Panel Data Estimation

Alfonso Novales
Departamento de Economía Cuantitativa
Universidad Complutense

November 2011

Contents

1 Panel data sets

Economic data sets that combine time series and cross sections are increasingly being available. Sometimes, they are created by a researcher that collects data on a given set of variables over a period of time for a set of countries. But often, they are produced because a cross section of individuals or firms are followed over time, and the values of some of their characteristics and decisions are collected in what is known as a Panel Data set. Examples of the latter are:

- National Longitudinal Surveys on Labor Market Experience (NLS) http://www.bls.gov/nls/nlsdoc.htm,
- Michigan Panel Study of Income Dynamics (PSID) http://psidonline.isr.umich.edu/ in which 8,000 families and 15,000 individuals, interviewed periodically from 1968 to the present.
- The Bank of Spain puts together the Encuesta Financiera de las Familias, http://www.bde.es/estadis/eff/eff.htm, a still short panel data on financial decisions.
- British Household Panel Survey (BHPS), http://www.iser.essex.ac.uk/ulsc/bhps, follows several thousand housegholds (over 5,000) anually, since 1991.
- German Socioeconomic Panel Data (GSOEP), http://dpls.dacc.wisc.edu/apdu/gsoep_cd_TOC.html,
- Medical Expenditure Panel Survey (MEPS), http://www.meps.ahrq.gov/
- Current Population Survey(CPS), http://www.census.gov/eps/, is a monthly survey of about 50,000 households. Each household is interviewed each month over a 4-month period, followed by a 8-month period without interviews, to be interviewed again afterwards. These are known as rotation panels.

A panel data has a cross section (N) and a time dimension (T). Depending on the type of panel Usually, the time dimension of the panel (T) is short, with a very large cross-sectional dimension (N). In that case, we search for consistency of estimates along the N-dimension. This is because panel data are usually oriented toward cross-section analysis, and heterogeneity across units is the central focus of the analysis. However, other possibilities also exist, like having relatively long time series for a short number of countries.

The general, linear panel data model is of the form:

$$y_{it} = x_{it}'\beta_i + z_i'\alpha + \varepsilon_{it},\ i = 1, 2, ..., N,\ t = 1, 2, ..., T$$

in which variables in vector x_{it} change over time and across individuals, while those in vector z_i change only across individuals while remaining constant over time for each individual. The specification above is generally designed for a large N, short T. The model above would then imply estimation of a large number of parameters, so it is usually assumed that coefficients are the same for all individuals, to allow for enough degrees of freedom. An example would estimate how family income, as well as the age and the level of education of the household head help affect family savings:

$$family\ savings_{it} = \alpha + \beta_1 income_{it} + \beta_2 age_{it} + \beta_3 educ_{it} + u_{it}$$

A panel data is very different of a SURE system of equations. In the latter, we have a set of equations with a different endogenous variable in each one of them. In a panel data we have always the same endogenous variable. We could see it as a system of equations for each time period, but it has a tight structure, that determines the correlation structure of the error term,as we will see later, contrary to what happens in a SURE system where we have to proceed by assumption. A panel data set is said to be *balanced* when all individuals are observed for the same number of time periods, while it is *unbalanced* when the opposite happens. If there is some self-selection, with individuals deciding when to be interviewed, or some systematic decision on when to interview subjects, then estimates may easily be biased. This requires some special treatment.

Some examples:

$$\begin{aligned}\ln(wage_{it}) &= \beta_0 + \beta_0 D91_t + \beta_2 D92_t + \beta_3 computer_{it} + \beta_4 \exp er_{it} + \\ &+ \beta_5 educ_{it} + \beta_6 female_i + u_{it}\end{aligned}$$

which is considered by Wooldridge (2002) to estimate the effect of computer usage (measured by hours of use in year t) on wages. The dummy variable $female_i$ is invariant through time, as it might be the case with the number of years of education ($educ_{it}$). Two dummy variables, invariant across the cross-section, are also included to allow for a time effect on wages. This specification allows for intercepts specific of each decision unit, while slope coefficients are assumed to be the same for each individual. We could also allow for cross effects by introducing the product of some explanatory variables like $computer_{it}$ and $female_i$.

A *difficulty* when working with panel data is that since we repeatedly observe the same units, it is usually no longer appropriate to assume that observations are independent, which may complicate the analysis in dynamic and nonlinear models. On the other hand, an *advantage* is that it allows us to deal with unobserved characteristics, and to identify certain facts at the individual level. Panel data are not only suitable to model why individuals behave differently, but also to model why a given unit behaves differently at different points in time. The double dimension structure of the panel data allows for testing hypothesis that could not be addressed in either a single cross-section or in a single set of time series: does consumption increase by 2% because everybody increases consumption by 2% or because half of the population increases consumption by 4%?. Ben-Porath (1973) observed that over time, 50% of women appear to be working at any time period. However, it is unclear whether these are always the same women or rather, each woman has a probability of 1/2 of being working at any time period. The two possibilities would have very different policy implications. Another typical example refers to the possibility of separating economies of scale from technological change. The former could be explored in a cross section, while the second is a proper hypothesis for tine series data, although then, the two effects would be confused. Usually constant returns to scales is assumed and then the time series data is used to test for technological change. A panel data can provide information on both issues at the same time.

