11 Panel Data Analysis

This chapter deals with econometric methods for panel data, i.e. observations z_{it}, available for cross-sectional units $i = 1 : n$ and time periods $t = 1 : T$. Typically, such data is analyzed from a cross-sectional perspective, i.e. in a setting where we think of n as being large relative to T. As opposed to cross-sectional analyses, the existence of the time dimension allows us to account for unit-specific unobserved heterogeneity that is fixed over time. As opposed to multivariate time series data, n is typically so large that it is impossible (and not desirable) to specify all parameters to be unit-specific, as done in VARs for example. Throughout this chapter, it is assumed that observations z_{it} are i.i.d. across i.

Under a dataset with large n and small T, the presence of unit-specific parameters leads to the incidental parameters problem (IPP); standard estimators are inconsistent unless we let $T \to \infty$, which is, however, a poor approximation of estimators' finite sample behavior with such datasets. The IPP is discussed in greater detail in Section 11.1.

Econometricians have come up with several ways to deal with the IPP. These can be split into three strands. "Random effects" methods (Section 11.2) propose assumptions that allow us to consistently estimate the remaining, homogeneous parameters by pushing unit-specific parameters into the residual. "Fixed effects" methods (Section 11.3) show how we can consistently estimate homogeneous parameters by using the time dimension to transform the model of interest and get rid of unit-specific parameters. Finally, "correlated random effects" methods (Section 11.4) propose consistent and efficient estimators by modeling unit-specific parameters and their relation to observed covariates. This is related to Bayesian methods for panel data, discussed in Section 11.5.

Note that the IPP arises analogously under the presence of time-specific parameters in an analysis where T is large and n is small, though this case is less often encountered in practice. The methods discussed in this chapter can analogously be applied in that case.

11.1 Incidental Parameters Problem

Suppose we are interested in estimating the following panel data regression:

$$y_{it} = \alpha + x_{it}'\beta + u_{it} \ , \quad \mathbb{E}[u_{it}x_{it}] = 0 \ , \ \ \mathbb{V}[u_{it}|x_{it}] = \sigma^2 \ ,$$

for observations $i = 1:n$ and $t = 1:T$. Omitting the explicit distinction between intercept and slope parameters, we can write this model also as

$$y_{it} = \tilde{x}_{it}'\tilde{\beta} + u_{it} \ , \quad \text{with} \quad \tilde{x}_{it} = (1, x_{it}')' \ , \ \ \tilde{\beta} = (\alpha, \beta')' \ .$$

Stacking all T regressions pertaining to the same unit i, we can write

$$y_i = \tilde{X}_i\tilde{\beta} + u_i \ ,$$

where y_i is a $T \times 1$ vector that stacks $\{y_{it}\}_{t=1}^T$, and analogously for the $T \times 1$ vector u_i and the $T \times k$ matrix $\tilde{X}_i$. This notation emphasizes the similarity to the cross-sectional case. As before, we are estimating n linear regressions. In contrast to before, the outcome variable y_i and error term u_i are vectors, not scalars, and $\tilde{X}_i$ is a matrix, not a vector.

We can estimate $\tilde{\beta}$ using OLS, by minimizing the sum of squared residuals

$$\min_{\tilde{\beta}} \sum_{i=1}^n \sum_{t=1}^T u_{it}^2 = \min_{\tilde{\beta}} \sum_{i=1}^n u_i'u_i = \min_{\tilde{\beta}} \sum_{i=1}^n (y_i - \tilde{X}_i\tilde{\beta})'(y_i - \tilde{X}_i\tilde{\beta}) \ .$$

This leads to the so-called "pooled" OLS (POLS) estimator

$$\hat{\tilde{\beta}}_{POLS} = \left(\sum_{i=1}^n \tilde{X}_i'\tilde{X}_i\right)^{-1} \sum_{i=1}^n \tilde{X}_i'y_i = \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it}y_{it} \ .$$

It is referred to as "pooled" because, relative to a purely cross-sectional analysis ($T = 1$), we do not make any particular use of the time dimension except that we "pool" all observations $i = 1:n$ and $t = 1:T$, treating the additional time periods in the same way as we would treat additional observations in the cross-sectional dimension.[1]

Suppose we are interested in the typical panel data-case where n is large and T is relatively

[1] The same POLS estimator arises as the ML estimator under the assumption $u_{it} \overset{i.i.d.}{\sim} N(0, \sigma^2)$. More generally, though, under $u_i \overset{i.i.d.}{\sim} N(0, \Sigma)$, we get $\hat{\tilde{\beta}}_{ML} = \left(\sum_{i=1}^n \tilde{X}_i'\Sigma^{-1}\tilde{X}_i\right)^{-1} \sum_{i=1}^n \tilde{X}_i'\Sigma^{-1}y_i$. Only if we assume that the errors are independent over time, i.e. $\Sigma = \sigma^2 I$, does the ML estimator coincide with the POLS estimator.

small. Then we would like to know the asymptotic behavior of our estimator under $N \to \infty$ and T fixed. Under the assumption that observations are i.i.d. in the cross-sectional dimension, we can apply the WLLN to get

$$\hat{\tilde{\beta}}_{POLS} - \tilde{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i'\tilde{X}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i'u_i \xrightarrow{p} \mathbb{E}[\tilde{X}_i'\tilde{X}_i]^{-1}\mathbb{E}[\tilde{X}_i'u_i].$$

Note that

$$\mathbb{E}[\tilde{X}_i'u_i] = \begin{bmatrix} \sum_{t=1}^{T}\mathbb{E}[\tilde{x}_{it,1}u_{it}] \\ \vdots \\ \sum_{t=1}^{T}\mathbb{E}[\tilde{x}_{it,k}u_{it}] \end{bmatrix} ,$$

and hence $\hat{\tilde{\beta}}_{POLS}$ is consistent provided that regressors in x_{it} and u_{it} are contemporaneously uncorrelated: $\mathbb{E}[x_{it}u_{it}] = 0 \; \forall \; t$. The regressors are allowed to be correlated with past (and future) u_{it}. This occurs, for example, if there is a feedback loop by which $y_{i,t-1}$ affects x_{it}.[2]

Note that we could also show consistency under $T \to \infty$ or $(n, T) \to \infty$ using the asymptotic analysis of time series data from Chapter 8. However, most panel data applications have a large n and small T dimension, and asymptotic analyses are supposed to provide useful guidance for an estimator's finite sample behavior. As a result, standard panel data asymptotics feature T fixed and $n \to \infty$. Unless otherwise specified, this is the type of asymptotics referred to in this chapter.

The above model is fully homogeneous; none of the parameters are specific to unit i, but all are common to all units i. Typically, this is seen as unattractive, because the data generating process is believed to differ across units i, with some units having a higher level of the outcome variable y_{it} than others, regardless of covariates x_{it} (i.e. they have a higher intercept α) or a stronger effect of some covariate $x_{it,k}$ on y_{it} than others (i.e. they have a higher β_k). At the other extreme is the fully heterogeneous specification:

$$y_{it} = \alpha_i + x_{it}'\beta_i + u_{it} = \tilde{x}_{it}'\tilde{\beta}_i + u_{it} \; , \quad \mathbb{E}[u_{it}x_{it}] = 0 \; , \;\; \mathbb{V}[u_{it}|x_{it}] = \sigma_i^2 \; .$$

[2] Using the usual argument from Chapter 3 but slightly differnt notation, we can also show asymptotic Normality:

$$\sqrt{n}(\hat{\tilde{\beta}}_{POLS} - \tilde{\beta}) \xrightarrow{d} N\left(0, \mathbb{E}[\tilde{X}_i'\tilde{X}_i]^{-1}\mathbb{E}[\tilde{X}_i'u_i(\tilde{X}_i'u_i)']\mathbb{E}[\tilde{X}_i'\tilde{X}_i]^{-1}\right) .$$

Under homoskedasticity and no autocorrelation, $\mathbb{E}[u_iu_i'|\tilde{X}_i] = \sigma^2 I$, and the asymptotic variance simplifies to $\sigma^2\mathbb{E}[\tilde{X}_i'\tilde{X}_i]^{-1}$, which we estimate as $\hat{\sigma}^2\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i'\tilde{X}_i\right]^{-1}$. However, we typically prefer the more generally valid heteroskedasticity- and autocorrelation (HAC)-robust estimator $\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i'\tilde{X}_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i'\hat{u}_i\hat{u}_i'\tilde{X}_i\right]\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i'\tilde{X}_i\right]^{-1}$.

