \bar{X}'\hat{\beta}_{fe}$$

where \bar{Y} and \bar{X} are averages from the full sample. This is the estimator reported by `xtreg`.

17.17 Estimation of Fixed Effects

For most applications researchers are interested in the coefficients β not the fixed effects u_i . But in some cases the fixed effects themselves are interesting. This arises when we want to measure the distribution of u_i to understand its heterogeneity. It also arises in the context of prediction. As discussed in Section 17.11 the fixed effects estimate \hat{u} is obtained by least squares applied to the regression (17.33). To find their solution, replace β in (17.33) with the least squares minimizer $\hat{\beta}_{fe}$ and apply least squares. Since this is the individual-specific intercept the solution is

$$\hat{u}_i = \frac{1}{T_i} \sum_{t \in S_i} (Y_{it} - X'_{it}\hat{\beta}_{fe}) = \bar{Y}_i - \bar{X}'_i\hat{\beta}_{fe}. \quad (17.51)$$

Alternatively, using (17.34) this is

$$\begin{aligned} \hat{u} &= (D'D)^{-1} D' (Y - X\hat{\beta}_{fe}) \\ &= \text{diag}\{T_i^{-1}\} \sum_{i=1}^N d_i \mathbf{1}'_i (Y_i - X_i\hat{\beta}_{fe}) \\ &= \sum_{i=1}^N d_i (\bar{Y}_i - \bar{X}'_i\hat{\beta}_{fe}) \\ &= (\hat{u}_1, \dots, \hat{u}_N)'. \end{aligned}$$

Thus the least squares estimates of the fixed effects can be obtained from the individual-specific means and does not require a regression with $N + k$ regressors.

If an intercept has been estimated (as discussed in the previous section) it should be subtracted from (17.51). In this case the estimated fixed effects are

$$\hat{u}_i = \bar{Y}_i - \bar{X}'_i\hat{\beta}_{fe} - \hat{\alpha}_{fe}. \quad (17.52)$$

With either estimator when the number of time series observations T_i is small \hat{u}_i will be an imprecise estimator of u_i . Thus calculations based on \hat{u}_i should be interpreted cautiously.

The fixed effects (17.52) may be obtained in Stata after `ivreg`, `fe` using the `predict u` command or after `areg` using the `predict d` command.

17.18 GMM Interpretation of Fixed Effects

We can also interpret the fixed effects estimator through the generalized method of moments.

Take the fixed effects model after applying the within transformation (17.21). We can view this as a system of T equations, one for each time period t . This is a multivariate regression model. Using the notation of Chapter 11 define the $T \times kT$ regressor matrix

$$\bar{\mathbf{X}}_i = \begin{pmatrix} \dot{X}'_{i1} & 0 & \cdots & 0 \\ \vdots & \dot{X}'_{i2} & & \vdots \\ 0 & 0 & \cdots & \dot{X}'_{iT} \end{pmatrix}. \quad (17.53)$$

If we treat each time period as a separate equation we have the kT moment conditions

$$\mathbb{E} \left[\bar{\mathbf{X}}'_i (\dot{\mathbf{Y}}_i - \dot{\mathbf{X}}_i \beta) \right] = 0.$$

This is an overidentified system of equations when $T \geq 3$ as there are k coefficients and kT moments. (However, the moments are collinear due to the within transformation. There are $k(T-1)$ effective moments.) Interpreting this model in the context of multivariate regression, overidentification is achieved by the restriction that the coefficient vector β is constant across time periods.

This model can be interpreted as a regression of $\dot{\mathbf{Y}}_i$ on $\dot{\mathbf{X}}_i$ using the instruments $\bar{\mathbf{X}}_i$. The 2SLS estimator using matrix notation is

$$\hat{\beta} = \left((\dot{\mathbf{X}}' \bar{\mathbf{X}}) (\bar{\mathbf{X}}' \bar{\mathbf{X}})^{-1} (\bar{\mathbf{X}}' \dot{\mathbf{X}}) \right)^{-1} \left((\dot{\mathbf{X}}' \bar{\mathbf{X}}) (\bar{\mathbf{X}}' \bar{\mathbf{X}})^{-1} (\bar{\mathbf{X}}' \dot{\mathbf{Y}}) \right).$$

Notice that

$$\begin{aligned} \bar{\mathbf{X}}' \bar{\mathbf{X}} &= \sum_{i=1}^n \begin{pmatrix} \dot{X}_{i1} & 0 & \cdots & 0 \\ \vdots & \dot{X}_{i2} & & \vdots \\ 0 & 0 & \cdots & \dot{X}_{iT} \end{pmatrix} \begin{pmatrix} \dot{X}'_{i1} & 0 & \cdots & 0 \\ \vdots & \dot{X}'_{i2} & & \vdots \\ 0 & 0 & \cdots & \dot{X}'_{iT} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \dot{X}_{i1} \dot{X}'_{i1} & 0 & \cdots & 0 \\ \vdots & \sum_{i=1}^n \dot{X}_{i2} \dot{X}'_{i2} & & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^n \dot{X}_{iT} \dot{X}'_{iT} \end{pmatrix}, \\ \bar{\mathbf{X}}' \dot{\mathbf{X}} &= \begin{pmatrix} \sum_{i=1}^n \dot{X}_{i1} \dot{X}'_{i1} \\ \vdots \\ \sum_{i=1}^n \dot{X}_{iT} \dot{X}'_{iT} \end{pmatrix}, \end{aligned}$$

and

$$\bar{\mathbf{X}}' \dot{\mathbf{Y}} = \begin{pmatrix} \sum_{i=1}^n \dot{X}_{i1} \dot{Y}_{i1} \\ \vdots \\ \sum_{i=1}^n \dot{X}_{iT} \dot{Y}_{iT} \end{pmatrix}.$$

Thus the 2SLS estimator simplifies to

$$\begin{aligned}\hat{\beta}_{2sls} &= \left(\sum_{t=1}^T \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right) \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right)^{-1} \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right) \right)^{-1} \\ &\quad \times \left(\sum_{t=1}^T \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right) \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right)^{-1} \left(\sum_{i=1}^n \dot{X}_{it} \dot{Y}_{it} \right) \right) \\ &= \left(\sum_{t=1}^T \sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right)^{-1} \left(\sum_{t=1}^T \sum_{i=1}^n \dot{X}_{it} \dot{Y}_{it} \right) \\ &= \hat{\beta}_{fe}\end{aligned}$$

the fixed effects estimator!

This shows that if we treat each time period as a separate equation with its separate moment equation so that the system is over-identified, and then estimate by GMM using the 2SLS weight matrix, the resulting GMM estimator equals the simple fixed effects estimator. There is no change by adding the additional moment conditions.

The 2SLS estimator is the appropriate GMM estimator when the equation error is serially uncorrelated and homoskedastic. If we use a two-step efficient weight matrix which allows for heteroskedasticity and serial correlation the GMM estimator is

$$\begin{aligned}\hat{\beta}_{gmm} &= \left(\sum_{t=1}^T \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right) \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \hat{e}_{it}^2 \right)^{-1} \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right) \right)^{-1} \\ &\quad \times \left(\sum_{t=1}^T \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \right) \left(\sum_{i=1}^n \dot{X}_{it} \dot{X}'_{it} \hat{e}_{it}^2 \right)^{-1} \left(\sum_{i=1}^n \dot{X}_{it} \dot{Y}_{it} \right) \right)\end{aligned}$$

where \hat{e}_{it} are the fixed effects residuals.

Notationally, this GMM estimator has been written for a balanced panel. For an unbalanced panel the sums over i need to be replaced by sums over individuals observed during time period t . Otherwise no changes need to be made.

17.19 Identification in the Fixed Effects Model

The identification of the slope coefficient β in fixed effects regression is similar to that in conventional regression but somewhat more nuanced.

It is most useful to consider the within-transformed equation, which can be written as $\dot{Y}_{it} = \dot{X}'_{it}\beta + \dot{\epsilon}_{it}$ or $\dot{Y}_i = \dot{X}_i\beta + \dot{\epsilon}_i$.

From regression theory we know that the coefficient β is the linear effect of \dot{X}_{it} on \dot{Y}_{it} . The variable \dot{X}_{it} is the deviation of the regressor from its individual-specific mean and similarly for \dot{Y}_{it} . Thus the fixed effects model does not identify the effect of the average level of X_{it} on the average level of Y_{it} , but rather the effect of the deviations in X_{it} on Y_{it} .

In any given sample the fixed effects estimator is only defined if $\sum_{i=1}^N \dot{X}'_i \dot{X}_i$ is full rank. The population analog (when individuals are i.i.d.) is

$$\mathbb{E} \left[\dot{X}'_i \dot{X}_i \right] > 0. \quad (17.54)$$

Equation (17.54) is the identification condition for the fixed effects estimator. It requires that the regressor matrix is full-rank in expectation after application of the within transformation. The regressors cannot contain any variable which does not have time-variation at the individual level nor a set of regressors whose time-variation at the individual level is collinear.

17.20 Asymptotic Distribution of Fixed Effects Estimator

In this section we present an asymptotic distribution theory for the fixed effects estimator in balanced panels. Unbalanced panels are considered in the following section.

We use the following assumptions.

Assumption 17.2

1. $Y_{it} = X'_{it}\beta + u_i + \varepsilon_{it}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$ with $T \geq 2$.
2. The variables (ε_i, X_i) , $i = 1, \dots, N$, are independent and identically distributed.
3. $\mathbb{E}[X_{is}\varepsilon_{it}] = 0$ for all $s = 1, \dots, T$.
4. $Q_T = \mathbb{E}[\dot{X}'_i \dot{X}_i] > 0$.
5. $\mathbb{E}[\varepsilon_{it}^4] < \infty$.
6. $\mathbb{E}\|X_{it}\|^4 < \infty$.

Given Assumption 17.2 we can establish asymptotic normality for $\hat{\beta}_{fe}$.

Theorem 17.2 Under Assumption 17.2, as $N \rightarrow \infty$, $\sqrt{N}(\hat{\beta}_{fe} - \beta) \xrightarrow{d} N(0, V_\beta)$ where $V_\beta = Q_T^{-1}\Omega_T Q_T^{-1}$ and $\Omega_T = \mathbb{E}[\dot{X}'_i \varepsilon_i \varepsilon'_i \dot{X}_i]$.

This asymptotic distribution is derived as the number of individuals N diverges to infinity while the time number of time periods T is held fixed. Therefore the normalization is \sqrt{N} rather than \sqrt{n} (though either could be used since T is fixed). This approximation is appropriate for the context of a large number of individuals. We could alternatively derive an approximation for the case where both N and T diverge to infinity but this would not be a stronger result. One way of thinking about this is that Theorem 17.2 does not require T to be large.

Theorem 17.2 may appear standard given our arsenal of asymptotic theory but in a fundamental sense it is quite different from any other result we have introduced. Fixed effects regression is effectively estimating $N + k$ coefficients – the k slope coefficients β plus the N fixed effects u – and the theory specifies that $N \rightarrow \infty$. Thus the number of estimated parameters is diverging to infinity at the same rate as sample size yet the estimator obtains a conventional mean-zero sandwich-form asymptotic distribution. In this sense Theorem 17.2 is new and special.

We now discuss the assumptions.

Assumption 17.2.2 states that the observations are independent across individuals i . This is commonly used for panel data asymptotic theory. An important implied restriction is that it means that we exclude from the regressors any serially correlated aggregate time series variation.

Assumption 17.2.3 imposes that X_{it} is strictly exogenous for ε_{it} . This is stronger than simple projection but is weaker than strict mean independence (17.18). It does not impose any condition on the individual-specific effects u_i .

Assumption 17.2.4 is the identification condition discussed in the previous section.

Assumptions 17.2.5 and 17.2.6 are needed for the central limit theorem.

We now prove Theorem 17.2. The assumptions imply that the variables $(\dot{X}_i, \varepsilon_i)$ are i.i.d. across i and have finite fourth moments. Thus by the WLLN

$$\frac{1}{N} \sum_{i=1}^N \dot{X}_i' \dot{X}_i \xrightarrow{p} \mathbb{E} [\dot{X}_i' \dot{X}_i] = \mathbf{Q}_T.$$

Assumption 17.2.3 implies

$$\mathbb{E} [\dot{X}_i' \varepsilon_i] = \sum_{t=1}^T \mathbb{E} [\dot{X}_{it} \varepsilon_{it}] = \sum_{t=1}^T \mathbb{E} [X_{it} \varepsilon_{it}] - \sum_{t=1}^T \sum_{j=1}^T \mathbb{E} [X_{ij} \varepsilon_{it}] = 0$$

so they are mean zero. Assumptions 17.2.5 and 17.2.6 imply that $\dot{X}_i' \varepsilon_i$ has a finite covariance matrix Ω_T . The assumptions for the CLT (Theorem 6.3) hold, thus

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \dot{X}_i' \varepsilon_i \xrightarrow{d} N(0, \Omega_T).$$

Together we find

$$\sqrt{N}(\hat{\beta}_{fe} - \beta) = \left(\frac{1}{N} \sum_{i=1}^N \dot{X}_i' \dot{X}_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \dot{X}_i' \varepsilon_i \right) \xrightarrow{d} \mathbf{Q}_T^{-1} N(0, \Omega_T) = N(0, \mathbf{V}_\beta)$$

as stated.

17.21 Asymptotic Distribution for Unbalanced Panels

In this section we extend the theory of the previous section to cover unbalanced panels under random selection. Our presentation is built on Section 17.1 of Wooldridge (2010).

Think of an unbalanced panel as a shortened version of an idealized balanced panel where the shortening is due to “missing” observations due to random selection. Thus suppose that the underlying (potentially latent) variables are $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT})'$ and $\mathbf{X}_i = (X_{i1}, \dots, X_{iT})'$. Let $\mathbf{s}_i = (s_{i1}, \dots, s_{iT})'$ be a vector of selection indicators, meaning that $s_{it} = 1$ if the time period t is observed for individual i and $s_{it} = 0$ otherwise. Then we can describe the estimators algebraically as follows.

Let $\mathbf{S}_i = \text{diag}(\mathbf{s}_i)$ and $\mathbf{M}_i = \mathbf{S}_i - \mathbf{s}_i(\mathbf{s}_i' \mathbf{s}_i)^{-1} \mathbf{s}_i'$, which is idempotent. The within transformations can be written as $\dot{\mathbf{Y}}_i = \mathbf{M}_i \mathbf{Y}_i$ and $\dot{\mathbf{X}}_i = \mathbf{M}_i \mathbf{X}_i$. They have the property that if $s_{it} = 0$ (so that time period t is missing) then the t^{th} element of $\dot{\mathbf{Y}}_i$ and the t^{th} row of $\dot{\mathbf{X}}_i$ are all zeros. The missing observations have been replaced by zeros. Consequently, they do not appear in matrix products and sums.

The fixed effects estimator of β based on the observed sample is

$$\hat{\beta}_{fe} = \left(\sum_{i=1}^N \dot{X}_i' \dot{X}_i \right)^{-1} \left(\sum_{i=1}^N \dot{X}_i' \dot{Y}_i \right).$$

Centered and normalized,

$$\sqrt{N}(\hat{\beta}_{fe} - \beta) = \left(\frac{1}{N} \sum_{i=1}^N \dot{X}_i' \dot{X}_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \dot{X}_i' \varepsilon_i \right).$$

Notationally this appears to be identical to the case of a balanced panel but the difference is that the within operator \mathbf{M}_i incorporates the sample selection induced by the unbalanced panel structure.