Panel data techniques have clear advantages in dealing with unobserved individual characteristics. Consider estimates of a Cobb-Douglas production function with data on a number of firms. Suppose the true model is,

$$y_{it} = \mu + x_{it}'\beta + m_i\beta_{k+1} + \varepsilon_{it}$$

where m_i is the management quality for firm i, which is assumed to be constant over time. The unobserved m_i variable is expected to be negatively correlated with the other explanatory variables, since a high quality management will possibly require a more efficient use of inputs. Therefore, excluding m_i from the estimation because of not being observable will bias estimates for the other parameters. With panel data, we can consider a firm specific effect, defined as $\alpha_i = \mu + m_i\beta_{k+1}$, and even hope to estimate its size, although it will be impossible to identify β_{k+1} by itself.

Similarly, a fixed time effect can be included in the model to capture the effect of all (observed and unobserved) variables that do not vary across the individual units. A final, more technical advantage, is that panel data models provide internal instruments for regressors that are endogenous, or are subject to measurement error. Usually, it can be argued that some transformations of the original variables are uncorrelated with the model's error term while being correlated with the explanatory variables themselves. This is interesting, since external instruments, which are often harder to justify, or for which data may be hard to find, may not be needed. For instance, if x_{it} is correlated with an omitted explanatory variable α_i (which will then be part of the error term), it can be argued that $x_{it} - \bar{x}_i$, where $\bar{x}_i$ is the time average for individual i, is uncorrelated with α_i and hence, it provides a valid instrument for x_{it}.

1.1 Estimation approaches

The individual or group time-invariant effects in z_i may be *observed*, like sex, race, location, or *unobserved*, like family specific characteristics, individual heterogeneity in skill or preferences, all of them being constant over time. If z_i is *observed* for all individuals, the model can be handled easily, as a standard regression model, to estimate vectors β and α in,

$$y_{it} = x_{it}'\beta_i + z_i'\alpha + \varepsilon_{it},\ i = 1, 2, ..., N,\ t = 1, 2, ..., T$$

which is identified by the standard condition,

$$E(\varepsilon_{it}/x_{it}, z_i) = 0 \forall i, t$$

This condition implies,

$$E(y_{it}/x_{it}, z_i) = x_{it}'\beta_i + z_i'\alpha,\ i = 1, 2, ..., N,\ t = 1, 2, ..., T$$

As in any regression model, this expectation is what we are interested on. Often, the error term ε_{it} is also assumed to be independent and identically distributed over individuals and over time, with mean zero and variance σ_ε^2.

Obviously, in the usually available short panels, the individual specific effects cannot be consistently estimated. Besides, the small number of observations would lead to a huge loss of precision. Hence, we need to collapse the linear combination of individual characteristics $z_i'\alpha$ into a single number, $z_i'\alpha = \alpha_i$,

$$y_{it} = x_{it}'\beta_i + \alpha_i + \varepsilon_{it},\ i = 1, 2, ..., N,\ t = 1, 2, ..., T$$

We then substitute our interest on the previous conditional expectation, $E(y_{it}/x_{it}, z_i)$, by a focus on:

$$E\left(y_t/x_{it}\right) = E\left(\alpha_i/x_{it}\right) + x_{it}'\beta$$

An important complication arises under standard estimation procedures when z_i is *unobservable.* Examples include the determination of wages on the basis of experience and education, with no observation of the productivity of the worker, or a study on health status of individuals with no data on usage of health services.Also, the determination of profits at the firm level lacking data on the quality of management. We cannot then compute the expectation conditioned on the values of these unobserved variables.

Here, there are two possibilities: if we are willing to accept the *Mean-independence* assumption, that the unobserved individual characteristics are independent of the variables in x_{it}: $E\left(\alpha_i/x_{it}\right) = \mu_i$, constant, we will have,

$$E\left(y_t/x_{it}\right) = \mu_i + x_{it}'\beta$$

and the model has an error term with two different components,

$$y_{it} = \mu_i + x_{it}'\beta + \left[\varepsilon_{it} + (\alpha_i - \mu_i)\right]$$

Under the Mean independence assumption, this specification does not pose serious estimation difficulties. This leads to the *Random Effects* model.