We can write $y_i = \tilde{X}_i\tilde{\beta}_i + u_i$. With $\tilde{\beta}_i$ specific to each unit i, we have n separate time series regressions, one for each unit i. Following the analysis in , we get

$$\hat{\tilde{\beta}}_{i,OLS} = (\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'y_i = \left(\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\sum_{t=1}^{T}\tilde{x}_{it}y_{it} \; .$$

In this case, we obviously do not have consistency unless $T \to \infty$.

With the fully homogeneous specification unattractive and the fully heterogeneous specification infeasible, researchers usually go for a compromise and let intercepts (and error term variances) be unit-specific:

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it} \; , \quad \mathbb{E}[u_{it}x_{it}] = 0 \; , \;\; \mathbb{V}[u_{it}|x_{it}] = \sigma^2 \; .$$

The term α_i is referred to as unit-specific heterogeneity or fixed effect. The latter terminology is somewhat unlucky, as there is also a fixed effects estimation method (see Section 11.3). In any case, the equation above makes clear that α_i contains all factors that affect y_{it}, that are not included in x_{it} and that are fixed over time (the time-varying factors are in u_{it}). Becuase we do not know the value of α_i, these factors are unobserved.

Suppose the above model is correctly specified and suppose we only had a cross-sectional dataset available, i.e. $T = 1$. Then we could estimate

$$y_{it} = \beta_0 + x_{it}'\beta + v_i \quad \text{for} \quad t = 1 \; ,$$

where $v_i = u_{it} + \alpha_i - \beta_0$. If the unobserved heterogeneity α_i is correlated with the covariates x_{it}, then our standard OLS estimator is biased and inconsistent.

Having a panel dataset, i.e. observing each observation i for $T > 1$, is a potential remedy, as it allows us to model α_i as a unit-specific intercept that is fixed over time. In line with that, we can write the above model as a regression with $k + n$ regressors; the k regressors in x_{it} and n unit-dummies:

$$y_{it} = x_{it}'\beta + \sum_{j=1}^{n} \mathbf{1}\{i = j\}\,\alpha_j + u_{it} = x_{it}^{*\prime}\beta^* + u_{it} \; ,$$

where $x_{it}^* = (x_{it}', \; \mathbf{1}\{i = 1\}, \; \mathbf{1}\{i = 2\}, ..., \; \mathbf{1}\{i = n\})'$ and $\beta^* = (\beta', \alpha_1, \alpha_2, ..., \alpha_n)'$. This leads to the following (pooled) OLS estimator for β^*:

$$\hat{\beta}^*_{OLS} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x^*_{it} x^{*\prime}_{it} \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} x^*_{it} y_{it} \ .$$

However, this estimator suffers from so-caled incidental parameters problem (IPP): because the number of regressors increases with n, the limit of $\frac{1}{n}\sum_{i=1}^{n} x^*_{it} x^{*\prime}_{it}$ as $n \to \infty$ is not well-defined, and, as a result, we cannot establish consistency of $\hat{\beta}^*_{OLS}$.[3]

To leverage the availability of panel data to remedy the inconsistency caused by unobserved heterogeneity α_i, we have to find a way to deal with such incidental parameters. Sections 11.2 to 11.5 present several approaches to do so. The random effects (RE) approach shows that, provided that α_i and x_{it} are uncorrelated, we can consistently estimate β by putting α_i into the error term. The fixed effects (FE) approach finds ways to construct a consistent estimator for β by transforming the original model equation above into one where α_i disappears. The correlated random effects (CRE) approach – closely related with Bayesian methods – shows that we can estimate β consistently by modeling the relation between α_i and x_{it} (for example by integrating out α_i).

11.2 Random Effects

By defining $v_{it} = u_{it} + \alpha_i - \beta_0$, we can write the model

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it} \quad \text{as} \quad y_{it} = \beta_0 + x'_{it}\beta + v_{it} \ .$$

Defining again $\tilde{x}_{it} = (1, x'_{it})'$ as well as $\tilde{\beta} = (\beta_0, \beta')'$, we can write this as

$$y_{it} = \tilde{x}'_{it}\tilde{\beta} + v_{it} \quad \text{or} \quad y_i = \tilde{X}_i\tilde{\beta} + v_i \ ,$$

where the $T \times 1$ vectors y_i and v_i and the $T \times k$ matrix X_i are defined analogously to before. Intuitively, we put the unit-specific intercepts α_i into the error term v_i and replaced it with a common intercept β_0. With this intercept included, $\mathbb{E}[v_i] = 0$ is guaranteed to hold. Define $\tilde{\alpha}_i = \alpha_i - \beta_0$ as the mean-zero unit-specific heterogeneity so that $v_i = u_i + \tilde{\alpha}_i$.

As with the fully homogeneous model in Section 11.1, we again have n linear regressions,

[3] Using a more stylized, but instructive example, in the Appendix it is shown analytically that the standard LS and ML estimator for σ^2 in the model $y_{it} = \alpha_i + u_{it}$, $u_{it} \overset{i.i.d.}{\sim} N(0, \sigma^2)$ is inconsistent due to the presence of α_i.

and estimating $\tilde{\beta}$ by OLS again leads to the POLS estimator

$$\hat{\tilde{\beta}}_{RE-OLS/POLS} = \left(\sum_{i=1}^{n} \tilde{X}_i' \tilde{X}_i\right)^{-1} \sum_{i=1}^{n} \tilde{X}_i' y_i \xrightarrow{p} \mathbb{E}[\tilde{X}_i' \tilde{X}_i]^{-1} \mathbb{E}[\tilde{X}_i' v_i] \ .$$

However, in contrast to the fully homogeneous model, here our error term v_{it} is not equal to the original error term u_{it}, but includes also the unit-specific heterogeneity α_i. This has two implications. First, to obtain consistency, we need to assume not only that x_{it} and u_{it} are contemporaneously uncorrelated, but that the same holds also for x_{it} and α_i. Typically, for this, we assume $\mathbb{E}[u_{it}|\tilde{x}_{it}, \tilde{\alpha}_i] = 0 \ \forall \ t$ and $\mathbb{E}[\tilde{\alpha}_i|\tilde{x}_{it}] = 0$. Second, we are obliged to use HAC-robust standard errors because

$$\Omega \equiv \mathbb{E}[v_i v_i'|\tilde{X}_i] = \mathbb{E}[(\alpha_i \iota + u_i)(\tilde{\alpha}_i \iota + u_i)'|\tilde{X}_i] = \mathbb{E}[\tilde{\alpha}_i^2 \iota\iota'|\tilde{X}_i] + \mathbb{E}[u_i u_i'|\tilde{X}_i]$$

is not diagonal.[4] This also means that the GLS estimator

$$\hat{\tilde{\beta}}_{RE-GLS} = \left(\sum_{i=1}^{n} \tilde{X}_i' \Omega^{-1} \tilde{X}_i\right)^{-1} \sum_{i=1}^{n} \tilde{X}_i' \Omega^{-1} y_i$$

is more efficient than the OLS estimator $\hat{\tilde{\beta}}_{RE-OLS}$. However, it is not clear whether the feasible GLS estimator, replacing Ω with an estimator $\hat{\Omega}$, is more efficient than $\hat{\tilde{\beta}}_{RE-OLS}$. The Appendix discusses the RE-GLS estimator further.