To derive a distribution theory for $\hat{\beta}_{fe}$ we need to be explicit about the stochastic nature of \mathbf{s}_i . That is, why are some time periods observed and some not? We could take several approaches:

1. We could treat \mathbf{s}_i as fixed (non-random). This is the easiest approach but the most unsatisfactory.
2. We could treat \mathbf{s}_i as random but independent of $(\mathbf{Y}_i, \mathbf{X}_i)$. This is known as “missing at random” and is a common assumption used to justify methods with missing observations. It is justified when the reason why observations are not observed is independent of the observations. This is appropriate, for example, in panel data sets where individuals enter and exit in “waves”. The statistical treatment is not substantially different from the case of fixed \mathbf{s}_i .
3. We could treat $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{s}_i)$ as jointly random but impose a condition sufficient for consistent estimation of β . This is the approach we take below. The condition turns out to be a form of mean independence. The advantage of this approach is that it is less restrictive than full independence. The disadvantage is that we must use a conditional mean restriction rather than uncorrelatedness to identify the coefficients.

The specific assumptions we impose are as follows.

Assumption 17.3

1. $Y_{it} = \mathbf{X}'_{it}\beta + u_i + \varepsilon_{it}$ for $i = 1, \dots, N$ with $T_i \geq 2$.
2. The variables $(\varepsilon_i, \mathbf{X}_i, \mathbf{s}_i)$, $i = 1, \dots, N$, are independent and identically distributed.
3. $\mathbb{E}[\varepsilon_{it} | \mathbf{X}_i, \mathbf{s}_i] = 0$.
4. $\mathbf{Q}_T = \mathbb{E}[\dot{\mathbf{X}}'_i \dot{\mathbf{X}}_i] > 0$.
5. $\mathbb{E}[\varepsilon_{it}^4] < \infty$.
6. $\mathbb{E} \|\mathbf{X}_{it}\|^4 < \infty$.

The primary difference with Assumption 17.2 is that we have strengthened strict exogeneity to strict mean independence. This imposes that the regression model is properly specified and that selection does not affect the mean of ε_{it} . It is less restrictive than full independence since \mathbf{s}_i can affect other moments of ε_{it} and more importantly does not restrict the joint dependence between \mathbf{s}_i and \mathbf{X}_i .

Given the above development it is straightforward to establish asymptotic normality.

Theorem 17.3 Under Assumption 17.3, as $N \rightarrow \infty$, $\sqrt{N}(\hat{\beta}_{fe} - \beta) \xrightarrow{d} N(0, V_\beta)$ where $V_\beta = \mathbf{Q}_T^{-1} \Omega_T \mathbf{Q}_T^{-1}$ and $\Omega_T = \mathbb{E}[\dot{\mathbf{X}}'_i \varepsilon_i \varepsilon'_i \dot{\mathbf{X}}_i]$.

We now prove Theorem 17.3. The assumptions imply that the variables $(\dot{\mathbf{X}}_i, \boldsymbol{\varepsilon}_i)$ are i.i.d. across i and have finite fourth moments. By the WLLN

$$\frac{1}{N} \sum_{i=1}^N \dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \xrightarrow{p} \mathbb{E} [\dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i] = \mathbf{Q}_T.$$

The random vectors $\dot{\mathbf{X}}_i' \boldsymbol{\varepsilon}_i$ are i.i.d. The matrix $\dot{\mathbf{X}}_i$ is a function of $(\mathbf{X}_i, \mathbf{s}_i)$ only. Assumption 17.3.3 and the law of iterated expectations implies

$$\mathbb{E} [\dot{\mathbf{X}}_i' \boldsymbol{\varepsilon}_i] = \mathbb{E} [\dot{\mathbf{X}}_i' \mathbb{E} [\boldsymbol{\varepsilon}_i | \mathbf{X}_i, \mathbf{s}_i]] = \mathbf{0}.$$

so that $\dot{\mathbf{X}}_i' \boldsymbol{\varepsilon}_i$ is mean zero. Assumptions 17.3.5 and 17.3.6 and the fact that \mathbf{s}_i is bounded implies that $\dot{\mathbf{X}}_i' \boldsymbol{\varepsilon}_i$ has a finite covariance matrix, which is Ω_T . The assumptions for the CLT hold, thus

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \dot{\mathbf{X}}_i' \boldsymbol{\varepsilon}_i \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Omega_T).$$

Together we obtain the stated result.

17.22 Heteroskedasticity-Robust Covariance Matrix Estimation

We have introduced two covariance matrix estimators for the fixed effects estimator. The classical estimator (17.36) is appropriate for the case where the idiosyncratic errors ε_{it} are homoskedastic and serially uncorrelated. The cluster-robust estimator (17.38) allows for heteroskedasticity and arbitrary serial correlation. In this and the following section we consider the intermediate case where ε_{it} is heteroskedastic but serially uncorrelated.

Assume that (17.18) and (17.26) hold but not necessarily (17.25). Define the conditional variances

$$\mathbb{E} [\varepsilon_{it}^2 | \mathbf{X}_i] = \sigma_{it}^2. \quad (17.55)$$

Then $\Sigma_i = \mathbb{E} [\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | \mathbf{X}_i] = \text{diag}(\sigma_{it}^2)$. The covariance matrix (17.24) can be written as

$$\mathbf{V}_{\text{fe}} = (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \sum_{t \in S_i} \dot{X}_{it} \dot{X}_{it}' \sigma_{it}^2 \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1}. \quad (17.56)$$

A natural estimator of σ_{it}^2 is $\hat{\varepsilon}_{it}^2$. Replacing σ_{it}^2 with $\hat{\varepsilon}_{it}^2$ in (17.56) and making a degree-of-freedom adjustment we obtain a White-type covariance matrix estimator

$$\hat{\mathbf{V}}_{\text{fe}} = \frac{n}{n - N - k} (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \sum_{t \in S_i} \dot{X}_{it} \dot{X}_{it}' \hat{\varepsilon}_{it}^2 \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1}.$$

Following the insight of White (1980) it may seem appropriate to expect $\hat{\mathbf{V}}_{\text{fe}}$ to be a reasonable estimator of \mathbf{V}_{fe} . Unfortunately this is not the case as discovered by Stock and Watson (2008). The problem is that $\hat{\mathbf{V}}_{\text{fe}}$ is a function of the individual-specific means $\bar{\varepsilon}_i$ which are negligible only if the number of time series observations T_i are large.

We can see this by a simple bias calculation. Assume that the sample is balanced and that the residuals are constructed with the true β . Then

$$\hat{\varepsilon}_{it} = \dot{\varepsilon}_{it} = \varepsilon_{it} - \frac{1}{T} \sum_{j=1}^T \varepsilon_{ij}.$$

Using (17.26) and (17.55)

$$\mathbb{E}[\hat{\varepsilon}_{it}^2 | \mathbf{X}_i] = \left(\frac{T-2}{T} \right) \sigma_{it}^2 + \frac{\bar{\sigma}_i^2}{T} \quad (17.57)$$

where $\bar{\sigma}_i^2 = T^{-1} \sum_{t=1}^T \sigma_{it}^2$. (See Exercise 17.10.) Using (17.57) and setting $k = 0$ we obtain

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{V}}_{\text{fe}} | \mathbf{X}] &= \frac{T}{T-1} (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \sum_{t \in S_i} \dot{X}_{it} \dot{X}_{it}' \mathbb{E}[\hat{\varepsilon}_{it}^2 | \mathbf{X}_i] \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \\ &= \left(\frac{T-2}{T-1} \right) \mathbf{V}_{\text{fe}} + \frac{1}{T-1} (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \bar{\sigma}_i^2 \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1}. \end{aligned}$$

Thus $\hat{\mathbf{V}}_{\text{fe}}$ is biased of order $O(T^{-1})$. Unless $T \rightarrow \infty$ this bias will persist as $N \rightarrow \infty$. $\hat{\mathbf{V}}_{\text{fe}}$ is unbiased in two contexts. The first is when the errors ε_{it} are homoskedastic. The second is when $T = 2$. (To show the latter requires some algebra so is omitted.)

To correct the bias for the case $T > 2$, Stock and Watson (2008) proposed the estimator

$$\tilde{\mathbf{V}}_{\text{fe}} = \left(\frac{T-1}{T-2} \right) \hat{\mathbf{V}}_{\text{fe}} - \frac{1}{T-1} \hat{\mathbf{B}}_{\text{fe}} \quad (17.58)$$

$$\hat{\mathbf{B}}_{\text{fe}} = (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \hat{\sigma}_i^2 \right) (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1}$$

$$\hat{\sigma}_i^2 = \frac{1}{T-1} \sum_{t=1}^T \hat{\varepsilon}_{it}^2. \quad (17.59)$$

You can check that $\mathbb{E}[\hat{\sigma}_i^2 | \mathbf{X}_i] = \bar{\sigma}_i^2$ and $\mathbb{E}[\tilde{\mathbf{V}}_{\text{fe}} | \mathbf{X}] = \mathbf{V}_{\text{fe}}$ so $\tilde{\mathbf{V}}_{\text{fe}}$ is unbiased for \mathbf{V}_{fe} . (See Exercise 17.11.)

Stock and Watson (2008) show that $\tilde{\mathbf{V}}_{\text{fe}}$ is consistent with T fixed and $N \rightarrow \infty$. In simulations they show that $\tilde{\mathbf{V}}_{\text{fe}}$ has excellent performance.

Because of the Stock-Watson analysis Stata no longer calculates the heteroskedasticity-robust covariance matrix estimator $\hat{\mathbf{V}}_{\text{fe}}$ when the fixed effects estimator is calculated using the `xtreg` command. Instead, the cluster-robust estimator $\hat{\mathbf{V}}_{\text{fe}}^{\text{cluster}}$ is reported when robust standard errors are requested. However, fixed effects is often implemented using the `areg` command which reports the biased estimator $\hat{\mathbf{V}}_{\text{fe}}$ if robust standard errors are requested. These leads to the practical recommendation that `areg` should be used with the `cluster(id)` option.

At present the corrected estimator (17.58) has not been programmed as a Stata option.

17.23 Heteroskedasticity-Robust Estimation – Unbalanced Case

A limitation with the bias-corrected robust covariance matrix estimator of Stock and Watson (2008) is that it was only derived for balanced panels. In this section we generalize their estimator to cover unbalanced panels.

The estimator is

$$\tilde{\mathbf{V}}_{\text{fe}} = (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \tilde{\Omega}_{\text{fe}} (\dot{\mathbf{X}}' \dot{\mathbf{X}})^{-1} \quad (17.60)$$

$$\tilde{\Omega}_{\text{fe}} = \sum_{i=1}^N \sum_{t \in S_i} \dot{X}_{it} \dot{X}_{it}' \left[\left(\frac{T_i \hat{\varepsilon}_{it}^2 - \hat{\sigma}_i^2}{T_i - 2} \right) \mathbb{1}_{\{T_i > 2\}} + \left(\frac{T_i \hat{\varepsilon}_{it}^2}{T_i - 1} \right) \mathbb{1}_{\{T_i = 2\}} \right]$$

where

$$\hat{\sigma}_i^2 = \frac{1}{T_i - 1} \sum_{t \in S_i} \hat{\varepsilon}_{it}^2.$$

To justify this estimator, as in the previous section make the simplifying assumption that the residuals are constructed with the true β . We calculate that

$$\mathbb{E}[\hat{\varepsilon}_{it}^2 | \mathbf{X}_i] = \left(\frac{T_i - 2}{T_i} \right) \sigma_{it}^2 + \frac{\bar{\sigma}_i^2}{T_i} \quad (17.61)$$

$$\mathbb{E}[\hat{\sigma}_i^2 | \mathbf{X}_i] = \bar{\sigma}_i^2. \quad (17.62)$$

You can show that under these assumptions, $\mathbb{E}[\tilde{\mathbf{V}}_{fe} | \mathbf{X}] = \mathbf{V}_{fe}$ and thus $\tilde{\mathbf{V}}_{fe}$ is unbiased for \mathbf{V}_{fe} . (See Exercise 17.12.)

In balanced panels the estimator $\tilde{\mathbf{V}}_{fe}$ simplifies to the Stock-Watson estimator (with $k = 0$).

17.24 Hausman Test for Random vs Fixed Effects

The random effects model is a special case of the fixed effects model. Thus we can test the null hypothesis of random effects against the alternative of fixed effects. The Hausman test is typically used for this purpose. The statistic is a quadratic in the difference between the fixed effects and random effects estimators. The statistic is

$$\begin{aligned} H &= (\hat{\beta}_{fe} - \hat{\beta}_{re})' \widehat{\text{var}}[\hat{\beta}_{fe} - \hat{\beta}_{re}]^{-1} (\hat{\beta}_{fe} - \hat{\beta}_{re}) \\ &= (\hat{\beta}_{fe} - \hat{\beta}_{re})' (\hat{\mathbf{V}}_{fe} - \hat{\mathbf{V}}_{re})^{-1} (\hat{\beta}_{fe} - \hat{\beta}_{re}) \end{aligned}$$

where both $\hat{\mathbf{V}}_{fe}$ and $\hat{\mathbf{V}}_{re}$ take the classical (non-robust) form.

The test can be implemented on a subset of the coefficients β . In particular this needs to be done if the regressors X_{it} contain time-invariant elements so that the random effects estimator contains more coefficients than the fixed effects estimator. In this case the test should be implemented only on the coefficients on the time-varying regressors.

An asymptotic $100\alpha\%$ test rejects if H exceeds the $1 - \alpha^{th}$ quantile of the χ_k^2 distribution where $k = \dim(\beta)$. If the test rejects this is evidence that the individual effect u_i is correlated with the regressors so the random effects model is not appropriate. On the other hand if the test fails to reject this evidence says that the random effects hypothesis cannot be rejected.

It is tempting to use the Hausman test to select whether to use the fixed effects or random effects estimator. One could imagine using the random effects estimator if the Hausman test fails to reject the random effects hypothesis and using the fixed effects estimator otherwise. This is not, however, a wise approach. This procedure – selecting an estimator based on a test – is known as a **pretest estimator** and is biased. The bias arises because the result of the test is random and correlated with the estimators.

Instead, the Hausman test can be used as a specification test. If you are planning to use the random effects estimator (and believe that the random effects assumptions are appropriate in your context) the Hausman test can be used to check this assumption and provide evidence to support your approach.

17.25 Random Effects or Fixed Effects?

We have presented the random effects and fixed effects estimators of the regression coefficients. Which should be used in practice? How should we view the difference?

The basic distinction is that the random effects estimator requires the individual error u_i to satisfy the conditional mean assumption (17.8). The fixed effects estimator does not require (17.8) and is robust to its violation. In particular, the individual effect u_i can be arbitrarily correlated with the regressors. On the other hand the random effects estimator is efficient under random effects (Assumption 17.1).

Current econometric practice is to prefer robustness over efficiency. Consequently, current practice is (nearly uniformly) to use the fixed effects estimator for linear panel data models. Random effects estimators are only used in contexts where fixed effects estimation is unknown or challenging (which occurs in many nonlinear models).

The labels “random effects” and “fixed effects” are misleading. These are labels which arose in the early literature and we are stuck with these labels today. In a previous era regressors were viewed as “fixed”. Viewing the individual effect as an unobserved regressor leads to the label of the individual effect as “fixed”. Today, we rarely refer to regressors as “fixed” when dealing with observational data. We view all variables as random. Consequently describing u_i as “fixed” does not make much sense and it is hardly a contrast with the “random effect” label since under either assumption u_i is treated as random. Once again, the labels are unfortunate but the key difference is whether u_i is correlated with the regressors.