However, in many applications it may be natural to believe that z_i and x_{it} will be correlated, so that $E\left(\alpha_i/x_{it}\right) = \mu_i + h(x_{it})$, and this dependence will be incorporated into the error term,

$$y_{it} = \mu_i + x_{it}'\beta + \left[\varepsilon_{it} + (\alpha_i - \{E\left(\alpha_i/x_{it}\right) - h(x_{it})\})\right]$$

This leads to the *Fixed Effects* model. The correlation between explanatory variables and the error term will then lead to inconsistent least-squares estimates, so whenever there is reason to believe that unobserved individual effects are correlated with the observed explanatory variables, we need to explore alternative estimation approaches.[1]

[1] We are usually interested in estimating the partial effects:

The estimation approach suggested depends on the assumptions on the correlations between ε_{it} and either $z_i\alpha$ or α_i.

- *Pooled regression*: Vector z_i contains only a constant term, the same for all individuals in the sample. Ordinary least squares estimates of the common parameters α and β in

$$y_{it} = \alpha + x_{it}'\beta + \varepsilon_{it}$$

using all the data on all the individuals for all time periods are then consistent and efficient.

- *Fixed effects*: If some z_i are unobserved, but correlated with some x_{it}, we have,

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

where $\alpha_i = z_i'\alpha$ captures all individual specific effects. The least squares estimator of β is biased and inconsistent, because of the omitted variable bias.

The *Fixed Effects* approach considers α_i as an individual-specific constant term in the regression. The term "fixed" does not refer to the individual effect being non-stochastic but rather, to being correlated with the variables in x_{it}. It will be impossible with this specification to distinguish between α_i and any other individual effect that is constant over time, so can just hope to identify a single individual-specific effect. The estimation approach in this situation will consist of transforming the data so as to get rid of the individual effects producing the inconsistent estimates. We can estimate constant individual specific effects, α_i, that can be treated in estimation as N unknown parameters, and the model is referred to as the *Fixed Effects model.* Because of these data transformations, we will have some difficulty in identifying the effects of time-invariant characteristics, like race or gender.

- *Random effects*: If the unobserved heterogeneity can be assumed to be uncorrelated with any other explanatory variable, and we assume that individual effects can be jointly considered as $z_i'\alpha = \mu + \alpha_i$, with $\alpha_i \sim [0, \sigma_\alpha^2]$, the model can then be written,

$$y_{it} = x_{it}'\beta + E(z_i'\alpha/x_{it}) + [z_i'\alpha - E(z_i'\alpha/x_{it})] + \varepsilon_{it} = x_{it}'\beta + \mu + (\alpha_i + \varepsilon_{it}) = x_{it}'\beta + \mu + u_{it}$$

$$\frac{\partial E[y_t/x_{jt}]}{\partial x_{jt}} = \beta_j, \; j = 1, 2, ..., k \;\; for\ all\ t$$

after correcting for individual characteristics. These marginal effects can be identified even if the conditional mean is not. For instance, it is possible to identify the effects on earnings of an additional year of schooling, controlling for individual effects, even though the individual effects and the conditional mean are not identified.

where $u_{it} = \alpha_i + \varepsilon_{it}$, with α_i being an individual specific element similar to ε_{it}, except for the fact that there is a single draw for α_i that enters the regression identically every period. Individual intercepts are then treated as draws from a distribution with mean α and variance σ_α^2. The essential *assumption* is that these draws are independent of the explanatory variables in x_{it}. The error term has then two components, a time invariant component, α_i, and the ε_{it} component, which is uncorrelated over time. It is sometimes referred to as *Random effects model or Error Components model*.

The presence of the α_i component in the error term induces necessarily some autocorrelation structure, even if the original error term in the model ε_{it} was independent over time and across individuals, since:

$$E(w_{it}.w_{is}) = E\left[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})\right] = \sigma_\alpha^2 \ if \ t \neq s \ and \ = \sigma_\alpha^2 + \sigma_\varepsilon^2 \ if \ t = s$$

2 The static linear model

2.1 Pooled OLS estimates

Consider the general panel data model,

$$y_{it} = z_i'\alpha + x_{it}'\beta + \varepsilon_{it}, \ i = 1, 2, ..., N; \ t = 1, 2, ..., T$$

where we assume that individual characteristics are either observable, or non-observable but uncorrelated with the variables in x_{it}.

Suppose that we are willing to make the crucial assumption:

$$E\left(z_i'\alpha / X_i\right) = \mu \ \forall i$$

Then,

$$y_{it} = \mu + x_{it}'\beta + \left[\varepsilon_{it} + \left(z_i'\alpha - E\left(z_i'\alpha/X_i\right)\right)\right], \ i = 1, 2, ..., N; \ t = 1, 2, ..., T$$

and we will have the same vector of coefficients across individuals or decision units. Here X_i includes both, the observable z_i and the x_{it} variables. We can then write the panel data model as the system:

$$\begin{aligned} y_{1t} &= x_{1t}'\beta + u_{1t}, \ t = 1, 2, ..., T_1 \\ y_{2t} &= x_{2t}'\beta + u_{2t}, \ t = 1, 2, ..., T_2 \\ & ... \\ y_{Nt} &= x_{Nt}'\beta + u_{Nt}, \ t = 1, 2, ..., T_N \end{aligned}$$

with error term: $u_{it} = \varepsilon_{it} + \left(z_i'\alpha - E\left(z_i'\alpha/X_i\right)\right)$, and we can think of the model as having a single regression with:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ ... \\ X_N \end{pmatrix} ; \; y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_N \end{pmatrix} ; \; \beta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

where each X_i matrix is T_ixk, while y_i is T_ix1.