Taking a step back, note that the parameter $\tilde{\alpha}_i$ is treated as a RV, despite the fact that the RE approach is frequentist, i.e. it regards (other) parameters $\tilde{\beta}$ as fixed and data $\{y_{it}, x_{it}\}_{i=1:n,t=1:T}$ as random. After all, we specify its mean and make assumptions about its correlation with covariates $\tilde{X}_i$.

[4] In particular, even if the errors u_{it} in the original model are homoskedastic and not autocorrelated, say $\mathbb{E}[u_i u_i'|\tilde{X}_i, \tilde{\alpha}_i] = \sigma_u^2 I_T$, and even if the variance of unit-specific heterogeneity is the same for all i and independent of $\tilde{X}_i$, $\mathbb{E}[\tilde{\alpha}_i^2|\tilde{X}_i] = \sigma_\alpha^2$, we obtain

$$\Omega = \sigma_\alpha^2 \iota\iota' + \sigma_u^2 I_T \ ,$$

i.e. a transformed model with autocorrelation. This autocorrelation arises because we push the same, time-constant α_i into the error term v_{it} at all periods t.

11.3 Fixed Effects

The RE-assumption that the (time-constant) unobserved factors affecting outcomes y_{it} are uncorrelated the observed factors affecting y_{it} is typically unattractive. The FE approach derives a consistent estimator for $\tilde{\beta}$ regardless of the correlation of α_i and x_{it}. FE estimators exploit the fact that we observe data over several time periods, while α_i is constant over time, in order to transform the original model,

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it} \;, \quad i = 1:n \;,\; t = 1:T \tag{11.1}$$

so that α_i disappears.

One way to do so is to take an average of Eq. (11.1) over time, $\bar{y}_i = \alpha_i + \bar{x}_i'\beta + \bar{u}_i$ – where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, and analogously for $\bar{x}_i$ and $\bar{u}_i$ –, and subtract these time-averages from the original model. This yields

$$\ddot{y}_{it} = \ddot{x}_{it}'\beta + \ddot{u}_{it} \;, \quad i = 1:n \;,\; t = 1:T \;,$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$, and analogously for $\ddot{x}_{it}$ and $\ddot{u}_{it}$. We call $\ddot{y}_{it}$ the demeaned variable y_{it}. Just as we can write the original model as

$$y_i = \alpha_i \iota + X_i\beta + u_i \;,$$

we can write the transformed model as

$$M_\iota y_i = M_\iota X_i \beta + M_\iota u_i \;,$$

where $M_\iota = I - P_\iota$ and $P_\iota = \iota(\iota'\iota)^{-1}\iota' = \frac{1}{T}\iota\iota'$ are projection matrices with the usual properties (see Section 3.1), and ι is a $T \times 1$ vector of ones. Estimating this transformed model by OLS yields the so-called "Within"-estimator

$$\hat{\beta}_W = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \ddot{x}_{it}\ddot{x}_{it}'\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} \ddot{x}_{it}\ddot{y}_{it} = \left(\sum_{i=1}^{n} X_i' M_\iota X_i\right)^{-1} \sum_{i=1}^{n} X_i' M_\iota y_i \;.$$

Another way to transform the original model so as to get rid of α_i is by taking first differences (FD) of Eq. (11.1), yielding

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} \;, \quad i = 1:n \;,\; t = 2:T \;,$$

or, equivalently,

$$y_i^\Delta = X_i^\Delta \beta + u_i^\Delta \ .$$

Estimating this transformed model by OLS yields the FD-estimator

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^n \sum_{t=2}^T \Delta x_{it} \Delta x_{it}' \right)^{-1} \sum_{i=1}^n \sum_{t=2}^T \Delta x_{it} \Delta y_{it} = \left(\sum_{i=1}^n X_i^{\Delta\prime} X_i^\Delta \right)^{-1} \sum_{i=1}^n X_i^{\Delta\prime} y_i^\Delta \ .$$

As usual, in order to form these OLS estimators, we need the respective denominator to be of full rank: $rank \left(\sum_{i=1}^n X_i' M X_i \right) = k$ and $rank \left(\sum_{i=1}^n X_i^{\Delta\prime} X_i^\Delta \right) = k$. These assumptions are the analogue of Assumption 3 from Chapter 3. In practical applications of the standard, cross-sectional linear regression, this assumption is not emphasized as we can mechanically satisfy it by choosing a set of covariates so as to avoid perfect multicollinearity. However, in the context of FE estimators, it is important to note that this assumption would be violated if there are regressors in the original covariates x_{it} that are fixed over time for all cross-sectional units, such as one's biological sex (at birth) or one's parental income at a specific age in childhood. This means that in FE regressions, we can only include variables that vary over time, at least for some cross-sectional units. By continuity, it also means that we get noisy estimators of coefficients multiplying regressors that vary little over time.

Assuming, as usual, that the model is correctly specified and that the observations are i.i.d. across i, we get the following probability limit for $\hat{\beta}_W$:

$$\hat{\beta}_W \xrightarrow{p} \beta + \mathbb{E}\left[\ddot{x}_{it} \ddot{x}_{it}' \right]^{-1} \mathbb{E}\left[\ddot{x}_{it} \ddot{u}_{it} \right] \ .$$

In contrast to the RE-estimator $\hat{\hat{\beta}}_{RE-OLS}$, consistency of the Within-estimator is unaffected by correlation between α_i and X_i, because α_i disappears altogether and is not just put into the error term as under RE. However, the Within-estimator requires not only contemporaneous exogeneity $\mathbb{E}[u_{it}|x_{it}] = 0 \ \forall \ t$ to hold, but strict exogeneity, i.e.

$$\mathbb{E}[u_{it}|X_i] = 0 \ \forall \ t \quad \Leftrightarrow \quad \mathbb{E}[u_{it}|x_{is}] = 0 \ \forall \ t, s \ .$$

In words, u_{it} and x_{is} need to be independent for all periods t, s, i.e. at all leads and lags. To see this, note that

$$\mathbb{E}\left[\sum_{t=1}^T \ddot{x}_{it} \ddot{u}_{it} \right] = \sum_{t=1}^T \mathbb{E}\left[\left(x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \right) \left(u_{it} - \frac{1}{T} \sum_{t=1}^T u_{it} \right) \right] = 0$$

needs to hold for consistency. Strict exogeneity is a much stronger assumption than contemporaneous exogeneity. It notably precludes any feedback-loops by which $y_{i,t-1}$ (or $y_{i,t-l}$ for some $l > 0$) affects x_{it}. A feedback loop from past y_{it} on x_{it} can often not be excluded, as for example in a study where y_{it} is the HIV infection rate in a region i and x_{it} includes condom sales, which likely depend on lagged y_{it}.

For $\hat{\beta}_{FD}$, we get

$$\hat{\beta}_{FD} \xrightarrow{p} \beta + \mathbb{E}\left[\sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}'\right]^{-1} \mathbb{E}\left[\sum_{t=2}^{T} \Delta x_{it} \Delta u_{it}\right] .$$

For consistency, it requires $(u_{it}, u_{i,t-1})$ and $(x_{it}', x_{i,t-1}')$ to be uncorrelated,[5] because

$$\mathbb{E}\left[\sum_{t=2}^{T} \Delta x_{it} \Delta u_{it}\right] = \sum_{t=2}^{T} \mathbb{E}\left[(x_{it} - x_{i,t-1})(u_{it} - u_{i,t-1})\right] = 0$$

is needed. This precludes any feedback-loops by which $y_{i,t-1}$ affects x_{it}, but not feedbacks from $y_{i,t-2}$ (or $y_{i,t-l}$ for some $l > 1$) on x_{it}.