17.26 Time Trends

In general we expect that economic agents will experience common shocks during the same time period. For example, business cycle fluctuations, inflation, and interest rates affect all agents in the economy. Therefore it is often desirable to include time effects in a panel regression model.

The simplest specification is a linear time trend

$$Y_{it} = X'_{it}\beta + \gamma t + u_i + \varepsilon_{it}.$$

For an introduction to time trends see Section 14.42. More flexible specifications (such as a quadratic) can also be used. For estimation it is appropriate to include the time trend t as an element of the regressor vector X_{it} and then apply fixed effects.

In some cases the time trends may be individual-specific. Series may be growing or declining at different rates. A linear time trend specification only extracts a common time trend. To allow for individual-specific time trends we need to include an interaction effect. This can be written as

$$Y_{it} = X'_{it}\beta + \gamma_i t + u_i + \varepsilon_{it}.$$

In a fixed effects specification the coefficients (γ_i, u_i) are treated as possibly correlated with the regressors. To eliminate them from the model we treat them as unknown parameters and estimate all by least squares. By the FWL theorem the estimator for β equals least squares of \hat{Y} on \hat{X} where their elements are the residuals from the least squares regressions on a linear time trend fit separately for each individual and variable.

17.27 Two-Way Error Components

In the previous section we discussed inclusion of time trends and individual-specific time trends. The functional forms imposed by linear time trends are restrictive. There is no economic reason to expect the “trend” of a series to be linear. Business cycle “trends” are cyclic. This suggests that it is desirable to be more flexible than a linear (or polynomial) specifications. In this section we consider the most flexible specification where the trend is allowed to take any arbitrary shape but will require that it is common rather than individual-specific.

The model we consider is the **two-way error component model**

$$Y_{it} = X'_{it}\beta + v_t + u_i + \varepsilon_{it}. \quad (17.63)$$

In this model u_i is an unobserved individual-specific effect, v_t is an unobserved time-specific effect, and ε_{it} is an idiosyncratic error.

The two-way model (17.63) can be handled either using random effects or fixed effects. In a random effects framework the errors v_t and u_i are modeled as in Assumption 17.1. When the panel is balanced the covariance matrix of the error vector $\mathbf{e} = \mathbf{v} \otimes \mathbf{1}_N + \mathbf{1}_T \otimes \mathbf{u} + \boldsymbol{\varepsilon}$ is

$$\text{var}[\mathbf{e}] = \Omega = (\mathbf{I}_T \otimes \mathbf{1}_N \mathbf{1}_N') \sigma_v^2 + (\mathbf{1}_T \mathbf{1}_T' \otimes \mathbf{I}_N) \sigma_u^2 + \mathbf{I}_n \sigma_\varepsilon^2. \quad (17.64)$$

When the panel is unbalanced a similar but cumbersome expression for (17.64) can be derived. This variance (17.64) can be used for GLS estimation of β .

More typically (17.63) is handled using fixed effects. The two-way within transformation subtracts both individual-specific means and time-specific means to eliminate both v_t and u_i from the two-way model (17.63). For a variable Y_{it} we define the time-specific mean as follows. Let S_t be the set of individuals i for which the observation t is included in the sample and let N_t be the number of these individuals. Then the time-specific mean at time t is

$$\tilde{Y}_t = \frac{1}{N_t} \sum_{i \in S_t} Y_{it}.$$

This is the average across all values of Y_{it} observed at time t .

For the case of balanced panels the **two-way within transformation** is

$$\ddot{Y}_{it} = Y_{it} - \bar{Y}_i - \tilde{Y}_t + \bar{Y} \quad (17.65)$$

where $\bar{Y} = n^{-1} \sum_{i=1}^N \sum_{t=1}^T Y_{it}$ is the full-sample mean. If Y_{it} satisfies the two-way component model

$$Y_{it} = v_t + u_i + \varepsilon_{it}$$

then $\bar{Y}_i = \bar{v} + u_i + \bar{\varepsilon}_i$, $\tilde{Y}_t = v_t + \bar{u} + \tilde{\varepsilon}_t$ and $\bar{Y} = \bar{v} + \bar{u} + \bar{\varepsilon}$. Hence

$$\begin{aligned} \ddot{Y}_{it} &= v_t + u_i + \varepsilon_{it} - (\bar{v} + u_i + \bar{\varepsilon}_i) - (v_t + \bar{u} + \tilde{\varepsilon}_t) + \bar{v} + \bar{u} + \bar{\varepsilon} \\ &= \varepsilon_{it} - \bar{\varepsilon}_i - \tilde{\varepsilon}_t + \bar{\varepsilon} = \ddot{\varepsilon}_{it} \end{aligned}$$

so the individual and time effects are eliminated.

The two-way within transformation applied to (17.63) yields

$$\ddot{Y}_{it} = \ddot{X}_{it}' \beta + \ddot{\varepsilon}_{it} \quad (17.66)$$

which is invariant to both v_t and u_i . The **two-way within estimator** is least squares applied to (17.66).

For the unbalanced case there are two computational approaches to implement the estimator. Both are based on the realization that the estimator is equivalent to including dummy variables for all time periods. Let τ_t be a set of T dummy variables where the t^{th} indicates the t^{th} time period. Thus the t^{th} element of τ_t is 1 and the remaining elements are zero. Set $\mathbf{v} = (v_1, \dots, v_T)'$ as the vector of time fixed effects. Notice that $v_t = \tau_t' \mathbf{v}$. We can write the two-way model as

$$Y_{it} = X_{it}' \beta + \tau_t' \mathbf{v} + u_i + \varepsilon_{it}. \quad (17.67)$$

This is the dummy variable representation of the two-way error components model.

Model (17.67) can be estimated by one-way fixed effects with regressors X_{it} and τ_t and coefficient vectors β and \mathbf{v} . This can be implemented by standard one-way fixed effects methods including `xtreg` or `areg` in Stata. This produces estimates of the slopes β as well as the time effects \mathbf{v} . To achieve identification one time dummy variable is omitted from τ_t so the estimated time effects are all relative to this

baseline time period. This is the most common method in practice to estimate a two-way fixed effects model. As the number of time periods is typically modest this is a computationally attractive approach.

The second computational approach is to eliminate the time effects by residual regression. This is done by the following steps. First, subtract individual-specific means for (17.67). This yields

$$\dot{Y}_{it} = \dot{X}'_{it}\beta + \dot{\tau}'_t\nu + \dot{\varepsilon}_{it}.$$

Second, regress \dot{Y}_{it} on $\dot{\tau}_t$ to obtain a residual \ddot{Y}_{it} and regress each element of \dot{X}_{it} on $\dot{\tau}_t$ to obtain a residual \ddot{X}_{it} . Third, regress \ddot{Y}_{it} on \ddot{X}_{it} to obtain the within estimator of β . These steps eliminate the fixed effects ν_t so the estimator is invariant to their value. What is important about this two-step procedure is that the second step is not a within transformation across the time index but rather standard regression.

If the two-way within estimator is used then the regressors X_{it} cannot include any time-invariant variables X_i or common time series variables X_t . Both are eliminated by the two-way within transformation. Coefficients are only identified for regressors which have variation both across individuals and across time.

If desired, the relevance of the time effects can be tested by an exclusion test on the coefficients ν . If the test rejects the hypothesis of zero coefficients then this indicates that the time effects are relevant in the regression model.

The fixed effects estimator of (17.63) is invariant to the values of ν_t and u_i , thus no assumptions need to be made concerning their stochastic properties.

To illustrate, the fourth column of Table 17.2 presents fixed effects estimates of the investment equation, augmented to included year dummy indicators, and is thus a two-way fixed effects model. In this example the coefficient estimates and standard errors are not greatly affected by the inclusion of the year dummy variables.

17.28 Instrumental Variables

Take the fixed effects model

$$Y_{it} = X'_{it}\beta + u_i + \varepsilon_{it}. \quad (17.68)$$

We say X_{it} is exogenous for ε_{it} if $\mathbb{E}[X_{it}\varepsilon_{it}] = 0$, and we say X_{it} is endogenous for ε_{it} if $\mathbb{E}[X_{it}\varepsilon_{it}] \neq 0$. In Chapter 12 we discussed several economic examples of endogeneity and the same issues apply in the panel data context. The primary difference is that in the fixed effects model we only need to be concerned if the regressors are correlated with the idiosyncratic error ε_{it} , as correlation between X_{it} and u_i is allowed.

As in Chapter 12 if the regressors are endogenous the fixed effects estimator will be biased and inconsistent for the structural coefficient β . The standard approach to handling endogeneity is to specify instrumental variables Z_{it} which are both relevant (correlated with X_{it}) yet exogenous (uncorrelated with ε_{it}).

Let Z_{it} be an $\ell \times 1$ instrumental variable where $\ell \geq k$. As in the cross-section case, Z_{it} may contain both included exogenous variables (variables in X_{it} that are exogenous) and excluded exogenous variables (variables not in X_{it}). Let \mathbf{Z}_i be the stacked instruments by individual and \mathbf{Z} be the stacked instruments for the full sample.

The dummy variable formulation of the fixed effects model is $Y_{it} = X'_{it}\beta + d'_i u + \varepsilon_{it}$ where d_i is an $N \times 1$ vector of dummy variables, one for each individual in the sample. The model in matrix notation for the full sample is

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{D}u + \boldsymbol{\varepsilon}. \quad (17.69)$$

Theorem 17.1 shows that the fixed effects estimator for β can be calculated by least squares estimation of (17.69). Thus the dummies \mathbf{D} should be viewed as included exogenous variables.

Consider 2SLS estimation of β using the instruments Z for X . Since D is an included exogenous variable it should also be used as an instrument. Thus 2SLS estimation of the fixed effects model (17.68) is algebraically 2SLS of the regression (17.69) of Y on (X, D) using the pair (Z, D) as instruments.

Since the dimension of D can be excessively large, as discussed in Section 17.11, it is advisable to use residual regression to compute the 2SLS estimator as we now describe.

In Section 12.12, we described several alternative representations for the 2SLS estimator. The fifth (equation (12.32)) shows that the 2SLS estimator for β equals

$$\hat{\beta}_{2sls} = \left(X' M_D Z (Z' M_D Z)^{-1} Z' M_D X \right)^{-1} \left(X' M_D Z (Z' M_D Z)^{-1} Z' M_D Y \right)$$

where $M_D = I_n - D(D'D)^{-1}D'$. The latter is the matrix within operator, thus $M_D Y = \dot{Y}$, $M_D X = \dot{X}$, and $M_D Z = \dot{Z}$. It follows that the 2SLS estimator is

$$\hat{\beta}_{2sls} = \left(\dot{X}' \dot{Z} (\dot{Z}' \dot{Z})^{-1} \dot{Z}' \dot{X} \right)^{-1} \left(\dot{X}' \dot{Z} (\dot{Z}' \dot{Z})^{-1} \dot{Z}' \dot{Y} \right).$$

This is convenient. It shows that the 2SLS estimator for the fixed effects model can be calculated by applying 2SLS to the within-transformed Y_{it} , X_{it} , and Z_{it} . The 2SLS residuals are $\hat{e} = \dot{Y} - \dot{X} \hat{\beta}_{2sls}$.

This estimator can be obtained using the Stata command `xtivreg fe`. It can also be obtained using the Stata command `ivregress` after making the within transformations.

The presentation above focused for clarity on the one-way fixed effects model. There is no substantial change in the two-way fixed effects model

$$Y_{it} = X'_{it} \beta + u_i + v_t + \varepsilon_{it}.$$

The easiest way to estimate the two-way model is to add $T - 1$ time-period dummies to the regression model and include these dummy variables as both regressors and instruments.

17.29 Identification with Instrumental Variables

To understand the identification of the structural slope coefficient β in the fixed effects model it is necessary to examine the reduced form equation for the endogenous regressors X_{it} . This is

$$X_{it} = \Gamma Z_{it} + W_i + \zeta_{it}$$

where W_i is a $k \times 1$ vector of fixed effects for the k regressors and ζ_{it} is an idiosyncratic error.

The coefficient matrix Γ is the linear effect of Z_{it} on X_{it} holding the fixed effects W_i constant. Thus Γ has a similar interpretation as the coefficient β in the fixed effects regression model. It is the effect of the variation in Z_{it} about its individual-specific mean on X_{it} .

The 2SLS estimator is a function of the within transformed variables. Applying the within transformation to the reduced form we find $\dot{X}_{it} = \Gamma \dot{Z}_{it} + \dot{\zeta}_{it}$. This shows that Γ is the effect of the within-transformed instruments on the regressors. If there is no time-variation in the within-transformed instruments or there is no correlation between the instruments and the regressors after removing the individual-specific means then the coefficient Γ will be either not identified or singular. In either case the coefficient β will not be identified.

Thus for identification of the fixed effects instrumental variables model we need

$$\mathbb{E} \left[\dot{Z}'_i \dot{Z}_i \right] > 0 \tag{17.70}$$

and

$$\text{rank}\left(\mathbb{E}\left[\dot{\mathbf{Z}}_i' \dot{\mathbf{X}}_i\right]\right) = k. \quad (17.71)$$

Condition (17.70) is the same as the condition for identification in fixed effects regression – the instruments must have full variation after the within transformation. Condition (17.71) is analogous to the relevance condition for identification of instrumental variable regression in the cross-section context but applies to the within-transformed instruments and regressors.

Condition (17.71) shows that to examine instrument validity in the context of fixed effects 2SLS it is important to estimate the reduced form equation using fixed effects (within) regression. Standard tests for instrument validity (F tests on the excluded instruments) can be applied. However, since the correlation structure of the reduced form equation is in general unknown it is appropriate to use a cluster-robust covariance matrix, clustered at the level of the individual.

17.30 Asymptotic Distribution of Fixed Effects 2SLS Estimator

In this section we present an asymptotic distribution theory for the fixed effects estimator. We provide a formal theory for the case of balanced panels and discuss an extension to the unbalanced case.

We use the following assumptions for balanced panels.

Assumption 17.4

1. $Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$ with $T \geq 2$.
2. The variables $(\varepsilon_i, \mathbf{X}_i, \mathbf{Z}_i)$, $i = 1, \dots, N$, are independent and identically distributed.
3. $\mathbb{E}[Z_{is}\varepsilon_{it}] = 0$ for all $s = 1, \dots, T$.
4. $\mathbf{Q}_{ZZ} = \mathbb{E}\left[\dot{\mathbf{Z}}_i' \dot{\mathbf{Z}}_i\right] > 0$.
5. $\text{rank}(\mathbf{Q}_{ZX}) = k$ where $\mathbf{Q}_{ZX} = \mathbb{E}\left[\dot{\mathbf{Z}}_i' \dot{\mathbf{X}}_i\right]$.
6. $\mathbb{E}[\varepsilon_{it}^4] < \infty$.
7. $\mathbb{E}\|\mathbf{X}_{it}\|^2 < \infty$.
8. $\mathbb{E}\|\mathbf{Z}_{it}\|^4 < \infty$.

Given Assumption 17.4 we can establish asymptotic normality for $\hat{\beta}_{2\text{sls}}$.

Theorem 17.4 Under Assumption 17.4, as $N \rightarrow \infty$, $\sqrt{N}(\hat{\beta}_{2\text{sls}} - \beta) \xrightarrow{d} N(0, \mathbf{V}_\beta)$ where

$$\mathbf{V}_\beta = (\mathbf{Q}'_{ZX} \Omega_{ZZ}^{-1} \mathbf{Q}_{ZX})^{-1} (\mathbf{Q}'_{ZX} \Omega_{ZZ}^{-1} \Omega_{Z\varepsilon} \Omega_{ZZ}^{-1} \mathbf{Q}_{ZX}) (\mathbf{Q}'_{ZX} \Omega_{ZZ}^{-1} \mathbf{Q}_{ZX})^{-1}$$

$$\Omega_{Z\varepsilon} = \mathbb{E}\left[\dot{\mathbf{Z}}_i' \varepsilon_i \varepsilon_i' \dot{\mathbf{Z}}_i\right].$$

The proof of the result is similar to Theorem 17.2 so is omitted. The key condition is Assumption 17.4.3, which states that the instruments are strictly exogenous for the idiosyncratic errors. The identification conditions are Assumptions 17.4.4 and 17.4.5, which were discussed in the previous section.