But the central point of this model is that the assumption we have made on $E\left(z_i'\alpha/X_i\right) = \mu$ is inappropriate in most panel data situations, in which the opposite will be likely to occur.

The pooled OLS estimator consists of applying OLS to the stacked y and X above:

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^{N} X_i'X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i'y_i\right) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}'x_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}'y_{it}\right)$$

The properties of the alternative estimators will depend on two things: a) the stochastic characteristics of the errror term in the original model, ε_{it}, and b) the relationship between the unobservable, ommitted individual characteristics, and the variables included in the model.

Regarding the first point, alternative possibilities are:

- The x_{it} are contemporaneously exogenous: $E(\varepsilon_{it}/x_{it}) = 0$
- A stronger assumption: The x_{it} are *strictly* exogenous: $E(\varepsilon_{it}/x_{is}) = 0 \; \forall t, s$

Strict exogeneity fails if $x_{it} = (1, y_{it-1})$, because: $E(\varepsilon_{it}/x_{i1,}x_{i2}, ..., x_{iT}) = E(\varepsilon_{it}/y_0, y_1, ..., y_{T-1}) = \varepsilon_{it}$

For the OLS estimator to be consistent we need lack of correlation between explanatory variables and error term, together with existence of second order moments of explanatory variables.

Consistency

The estimator is consistent for $N \to \infty$ under conditions: i) $E(x_{it}'u_{it}) = 0_k, t = 1, 2, ..., T$, ii) $rank\left[E(\sum_{t=1}^{T} x_{it}'x_{it})\right] = k$,with an asymptotic probability distribution:

$$\sqrt{N}\left(\hat{\beta}_{POLS} - \beta\right) \underset{d}{\rightarrow} N\left(0, A^{-1}BA^{-1}\right)$$

where $A = E(X_i'X_i)$ is estimated by $\hat{A} = N^{-1}\sum_{i=1}^{N} X_i'X_i$, and $B = Var(X_i'u_i) = E(X_i'u_iu_i'X_i)$ is estimated by $\hat{B} = N^{-1}\sum_{i=1}^{N} X_i'\hat{u}_i\hat{u}_i'X_i$, [2] so that

[2] This is the generalization of the standard variance-covariance matrix for the OLS estimator: $\Sigma(\hat{\beta}_{OLS}) = (X'X)^{-1}(X'\Sigma X)(X'X)^{-1}$

the covariance matrix is estimated by:[3]

$$\Sigma(\hat{\beta}_{POLS}) = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}X_i'X_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}X_i'\hat{u}_i\hat{u}_i'X_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}X_i'X_i\right)^{-1} =$$
$$\left(\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}x_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\left(\sum_{t=1}^{T}x_{it}\hat{u}_{it}\right)\sum_{i=1}^{N}\left(\sum_{t=1}^{T}x_{it}\hat{u}_{it}\right)'\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}x_{it}'\right)^{-1}$$

Efficiency

It will not be an efficient estimator, because the structure of the error term induces autocorrelation: $E(u_{it}u_{is}) = \sigma_u^2$ when $t \neq s$. The variance-covariance matrix above incorporates the fact that unobserved individual characteristics introduce autocorrelation in the error term. The practical consequence of the described autocorrelation is that with the panel data we have less information than with NT independent observations.

Unobservable individual effects We now suppose that some of the individual effects are not observable, and we include them into a single variable α_i. Let us denote $u_{it} = \varepsilon_{it} + \alpha_i,\ i = 1, 2, ..., N;\ t = 1, 2, ..., T$. That would have two implications:

1. if any of the unobservables in α_i is correlated with any of the x_{it} variables, then condition $i)$ above will no longer hold, and the pooled least squares estimate will be biased and inconsistent,

2. estimating by pooled least-squares we have that the presence of individual effects in the error term introduces a specific form of autocorrelation, because error terms corresponding to a same individual will be correlated with each other:

$$E(u_{it}u_{is}) = \sigma_\alpha^2,\ t \neq s$$

The estimate of the variance-covariance matrix proposed in White (19xx) is robust against possible cross correlation among error terms across equations in the same time period, or against a different variance for the error term in each equation (time-varying variances). The conditional variance is also allowed to