For the asymptotic variance of the Within-estimator, using standard calculations we get $\sqrt{n}(\hat{\beta}_W - \beta) \xrightarrow{d} N(0, V_W)$ with

$$V_W = \mathbb{E}\left[\sum_{t=1}^{T} \ddot{x}_{it} \ddot{x}_{it}'\right]^{-1} \mathbb{E}\left[\left(\sum_{t=1}^{T} \ddot{x}_{it} \ddot{u}_{it}\right)\left(\sum_{s=1}^{T} \ddot{x}_{is} \ddot{u}_{is}\right)'\right] \mathbb{E}\left[\sum_{t=1}^{T} \ddot{x}_{it} \ddot{x}_{it}'\right]^{-1} ,$$

and for the FD-estimator we get $\sqrt{n}(\hat{\beta}_{FD} - \beta) \xrightarrow{d} N(0, V_{FD})$ with

$$V_{FD} = \mathbb{E}\left[\sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}'\right]^{-1} \mathbb{E}\left[\left(\sum_{t=2}^{T} \Delta x_{it} \Delta u_{it}\right)\left(\sum_{s=2}^{T} \Delta x_{is} \Delta u_{is}\right)'\right] \mathbb{E}\left[\sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}'\right]^{-1} .$$

Replacing expectations with sample means and $\ddot{u}_{it}$ and Δu_{it} with $\hat{\ddot{u}}_{it}$ and $\widehat{\Delta u_{it}}$, respectively, yields the HAC-robust variance estimators $\hat{V}_W$ and $\hat{V}_{FD}$.[6] [7] [8]

Even if strict exogeneity is satisfied, the consistency of FE estimators comes at an efficiency

[5] That is, $\mathbb{E}[u_{it}|x_{it}] = 0$, $\mathbb{E}[u_{it}|x_{i,t-1}] = 0$ and $\mathbb{E}[u_{i,t-1}|x_{i,t}] = 0 \; \forall \; t$.

[6] Note that the respective term in the middle can be written out further, analogously as done in Section 8.4.2 in the context of autocorrelated errors in time series regressions.

[7] Under homoskedasticity, these formulas simplify, but $\sigma^2 = \mathbb{V}[u_{it}]$ appears, which needs to be estimated. Under the presence of incidental parameters α_i, the standard estimator $\hat{\sigma}_u^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \hat{u}_{it}^2$ is inconsistent, because $\hat{u}_{it} = y_{it} - \hat{\alpha}_i - x_{it}\hat{\beta}$, and $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta}$ is inconsistent. As a result, in practice one uses HAC-robust variances.

[8] Note also that for the FD estimator we need to use an autocorrelation-robust variance because $\Delta u_{it} = u_{it} - u_{i,t-1}$ is obviously autocorrelated.

loss compared to the RE-GLS (and RE-OLS) estimator(s). This is easiest seen in the FD-transformation, in which we loose the n observations pertaining to the first time period $t = 1$. In addition, first-differenced covariates have a lower variance than the original covariates measured in levels. The efficiency loss of the Within-estimator is somewhat more subtle. It arises because the Within-estimator only exploits variation across time and disregards the time-constant variation across cross-sectional units.[9] As a result, if the core RE assumption of X_i and α_i being uncorrelated is indeed satisfied, we prefer the RE-estimators. If instead it is violated, we of course prefer the less efficient but consistent FE estimators. The Hausman test helps us in this choice. It compares $\hat{\beta}_{RE-GLS}$ and $\hat{\beta}_W$, testing the hypothesis $\mathscr{H}_0 : \delta = \beta_{RE-GLS} - \beta_W = 0$ using the test statistic

$$n\left(\hat{\beta}_W - \hat{\beta}_{RE-GLS}\right)'\left(A\mathbb{V}[\hat{\beta}_W] - A\mathbb{V}[\hat{\beta}_{RE-GLS}]\right)^{-1}\left(\hat{\beta}_W - \hat{\beta}_{RE-GLS}\right) \xrightarrow{d} \chi^2_k ,$$

where $A\mathbb{V}[\cdot]$ denotes the asymptotic variance. If $\mathscr{H}_0$ is accepted, then the difference between $\hat{\beta}_{RE-GLS}$ and $\hat{\beta}_W$ is deemed small enough to suggest that both are consistent, i.e. that X_i and α_i are indeed uncorrelated, and therefore we should use the more efficient $\hat{\beta}_{RE,GLS}$.

To sum up, the FE estimators work under arbitrary correlation between the unobserved heterogeneity α_i and covariates X_i, but they cannot deal with time-constant regressors and their consistency is paid for by an efficiency loss relative to RE estimators. Most importantly, their consistency requires strict exogeneity, a much stronger assumption than contemporaneous exogeneity of covariates and error terms. In practice, this assumption is often deemed violated. In particular, for dynamic panels, where x_{it} contains lagged y_{it}, it is violated by construction, because, for example, $y_{i,t-1}$ is a function of $u_{i,t-1}$. Three alternative approaches can be used as potential remedies: FE-IV estimation, bias correction – discussed below – and CRE approaches – discussed in Section 11.4.

FE-IV Estimation Suppose we relax the strict exogeneity assumption,

$$\mathbb{E}[u_{it}|X_i, \alpha_i] = \mathbb{E}[u_{it}|x_{i1}, ..., x_{iT}, \alpha_i] = 0 ,$$

to sequential exogeneity,

$$\mathbb{E}[u_{it}|x_{i1}, ..., x_{it}, \alpha_i] = 0 .$$

This allows for feedback loops by which $y_{i,t-l}$ for some $l > 0$ affects x_{it}, which renders $u_{i,t-l}$ and x_{it} correlated. This allows for feedback loops by which lagged y_{it} affect x_{it}, rendering

[9] In the Appendix, the RE-GLS estimator is derived as the optimal weighted average of the Within-estimator and the Between-estimator. The latter does the opposite of the Within-estimator; it disregards the variation across time and only uses the time-constant variation across cross-sectional units. This underscores that the RE-GLS estimator is more efficient than the FE-Within estimator.

x_{it} correlated with lagged u_{it}. To be specific, let x_{it} be composed of $y_{i,t-1}$ and $\tilde{x}_{it}$, where the latter satisfies strict exogeneity. The first-differenced specification is then

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} = \beta_1 \Delta y_{i,t-1} + \Delta \tilde{x}_{it}'\beta_{-1} + \Delta u_{it} \ .$$

We know $\hat{\beta}_W$ and $\hat{\beta}_{FD}$ are inconsistent because x_{it} is not strictly – but only sequentially – exogenous due to the presence of $y_{i,t-1}$.

In this environment, Anderson and Hsiao (1981) propose to use $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$. Under sequential exogeneity, instrument-exogeneity is satisfied:

$$\mathbb{E}[y_{is}\Delta u_{it}] = 0 \quad \text{for } s \leq t-2 \ .$$ [10]

Instrument-relevance is typically satisfied, too, as first differences tend to be correlated with past (and future) levels. Going a step further, Arellano and Bond (1991) suggest using not only $y_{i,t-2}$, but $(y_{i,t-2}, ..., y_{i1})$.[11]

Using similar reasoning, other approaches use sequential exogeneity to circumvent FE methods altogether rather than to save their consistency. For example, Blundell and Bond (1998) start from the original specification

$$y_{it} = x_{it}'\beta + \alpha_i + u_{it} \ ,$$

where the correlation between α_i and x_{it} is suspected to be due to $y_{i,t-1}$, contained in x_{it}. As under RE, they put α_i into the error term $v_{it} = \alpha_i + u_{it}$. To avoid bias and inconsistency, they use $\Delta y_{i,t-1}$ as an IV for $y_{i,t-1}$. Provided that the model is correctly specified, $\Delta y_{i,t-1}$ is not a function of α_i.