The theorem is stated for balanced panels. For unbalanced panels we can modify the theorem as in Theorem 17.3 by adding the selection indicators \mathbf{s}_i and replacing Assumption 17.4.3 with $\mathbb{E}[\varepsilon_{it} | \mathbf{Z}_i, \mathbf{s}_i] = 0$, which states that the idiosyncratic errors are mean independent of the instruments and selection.

If the idiosyncratic errors ε_{it} are homoskedastic and serially uncorrelated then the covariance matrix simplifies to

$$\mathbf{V}_\beta = (\mathbf{Q}'_{ZX} \Omega_{ZZ}^{-1} \mathbf{Q}_{ZX})^{-1} \sigma_\varepsilon^2.$$

In this case a classical homoskedastic covariance matrix estimator can be used. Otherwise a cluster-robust covariance matrix estimator can be used, and takes the form

$$\begin{aligned} \widehat{\mathbf{V}}_{\widehat{\beta}} &= \left(\dot{\mathbf{X}}' \dot{\mathbf{Z}} (\dot{\mathbf{Z}}' \dot{\mathbf{Z}})^{-1} \dot{\mathbf{Z}}' \dot{\mathbf{X}} \right)^{-1} \left(\dot{\mathbf{X}}' \dot{\mathbf{Z}} (\dot{\mathbf{Z}}' \dot{\mathbf{Z}})^{-1} \left(\sum_{i=1}^N \dot{\mathbf{Z}}_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' \dot{\mathbf{Z}}_i \right) \right. \\ &\quad \left. \times (\dot{\mathbf{Z}}' \dot{\mathbf{Z}})^{-1} (\dot{\mathbf{Z}}' \dot{\mathbf{X}}) \left(\dot{\mathbf{X}}' \dot{\mathbf{Z}} (\dot{\mathbf{Z}}' \dot{\mathbf{Z}})^{-1} \dot{\mathbf{Z}}' \dot{\mathbf{X}} \right)^{-1} \right). \end{aligned}$$

As for the case of fixed effects regression, the heteroskedasticity-robust covariance matrix estimator is not recommended due to bias when T is small, and a bias-corrected version has not been developed.

The Stata command `xtivreg, fe` by default reports the classical homoskedastic covariance matrix estimator. To obtain a cluster-robust covariance matrix use option `vce(robust)` or `vce(cluster id)`.

17.31 Linear GMM

Consider the just-identified 2SLS estimator. It solves the equation $\dot{\mathbf{Z}}' (\dot{\mathbf{Y}} - \dot{\mathbf{X}}\beta) = 0$. These are sample analogs of the population moment condition $\mathbb{E}[\dot{\mathbf{Z}}_i' (\dot{\mathbf{Y}}_i - \dot{\mathbf{X}}_i\beta)] = 0$. These population conditions hold at the true β because $\dot{\mathbf{Z}}' \mathbf{u} = \mathbf{Z}' \mathbf{M} \mathbf{D} \mathbf{u} = 0$ as \mathbf{u} lies in the null space of \mathbf{D} , and $\mathbb{E}[\dot{\mathbf{Z}}_i' \varepsilon] = 0$ is implied by Assumption 17.4.3.

The population orthogonality conditions hold in the overidentified case as well. In this case an alternative to 2SLS is GMM. Let $\widehat{\mathbf{W}}$ be an estimator of $\mathbf{W} = \mathbb{E}[\dot{\mathbf{Z}}_i' \varepsilon_i \varepsilon_i' \dot{\mathbf{Z}}_i]$, for example

$$\widehat{\mathbf{W}} = \frac{1}{N} \sum_{i=1}^N \dot{\mathbf{Z}}_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' \dot{\mathbf{Z}}_i \quad (17.72)$$

where $\widehat{\varepsilon}_i$ are the 2SLS fixed effects residuals. The GMM fixed effects estimator is

$$\widehat{\beta}_{\text{gmm}} = \left(\dot{\mathbf{X}}' \dot{\mathbf{Z}} \widehat{\mathbf{W}}^{-1} \dot{\mathbf{Z}}' \dot{\mathbf{X}} \right)^{-1} \left(\dot{\mathbf{X}}' \dot{\mathbf{Z}} \widehat{\mathbf{W}}^{-1} \dot{\mathbf{Z}}' \dot{\mathbf{Y}} \right). \quad (17.73)$$

The estimator (17.73)-(17.72) does not have a Stata command but can be obtained by generating the within transformed variables $\dot{\mathbf{X}}$, $\dot{\mathbf{Z}}$ and $\dot{\mathbf{Y}}$, and then estimating by GMM a regression of $\dot{\mathbf{Y}}$ on $\dot{\mathbf{X}}$ using $\dot{\mathbf{Z}}$ as instruments using a weight matrix clustered by individual.

17.32 Estimation with Time-Invariant Regressors

One of the disappointments with the fixed effects estimator is that it cannot estimate the effect of regressors which are time-invariant. They are not identified separately from the fixed effect and are eliminated by the within transformation. In contrast, the random effects estimator allows for time-invariant

regressors but does so only by assuming strict exogeneity which is stronger than typically desired in economic applications.

It turns out that we can consider an intermediate case which maintains the fixed effects assumptions for the time-varying regressors but uses stronger assumptions on the time-invariant regressors. For our exposition we will denote the time-varying regressors by the $k \times 1$ vector X_{it} and the time-invariant regressors by the $\ell \times 1$ vector Z_i .

Consider the linear regression model

$$Y_{it} = X'_{it}\beta + Z'_i\gamma + u_i + \varepsilon_{it}.$$

At the level of the individual this can be written as

$$Y_i = X_i\beta + Z_i\gamma + \mathbf{1}_i u_i + \varepsilon_i$$

where $Z_i = \mathbf{1}_i Z'_i$. For the full sample in matrix notation we can write this as

$$Y = X\beta + Z\gamma + \mathbf{u} + \varepsilon. \quad (17.74)$$

We maintain the assumption that the idiosyncratic errors ε_{it} are uncorrelated with both X_{it} and Z_i at all time horizons:

$$\mathbb{E}[X_{it}\varepsilon_{it}] = 0 \quad (17.75)$$

$$\mathbb{E}[Z_i\varepsilon_{it}] = 0. \quad (17.76)$$

In this section we consider the case where Z_i is uncorrelated with the individual-level error u_i , thus

$$\mathbb{E}[Z_i u_i] = 0, \quad (17.77)$$

but the correlation of X_{it} and u_i is left unrestricted. In this context we say that Z_i is exogenous with respect to the fixed effect u_i while X_{it} is endogenous with respect to u_i . Note that this is a different type of endogeneity than considered in the sections on instrumental variables: there endogeneity meant correlation with the idiosyncratic error ε_{it} . Here endogeneity means correlation with the fixed effect u_i .

We consider estimation of (17.74) by instrumental variables and thus need instruments which are uncorrelated with the error $u_i + \varepsilon_{it}$. The time-invariant regressors Z_i satisfy this condition due to (17.76) and (17.77), thus

$$\mathbb{E}[Z'_i(Y_i - X_i\beta - Z_i\gamma)] = 0.$$

While the time-varying regressors X_{it} are correlated with u_i the within transformed variables \dot{X}_{it} are uncorrelated with $u_i + \varepsilon_{it}$ under (17.75), thus

$$\mathbb{E}[\dot{X}'_i(Y_i - X_i\beta - Z_i\gamma)] = 0.$$

Therefore we can estimate (β, γ) by instrumental variable regression using the instrument set (\dot{X}, Z) . Specifically, regression of Y on X and Z treating X as endogenous, Z as exogenous, and using the instrument \dot{X} . Write this estimator as $(\hat{\beta}, \hat{\gamma})$. This can be implemented using the Stata `ivregress` command after constructing the within transformed \dot{X} .

This instrumental variables estimator is algebraically equal to a simple two-step estimator. The first step $\hat{\beta} = \hat{\beta}_{fe}$ is the fixed effects estimator. The second step sets $\hat{\gamma} = (Z'Z)^{-1}(Z'\hat{u})$, the least squares coefficient from the regression of the estimated fixed effect \hat{u}_i on Z_i . To see this equivalence observe that the instrumental variables estimator solves the sample moment equations

$$\dot{X}'(Y - X\beta - Z\gamma) = 0 \quad (17.78)$$

$$Z'(Y - X\beta - Z\gamma) = 0. \quad (17.79)$$

Notice that $\dot{\mathbf{X}}_i' \mathbf{Z}_i = \dot{\mathbf{X}}_i' \mathbf{1}_i \mathbf{Z}_i' = 0$ so $\dot{\mathbf{X}}' \mathbf{Z} = 0$. Thus (17.78) is the same as $\dot{\mathbf{X}}' (\mathbf{Y} - \mathbf{X}\beta) = 0$ whose solution is $\hat{\beta}_{fe}$. Plugging this into the left-side of (17.79) we obtain

$$\mathbf{Z}' (\mathbf{Y} - \mathbf{X}\hat{\beta}_{fe} - \mathbf{Z}\gamma) = \mathbf{Z}' (\bar{\mathbf{Y}} - \bar{\mathbf{X}}\hat{\beta}_{fe} - \mathbf{Z}\gamma) = \mathbf{Z}' (\hat{\mathbf{u}} - \mathbf{Z}\gamma)$$

where $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$ are the stacked individual means $\mathbf{1}_i \bar{Y}_i$ and $\mathbf{1}_i \bar{X}_i'$. Set equal to 0 and solving we obtain the least squares estimator $\hat{\gamma} = (\mathbf{Z}' \mathbf{Z})^{-1} (\mathbf{Z}' \hat{\mathbf{u}})$ as claimed. This equivalence was first observed by Hausman and Taylor (1981).

For standard error calculation it is recommended to estimate (β, γ) jointly by instrumental variable regression and use a cluster-robust covariance matrix clustered at the individual level. Classical and heteroskedasticity-robust estimators are misspecified due to the individual-specific effect u_i .

The estimator $(\hat{\beta}, \hat{\gamma})$ is a special case of the Hausman-Taylor estimator described in the next section. (For an unknown reason the above estimator cannot be estimated using Stata's `xthtaylor` command.)

17.33 Hausman-Taylor Model

Hausman and Taylor (1981) consider a generalization of the previous model. Their model is

$$Y_{it} = X_{1it}' \beta_1 + X_{2it}' \beta_2 + Z_{1i}' \gamma_1 + Z_{2i}' \gamma_2 + u_i + \varepsilon_{it}$$

where X_{1it} and X_{2it} are time-varying and Z_{1i} and Z_{2i} are time-invariant. Let the dimensions of X_{1it} , X_{2it} , Z_{1i} , and Z_{2i} be k_1 , k_2 , ℓ_1 , and ℓ_2 , respectively.

Write the model in matrix notation as

$$\mathbf{Y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{Z}_1 \gamma_1 + \mathbf{Z}_2 \gamma_2 + \mathbf{u} + \boldsymbol{\varepsilon}. \quad (17.80)$$

Let $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ denote conformable matrices of individual-specific means and let $\dot{\mathbf{X}}_1 = \mathbf{X}_1 - \bar{\mathbf{X}}_1$ and $\dot{\mathbf{X}}_2 = \mathbf{X}_2 - \bar{\mathbf{X}}_2$ denote the within-transformed variables.

The Hausman-Taylor model assumes that all regressors are uncorrelated with the idiosyncratic error ε_{it} at all time horizons and that X_{1it} and Z_{1i} are exogenous with respect to the fixed effect u_i so that

$$\mathbb{E}[X_{1it} u_i] = 0$$

$$\mathbb{E}[Z_{1i} u_i] = 0.$$

The regressors X_{2it} and Z_{2i} , however, are allowed to be correlated with u_i .

Set $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{Z}_1, \mathbf{Z}_2)$ and $\beta = (\beta_1, \beta_2, \gamma_1, \gamma_2)$. The assumptions imply the following population moment conditions

$$\mathbb{E}[\dot{\mathbf{X}}_1' (\mathbf{Y} - \mathbf{X}\beta)] = 0$$

$$\mathbb{E}[\dot{\mathbf{X}}_2' (\mathbf{Y} - \mathbf{X}\beta)] = 0$$

$$\mathbb{E}[\bar{\mathbf{X}}_1' (\mathbf{Y} - \mathbf{X}\beta)] = 0$$

$$\mathbb{E}[\mathbf{Z}_1' (\mathbf{Y} - \mathbf{X}\beta)] = 0.$$

There are $2k_1 + k_2 + \ell_1$ moment conditions and $k_1 + k_2 + \ell_1 + \ell_2$ coefficients. Identification requires $k_1 \geq \ell_2$: that there are at least as many exogenous time-varying regressors as endogenous time-invariant regressors. (This includes the model of the previous section where $k_1 = \ell_2 = 0$.)

Given the moment conditions the coefficients $\beta = (\beta_1, \beta_2, \gamma_1, \gamma_2)$ can be estimated by 2SLS regression of (17.80) using the instruments $\mathbf{Z} = (\dot{\mathbf{X}}_1, \dot{\mathbf{X}}_2, \bar{\mathbf{X}}_1, \mathbf{Z}_1)$ or equivalently $\mathbf{Z} = (\mathbf{X}_1, \dot{\mathbf{X}}_2, \bar{\mathbf{X}}_1, \mathbf{Z}_1)$. This is 2SLS regression treating \mathbf{X}_1 and \mathbf{Z}_1 as exogenous and \mathbf{X}_2 and \mathbf{Z}_2 as endogenous using the excluded instruments $\dot{\mathbf{X}}_2$ and $\bar{\mathbf{X}}_1$.

It is recommended to use cluster-robust covariance matrix estimation clustered at the individual level. Neither conventional nor heteroskedasticity-robust covariance matrix estimators should be used as they are misspecified due to the individual-specific effect u_i .

When the model is just-identified the estimators simplify as follows. $\hat{\beta}_1$ and $\hat{\beta}_2$ are the fixed effects estimator. $\hat{\gamma}_1$ and $\hat{\gamma}_2$ equal the 2SLS estimator from a regression of \hat{u}_i on Z_{1i} and Z_{2i} using \bar{X}_{1i} as an instrument for Z_{2i} . (See Exercise 17.14.)

When the model is over-identified the equation can also be estimated by GMM with a cluster-robust weight matrix using the same equations and instruments.

This estimator with cluster-robust standard errors can be calculated using the Stata `ivregress cluster(id)` command after constructing the transformed variables $\dot{\mathbf{X}}_2$ and $\bar{\mathbf{X}}_1$.

The 2SLS estimator described above corresponds with the Hausman and Taylor (1981) estimator in the just-identified case with a balanced panel.