[3] Since $X_i = \begin{pmatrix} x_{i11} & ... & x_{i1T} \\ ... & ... & ... \\ x_{ik1} & ... & x_{ikT} \end{pmatrix}$, then: $\sum_{i=1}^{N} X_i'X_i = \sum_{i=1}^{N}\begin{pmatrix} \sum_{t=1}^{T} x_{i1t}^2 & ... & \sum_{t=1}^{T} x_{i1t}x_{ikt} \\ ... & ... & ... \\ \sum_{t=1}^{T} x_{ikt}x_{i1t} & ... & \sum_{t=1}^{T} x_{ikt}^2 \end{pmatrix}$,the same kxk matrix we obtain from adding up over $i = 1, 2, ..., N$ and over time the kxk matrices of products: $x_{it}x_{it}' = \begin{pmatrix} x_{i1t} \\ ... \\ x_{ikt} \end{pmatrix}\begin{pmatrix} x_{i1t} & ... & x_{ikt} \end{pmatrix}$

depend on X_i arbitrarily. However, it does not take into account the possible autocorrelation of the error term, as it will be the case if we estimate by Pooled least-squares the *Fixed Effects* model. This is taken into account in the estimate proposed above. Alternatively, we can follow the Newey-West approach to obtain a *panel-robust estimate of the variance-covariance matrix*:

$$\begin{aligned}\Sigma(\hat{\beta}_{POLS}) &= \left(\sum_{i=1}^{N} X_i' X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' \hat{u}_i \hat{u}_i' X_i\right) \left(\sum_{i=1}^{N} X_i' X_i\right)^{-1} = \\ & \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} (\hat{u}_{it}x_{it})(\hat{u}_{is}x_{is})'\right) \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1}\end{aligned}$$

If the conditional covariance of u_{it} is independent of x_{is} for all s, then,

$$\Sigma(\hat{\beta}_{POLS}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1} \left[\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N} \hat{u}_{it}\hat{u}_{is}\right) x_{it}x_{is}'\right] \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1}$$

We need to be aware of the fact that the term 'robust' applied to the variance-covariance matrix produced by some statistical packages may refer to just the correction for heteroskedasticity. However, in many relevant cases, the important effect in panel data is the autocorrelation induced by the repeated observations in a same individual.

Example: Using the Cornwell-Ruport (1988) data set, Green (6ed.), p187, example 9.1, show estimates of the returns to schooling by an equation in which logged wages are explained by working experience, their squared value, weeks worked, years of education, and a set of dummy variables to represent whether a given worker: has a blue collar occupation, works in manufacturing industry, resides in the south, resides in an SMSA, is married, the wage is set by a union contract, is a female, is black. The sample is made of 595 workers, which are followed over a 7-year period, 1976-1982. Each year of education is estimated to increase wages by 5.67%. OLS standard errors are in this example of similar size to White's robust standard errors, while both of them are about half size of the Panel robust standard errors. It means that ignoring the within-group correlations in this case matters a lot, substantially affecting inference through the implied autocorrelation of the error term.

The model can also be estimated using individual sample means, for a sample of 595 observations. We will still have the inconsistency of least-squares estimates in the fixed effects model, but the within-group autocorrelation now disappears. Table 9.2 in Green (6ed.) shows similar coefficient estimates. White's robust standard errors are now similar to the Panel robust standard errors for the whole Panel data sample.

2.1.1 Hypothesis testing

Linear hypothesis of the form: $H_0 : R\beta = r$ can be tested by the usual Wald statistic:

$$W = (R\beta - r)' \left(R\Sigma(\hat{\beta}_{POLS})R' \right) (R\beta = r)$$

that obeys a chi-square distribution with q degrees of freedom, q being the number of rows in R and r (the number of independent restrictions being tested).

2.2 Generalized pooled least squares estimation

When we have some structure on the form of the conditional covariance matrix of u_i, we can prefer to use GLS estimation, in search of improved efficiency. Since we use random sampling, the unconditional covariance matrix should be the same for each observation unit: $\Omega = E(u_i u_i')$, a TxT matrix. As usual, the numerical values of the elements in the variance-covariance will be unknown, and we will have to estimate them first, then moving into what is usually known as Feasible GLS estimation ($FGLS$).

Remember we have one equation for each time period, with N observations in each equation. It is important to bear in mind that consistency of GLS estimator needs of a stronger condition on lack of correlation between explanatory variables and error terms. Now, each element in X_i must be uncorrelated with u_i [Wooldridge (2002)]. This is because for consistency we now need $p\lim \left(\frac{1}{N} \sum_{i=1}^{N} X_i' \Omega u_i \right) = 0$. A typical case when this will not hold is in dynamic panel data estimates under autocorrelation of the error term.

To construct the GLS estimator, we would follow the standard practice of pre-multiplying the equation by $\Omega^{-1/2}$, and:

$$\hat{\beta}_{GLS} = \left(\sum_{i=1}^{N} X_i' \Omega X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \Omega y_i \right)$$

The reason to need a more strict condition on lack of correlation to show consistency is that we now need: $E(X_i' \Omega u_i) = 0_k$.The asymptotic distribution is:

$$\sqrt{N} \left(\hat{\beta}_{GLS} - \beta \right) \underset{d}{\rightarrow} N \left(0, A^{-1} B A^{-1} \right)$$

where $A = E(X_i' \Omega^{-1} X_i)$ and $B = E(X_i' \Omega^{-1} u_i u_i' \Omega^{-1} X_i)$ which are estimated by using a consistent estimate $\hat{\Omega}$ of Ω , computed using the residuals from a first-step set of consistent, but inefficient least squares regressions.