Another IV-based approach is proposed by Ahn and Schmidt (1995). It uses the uncorrelatedness of α_i and u_{it} and assumes that u_{it} is serially uncorrelated to form moment restrictions of the form

$$0 = \mathbb{E}[(\alpha_i + u_{i\tau})\Delta u_{i,t-1}] = \mathbb{E}[(y_{i\tau} - x_{i\tau}'\beta)(\Delta y_{i,t-1} - \Delta x_{i,t-1}'\beta)] \ , \quad \text{for } \tau \geq t \ .$$

The resulting GMM estimators are consistent even if strict exogeneity is violated.

[10] That is, one would use $z_{it} = (\Delta\tilde{x}_{it}', y_{i,t-2})'$ as an instrument for Δx_{it}, assuming that strict exogeneity holds between u_{it} and $\tilde{x}_{it}$. To develop an understanding of these methods, ignore $\tilde{x}_{it}$ and assume that $x_{it} = y_{i,t-1}$.

[11] Note that in this case the number of IVs changes for different observations; ignoring the regressions for $t=1$, at $t=2$ we have y_{i1} as an IV for Δy_{i2}, at $t=3$ we have $(y_{i1}, y_{i2})'$ as an IV for Δy_{i3}, etc. For this reason, one has to implement the IV estimation via GMM. See Section 6.3 and Section 6.4.

In sum, FE-IV methods lead to consistent estimators even if strict exogeneity is violated, but sequential exogeneity is satisfied. However, they suffer from weak IV issues because the correlation between a variable like $y_{i,t-2}$ and $\Delta y_{i,t-1}$ is generally weak, as first-differencing removes a lot of variation.

Bias Correction Another approach to deal with the inconsistency of FE estimators under failure of strict exogeneity is to derive the asymptotic bias, approximate it for finite samples and remove it from $\hat{\beta}_{FD}$ or $\hat{\beta}_W$. This leads to consistent estimators and avoids the inefficiency of IV-based approaches. However, the issue here is that the asymptotic bias may not be a good characterization of the finite sample bias.

11.4 Correlated Random Effects

FE approaches transform the original model of interest to simply get rid of α_i under arbitrary correlation between α_i and X_i. Under some conditions, they are consistent, but they always incur an efficiency loss relative to the RE estimators that can be used if α_i and X_i are not correlated. CRE approaches aim at averting some of that efficiency loss by modeling the correlation between α_i and X_i. For some models and under some assumptions on that correlation, we get the Within estimator (see Appendix), but in other settings CRE approaches can lead to more efficient estimators which are consistent even if strict exogeneity fails.

Under the umbrella of CRE fall a lot of different methods. Some apply minimum distance techniques, some simply plug in the supposed functional form of $\alpha_i = f(X_i)$ to transform the original model equation (see Appendix), while others use Quasi-Maximum Likelihood Estimation (QMLE) and relate to Bayesian variants of CRE (discussed in Section 11.5). In this exposition, we focus on QMLE.

Suppose $y_{it} = \alpha_i + x_{it}'\beta + u_{it}$ with $u_{it} \sim N(0, \sigma_u^2)$ homoskedastic for simplicity. We can construct the likelihood using cross-sectional independence as

$$\begin{aligned}\mathcal{L}(\beta, \sigma_u^2, \{\alpha_i\}_i | Y, X) &\equiv p(Y|X, \{\alpha_i\}_i, \beta, \sigma_u^2) \\ &= \prod_{i=1}^{n} p(y_i | X_i, \alpha_i, \beta, \sigma_u^2) \ ,\end{aligned}$$

where we could further write out $p(y_i|X_i, \alpha_i, \beta, \sigma_u^2) = \prod_{t=1}^{T} p(y_{it}|x_{it}, \alpha_i, \beta, \sigma_u^2)$. The problem with this likelihood is that the incidental parameters $\{\alpha_i\}_i$ appear in it, and their presence renders the estimation of (β, σ_u^2) inconsistent. The CRE-QMLE approach constructs a quasi-likelihood by integrating out $\{\alpha_i\}_i$ using some density, which takes into account their (potential) correlation with $\{X_i\}_i$. For example, we could assume

$$\alpha_i | X_i, \gamma, \sigma_\alpha^2 \sim N(\bar{X}_i' \gamma, \sigma_\alpha^2) \ ,$$

where $\bar{X}_i$ is the $k \times 1$ of time-averages of covariates in X_i.[12] Because this distributional assumption is not an integral part of the model of interest and because we regard it only as an approximation, we speak of quasi-ML estimation. Using this density, we can form

$$\begin{aligned}
\mathcal{L}_{CRE}(\beta, \sigma_u^2, \gamma, \sigma_\alpha^2 | Y, X) &\equiv p(Y | X, \beta, \sigma_u^2, \gamma, \sigma_\alpha^2) \\
&= \int p(Y, \{\alpha_i\}_i | X, \beta, \sigma_u^2, \gamma, \sigma_\alpha^2) d\{\alpha_i\}_i \\
&= \int p(Y | X, \{\alpha_i\}_i, \beta, \sigma_u^2) p(\{\alpha_i\}_i | X, \gamma, \sigma_\alpha^2) d\{\alpha_i\}_i \\
&= \prod_{i=1}^n \left\{ \int p(y_i | X_i, \alpha_i, \beta, \sigma_u^2) p(\alpha_i | X_i, \gamma, \sigma_\alpha^2) d\alpha_i \right\} \ .
\end{aligned}$$

Essentially, we replace the incidental parameters $\{\alpha_i\}_i$ with the two parameters γ and σ_α^2 that summarize their behavior. Once $\{\alpha_i\}_i$ is integrated out in that way, we can apply standard ML theory (extremum estimation theory; see Chapter 5) to analyze the asymptotic properties of $\hat{\theta}$, the MLE for $\theta = (\beta', \sigma_u^2, \gamma', \sigma_\alpha^2)'$. The resulting estimators perform very well even if the density $\alpha_i | X_i$ is misspecified. In fact, Liu, Moon & Schorfheide (2020) show in the context of the dynamic panel data model that the resulting estimators derived under the assumption $\alpha_i | y_{i0} \sim N(\gamma y_{i0}, \sigma_\alpha^2)$ are consistent as long as the distribution of α_i is indeed only a function of y_{i0} and not other aspects of the covariates, $Y_{1:T-1}$ in this case. An additional important advantage of the QMLE is that we can integrate out alpha in many settings where subtracting it as under FE approaches is not possible.

Solving the above integral and actually deriving $\mathcal{L}_{CRE}(\beta, \sigma_u^2, \gamma, \sigma_\alpha^2 | Y, X)$ is easier than it seems. The integral of interest appears as the denominator in the following Bayes' formula:

$$p(\alpha_i | y_i, X_i, \beta, \sigma_u^2, \gamma, \sigma_\alpha^2) = \frac{p(y_i | X_i, \alpha_i, \beta, \sigma_u^2) p(\alpha_i | X_i, \gamma, \sigma_\alpha^2)}{\int p(y_i | X_i, \alpha_i, \beta, \sigma_u^2) p(\alpha_i | X_i, \gamma, \sigma_\alpha^2) d\alpha_i} \ .$$

It is the MDD obtained in the Bayesian estimation of α_i conditional on the parameters (β, σ_u^2) and the hyperparameters $(\gamma, \sigma_\alpha^2)$. Equipped with the knowledge from Section 4.5, we know how to derive it. First, we derive the conditional posterior of α_i on the LHS using proportionality of posterior and likelihood times prior. Then we invert Bayes' formula, dividing the product of likelihood and prior by the posterior, and cancel all terms that involve

[12] Under a dynamic panel data model, $y_{it} = \alpha_i + \rho y_{i,t-1} + u_{it}$, it is common to let α_i depend on the initialization $y_{i,0}$ rather than the time-average of the covariate $y_{i,t-1}$.