Hausman and Taylor derived their estimator under the stronger assumption that the errors ε_{it} and u_i are strictly mean independent and homoskedastic and consequently proposed a GLS-type estimator which is more efficient when these assumptions are correct. Define $\Omega = \text{diag}(\Omega_i)$ where $\Omega_i = \mathbf{I}_i + \mathbf{1}_i \mathbf{1}_i' \sigma_u^2 / \sigma_\varepsilon^2$ and σ_ε^2 and σ_u^2 are the variances of the error components ε_{it} and u_i . Define as well the transformed variables $\tilde{\mathbf{Y}} = \Omega^{-1/2} \mathbf{Y}$, $\tilde{\mathbf{X}} = \Omega^{-1/2} \mathbf{X}$ and $\tilde{\mathbf{Z}} = \Omega^{-1/2} \mathbf{Z}$. The Hausman-Taylor estimator is

$$\begin{aligned} \hat{\beta}_{\text{ht}} &= \left(\mathbf{X}' \Omega^{-1} \mathbf{Z} (\mathbf{Z}' \Omega^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \Omega^{-1} \mathbf{X} \right)^{-1} \left(\mathbf{X}' \Omega^{-1} \mathbf{Z} (\mathbf{Z}' \Omega^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \Omega^{-1} \mathbf{Y} \right) \\ &= \left(\tilde{\mathbf{X}}' \tilde{\mathbf{Z}} (\tilde{\mathbf{Z}}' \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}' \tilde{\mathbf{X}} \right)^{-1} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{Z}} (\tilde{\mathbf{Z}}' \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}' \tilde{\mathbf{Y}} \right). \end{aligned}$$

Recall from (17.47) that $\Omega_i^{-1/2} = \mathbf{M}_i + \rho_i \mathbf{P}_i$ where ρ_i is defined in (17.46). Thus

$$\begin{aligned} \tilde{Y}_i &= Y_i - (1 - \rho_i) \bar{Y}_i \\ \tilde{X}_{1i} &= X_{1i} - (1 - \rho_i) \bar{X}_{1i} \\ \tilde{X}_{2i} &= X_{2i} - (1 - \rho_i) \bar{X}_{2i} \\ \tilde{Z}_{1i} &= \rho_i Z_{1i} \\ \tilde{Z}_{2i} &= \rho_i Z_{2i} \\ \tilde{\dot{X}}_{1i} &= \dot{X}_{1i} \\ \tilde{\dot{X}}_{2i} &= \dot{X}_{2i}. \end{aligned}$$

It follows that the Hausman-Taylor estimator can be calculated by 2SLS regression of \tilde{Y}_i on $(\tilde{X}_{1i}, \tilde{X}_{2i}, \rho_i Z_{1i}, \rho_i Z_{2i})$ using the instruments $(\dot{X}_{1i}, \dot{X}_{2i}, \rho_i \bar{X}_{1i}, \rho_i Z_{2i})$.

When the panel is balanced the coefficients ρ_i all equal and scale out from the instruments. Thus the estimator can be calculated by 2SLS regression of \tilde{Y}_i on $(\tilde{X}_{1i}, \tilde{X}_{2i}, Z_{1i}, Z_{2i})$ using the instruments $(\dot{X}_{1i}, \dot{X}_{2i}, \bar{X}_{1i}, Z_{2i})$.

In practice ρ_i is unknown. It can be estimated as in (17.48) with the modification that the error variance is estimated from the untransformed 2SLS regression. Under the homoskedasticity assumptions used by Hausman and Taylor the estimator $\hat{\beta}_{\text{ht}}$ has a classical asymptotic covariance matrix. When these assumptions are relaxed the covariance matrix can be estimated using cluster-robust methods.

The Hausman-Taylor estimator with cluster-robust standard errors can be implemented in Stata by the command `xthtaylor vce(robust)`. This Stata command, for an unknown reason, requires that there is at least one exogenous time-invariant variable ($\ell_1 \geq 1$) and at least one exogenous time-varying variable ($k_1 \geq 1$), even when the model is identified. Otherwise, the estimator can be implemented using the instrumental variable method described above.

The Hausman-Taylor estimator was refined by Amemiya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989) who proposed more efficient versions using additional instruments which are valid under stronger orthogonality conditions. The observation that in the unbalanced case the instruments should be weighted by ρ_i was made by Gardner (1998).

In the over-identified case it is unclear if it is preferred to use the simpler 2SLS estimator $\hat{\beta}_{2sls}$ or the GLS-type Hausman-Taylor estimator $\hat{\beta}_{ht}$. The advantages of $\hat{\beta}_{ht}$ are that it is asymptotically efficient under their stated homoskedasticity and serial correlation conditions and that there is an available program in Stata. The advantage of $\hat{\beta}_{2sls}$ is that it is much simpler to program (if doing so yourself), may have better finite sample properties (because it avoids variance-component estimation), and is the natural estimator from the the modern GMM viewpoint.

To illustrate, the final column of Table 17.2 contains Hausman-Taylor estimates of the investment model treating Q_{it-1} , D_{it-1} , and T_i as endogenous for u_i and CF_{it-1} and the industry dummies as exogenous. Relative to the fixed effects models this allows estimation of the coefficients on the trading indicator T_i . The most interesting change relative to the previous estimates is that the coefficient on the trading indicator T_i doubles in magnitude relative to the random effects estimate. This is consistent with the hypothesis that T_i is correlated with the fixed effect and hence the random effects estimate is biased.

17.34 Jackknife Covariance Matrix Estimation

As an alternative to asymptotic inference the delete-cluster jackknife can be used for covariance matrix calculation. In the context of fixed effects estimation the delete-cluster estimators take the form

$$\hat{\beta}_{(-i)} = \left(\sum_{j \neq i} \dot{\mathbf{X}}_j' \dot{\mathbf{X}}_j \right)^{-1} \left(\sum_{j \neq i} \dot{\mathbf{X}}_j' \dot{\mathbf{Y}}_j \right) = \hat{\beta}_{fe} - \left(\sum_{i=1}^N \dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \right)^{-1} \dot{\mathbf{X}}_i' \tilde{\mathbf{e}}_i.$$

where

$$\begin{aligned} \tilde{\mathbf{e}}_i &= \left(\mathbf{I}_i - \dot{\mathbf{X}}_i \left(\dot{\mathbf{X}}_i' \dot{\mathbf{X}}_i \right)^{-1} \dot{\mathbf{X}}_i' \right)^{-1} \hat{\mathbf{e}}_i \\ \hat{\mathbf{e}}_i &= \dot{\mathbf{Y}}_i - \dot{\mathbf{X}}_i \hat{\beta}_{fe}. \end{aligned}$$

The delete-cluster jackknife estimator of the variance of $\hat{\beta}_{fe}$ is

$$\begin{aligned} \hat{\mathbf{V}}_{\hat{\beta}}^{\text{jack}} &= \frac{N-1}{N} \sum_{i=1}^N \left(\hat{\beta}_{(-i)} - \bar{\beta} \right) \left(\hat{\beta}_{(-i)} - \bar{\beta} \right)' \\ \bar{\beta} &= \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{(-i)}. \end{aligned}$$

The delete-cluster jackknife estimator $\hat{\mathbf{V}}_{\hat{\beta}}^{\text{jack}}$ is similar to the cluster-robust covariance matrix estimator.

For parameters which are functions $\hat{\theta}_{fe} = r(\hat{\beta}_{fe})$ of the fixed effects estimator the delete-cluster jack-

knife estimator of the variance of $\hat{\theta}_{fe}$ is

$$\begin{aligned}\hat{V}_{\hat{\theta}}^{\text{jack}} &= \frac{N-1}{N} \sum_{i=1}^N \left(\hat{\theta}_{(-i)} - \bar{\theta} \right) \left(\hat{\theta}_{(-i)} - \bar{\theta} \right)' \\ \hat{\theta}_{(-i)} &= r(\hat{\beta}_{(-i)}) \\ \bar{\theta} &= \frac{1}{N} \sum_{i=1}^N \hat{\theta}_{(-i)}.\end{aligned}$$

The estimator $\hat{V}_{\hat{\theta}}^{\text{jack}}$ is similar to the delta-method cluster-robust covariance matrix estimator for $\hat{\theta}$.

As in the context of i.i.d. samples one advantage of the jackknife covariance matrix estimator is that it does not require the user to make a technical calculation of the asymptotic distribution. A downside is an increase in computation cost as N separate regressions are effectively estimated. This can be particularly costly in micro panels which have a large number N of individuals.

In Stata jackknife standard errors for fixed effects estimators are obtained by using either `xtreg fe vce(jackknife)` or `areg absorb(id) cluster(id) vce(jackknife)` where `id` is the cluster variable. For the fixed effects 2SLS estimator use `xtivreg fe vce(jackknife)`.

17.35 Panel Bootstrap

Bootstrap methods can also be applied to panel data by a straightforward application of the pairs cluster bootstrap which samples entire individuals rather than single observations. In the context of panel data we call this the panel nonparametric bootstrap.

The **panel nonparametric bootstrap** samples N individual histories (Y_i, X_i) to create the bootstrap sample. Fixed effects (or any other estimation method) is applied to the bootstrap sample to obtain the coefficient estimates. By repeating B times, bootstrap standard errors for coefficients estimates, or functions of the coefficient estimates, can be calculated. Percentile-type and percentile-t confidence intervals can be calculated. The BC_a interval requires an estimator of the acceleration coefficient a which is a scaled jackknife estimate of the third moment of the estimator. In panel data the delete-cluster jackknife should be used for estimation of a .

In Stata, to obtain bootstrap standard errors and confidence intervals use either `xtreg, vce(bootstrap, reps(#))` or `areg, absorb(id) cluster(id) vce(bootstrap, reps(#))` where `id` is the cluster variable and `#` is the number of bootstrap replications. For the fixed effects 2SLS estimator use `xtivreg, fe vce(bootstrap, reps(#))`.

17.36 Dynamic Panel Models

The models so far considered in this chapter have been static with no dynamic relationships. In many economic contexts it is natural to expect that behavior and decisions are dynamic, explicitly depending on past behavior. In our investment equation, for example, economic models predict that a firm's investment in any given year will depend on investment decisions from previous years. These considerations lead us to consider explicitly dynamic models.

The workhorse dynamic model in a panel framework is the p^{th} -order autoregression with regressors and a one-way error component structure. This is

$$Y_{it} = \alpha_1 Y_{i,t-1} + \cdots + \alpha_p Y_{i,t-p} + X'_{it} \beta + u_i + \varepsilon_{it}. \quad (17.81)$$

where α_j are the autoregressive coefficients, X_{it} is a k vector of regressors, u_i is an individual effect, and ε_{it} is an idiosyncratic error. It is conventional to assume that the errors u_i and ε_{it} are mutually

independent and the ε_{it} are serially uncorrelated and mean zero. For the present we will assume that the regressors X_{it} are strictly exogenous (17.17). In Section 17.41 we discuss predetermined regressors.

For many illustrations we will focus on the AR(1) model

$$Y_{it} = \alpha Y_{i,t-1} + u_i + \varepsilon_{it} \quad (17.82)$$

The dynamics should be interpreted individual-by-individual. The coefficient α in (17.82) equals the first-order autocorrelation. When $\alpha = 0$ the series is serially uncorrelated (conditional on u_i). $\alpha > 0$ means Y_{it} is positively serially correlated. $\alpha < 0$ means Y_{it} is negatively serially correlated. An autoregressive unit root holds when $\alpha = 1$, which means that Y_{it} follows a random walk with possible drift. Since u_i is constant for a given individual it should be treated as an individual-specific intercept. The idiosyncratic error ε_{it} plays the role of the error in a standard time series autoregression.

If $|\alpha| < 1$ the model (17.82) is stationary. By standard autoregressive backwards recursion we calculate that

$$Y_{it} = \sum_{j=0}^{\infty} \alpha^j (u_i + \varepsilon_{i,t-j}) = (1 - \alpha)^{-1} u_i + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,t-j}. \quad (17.83)$$

Thus conditional on u_i the mean and variance of Y_{it} are $(1 - \alpha)^{-1} u_i$ and $(1 - \alpha^2)^{-1} \sigma_\varepsilon^2$, respectively. The k^{th} autocorrelation (conditional on u_i) is α^k . Notice that the effect of cross-section variation in u_i is to shift the mean but not the variance or serial correlation. This implies that if we view time series plots of Y_{it} against time for a set of individuals i , the series Y_{it} will appear to have different means but have similar variances and serial correlation.

As with the case with time series data, serial correlation (large α) can proxy for other factors such as time trends. Thus in applications it will often be useful to include time effects to eliminate spurious serial correlation.

17.37 The Bias of Fixed Effects Estimation

To estimate the panel autoregression (17.81) it may appear natural to use the fixed effects (within) estimator. Indeed, the within transformation eliminates the individual effect u_i . The trouble is that the within operator induces correlation between the AR(1) lag and the error. The result is that the within estimator is inconsistent for the coefficients when T is fixed. A thorough explanation appears in Nickell (1981). We describe the basic problem in this section focusing on the AR(1) model (17.82).

Applying the within operator to (17.82) we obtain

$$\dot{Y}_{it} = \alpha \dot{Y}_{i,t-1} + \dot{\varepsilon}_{it}$$

for $t \geq 2$. As expected the individual effect is eliminated. The difficulty is that $\mathbb{E}[\dot{Y}_{i,t-1} \dot{\varepsilon}_{it}] \neq 0$ because both $\dot{Y}_{i,t-1}$ and $\dot{\varepsilon}_{it}$ are functions of the entire time series.

To see this clearly in a simple example, suppose we have a balanced panel with $T = 3$. There are two observed pairs $(Y_{it}, Y_{i,t-1})$ per individual so the within estimator equals the differenced estimator. Applying the differencing operator to (17.82) for $t = 3$ we find

$$\Delta Y_{i3} = \alpha \Delta Y_{i2} + \Delta \varepsilon_{i3}. \quad (17.84)$$

Because of the lagged dependent variable and differencing there is effectively one observation per individual. Notice that the individual effect has been eliminated.

The fixed effects estimator of α is equal to the least squares estimator applied to (17.84), which is

$$\hat{\alpha}_{fe} = \left(\sum_{i=1}^N \Delta Y_{i2}^2 \right)^{-1} \left(\sum_{i=1}^N \Delta Y_{i2} \Delta Y_{i3} \right) = \alpha + \left(\sum_{i=1}^N \Delta Y_{i2}^2 \right)^{-1} \left(\sum_{i=1}^N \Delta Y_{i2} \Delta \varepsilon_{i3} \right).$$

The differenced regressor and error are negatively correlated. Indeed

$$\begin{aligned}\mathbb{E}[\Delta Y_{i2} \Delta \varepsilon_{i3}] &= \mathbb{E}[(Y_{i2} - Y_{i1})(\varepsilon_{i3} - \varepsilon_{i2})] \\ &= \mathbb{E}[Y_{i2} \varepsilon_{i3}] - \mathbb{E}[Y_{i1} \varepsilon_{i3}] - \mathbb{E}[Y_{i2} \varepsilon_{i2}] + \mathbb{E}[Y_{i1} \varepsilon_{i2}] \\ &= 0 - 0 - \sigma_\varepsilon^2 + 0 \\ &= -\sigma_\varepsilon^2.\end{aligned}$$

Using the variance formula for AR(1) models (assuming $|\alpha| < 1$) we calculate that $\mathbb{E}[(\Delta Y_{i2})^2] = 2\sigma_\varepsilon^2/(1 + \alpha)$. It follows that the probability limit of the fixed effects estimator $\hat{\alpha}_{fe}$ of α in (17.84) is

$$\text{plim}_{N \rightarrow \infty} (\hat{\alpha}_{fe} - \alpha) = \frac{\mathbb{E}[\Delta Y_{i2} \Delta \varepsilon_{i3}]}{\mathbb{E}[(\Delta Y_{i2})^2]} = -\frac{1 + \alpha}{2}. \quad (17.85)$$

It is typical to call (17.85) the “bias” of $\hat{\alpha}_{fe}$, though it is technically a probability limit.