In most applications, it is natural to assume that: $E(X_i' \Omega^{-1} u_i u_i' \Omega^{-1} X_i) = E(X_i' \Omega^{-1} X_i)$, which implies $B = A$ and hence, the asymptotic variance of $\hat{\beta}_{GLS}$ becomes: $\Sigma(\hat{\beta}_{GLS}) = A^{-1}/N = \left[E(X_i' \Omega^{-1} X_i) \right]^{-1} /N$, which can be estimated by: $\hat{\Sigma}(\hat{\beta}_{GLS}) = (\sum_{i=1}^{N} X_i' \Omega^{-1} X_i)$. This assumption essentially requires conditional homoskedasticity (constant conditional variances and covariances), *i.e.*, that the expectation $E(u_i'\Omega^{-1}X_i)$ does not depend on X_i.

3 The Fixed Effects model

This model embeds the idea that all the *unobservable* individual effects for each observation are aggregated in a single term α_i.Under the assumption that:

$$E(\alpha_i/X_i) = h(X_i),$$

is constant over time, that constant being the *Fixed individual Effect*, each individual effect α_i can be treated as an unknown parameter to be estimated, and we get a linear regression model in which the intercept is allowed to vary across individuals,

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

with

$$u_{it} = \varepsilon_{it} + (\alpha_i - h(X_i)), \ with\ \varepsilon_{it} \sim i., i.d.(0, \sigma_\varepsilon^2)$$

The model will usually imply a rather large number N of regressors which it could lead to a noticeable loss of precision. It can be implemented in a simpler way by taking into account that individual effects disappear if we apply the *Within transformation*, to transform the data in deviations with respect to individual means. Taking averages in the previous equation: $\bar{u}_i = \bar{\varepsilon}_i + (\alpha_i - h(X_i))$, so that $u_{it} - \bar{u}_i = \varepsilon_{it} - \bar{\varepsilon}_i$, and:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i),\ i = 1, 2, ..., N;\ t = 1, 2, ..., T$$

Applying least squares to this model, we get the *Within estimator* of the *Fixed Effects* model,

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i) \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(y_{it} - \bar{y}_i);$$

The estimator will be consistent as $N \to \infty$ if $E[(x_{it} - \bar{x}_i)\varepsilon_{it}] = 0$.This will hold if x_{it} is uncorrelated with ε_{it} and $\bar{x}_i$ has no correlation with the error term. These are implied by *strict exogeneity* of the regressors:

$$E(x_{it}\varepsilon_{is}) = 0\ \forall t, s$$

Strict exogeneity precludes the inclusion in x_{it} of lagged dependent variables or variables that depend upon the history of y_{it}. For instance, explaining labour supply of an individual, we may want to include as a regressor years of experience, but experience will clearly depend upon the person's labour history.

By applying the *Within* transformation, the individual specific constant characteristics will have dropped from the model. Individual effects can later be recovered by,

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i{}'\hat{\beta}_{FE},\ i = 1, 2, ..., N$$

which are unbiased, but will not be consistent if just the cross-section dimension tends to infinity. For consistency we will need $T \to \infty$. The reason that these are not consistent as $N \to \infty$ if T is short is that that leaves us with a very limited amount of information to estimate each individual fixed effect, and x and y averages do not converge to any well defined limit as the number of individuals increases. This is an interesting situation in which it is possible to estimate the β coefficients consistently, even if the α_i cannot be estimated consistently because of a short time dimension.

As we can see, we can just recover a single α_i variable for each individual, which is the reason why the specific effects for a same infdividual need to be considered as aggregated in a single variable α_i.

If there are some *observed* individual effects z_i, their joint influence can be recovered by regression,

$$(\bar{y}_i - \bar{x}_i{}'\hat{\beta}_{FE}) = z_i'\gamma + \left[\alpha_i + \bar{\varepsilon}_i - \bar{x}_i{}'\left(\hat{\beta}_{FE} - \beta\right)\right]$$

leading to consistent estimates of γ if each variable in vector z is uncorrelated with ε_{it} and with α_i. As in the case of the unobservable, time invariant individual effects, the estimated coefficient will not be very reliable with a short time dimension T .

The variance-covariance matrix,

$$Var(\hat{\beta}_{FE}) = \sigma_\epsilon^2 \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i)\right)^{-1}$$

assumes that individual effects are independent across individuals and time. Unless T is large, this will underestimate the true variance. The reason is that the error covariance matrix in the transformed regression is singular (since the T transformed errors for each individual add up to zero), and the variance of $\varepsilon_{it} - \bar{\varepsilon}_i$ is $(T-1)/T\sigma_\varepsilon^2$, rather than σ_ε^2. If, for instance, $T = 3$, then the variance of $\varepsilon_{it} - \bar{\varepsilon}_i$ will be $\frac{2\sigma_\varepsilon^2}{3}$. A consistent estimator for σ_ε^2 can obtained from the Within groups estimation,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}_{FE})^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)'\hat{\beta}_{FE}\right]^2$$

although the appropriate number of degrees of freedom would be $N(T-1)-k$, and we will have to introduce a correction factor.