$\{\alpha_i\}_i$. We obtain

$$
\begin{aligned}
p(\alpha_i|y_i,\cdot) &\propto p(y_i \mid \alpha_i,\cdot)p(\alpha_i|\cdot) \\
&= (2\pi\sigma_u^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma_u^2}(y_i-\iota\alpha_i-X_i\beta)'(y_i-\iota\alpha_i-X_i\beta)\right\}(2\pi\sigma_\alpha^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma_\alpha^2}(\alpha_i-\bar{X}_i'\gamma)^2\right\} \\
&\propto exp\left\{-\frac{1}{2\sigma_u^2}(\alpha_i^2\iota'\iota-2\alpha_i\iota'(y_i-X_i\beta)-\frac{1}{2\sigma_\alpha^2}(\alpha_i^2-2\alpha_i\bar{X}_i'\gamma)\right\} \ .
\end{aligned}
$$

This lets us deduce that

$$
\alpha_i|y_i,\cdot \sim N\left(\bar{\mu}_i,\bar{\sigma}_i^2\right) \ , \quad \bar{\sigma}_i^2 = \left(\frac{T}{\sigma_u^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1} \ , \ \bar{\mu}_i = \bar{\sigma}_i^2\left(\frac{1}{\sigma_u^2}\iota'(y_i-X_i\beta)+\frac{1}{\sigma_\alpha^2}\bar{X}_i'\gamma\right) \ ,
$$

where it is used that $\iota'\iota = T$. For the MDD, we then get

$$
\begin{aligned}
p(y_i|\cdot) &= \frac{p(y_i \mid \alpha_i,\cdot)p(\alpha_i|\cdot)}{p(\alpha_i|y_i,\cdot)} \\
&= \frac{(2\pi\sigma_u^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma_u^2}(y_i-\iota\alpha_i-X_i\beta)'(y_i-\iota\alpha_i-X_i\beta)\right\}(2\pi\sigma_\alpha^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma_\alpha^2}(\alpha_i-\bar{X}_i'\gamma)^2\right\}}{(2\pi\bar{\sigma}_i^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\bar{\sigma}_i^2}(\alpha_i-\bar{\mu}_i)^2\right\}} \\
&= (2\pi)^{-\frac{T}{2}}(\sigma_u^2)^{-\frac{T}{2}}(\sigma_\alpha^2)^{-\frac{1}{2}}(\bar{\sigma}_i^2)^{\frac{1}{2}} exp\left\{-\frac{1}{2\sigma_u^2}(y_i-X_i\beta)'(y_i-X_i\beta)-\frac{1}{2\sigma_\alpha^2}(\bar{X}_i'\gamma)^2+\frac{1}{2\bar{\sigma}_i^2}\bar{\mu}_i^2\right\} \\
&= (2\pi)^{-\frac{T}{2}}(\sigma_u^2)^{-\frac{T}{2}}(\sigma_\alpha^2)^{-\frac{1}{2}}\left(\frac{T}{\sigma_u^2}+\frac{1}{\sigma_\alpha^2}\right)^{-\frac{1}{2}} \\
&\quad \cdot \ exp\left\{-\frac{1}{2}\left(\frac{(y_i-X_i\beta)'(y_i-X_i\beta)}{\sigma_u^2}+\frac{(\bar{X}_i'\gamma)^2}{\sigma_\alpha^2}+\left(\frac{T}{\sigma_u^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}\left(\frac{\iota'(y_i-X_i\beta)}{\sigma_u^2}+\frac{\left(\bar{X}_i'\gamma\right)}{\sigma_\alpha^2}\right)^2\right)\right.
\end{aligned}
$$

and so

$$
\begin{aligned}
\mathcal{L}_{CRE}(\beta,\sigma_u^2,\gamma,\sigma_\alpha^2|Y,X) = (2\pi)^{-\frac{nT}{2}}(\sigma_u^2)^{-\frac{nT}{2}}(\sigma_\alpha^2)^{-\frac{n}{2}}&\left(\frac{T}{\sigma_u^2}+\frac{1}{\sigma_\alpha^2}\right)^{-\frac{n}{2}} \\
\cdot \ exp&\left\{-\frac{1}{2}\sum_{i=1}^n\left(\frac{(y_i-X_i\beta)'(y_i-X_i\beta)}{\sigma_u^2}+\frac{(\bar{X}_i'\gamma)^2}{\sigma_\alpha^2}+\right.\right. \\
&\left.\left.\cdot \ \left(\frac{T}{\sigma_u^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}\left(\frac{\iota'(y_i-X_i\beta)}{\sigma_u^2}+\frac{\left(\bar{X}_i'\gamma\right)}{\sigma_\alpha^2}\right)^2\right)\right\} \ .
\end{aligned}
$$

Because this likelihood is then used to estimate the hyperparameters γ and σ_α^2, this is an empirical Bayes estimation approach, i.e. we determine an "objective" prior for α_i so as to fit the data best. From Section 4.5, we know that choosing hyperparameters to maximize the

MDD (optimally) exploits the bias-variance-trade-off and leads to efficient estimators. Note that in the context of CRE-QMLE, the priors for all α_i share the same hyperparameters, which are estimated using data for all cross-sectional units i.

Estimating Heterogeneous Coefficients The CRE-QMLE approach allows us not only to obtain efficient and oftentimes consistent estimators for β and σ_u^2, it also motivates estimators for the incidental parameters $\{\alpha_i\}_i$ that improve upon the standard estimator

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} - x_{it}'\hat{\beta} \ ,$$

where $\hat{\beta}$ is some consistent estimator of β. The problem with the estimator $\hat{\alpha}_i$ is that it is very noisy if T is small. The conditional posterior mean for α_i,

$$\bar{\mu}_i = \left(\frac{T}{\sigma_u^2} + \frac{1}{\sigma_\alpha^2}\right)^{-1} \left(\frac{1}{\sigma_u^2}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta) + \frac{1}{\sigma_\alpha^2}\bar{X}_i'\gamma\right) \ ,$$

performs better.[13] In the context of the frequentist QMLE approach, we would replace the unknown parameters $(\sigma_u^2, \sigma_\alpha^2, \gamma)$ and β with consistent estimators (e.g. QMLE).[14] As a result, $\bar{\mu}_i$ shrinks $\hat{\alpha}_i$ to the mean function $\bar{X}_i'\gamma$ that is common among all cross-sectional units, whereby replacing γ with $\hat{\gamma}_{QMLE}$ means that we determine this function so as to obtain an optimal data fit across the whole cross-section. Under the alternative prior $\alpha_i \sim N(\mu, \sigma_\alpha^2)$, we would shrink $\hat{\alpha}_i$ to some common mean value $\hat{\mu}$. This reduces the noise in the estimators $\{\hat{\alpha}_i\}_i$, in particular for observations i associated with extreme values of α_i (see Liu et al. (2020); Botosaru et al. (2023)).

11.5 Bayesian Analysis of Panel Data

The CRE-QMLE approach has a Bayesian flavor because it specifies a (conditional) prior for α_i and derives the (conditional) MDD where α_i has been integrated out. A fully Bayesian approach to analyzing the panel data model $y_{it} = \alpha_i + x_{it}'\beta + u_{it}$, $u_{it} \sim N(0, \sigma_u^2)$ would specify priors for β and σ_u^2 as well, and it would specify hyperpriors for the auxiliary parameters γ and σ_α^2 which appear in the prior for α_i. For example, one could choose a Normal-Inverse

[13] Note that $\iota'(y_i - X_i\beta) = \sum_{t=1}^{T} y_{it} - x_{it}'\beta$.

[14] Liu et al. (2020) show that one can motivate this estimator using "Tweedie's formula", an old concept from statistics. It derives the conditional posterior mean as the MLE $\hat{\alpha}_i$ plus a correction term which accounts for the shrinkage.