The bias found in (17.85) is large. For $\alpha = 0$ the bias is $-1/2$ and increases towards 1 as $\alpha \rightarrow 1$. Thus for any $\alpha < 1$ the probability limit of $\hat{\alpha}_{fe}$ is negative! This is extreme bias.

Now take the case $T > 3$. From Nickell’s (1981) expressions and some algebra, we can calculate that the probability limit of the fixed effects estimator for $|\alpha| < 1$ is

$$\text{plim}_{N \rightarrow \infty} (\hat{\alpha}_{fe} - \alpha) = \frac{1 + \alpha}{\frac{2\alpha}{1 - \alpha} - \frac{T - 1}{1 - \alpha^{T-1}}}. \quad (17.86)$$

It follows that the bias is of order $O(1/T)$.

It is often asserted that it is okay to use fixed effects if T is sufficiently large, e.g. $T \geq 30$. However, from (17.86) we can calculate that for $T = 30$ the bias of the fixed effects estimator is -0.056 when $\alpha = 0.5$ and the bias is -0.15 when $\alpha = 0.9$. For $T = 60$ and $\alpha = 0.9$ the bias is -0.05 . These magnitudes are unacceptably large. This includes the longer time series encountered in macro panels. Thus the Nickell bias problem applies to both micro and macro panel applications.

The conclusion from this analysis is that the fixed effects estimator should not be used for models with lagged dependent variables even if the time series dimension T is large.

17.38 Anderson-Hsiao Estimator

Anderson and Hsiao (1982) made an important breakthrough by showing that a simple instrumental variables estimator is consistent for the parameters of (17.81).

The method first eliminates the individual effect u_i by first-differencing (17.81) for $t \geq p + 1$

$$\Delta Y_{it} = \alpha_1 \Delta Y_{i,t-1} + \alpha_2 \Delta Y_{i,t-2} + \cdots + \alpha_p \Delta Y_{i,t-p} + \Delta X'_{it} \beta + \Delta \varepsilon_{it}. \quad (17.87)$$

This eliminates the individual effect u_i . The challenge is that first-differencing induces correlation between $\Delta Y_{i,t-1}$ and $\Delta \varepsilon_{it}$:

$$\mathbb{E}[\Delta Y_{i,t-1} \Delta \varepsilon_{it}] = \mathbb{E}[(Y_{i,t-1} - Y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})] = -\sigma_\varepsilon^2.$$

The other regressors are not correlated with $\Delta \varepsilon_{it}$. For $s > 1$, $\mathbb{E}[\Delta Y_{i,t-s} \Delta \varepsilon_{it}] = 0$, and when X_{it} is strictly exogenous $\mathbb{E}[\Delta X_{it} \Delta \varepsilon_{it}] = 0$.

The correlation between $\Delta Y_{i,t-1}$ and $\Delta \varepsilon_{it}$ is endogeneity. One solution to endogeneity is to use an instrument. Anderson-Hsiao pointed out that $Y_{i,t-2}$ is a valid instrument because it is correlated with $\Delta Y_{i,t-1}$ yet uncorrelated with $\Delta \varepsilon_{it}$.

$$\mathbb{E}[Y_{i,t-2} \Delta \varepsilon_{it}] = \mathbb{E}[Y_{i,t-2} \varepsilon_{it}] - \mathbb{E}[Y_{i,t-2} \varepsilon_{i,t-1}] = 0. \quad (17.88)$$

The Anderson-Hsiao estimator is IV using $Y_{i,t-2}$ as an instrument for $\Delta Y_{i,t-1}$. Equivalently, this is IV using the instruments $(Y_{i,t-2}, \dots, Y_{i,t-p-1})$ for $(\Delta Y_{i,t-1}, \dots, \Delta Y_{i,t-p})$. The estimator requires $T \geq p+2$.

To show that this estimator is consistent, for simplicity assume we have a balanced panel with $T = 3$, $p = 1$, and no regressors. In this case the Anderson-Hsiao IV estimator is

$$\hat{\alpha}_{iv} = \left(\sum_{i=1}^N Y_{i1} \Delta Y_{i2} \right)^{-1} \left(\sum_{i=1}^N Y_{i1} \Delta Y_{i3} \right) = \alpha + \left(\sum_{i=1}^N Y_{i1} \Delta Y_{i2} \right)^{-1} \left(\sum_{i=1}^N Y_{i1} \Delta \varepsilon_{i3} \right).$$

Under the assumption that ε_{it} is serially uncorrelated, (17.88) shows that $\mathbb{E}[Y_{i1} \Delta \varepsilon_{i3}] = 0$. In general, $\mathbb{E}[Y_{i1} \Delta Y_{i2}] \neq 0$. As $N \rightarrow \infty$

$$\hat{\alpha}_{iv} \xrightarrow{p} \alpha - \frac{\mathbb{E}[Y_{i1} \Delta \varepsilon_{i3}]}{\mathbb{E}[Y_{i1} \Delta Y_{i2}]} = \alpha.$$

Thus the IV estimator is consistent for α .

The Anderson-Hsiao IV estimator relies on two critical assumptions. First, the validity of the instrument (uncorrelatedness with the equation error) relies on the assumption that the dynamics are correctly specified so that ε_{it} is serially uncorrelated. For example, many applications use an AR(1). If instead the true model is an AR(2) then $Y_{i,t-2}$ is not a valid instrument and the IV estimates will be biased. Second, the relevance of the instrument (correlatedness with the endogenous regressor) requires $\mathbb{E}[Y_{i1} \Delta Y_{i2}] \neq 0$. This turns out to be problematic and is explored further in Section 17.40. These considerations suggest that the validity and accuracy of the estimator are likely to be sensitive to these unknown features.

17.39 Arellano-Bond Estimator

The orthogonality condition (17.88) is one of many implied by the dynamic panel model. Indeed, all lags $Y_{i,t-2}, Y_{i,t-3}, \dots$ are valid instruments. If $T > p+2$ these can be used to potentially improve estimation efficiency. This was first pointed out by Holtz-Eakin, Newey, and Rosen (1988) and further developed by Arellano and Bond (1991).

Using these extra instruments has a complication that there are a different number of instruments for each time period. The solution is to view the model as a system of T equations as in Section 17.18.

It will be useful to first write the model in vector notation. Stack the differenced regressors $(\Delta Y_{i,t-1}, \dots, \Delta Y_{i,t-p}, \Delta X'_{it})$ into a matrix $\Delta \mathbf{X}_i$ and the coefficients into a vector θ . We can write (17.87) as $\Delta \mathbf{Y}_i = \Delta \mathbf{X}_i \theta + \Delta \varepsilon_i$. Stacking all N individuals this can be written as $\Delta \mathbf{Y} = \Delta \mathbf{X} \theta + \Delta \varepsilon$.

For period $t = p+2$ we have $p+k$ valid instruments $[Y_{i1}, \dots, Y_{ip}, \Delta X'_{i,p+2}]$. For period $t = p+3$ there are $p+1+k$ valid instruments $[Y_{i1}, \dots, Y_{ip+1}, \Delta X'_{i,p+3}]$. For period $t = p+4$ there are $p+2+k$ instruments. In general, for any $t \geq p+2$ there are $t-2$ instruments $[Y_{i1}, \dots, Y_{i,t-2}, \Delta X'_{it}]$. Similarly to (17.53) we can define the instrument matrix for individual i as

$$\mathbf{Z}_i = \begin{bmatrix} [Y_{i1}, \dots, Y_{ip}, \Delta X'_{i,p+2}] & 0 & 0 \\ 0 & [Y_{i1}, \dots, Y_{ip+1}, \Delta X'_{i,p+3}] & 0 \\ & & \ddots \\ 0 & 0 & [Y_{i1}, Y_{i2}, \dots, Y_{i,T-2}, \Delta X'_{iT}] \end{bmatrix}. \quad (17.89)$$

This is $(T-p-1) \times \ell$ where $\ell = k(T-p-1) + ((T-2)(T-1) - (p-2)(p-1))/2$. This instrument matrix consists of all lagged values $Y_{i,t-2}, Y_{i,t-3}, \dots$ which are available in the data set plus the differenced strictly exogenous regressors.

The ℓ moment conditions are

$$\mathbb{E}[\mathbf{Z}'_i (\Delta \mathbf{Y}_i - \Delta \mathbf{X}_i \alpha)] = 0. \quad (17.90)$$

If $T > p + 2$ then $\ell > p$ and the model is overidentified. Define the $\ell \times \ell$ covariance matrix for the moment conditions

$$\Omega = \mathbb{E} [\mathbf{Z}'_i \Delta \boldsymbol{\varepsilon}_i \Delta \boldsymbol{\varepsilon}'_i \mathbf{Z}_i].$$

Let \mathbf{Z} denote \mathbf{Z}_i stacked into a $(T - p - 1)N \times \ell$ matrix. The efficient GMM estimator of α is

$$\hat{\alpha}_{\text{gmm}} = (\Delta \mathbf{X}' \mathbf{Z} \Omega^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{Z} \Omega^{-1} \mathbf{Z}' \Delta \mathbf{Y}).$$

If the errors ε_{it} are conditionally homoskedastic then

$$\Omega = \mathbb{E} [\mathbf{Z}'_i \mathbf{H} \mathbf{Z}_i] \sigma_\varepsilon^2$$

where \mathbf{H} is given in (17.31). In this case set

$$\hat{\Omega}_1 = \sum_{i=1}^N \mathbf{Z}'_i \mathbf{H} \mathbf{Z}_i$$

as a (scaled) estimate of Ω . Under these assumptions an asymptotically efficient GMM estimator is

$$\hat{\alpha}_1 = (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \Delta \mathbf{Y}). \quad (17.91)$$

Estimator (17.91) is known as the **one-step Arellano-Bond GMM estimator**.

Under the assumption that the error ε_{it} is homoskedastic and serially uncorrelated, a classical covariance matrix estimator for $\hat{\alpha}_1$ is

$$\hat{\mathbf{V}}_1^0 = (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2 \quad (17.92)$$

where $\hat{\sigma}_\varepsilon^2$ is the sample variance of the one-step residuals $\hat{\boldsymbol{\varepsilon}}_i = \Delta \mathbf{Y}_i - \Delta \mathbf{X}_i \hat{\alpha}$. A covariance matrix estimator which is robust to violation of these assumptions is

$$\hat{\mathbf{V}}_1 = (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \hat{\Omega}_2 \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \Delta \mathbf{X}) (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_1^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} \quad (17.93)$$

where

$$\hat{\Omega}_2 = \sum_{i=1}^N \mathbf{Z}'_i \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}'_i \mathbf{Z}_i$$

is a (scaled) cluster-robust estimator of Ω using the one-step residuals.

An asymptotically efficient two-step GMM estimator which allows heteroskedasticity is

$$\hat{\alpha}_2 = (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \Delta \mathbf{Y}). \quad (17.94)$$

Estimator (17.94) is known as the **two-step Arellano-Bond GMM estimator**. An appropriate robust covariance matrix estimator for $\hat{\alpha}_2$ is

$$\hat{\mathbf{V}}_2 = (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \hat{\Omega}_3 \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \Delta \mathbf{X}) (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1} \quad (17.95)$$

where

$$\hat{\Omega}_3 = \sum_{i=1}^N \mathbf{Z}'_i \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}'_i \mathbf{Z}_i$$

is a (scaled) cluster-robust estimator of Ω using the two-step residuals $\hat{\boldsymbol{\varepsilon}}_i = \Delta \mathbf{Y}_i - \Delta \mathbf{X}_i \hat{\alpha}_2$. Asymptotically, $\hat{\mathbf{V}}_2$ is equivalent to

$$\tilde{\mathbf{V}}_2 = (\Delta \mathbf{X}' \mathbf{Z} \hat{\Omega}_2^{-1} \mathbf{Z}' \Delta \mathbf{X})^{-1}. \quad (17.96)$$

The GMM estimator can be iterated until convergence to produce an iterated GMM estimator.

The advantage of the Arellano-Bond estimator over the Anderson-Hsiao estimator is that when $T > p + 2$ the additional (overidentified) moment conditions reduce the asymptotic variance of the estimator and stabilize its performance. The disadvantage is that when T is large using the full set of lags as instruments may cause a “many weak instruments” problem. The advised compromise is to limit the number of lags used as instruments.

The advantage of the one-step Arellano-Bond estimator is that the weight matrix $\hat{\Omega}_1$ does not depend on residuals and is therefore less random than the two-step weight matrix $\hat{\Omega}_2$. This can result in better performance by the one-step estimator in small to moderate samples especially when the errors are approximately homoskedastic. The advantage of the two-step estimator is that it achieves asymptotic efficiency allowing for heteroskedasticity and is thus expected to perform better in large samples with non-homoskedastic errors.

To summarize, the Arellano-Bond estimator applies GMM to the first-differenced equation (17.87) using a set of available lags $Y_{i,t-2}, Y_{i,t-3}, \dots$ as instruments for $\Delta Y_{i,t-1}, \dots, \Delta Y_{i,t-p}$.

The Arellano-Bond estimator may be obtained in Stata using either the `xtabond` or `xtdpd` command. The default setting is the one-step estimator (17.91) and non-robust standard errors (17.92). For the two-step estimator and robust standard errors use the `twostep vce(robust)` options. Reported standard errors in Stata are based on Windmeijer’s (2005) finite-sample correction to the asymptotic estimator (17.96). The robust covariance matrix (17.95) nor the iterated GMM estimator are implemented.

17.40 Weak Instruments

Blundell and Bond (1998) pointed out that the Anderson-Hsiao and Arellano-Bond estimators suffer from weak instruments. This can be seen easiest in the AR(1) model with the Anderson-Hsiao estimator which uses $Y_{i,t-2}$ as an instrument for $\Delta Y_{i,t-1}$. The reduced form equation for $\Delta Y_{i,t-1}$ is

$$\Delta Y_{i,t-1} = Y_{i,t-2}\gamma + v_{it}.$$

The reduced form coefficient γ is defined by projection. Using $\Delta Y_{i,t-1} = (\alpha - 1)Y_{i,t-2} + u_i + \varepsilon_{i,t-1}$ and $\mathbb{E}[Y_{i,t-2}\varepsilon_{i,t-1}] = 0$ we calculate that

$$\gamma = \frac{\mathbb{E}[Y_{i,t-2}\Delta Y_{i,t-1}]}{\mathbb{E}[Y_{i,t-2}^2]} = (\alpha - 1) + \frac{\mathbb{E}[Y_{i,t-2}u_i]}{\mathbb{E}[Y_{i,t-2}^2]}.$$

Assuming stationarity so that (17.83) holds,

$$\mathbb{E}[Y_{i,t-2}u_i] = \mathbb{E}\left[\left(\frac{u_i}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,t-2-j}\right)u_i\right] = \frac{\sigma_u^2}{1-\alpha}$$

and

$$\mathbb{E}[Y_{i,t-2}^2] = \mathbb{E}\left[\left(\frac{u_i}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,t-2-j}\right)^2\right] = \frac{\sigma_u^2}{(1-\alpha)^2} + \frac{\sigma_\varepsilon^2}{(1-\alpha^2)}$$

where $\sigma_u^2 = \mathbb{E}[u_i^2]$ and $\sigma_\varepsilon^2 = \mathbb{E}[\varepsilon_{it}^2]$. Using these expressions and a fair amount of algebra, Blundell and Bond (1998) found that the reduced form coefficient equals

$$\gamma = (\alpha - 1) \left(\frac{k}{k + \sigma_u^2 / \sigma_\varepsilon^2} \right) \quad (17.97)$$

where $k = (1 - \alpha) / (1 + \alpha)$.