A panel-robust estimate of the variance-covariance matrix is,

$$\Sigma(\hat{\beta}_{POLS}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'\right)^{-1}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}(\varepsilon_{it}-\bar{\varepsilon}_i)(\varepsilon_{is}-\bar{\varepsilon}_i)'(x_{is}-\bar{x}_i)(x_{it}-\bar{x}_i)'\right].$$
$$\cdot\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'\right)^{-1}$$

A variance for individual effects can be obtained from:

$$Var(\hat{\alpha}_i) = \frac{\sigma_\epsilon^2}{T} + \bar{x}_i' Var(\hat{\beta}_{FE})\bar{x}_i$$

showing that estimates of individual effects α_i are *inconsistent*, since even though $Var(\hat{\beta}_{FE})$ converges to zero with N, that is not the case with the first term in $Var(\hat{\alpha}_i)$. This is because of estimating each individual effect with a small number T of observations.

Defining N individual dummy variables ($D_{ij} = 1$,if $i = j, j = 1, 2, ..., N$, and $D_{ij} = 0$ otherwise) the model can also be written,

$$y_{it} = \sum_{j=1}^{N} \alpha_j D_{ij} + x_{it}'\beta + \varepsilon_{it}$$

which is known as the *least squares dummy variable (LSDV) estimator*. As mentioned above, a limitation of this model is that all time invariant, unobservable individual effects get confused with each other in a single α_i variable for each individual, and we are just able to estimate their aggregate influence over y_{it}.

3.0.1 Testing the significance of the group effects

Even though we can use the above results to tests for significance of either one of the individual effects, the natural hypothesis is to test that they are all equal to each other. If that is the case, the restricted model leads to the pooled least squares estimate, and we have an F-test,

$$F(N-1, NT-N-k) = \frac{(R^2_{LSDV} - R^2_{POLS})/(N-1)}{(1-R^2_{LSDV})/(NT-N-k)}$$

The correction for the F-test comes from the fact that in the Pooled OLS estimator we have $NT - k - 1$ coefficients, while in the $LSDV$ estimator we estimated $NT - N - k$ coefficients, with a difference of $N - 1$.

3.0.2 Fixed time effects

The model can be extended to accommodate fixed time effects through time dummy variables. However, to avoid perfect collinearity, we should just include $T - 1$ of the possible time effects. Alternatively, we can specify the model,

$$y_{it} = x'_{it}\beta + \mu + \alpha_i + \delta_t + \varepsilon_{it},\ with\ \sum_{i=1}^{N} \alpha_i = \sum_{t=1}^{T} \delta_t = 0$$

Least-squares estimates of the slopes β can be obtained by a regression of y^*_{it} on vector x^*_{it}, with,

$$y^*_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}, \bar{y}_t = N^{-1} \sum_{t=1}^{N} y_{it}, \bar{y} = N^{-1}T^{-1} \sum_{t=1}^{N} \sum_{t=1}^{T} y_{it}$, and similar expressions apply to vector x.

Once we have estimates for the vector β, we can recover estimates for the remaining parameters from,

$$\begin{aligned} \hat{\mu} &= \bar{y} - \bar{x}'\hat{\beta} \\ \hat{\alpha}_i &= (\bar{y}_i - \bar{y}) - (\bar{x}_i - \bar{x})'\hat{\beta} \\ \delta_t &= (\bar{y}_t - \bar{y}) - (\bar{x}_t - \bar{x})'\hat{\beta} \end{aligned}$$

The variance-covariance matrix is obtained from the standard cross-moment product of transformed explanatory variables, with an estimate of $\hat{\sigma}^2_\varepsilon$ being obtained from $\hat{\sigma}^2_\varepsilon = RSS/[NT-(N-1)-(T-1)-k-1]$. As we will see below, there are more general models allowing for time effects.

Example: Green (6ed.), ex. 9.4, estimates the model in the previous example, for logged wages, with a constant intercept and $T-1$ time dummies. The constant individual characteristics: education, sex (female-dummy) and race (white-dummy), need to be dropped now, so that we lose the main interest of estimating the returns to education. Pooled least squares estimates are obtained for an initial specification that includes a single, common intercept and no time dummies. A second model includes again a single intercept but also time dummies. A third specification allows allowing for individual specific intercepts and no fixed time effects, while a final model allows for both, fixed time and individual characteristics. In this final specification, we need to drop an additional time dummy variable, because the Experience variable is a natural time trend. The significance of individual effects and/or fixed time effects can now be tested by comparing the Residual Sums of Squares of appropriately chosen specifications. Green also suggests comparing the conventional estimate and the robust estimate, the latter with data in group mean deviations form, of the variance-covariance matrices as a specification test for the individual effects model. If the specification is correct, there should not be any heterogeneity in the error term and hence, not heteroscedasticity or autocorrelation left. In the example, robust standard errors are of the order of 20 times as large as the conventional ones, clearly pointing out to misspecification errors.