Gamma prior for (β, σ_u^2) and uniform/improper priors for $(\gamma, \sigma_\alpha^2)$. This leads to the posterior

$$p\left(\{\alpha_i\}_i, \beta, \sigma_u^2, \gamma, \sigma_\alpha^2 | Y, X\right) \propto p\left(Y|X, \{\alpha_i\}_i, \beta, \sigma_u^2, \gamma, \sigma_\alpha^2\right) \; p(\beta|\sigma_u^2) \, p(\sigma_u^2) \prod_{i=1}^n p(\alpha_i|\gamma, \sigma_\alpha^2) \; .$$

It can be obtained numerically using Gibbs sampling (see Section 7.2), as it is easy to derive the posteriors of each of the five parameter-groups $\{\alpha_i\}_i$, β, σ_u^2, γ and σ_α^2 conditioning on the others.[15]

Appendix

Illustration of IPP Suppose we want to estimate σ^2 (and α_i) in the model

$$y_{it} = \alpha_i + u_{it} \; , \quad u_{it} \sim N(0, \sigma^2) \; .$$

Note that we can write this as $y_i = \iota\alpha_i + u_i$, where $u_i \sim N(0, \sigma^2 I_T)$ and ι is a $T \times 1$ vector of ones. We get the OLS and ML estimators $\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^T y_{it}$ and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_{it} - \hat{\alpha}_i)^2 = \frac{1}{n}\sum_{i=1}^n \left(\frac{1}{T} y_i' M y_i\right) \; ,$$

where $M = I_T - \iota(\iota'\iota)^{-1}\iota' = I_T - \frac{1}{T}\iota\iota'$.

We have $\hat{\alpha}_i \sim N(\alpha_i, \frac{1}{T}\sigma^2)$, which is inconsistent unless $T \to \infty$. Moreover, the presence of $\hat{\alpha}_i$ leads to an inconsistent estimator for the homoegeneous parameter σ^2 as well. To show this, first note that $MY_i = M(\iota\alpha_i + u_i) = Mu_i$. Therefore,

$$\frac{1}{T} y_i' M y_i = \frac{1}{T} u_i' M u_i = \frac{1}{T} u_i' u_i - (\frac{1}{T} L' u_i)^2 = \frac{1}{T}\sum_{t=1}^T u_{it}^2 - (\frac{1}{T}\sum_{t=1}^T u_{it})^2 \; .$$

Using this result, we get

$$\hat{\sigma}^2 \xrightarrow{p} \mathbb{E}[\frac{1}{T} Y_i' M Y_i] = \sigma^2 - \frac{1}{T^2}\sum_{t=1}^T \sum_s \mathbb{E}[u_{it} u_{is}] = \left(1 - \frac{1}{T}\right)\sigma^2 \text{ as } n \to \infty \; .$$

In his example, correcting the bias is easy. However, this is not true in general because often the asymptotic bias cannot be derived.

[15] The first is derived above. The posteriors for (β, σ_u^2) and $(\gamma, \sigma_\alpha^2)$ conditional on $\{\alpha_i\}_i$ are both Normal-Inverse Gamma, with standard derivations.

RE-GLS Estimation The GLS estimator

$$\hat{\tilde{\beta}}_{RE-GLS} = \left(\sum_{i=1}^{n} \tilde{X}_i' \Omega^{-1} \tilde{X}_i \right)^{-1} \sum_{i=1}^{n} \tilde{X}_i' \Omega^{-1} y_i$$

is consistent and asymptotically Normal under the same assumptions as those needed for consistency and asmyptotic Normality of $\hat{\tilde{\beta}}_{RE-OLS}$. If indeed $\mathbb{E}[v_i v_i' | \tilde{X}_i] = \Omega$, then

$$\mathbb{V}[\hat{\tilde{\beta}}_{RE-GLS} | \tilde{X}_i] = \left(\sum_{i=1}^{n} \tilde{X}_i' \Omega^{-1} \tilde{X}_i \right)^{-1} ,$$

whereas this expression is somewhat more involved if $\mathbb{E}[v_i v_i' | \tilde{X}_i]$ is a function of $\tilde{X}_i$. Relatedly, note that $\mathbb{E}[v_i v_i' | \tilde{X}_i] = \Omega$ implies homoskedasticity. Hence, the motivation for using GLS is different than under a cross-sectional regression with heteroskedasticity (see Section 3.4). Here we use GLS because of autocorrelation in v_{it} induced by the presence of the time-invariant α_i.

Because we do not know Ω, the RE-GLS estimator is infeasible. A feasible version replaces Ω with an estimator $\hat{\Omega}$. Assuming homoskedasticity of the original errors, $\mathbb{E}[u_i u_i' | \tilde{X}_i, \tilde{\alpha}_i] = \sigma_u^2 I_T$, as well as a constant variance of unit-specific heterogeneity across i, $\mathbb{E}[\tilde{\alpha}_i^2 | \tilde{X}_i] = \sigma_\alpha^2$, we obtain

$$\Omega = \sigma_\alpha^2 \iota \iota' + \sigma_u^2 I_T .$$

As a result, estimating Ω involves estimating σ_u^2 and σ_α^2. We can use

$$\hat{\sigma}_v^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{v}_{it}^2 \xrightarrow{p} \sigma_v^2 \equiv \sigma_u^2 + \sigma_\alpha^2 = \mathbb{E}[v_{it}^2] ,$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{n\frac{T(T-1)}{2}} \sum_{i=1}^{n} \sum_{t,s=1,t\neq s} \hat{v}_{it} \hat{v}_{is} \xrightarrow{p} \sigma_\alpha^2 = \mathbb{E}[v_{it} v_{is}] \quad \text{for } s \neq t .$$

Note that these two estimators condition on an estimate $\hat{\tilde{\beta}}$ because $\hat{v}_{it} = (y_{it} - \tilde{x}_{it}' \hat{\tilde{\beta}})$.

To perform a two-step GLS estimation, one would take $\hat{v}_{it} = y_{it} - x_{it}' \hat{\tilde{\beta}}_{RE,OLS}$ to be the estimated error terms based on a preliminary RE-OLS regression. Instead, one can also iterate on the expression for $\hat{\tilde{\beta}}_{RE,GLS} | \ (\sigma_\alpha^2, \sigma_u^2)$ and these two expressions for $\hat{\sigma}_v^2 | \ \tilde{\beta}$ and $\hat{\sigma}_\alpha^2 | \tilde{\beta}$ until convergence. This latter version of the GLS estimator is also the MLE under the assumption that $v_i | \tilde{X}_i \sim N(0, \Omega)$. To derive it, we use the conditional likelihood $p(Y|X, \theta) = \prod_{i=1}^{n} p(y_i | \tilde{X}_i, \theta)$, constructed by relying cross-sectional independence.[16] Bayesian estimation

[16]Thereby, θ contains $\tilde{\beta}$, σ_α^2 and σ_u^2 under (or, more generally, if the homoskedasticity assumption **??** does

of the RE model uses this same likelihood and augments it with prior specifications for $\tilde{\beta}|(\sigma_u^2, \sigma_\alpha^2)$ and $\sigma_u^2|\tilde{\beta}$ and $\sigma_\alpha^2|\tilde{\beta}$.

On the Relation between the RE-GLS and Within Estimators The Within-estimator is the (pooled) OLS estimator for the transformed regression equation

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

where $\ddot{y}_{it} = y_{it} - \bar{y}_i = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and same for $\ddot{x}_{it}$ and $\ddot{u}_{it}$:

$$\hat{\beta}_W = \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{y}_{it} = \left(\frac{1}{n}\sum_{i=1}^{n}X_i'MX_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}X_i'My_i \ ,$$

where $M = I_T - P$ and $P = \iota(\iota'\iota)^{-1}\iota' = \frac{1}{T}\iota\iota'$. By time-demeaning observations for each i, $\hat{\beta}_W$ only exploits variation across time and subtracts the time-constant variation across cross-sectional units i. More generally, it removes a substantial amount of information about coefficients associated with regressors that move little over time.