The Anderson-Hsiao instrument $Y_{i,t-2}$ is weak if γ is close to zero. From (17.97) we see that $\gamma = 0$ when either $\alpha = 1$ (a unit root) or $\sigma_u^2/\sigma_\varepsilon^2 = \infty$ (the idiosyncratic effect is small relative to the individual-specific effect). In either case the coefficient α is not identified. We know from our earlier study of the weak instruments problem (Section 12.36) that when γ is close to zero then α is weakly identified and the estimators will perform poorly. This means that when the autoregressive coefficient α is large or the individual-specific effect dominates the idiosyncratic effect these estimators will be weakly identified, have poor performance, and conventional inference methods will be misleading. Since the value of α and the relative variances are unknown *a priori* this means that we should generically treat this class of estimators as weakly identified.

An alternative estimator which has improved performance is discussed in Section 17.42.

17.41 Dynamic Panels with Predetermined Regressors

The assumption that regressors are strictly exogenous is restrictive. A less restrictive assumption is that the regressors are predetermined. Dynamic panel methods can be modified to handle predetermined regressors by using their lags as instruments.

Definition 17.2 The regressor X_{it} is **predetermined** for the error ε_{it} if

$$\mathbb{E}[X_{i,t-s}\varepsilon_{it}] = 0 \quad (17.98)$$

for all $s \geq 0$.

The difference between strictly exogenous and predetermined regressors is that for the former (17.98) holds for all s not just $s \geq 0$. One way of interpreting a regression model with predetermined regressors is that the model is a projection on the complete past history of the regressors.

Under (17.98), leads of X_{it} can be correlated with ε_{it} , that is $\mathbb{E}[X_{i,t+s}\varepsilon_{it}] \neq 0$ for $s \geq 1$, or equivalently X_{it} can be correlated with lags of ε_{ij} , that is $\mathbb{E}[X_{it}\varepsilon_{i,t-s}] \neq 0$ for $s \geq 1$. This means that X_{it} can respond dynamically to past values of Y_{it} , as in, for example, an unrestricted vector autoregression.

Consider the differenced equation (17.87)

$$\Delta Y_{it} = \alpha_1 \Delta Y_{i,t-1} + \alpha_2 \Delta Y_{i,t-2} + \cdots + \alpha_p \Delta Y_{i,t-p} + \Delta X'_{it} \beta + \Delta \varepsilon_{it}.$$

When the regressors are predetermined but not strictly exogenous, X_{it} and ε_{it} are uncorrelated but ΔX_{it} and $\Delta \varepsilon_{it}$ are correlated. To see this,

$$\begin{aligned} \mathbb{E}[\Delta X_{it} \Delta \varepsilon_{it}] &= \mathbb{E}[X_{it} \varepsilon_{it}] - \mathbb{E}[X_{i,t-1} \varepsilon_{it}] - \mathbb{E}[X_{it} \varepsilon_{i,t-1}] + \mathbb{E}[X_{i,t-1} \varepsilon_{i,t-1}] \\ &= -\mathbb{E}[X_{it} \varepsilon_{i,t-1}] \neq 0. \end{aligned}$$

This means that if we treat ΔX_{it} as exogenous the coefficient estimates will be biased.

To solve the correlation problem we can use instruments for ΔX_{it} . A valid instrument is $X_{i,t-1}$ because it is generally correlated with ΔX_{it} yet uncorrelated with $\Delta \varepsilon_{it}$. Indeed, for any $s \geq 1$

$$\mathbb{E}[X_{i,t-s} \Delta \varepsilon_{it}] = \mathbb{E}[X_{i,t-s} \varepsilon_{it}] - \mathbb{E}[X_{i,t-s} \varepsilon_{i,t-1}] = 0.$$

Consequently, Arellano and Bond (1991) recommend the instrument set $(X_{i1}, X_{i2}, \dots, X_{i,t-1})$. When the number of time periods is large it is advised to limit the number of instrument lags to avoid the many weak instruments problem.

Algebraically, GMM estimation is the same as the estimators described in Section 17.39, except that the instrument matrix (17.89) is modified to

$$\mathbf{Z}_i = \begin{bmatrix} [Y_{i1}, \dots, Y_{ip}, X'_{i1}, \dots, X'_{i,p+1}] & 0 & 0 \\ 0 & [Y_{i1}, \dots, Y_{i,p+1}, X'_{i1}, \dots, X'_{i,p+2}] & 0 \\ & & \ddots \\ 0 & 0 & [Y_{i1}, \dots, Y_{i,T-2}, X'_{i1}, \dots, X'_{i,T-1}] \end{bmatrix}. \quad (17.99)$$

To understand how the model is identified we examine the reduced form equation for the regressor. For $t = p + 2$ and using the first lag as an instrument the reduced form is

$$\Delta X_{it} = \gamma_1 Y_{i,t-2} + \Gamma_2 X_{i,t-1} + \zeta_{it}.$$

The model is identified if Γ_2 is full rank. This is valid (in general) when X_{it} is stationary. Identification fails, however, when X_{it} has a unit root. This indicates that the model will be weakly identified when the predetermined regressors are highly persistent.

The method generalizes to handle multiple lags of the predetermined regressors. To see this, write the model explicitly as

$$Y_{it} = \alpha_1 Y_{i,t-1} + \dots + \alpha_p Y_{i,t-p} + X'_{it} \beta_1 + \dots + X'_{i,t-q} \beta_q + u_i + \varepsilon_{it}.$$

In first differences the model is

$$\Delta Y_{it} = \alpha_1 \Delta Y_{i,t-1} + \dots + \alpha_p \Delta Y_{i,t-p} + \Delta X'_{it} \beta_1 + \dots + \Delta X'_{i,t-q} \beta_q + \Delta \varepsilon_{it}.$$

A sufficient set of instruments for the regressors are $(X_{i,t-1}, \Delta X_{i,t-1}, \dots, \Delta X_{i,t-q})$ or equivalently $(X_{i,t-1}, X_{i,t-2}, \dots, X_{i,t-q-1})$.

In many cases it is more reasonable to assume that $X_{i,t-1}$ is predetermined but not X_{it} , because X_{it} and ε_{it} may be endogenous. This, for example, is the standard assumption in vector autoregressions. In this case the estimation method is modified to use the instruments $(X_{i,t-2}, X_{i,t-3}, \dots, X_{i,t-q-1})$. While this weakens the exogeneity assumption it also weakens the instrument set as now the reduced form uses the second lag $X_{i,t-2}$ to predict ΔX_{it} .

The advantage obtained by treating a regressor as predetermined (rather than strictly exogenous) is that it is a substantial relaxation of the dynamic assumptions. Otherwise the parameter estimates will be inconsistent due to endogeneity.

The major disadvantage of treating a regressor as predetermined is that it substantially reduces the strength of identification especially when the predetermined regressors are highly persistent.

In Stata the `xtabond` command by default treats independent regressors as strictly exogenous. To treat the regressors as predetermined use the option `pre`. By default all regressor lags are used as instruments, but the number can be limited if specified.

17.42 Blundell-Bond Estimator

Arellano and Bover (1995) and Blundell and Bond (1998) introduced a set of orthogonality conditions which reduce the weak instrument problem discussed in the Section 17.40 and improve performance in finite samples.

Consider the levels AR(1) model with no regressors (17.82). Recall, least squares (pooled) regression is inconsistent because the regressor $Y_{i,t-1}$ is correlated with the error u_i . This raises the question: Is there an instrument Z_{it} which solves this problem in the sense that Z_{it} is correlated with $Y_{i,t-1}$ yet

uncorrelated with $u_{it} + \varepsilon_{it}$? Blundell-Bond propose the instrument $\Delta Y_{i,t-1}$. Clearly, $\Delta Y_{i,t-1}$ and $Y_{i,t-1}$ are correlated so $\Delta Y_{i,t-1}$ satisfies the relevance condition. Also, $\Delta Y_{i,t-1}$ is uncorrelated with the idiosyncratic error ε_{it} when the latter is serially uncorrelated. Thus the key to the Blundell-Bond instrument is whether or not

$$\mathbb{E}[\Delta Y_{i,t-1} u_i] = 0. \quad (17.100)$$

Blundell and Bond (1998) show that a sufficient condition for (17.100) is

$$\mathbb{E}\left[\left(Y_{i1} - \frac{u_i}{1-\alpha}\right) u_i\right] = 0. \quad (17.101)$$

Recall that $u_i/(1-\alpha)$ is the conditional mean of Y_{i1} under stationarity. Condition (17.101) states that the deviation of the initial condition Y_{i1} from this conditional mean is uncorrelated with the individual effect u_i . Condition (17.101) is implied by stationarity but is somewhat weaker.

To see that (17.101) implies (17.100), by applying recursion to (17.87) we find that

$$\Delta Y_{i,t-1} = \alpha^{t-3} \Delta Y_{i2} + \sum_{j=0}^{t-3} \alpha^j \Delta \varepsilon_{i,t-1-j}.$$

Also,

$$\Delta Y_{i2} = (\alpha - 1) Y_{i1} + u_i + \varepsilon_{i2} = (\alpha - 1) \left(Y_{i1} - \frac{u_i}{1-\alpha} \right) + \varepsilon_{i2}.$$

Hence

$$\begin{aligned} \mathbb{E}[\Delta Y_{i,t-1} u_i] &= \mathbb{E}\left[\left(\alpha^{t-3} (\alpha - 1) \left(Y_{i1} - \frac{u_i}{1-\alpha} \right) + \alpha^{t-3} \varepsilon_{i2} + \sum_{j=0}^{t-3} \alpha^j \Delta \varepsilon_{i,t-1-j} \right) u_i\right] \\ &= \alpha^{t-3} (\alpha - 1) \mathbb{E}\left[\left(Y_{i1} - \frac{u_i}{1-\alpha} \right) u_i\right] \\ &= 0 \end{aligned}$$

under (17.101), as claimed.

Now consider the full model (17.81) with predetermined regressors. Consider the assumption that the regressors have constant correlation with the individual effect

$$\mathbb{E}[X_{it} u_i] = \mathbb{E}[X_{is} u_i]$$

for all s . This implies

$$\mathbb{E}[\Delta X_{it} u_i] = 0 \quad (17.102)$$

which means that the differenced predetermined regressors ΔX_{it} can also be used as instruments for the level equation.

Using (17.100) and (17.102) Blundell and Bond propose the following moment conditions for GMM estimation

$$\mathbb{E}[\Delta Y_{i,t-1} (Y_{it} - \alpha_1 Y_{i,t-1} - \dots - \alpha_p Y_{i,t-p} - X'_{it} \beta)] = 0 \quad (17.103)$$

$$\mathbb{E}[\Delta X_{i,t} (Y_{it} - \alpha_1 Y_{i,t-1} - \dots - \alpha_p Y_{i,t-p} - X'_{it} \beta)] = 0 \quad (17.104)$$

for $t = p+2, \dots, T$. Notice that these are for the levels (undifferenced) equation while the Arellano-Bond (17.90) moments are for the differenced equation (17.87). We can write (17.103)-(17.104) in vector notation if we set $\mathbf{Z}_{2i} = \text{diag}(\Delta Y_{i2}, \dots, \Delta Y_{iT-1}, \Delta X_{i3}, \dots, \Delta X_{iT})$. Then (17.103)-(17.104) equals

$$\mathbb{E}[\mathbf{Z}_{2i} (\mathbf{Y}_i - \mathbf{X}_i \theta)] = 0. \quad (17.105)$$

Blundell and Bond proposed combining the ℓ Arellano-Bond moments with the levels moments. This can be done by stacking the moment conditions (17.90) and (17.105). Recall from Section 17.39 the variables ΔY_i , ΔX_i , and Z_i . Define the stacked variables $\bar{Y}_i = (\Delta Y_i', Y_i')'$, $\bar{X}_i = (\Delta X_i', X_i')'$ and $\bar{Z}_i = \text{diag}(Z_i, Z_{2i})$. The stacked moment conditions are

$$\mathbb{E}[\bar{Z}_i(\bar{Y}_i - \bar{X}_i\theta)] = 0.$$

The Blundell-Bond estimator is found by applying GMM to this equation. They call this a systems GMM estimator. Let \bar{Y} , \bar{X} , and \bar{Z} denote \bar{Y}_i , \bar{X}_i , and \bar{Z}_i stacked into matrices. Define $\bar{H} = \text{diag}(\mathbf{H}, \mathbf{I}_{T-2})$ where \mathbf{H} is from (17.31) and set

$$\hat{\Omega}_1 = \sum_{i=1}^N \bar{Z}_i' \bar{H} \bar{Z}_i.$$

The **Blundell-Bond one-step GMM estimator** is

$$\hat{\theta}_1 = (\bar{X}' \bar{Z} \hat{\Omega}_1^{-1} \bar{Z}' \bar{X})^{-1} (\bar{X}' \bar{Z} \hat{\Omega}_1^{-1} \bar{Z}' \bar{Y}). \quad (17.106)$$

The systems residuals are $\hat{\epsilon}_i = \bar{Y}_i - \bar{X}_i \hat{\theta}_1$. A robust covariance matrix estimator is

$$\hat{V}_1 = (\bar{X}' \bar{Z} \hat{\Omega}_1^{-1} \bar{Z}' \bar{X})^{-1} (\bar{X}' \bar{Z} \hat{\Omega}_1^{-1} \bar{Z}' \hat{\Omega}_2 \bar{Z} \hat{\Omega}_1^{-1} \bar{Z}' \bar{X}) (\bar{X}' \bar{Z} \hat{\Omega}_1^{-1} \bar{Z}' \bar{X})^{-1} \quad (17.107)$$

where

$$\hat{\Omega}_2 = \sum_{i=1}^N \bar{Z}_i' \hat{\epsilon}_i \hat{\epsilon}_i' \bar{Z}_i.$$

The **Blundell-Bond two-step GMM estimator** is

$$\hat{\theta}_2 = (\bar{X}' \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \bar{X})^{-1} (\bar{X}' \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \bar{Y}). \quad (17.108)$$

The two-step systems residuals are $\hat{\epsilon}_i = \bar{Y}_i - \bar{X}_i \hat{\theta}_2$. A robust covariance matrix estimator is

$$\hat{V}_2 = (\bar{X}' \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \bar{X})^{-1} (\bar{X}' \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \hat{\Omega}_3 \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \bar{X}) (\bar{X}' \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \bar{X})^{-1} \quad (17.109)$$

where

$$\hat{\Omega}_3 = \sum_{i=1}^N \bar{Z}_i' \hat{\epsilon}_i \hat{\epsilon}_i' \bar{Z}_i.$$

Asymptotically, \hat{V}_2 is equivalent to

$$\tilde{V}_2 = (\bar{X}' \bar{Z} \hat{\Omega}_2^{-1} \bar{Z}' \bar{X})^{-1}. \quad (17.110)$$

The GMM estimator can be iterated until convergence to produce an iterated GMM estimator.

Simulation experiments reported in Blundell and Bond (1998) indicate that their systems GMM estimator performs substantially better than the Arellano-Bond estimator, especially when α is close to one or the variance ratio $\sigma_u^2/\sigma_\varepsilon^2$ is large. The explanation is that the orthogonality condition (17.103) does not suffer the weak instrument problem in these cases.

The advantage of the Blundell-Bond estimator is that the added orthogonality condition (17.103) greatly improves performance relative to the Arellano-Bond estimator when the latter is weakly identified. A disadvantage of the Blundell-Bond estimator is that their orthogonality condition is justified by a stationarity condition (17.101) and violation of the latter may induce estimation bias.

The advantages and disadvantages of the one-step versus two-step Blundell-Bond estimators are the same as described for the Arellano-Bond estimator as described in Section 17.39. Also as described there

when T is large it may be desired to limit the number of lags to use as instruments in order to avoid the many weak instruments problem.

The Blundell-Bond estimator may be obtained in Stata using either the `xtdpdsys` or `xtdpd` command. The default setting is the one-step estimator (17.106) and non-robust standard errors. For the two-step estimator and robust standard errors use the `twostep vce(robust)` options. Stata standard errors are Windmeijer's (2005) finite-sample correction to the asymptotic estimate (17.110). The robust covariance matrix estimator (17.109) nor the iterated GMM estimator are implemented.

17.43 Forward Orthogonal Transformation

Arellano and Bover (1995) proposed an alternative transformation which eliminates the individual-specific effect and may have advantages in dynamic panel models. The **forward orthogonal transformation** is

$$Y_{it}^* = c_{it} \left(Y_{it} - \frac{1}{T_i - t} (Y_{i,t+1} + \dots + Y_{iT_i}) \right) \quad (17.111)$$

where $c_{it}^2 = (T_i - t) / (T_i - t + 1)$. This can be applied to all but the final observation (which is lost). Essentially, Y_{it}^* subtracts from Y_{it} the average of the remaining values and then rescales so that the variance is constant under the assumption of homoskedastic errors. The transformation (17.111) was originally proposed for time-series observations by Hayashi and Sims (1983).

At the level of the individual this can be written as $\mathbf{Y}_i^* = \mathbf{A}_i \mathbf{Y}_i$ where \mathbf{A}_i is the $(T_i - 1) \times T_i$ orthogonal deviation operator

$$\mathbf{A}_i = \text{diag} \left(\sqrt{\frac{T_i - 1}{T_i}}, \dots, \sqrt{\frac{1}{2}} \right) \begin{bmatrix} 1 & -\frac{1}{T_i - 1} & -\frac{1}{T_i - 1} & \dots & -\frac{1}{T_i - 1} & -\frac{1}{T_i - 1} & -\frac{1}{T_i - 1} \\ 0 & 1 & -\frac{1}{T_i - 2} & \dots & -\frac{1}{T_i - 2} & -\frac{1}{T_i - 2} & -\frac{1}{T_i - 2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

Important properties of the matrix \mathbf{A}_i are that $\mathbf{A}_i \mathbf{1}_i = 0$ (so it eliminates individual effects), $\mathbf{A}_i' \mathbf{A}_i = \mathbf{M}_i$, and $\mathbf{A}_i \mathbf{A}_i' = \mathbf{I}_{T_i - 1}$. These can be verified by direct multiplication.

Applying the transformation \mathbf{A}_i to (17.81) we obtain

$$Y_{it}^* = \alpha_1 Y_{i,t-1}^* + \dots + \alpha_p Y_{i,t-p}^* + X_{it}^{*'} \beta + \varepsilon_{it}^*. \quad (17.112)$$

for $t = p + 1, \dots, T - 1$. This is equivalent to first differencing (17.87) when $T = 3$ but differs for $T > 3$.

What is special about the transformed equation (17.112) is that under the assumption that ε_{it} are serially uncorrelated and homoskedastic the error vector $\boldsymbol{\varepsilon}_i^*$ has variance $\sigma_\varepsilon^2 \mathbf{A}_i \mathbf{A}_i' = \sigma_\varepsilon^2 \mathbf{I}_{T_i - 1}$. This means that $\boldsymbol{\varepsilon}_i^*$ has the same covariance structure as $\boldsymbol{\varepsilon}_i$. Thus the orthogonal transformation operator eliminates the fixed effect while preserving the covariance structure. This is in contrast to (17.87) which has serially correlated errors $\Delta \varepsilon_{it}$.

The transformed error ε_{it}^* is a function of $\varepsilon_{it}, \varepsilon_{it+1}, \dots, \varepsilon_{iT}$. Thus valid instruments are $Y_{it-1}, Y_{it-2}, \dots$. Using the instrument matrix \mathbf{Z}_i from (17.89) in the case of strictly exogenous regressors or (17.99) with predetermined regressors the ℓ moment conditions can be written using matrix notation as

$$\mathbb{E} [\mathbf{Z}_i' (\mathbf{Y}_i^* - \mathbf{X}_i^* \theta)] = 0. \quad (17.113)$$

Define the $\ell \times \ell$ covariance matrix

$$\Omega = \mathbb{E} [\mathbf{Z}_i' \boldsymbol{\varepsilon}_i^* \boldsymbol{\varepsilon}_i^{*'} \mathbf{Z}_i].$$

If the errors ε_{it} are conditionally homoskedastic then $\Omega = \mathbb{E} [Z_i' Z_i] \sigma_\varepsilon^2$. Thus an asymptotically efficient GMM estimator is 2SLS applied to the orthogonalized equation using Z_i as an instrument. In matrix notation,

$$\hat{\theta}_1 = \left(X^{*'} Z (Z' Z)^{-1} Z' X^* \right)^{-1} Y^*.$$

This is the one-step GMM estimator.

Given the residuals $\hat{\varepsilon}_i = Y_i^* - X_i^* \hat{\theta}_1$ the two-step GMM estimator which is robust to heteroskedasticity and arbitrary serial correlation is

$$\hat{\theta}_2 = \left(X^{*'} Z \hat{\Omega}_2^{-1} Z' X^* \right)^{-1} \left(X^{*'} Z \hat{\Omega}_2^{-1} Z' Y^* \right)$$

where

$$\hat{\Omega}_2 = \sum_{i=1}^N Z_i' \hat{\varepsilon}_i \hat{\varepsilon}_i' Z_i.$$

Standard errors for $\hat{\theta}_1$ and $\hat{\theta}_2$ can be obtained using cluster-robust methods.

Forward orthogonalization may have advantages over first differencing. First, the equation errors in (17.112) have a scalar covariance structure under i.i.d. idiosyncratic errors which is expected to improve estimation precision. It also implies that the one-step estimator is 2SLS rather than GMM. Second, while there has not been a formal analysis of the weak instrument properties of the estimators after forward orthogonalization it appears that if $T > p + 2$ the method is less affected by weak instruments than first differencing. The disadvantages of forward orthogonalization are that it treats early observations asymmetrically from late observations, it is less thoroughly studied than first differencing, and is not available with several popular estimation methods.

The Stata command `xtdpd` includes forward orthogonalization as an option but not when levels (Blundell-Bond) instruments are included or if there are gaps in the data. An alternative is the downloadable Stata package `xtabond2`.

17.44 Empirical Illustration

We illustrate the dynamic panel methods with the investment model (17.3). Estimates from two models are presented in Table 17.3. Both are estimated by Blundell-Bond two-step GMM with lags 2 through 6 as instruments, a cluster-robust weight matrix, and clustered standard errors.

The first column presents estimates of an AR(2) model. The estimates show that the series has a moderate amount of positive serial correlation but appears to be well modeled as an AR(1) as the AR(2) coefficient is close to zero. This pattern of serial correlation is consistent with the presence of investment projects which span two years.

The second column presents estimates of the dynamic version of the investment regression (17.3) excluding the trading indicator. Two lags are included of the dependent variable and each regressor. The regressors are treated as predetermined in contrast to the fixed effects regressions which treated the regressors as strictly exogenous. The regressors are not contemporaneous with the dependent variable but lagged one and two periods. This is done so that they are valid predetermined variables. Contemporaneous variables are likely endogenous so should not be treated as predetermined.

The estimates in the second column of Table 17.3 complement the earlier results. The evidence shows that investment has a moderate degree of serial dependence, is positively related to the first lag of Q , and is negatively related to lagged debt. Investment appears to be positively related to change in cash flow, rather than the level. Thus an increase in cash flow in year $t - 1$ leads to investment in year t .

Table 17.3: Estimates of Dynamic Investment Equation

	AR(2)	AR(2) with Regressors
I_{it-1}	0.3191 (0.0172)	0.2519 (0.0220)
I_{it-2}	0.0309 (0.0112)	0.0137 (0.0125)
Q_{it-1}		0.0018 (0.0007)
Q_{it-2}		-0.0000 (0.0003)
D_{it-1}		-0.0154 (0.0058)
D_{it-2}		-0.0043 (0.0054)
CF_{it-1}		0.0400 (0.0091)
CF_{it-2}		-0.0290 (0.0051)

Two-step GMM estimates. Cluster-robust standard errors in parenthesis.

All regressions include time effects. GMM instruments include lags 2 through 6.

17.45 Exercises

Exercise 17.1

(a) Show (17.11) and (17.12).

(b) Show (17.13).

Exercise 17.2 Is $\mathbb{E}[\varepsilon_{it} | X_{it}] = 0$ sufficient for $\hat{\beta}_{fe}$ to be unbiased for β ? Explain why or why not.

Exercise 17.3 Show that $\text{var}[\dot{X}_{it}] \leq \text{var}[X_{it}]$.

Exercise 17.4 Show (17.24).

Exercise 17.5 Show (17.28).

Exercise 17.6 Show that when $T = 2$ the differenced estimator equals the fixed effects estimator.

Exercise 17.7 In Section 17.14 it is described how to estimate the individual-effect variance σ_u^2 using the between residuals. Develop an alternative estimator of σ_u^2 only using the fixed effects error variance $\hat{\sigma}_\varepsilon^2$ and the levels error variance $\hat{\sigma}_\varepsilon^2 = n^{-1} \sum_{i=1}^N \sum_{t \in S_i} \hat{e}_{it}^2$ where $\hat{e}_{it} = Y_{it} - X'_{it} \hat{\beta}_{fe}$ are computed from the levels variables.

Exercise 17.8 Verify that $\hat{\sigma}_\varepsilon^2$ defined in (17.37) is unbiased for σ_ε^2 under (17.18), (17.25) and (17.26).

Exercise 17.9 Develop a version of Theorem 17.2 for the differenced estimator $\hat{\beta}_\Delta$. Can you weaken Assumption 17.2.3? State an appropriate version which is sufficient for asymptotic normality.

Exercise 17.10 Show (17.57).

Exercise 17.11

- (a) For $\hat{\sigma}_i^2$ defined in (17.59) show $\mathbb{E}[\hat{\sigma}_i^2 | \mathbf{X}_i] = \bar{\sigma}_i^2$.
- (b) For $\tilde{\mathbf{V}}_{fe}$ defined in (17.58) show $\mathbb{E}[\tilde{\mathbf{V}}_{fe} | \mathbf{X}] = \mathbf{V}_{fe}$.

Exercise 17.12

- (a) Show (17.61).
- (b) Show (17.62).
- (c) For $\tilde{\mathbf{V}}_{fe}$ defined in (17.60) show $\mathbb{E}[\tilde{\mathbf{V}}_{fe} | \mathbf{X}] = \mathbf{V}_{fe}$.

Exercise 17.13 Take the fixed effects model $Y_{it} = X_{it}\beta_1 + X_{it}^2\beta_2 + u_i + \varepsilon_{it}$. A researcher estimates the model by first obtaining the within transformed \dot{Y}_{it} and \dot{X}_{it} and then regressing \dot{Y}_{it} on \dot{X}_{it} and \dot{X}_{it}^2 . Is the correct estimation method? If not, describe the correct fixed effects estimator.

Exercise 17.14 In Section 17.33 verify that in the just-identified case the 2SLS estimator $\hat{\beta}_{2sls}$ simplifies as claimed: $\hat{\beta}_1$ and $\hat{\beta}_2$ are the fixed effects estimator. $\hat{\gamma}_1$ and $\hat{\gamma}_2$ equal the 2SLS estimator from a regression of $\hat{\mathbf{u}}$ on \mathbf{Z}_1 and \mathbf{Z}_2 using $\bar{\mathbf{X}}_1$ as an instrument for \mathbf{Z}_2 .

Exercise 17.15 In this exercise you will replicate and extend the empirical work reported in Arellano and Bond (1991) and Blundell and Bond (1998). Arellano-Bond gathered a dataset of 1031 observations from an unbalanced panel of 140 U.K. companies for 1976-1984 and is in the datafile AB1991 on the textbook webpage. The variables we will be using are log employment (N), log real wages (W), and log capital (K). See the description file for definitions.

- (a) Estimate the panel AR(1) $K_{it} = \alpha K_{it-1} + u_i + v_t + \varepsilon_{it}$ using Arellano-Bond one-step GMM with clustered standard errors. Note that the model includes year fixed effects.
- (b) Re-estimate using Blundell-Bond one-step GMM with clustered standard errors.
- (c) Explain the difference in the estimates.

Exercise 17.16 This exercise uses the same dataset as the previous question. Blundell and Bond (1998) estimated a dynamic panel regression of log employment N on log real wages W and log capital K . The following specification¹ used the Arellano-Bond one-step estimator, treating $W_{i,t-1}$ and $K_{i,t-1}$ as predetermined.

$$N_{it} = \begin{matrix} .7075 \\ (.0842) \end{matrix} N_{i,t-1} - \begin{matrix} .7088 \\ (.1171) \end{matrix} W_{it} + \begin{matrix} .5000 \\ (.1113) \end{matrix} W_{i,t-1} + \begin{matrix} .4660 \\ (.1010) \end{matrix} K_{it} - \begin{matrix} .2151 \\ (.0859) \end{matrix} K_{i,t-1}. \quad (17.114)$$

This equation also included year dummies and the standard errors are clustered.

¹Blundell and Bond (1998), Table 4, column 3.

- (a) Estimate (17.114) using the Arellano-Bond one-step estimator treating W_{it} and K_{it} as strictly exogenous.
- (b) Estimate (17.114) treating $W_{i,t-1}$ and $K_{i,t-1}$ as predetermined to verify the results in (17.114). What is the difference between the estimates treating the regressors as strictly exogenous versus predetermined?
- (c) Estimate the equation using the Blundell-Bond one-step systems GMM estimator.
- (d) Interpret the coefficient estimates viewing (17.114) as a firm-level labor demand equation.
- (e) Describe the impact on the standard errors of the Blundell-Bond estimates in part (c) if you forget to use clustering. (You do not have to list all the standard errors, but describe the magnitude of the impact.)

Exercise 17.17 Use the datafile Invest1993 on the textbook webpage. You will be estimating the panel AR(1) $D_{it} = \alpha D_{i,t-1} + u_i + \varepsilon_{it}$ for $D = \text{debt}/\text{assets}$ (this is *debt*_a in the datafile). See the description file for definitions.

- (a) Estimate the model using Arellano-Bond twostep GMM with clustered standard errors.
- (b) Re-estimate using Blundell-Bond twostep GMM.
- (c) Experiment with your results, trying twostep versus onestep, AR(1) versus AR(2), number of lags used as instruments, and classical versus robust standard errors. What makes the most difference for the coefficient estimates? For the standard errors?

Exercise 17.18 Use the datafile Invest1993 on the textbook webpage. You will be estimating the model

$$D_{it} = \alpha D_{i,t-1} + \beta_1 I_{i,t-1} + \beta_2 Q_{i,t-1} + \beta_3 CF_{i,t-1} + u_i + \varepsilon_{it}.$$

The variables are *debt*_a, *inva*, *vala*, and *cfa* in the datafile. See the description file for definitions.

- (a) Estimate the above regression using Arellano-Bond two-step GMM with clustered standard errors treating all regressors as predetermined.
- (b) Re-estimate using Blundell-Bond two-step GMM treating all regressors as predetermined.
- (c) Experiment with your results, trying two-step versus one-step, number of lags used as instruments, and classical versus robust standard errors. What makes the most difference for the coefficient estimates? For the standard errors?