Its counterpart is the Between-estimator, which exploits only the time-constant variation across cross-sectional units i. It is the OLS estimator for the transformed regression

$$\bar{y}_i = \bar{x}_i'\beta + \alpha_i + \bar{u}_i \ ,$$

where α_i is put into the error term:

$$\hat{\beta}_B = \left(\frac{1}{n}\sum_{i=1}^{n}\bar{x}_i\bar{x}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\bar{x}_i\bar{y}_i = \left(\frac{1}{n}\sum_{i=1}^{n}X_i'PX_i\right)^{-1}\frac{1}{n}\sum X_i'Py_i \ .$$

This estimator is rarely used in practice, as it is only useful if the variation over time is not helpful for identification.

The GLS estimator of β can be derived as the weighted average of the Within and Between estimator.[17] The weights are chosen optimally in the sense that they are given by the

not hold, it contains $\tilde{\beta}$ and Ω). The (joint) MLE for $\tilde{\beta}$ and Ω cannot be obtained analytically, because one only gets expressions for $\hat{\tilde{\beta}}|\Omega$ – this is the above expression for $\hat{\tilde{\beta}}_{RE,GLS}$ – and $\hat{\Omega}|\tilde{\beta}$ – under homoskedasticity these are the above two expressions for $\hat{\sigma}_v^2|\tilde{\beta}$ and $\hat{\sigma}_\alpha^2|\tilde{\beta}$. Instead, it can be obtained by iterating on these two conditional estimators until convergence (see Section 7.1) (or by brute force numerical optimization).

[17] Note that, in contrast to before, here we drop the intercept in the RE-specification. In our notation, we estimate $y_i = X_i\beta + v_i$ instead of $y_i = \tilde{X}_i\tilde{\beta} + v_i$. This is done to align with the FE-Within-specification, which cannot contain an intercept, as it would disappear when subtracting time-averages.

respective of each estimator's variance:

$$\hat{\beta}_{RE,GLS} = \left(\mathbb{V}[\hat{\beta}_B|X]^{-1} + \mathbb{V}[\hat{\beta}_W|X]^{-1}\right)^{-1} \ \left(\mathbb{V}[\hat{\beta}_B|X]^{-1}\hat{\beta}_B + \mathbb{V}[\hat{\beta}_w|X]^{-1}\hat{\beta}_W\right) \ .$$

Note that there are no covariance-terms because $\hat{\beta}_B$ and $\hat{\beta}_W$ are uncorrelated under the RE-assumptions. To see this, note that $y_i = X_i'\beta + \iota\alpha_i + u_i$ and $M\iota = 0$. This allows us to write

$$\hat{\beta}_B = \beta + \left(\frac{1}{n}\sum_{i=1}^n X_i'PX_i\right)^{-1} \frac{1}{n}\sum_{i=1}^n X_i'P(\iota\alpha_i + u_i) \ , \quad \hat{\beta}_W = \beta + \left(\frac{1}{n}\sum_{i=1}^n X_i'MX_i\right)^{-1} \frac{1}{n}\sum_{i=1}^n X_i'Mu_i \ .$$

The two are uncorrelated because

$$\mathbb{E}[P(\iota\alpha_i + u_i)u_i'M'|X] = P\iota\mathbb{E}[\alpha_i u_i'|X]M + P\mathbb{E}[u_i u_i'|X]M = P\cdot\sigma^2 I\cdot M = 0 \ ,$$

provided that $\mathbb{E}[\alpha_i u_i'|X] = 0$.

We get

$$\begin{aligned}\mathbb{V}[\hat{\beta}_B|X] &= \left(\frac{1}{n}\sum_{i=1}^n X_i'PX_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n X_i'P(\sigma_\alpha^2\iota\iota' + \sigma_u^2 I)PX_i\right)\left(\frac{1}{n}\sum_{i=1}^n X_i'PX_i\right)^{-1} \\ &= \left(T\sigma_\alpha^2 + \sigma_u^2\right)\left(\frac{1}{n}\sum_{i=1}^n X_i'PX_i\right)^{-1} \ ,\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}[\hat{\beta}_W|X] &= \left(\frac{1}{n}\sum_{i=1}^n X_i'MX_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n X_i'M(\sigma_u^2 I)MX_i\right)\left(\frac{1}{n}\sum_{i=1}^n X_i'MX_i\right)^{-1} \\ &= \sigma_u^2\left(\frac{1}{n}\sum_{i=1}^n X_i'MX_i\right)^{-1} \ ,\end{aligned}$$

Therefore,

$$
\begin{aligned}
\hat{\beta}_{RE,GLS} &= \left[\frac{1}{T\sigma_\alpha^2+\sigma_u^2}\left(\frac{1}{n}\sum_{i=1}^n X_i'PX_i\right) + \frac{1}{\sigma_u^2}\left(\frac{1}{N}\sum_{i=1}^n X_i'MX_i\right)\right]^{-1} \\
&\quad \cdot \left[\frac{1}{T\sigma_\alpha^2+\sigma_u^2}\left(\frac{1}{n}\sum_{i=1}^n X_i'Py_i\right) + \frac{1}{\sigma_u^2}\left(\frac{1}{n}\sum_{i=1}^n X_i'My_i\right)\ \right] \\
&= \left[\frac{1}{n}\sum_{i=1}^n X_i'\Omega^{-1}X_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^n X_i'\Omega^{-1}y_i\ \right]\ ,
\end{aligned}
$$

where the last equality follows because $\left(\frac{1}{T\sigma_\alpha^2+\sigma_u^2}P + \frac{1}{\sigma_u^2}M\right)\Omega = I$ since $\Omega = \sigma_\alpha^2\iota\iota' + \sigma_u^2 I$.

Within-Estimator as CRE-Estimator Suppose $y_i = \beta x_i + \iota\alpha_i + u_i$, with β being a scalar (i.e. x_{it} containing a single regressor) and we assume $\alpha_i = \delta\bar{x}_i + \eta_i$, with $\bar{x}_i = \frac{1}{T}\sum_{i=1}^n x_{it}$. As shown in the following, we can then derive the Within-estimator as a CRE estimator.

We get

$$y_i = \beta x_i + \delta\iota\bar{x}_i + v_i = \beta\left(x_i - \iota\bar{x}_i\right) + (\delta+\beta)\iota\bar{x}_i + v_i = Z_i\gamma + v_i\ ,$$

where $\gamma = (\beta, \beta+\delta)'$ and $Z_i = [(x_i - \iota\bar{x}_i)\ ,\ \iota\bar{x}_i]$. The resulting OLS estimator for γ is a CRE estimator, because it is derived using an assumption on the nature of the correlation between α_i and x_i. The implied CRE estimator for β, $\hat{\beta} = \hat{\gamma}_1$ is equal to the Within-estimator. To see this, note that

$$\hat{\gamma} = \left(\frac{1}{n}\sum_{i=1}^n Z_i'Z_i\right)^{-1}\frac{1}{n}Z_i'y_i = \left(\frac{1}{n}\sum_{i=1}^n [(P_\iota+M_\iota)Z_i]'[(P_\iota+M_\iota)Z_i]\right)^{-1}\frac{1}{n}\sum_{i=1}^n [(P_\iota+M_\iota)Z_i]'y_i\ ,$$

whereby $P_\iota Z_i = [0\ ,\ \iota\bar{x}_i]$ and $M_\iota Z_i = [(x_i - \iota\bar{x}_i)\ ,\ 0] = [M_\iota x_i\ ,\ 0]$. As a result, the first element of $\hat{\gamma}$ is equal to

$$\hat{\gamma}_1 = \left(\frac{1}{n}\sum_{i=1}^n (M_\iota x_i)'(M_\iota x_i)\right)^{-1}\frac{1}{n}\sum_{i=1}^n (M_\iota x_i)'y_i = \hat{\beta}_W\ .$$