4 Within and between estimators

The original Panel data specification,

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

can be written in terms of group means,

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

and in deviations from group means:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\,\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

All three models could be consistently estimated (although possibly not efficiently) by least-squares. Consider the overall second order matrices,

$$S_{xx}^{total} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})(x_{it} - \bar{x})'\,;\; S_{xy}^{total} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})(y_{it} - \bar{y})'$$

the within group matrices,

$$S_{xx}^{within} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\,;\; S_{xy}^{within} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)'$$

and the between-groups matrices,

$$\begin{aligned} S_{xx}^{between} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' = T\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\,; \\ S_{xy}^{between} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})' = T\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})' \end{aligned}$$

Notice that:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(\bar{x}_i - \bar{x})' = \sum_{i=1}^{N}\left(\sum_{t=1}^{T}(x_{it} - \bar{x}_i)\right)(\bar{x}_i - \bar{x})' = 0$$

because the inside bracket is equal to zero. Therefore, we have,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})(x_{it} - \bar{x})' = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + T\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

so that,

$$S_{xx}^{total} = S_{xx}^{within} + S_{xx}^{between},\; S_{xy}^{total} = S_{xy}^{within} + S_{xy}^{between},$$

4.1 The Within groups estimator

The Within-groups estimator is defined,

$$\hat{\beta}^{within} = \left[S_{xx}^{within}\right]^{-1} S_{xy}^{within} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\bar{x}_i\right)\left(x_{it}-\bar{x}_i\right)'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\bar{x}_i\right)\left(y_{it}-\bar{y}_i\right)\right)$$

so that it is the OLS estimator in the model,

$$y_{it}-\bar{y}_i = \left(x_{it}-\bar{x}_i\right)'\beta + \left(\varepsilon_{it}-\bar{\varepsilon}_i\right),\ 1,2,...,N$$

where the possible individual specific intercepts have cancelled out. For that reason, it yields consistent estimates of the panel data model under the *Fixed Effects* assumption, whereas the Pooled OLS and the Between estimator that we are about to see, do not. The *Within Groups* estimator is the same as the *Fixed Effects* estimator and the *Least-Squares Dummy Variable* estimator that we saw above. It can also be thought of as estimating regressions from dependent and time-varying independent variables on individual dummies and estimating a regression between the residuals from these auxiliary regressions. Of course, the limitation of this approach is the impossibility to estimate the coefficients of time-invariant individual characteristics like race and gender.

4.2 The Between groups estimator

The Between groups estimator above is obtained applying least squares to the data averaged for each individual, in deviations from the global sample average,

$$\bar{y}_i-\bar{y} = \left(\bar{x}_i-\bar{x}\right)'\beta + \left(\alpha_i+\bar{\varepsilon}_i\right),\ 1,2,...,N$$

so that,

$$\hat{\beta}^{between} = \left(\sum_{i=1}^{N}(\bar{x}_i-\bar{x})'(\bar{x}_i-\bar{x})\right)^{-1}\left(\sum_{i=1}^{N}(\bar{x}_i-\bar{x})'(\bar{y}_i-\bar{y})\right) = \left[S_{xx}^{between}\right]^{-1} S_{xy}^{between}$$

This estimator is a cross section regression with N data points.The *Between* groups estimator uses just the cross-sectional variation in the data, while the pooled OLS estimator uses variation both over time and across individuals. The *Between* groups estimator uses only information on how each individual differs from the global average, ignoring the variation over time for each individual in the sample.

An interesting feature of the *Between* estimator is that it tends to reduce the effect of measurement errors, since it uses time averages. It would be consistent with $T \to \infty$ but that is un unlikely condition in most panel data sets.

Strong exogeneity is needed for consistency, since we need the individual means $\bar{x}_i$ to be uncorrelated with α_i. Sufficient, although not necessary conditions for consistency are: $E(\alpha_i x_{it}) = 0\forall t$, and $E(\varepsilon_{it}x_{is}) = 0\forall s,t$. These are of course very strict assumptions. The problem is that the transformation in differences with respect to group or individual means does not solve the issue of the possible correlation between unobserved individual characteristics and observed explanatory variables: $E(\alpha_i/X_i) = h(X_i)$. Mundlak (1978) analyzes the case when it can be assumed that such expectation is a function of the group means: $E(\alpha_i/X_i) = \bar{x}_i'\ \gamma$. We would then have: $y_{it} = \mu + x_{it}'\beta + \bar{x}_i'\gamma + \varepsilon_{it}$, and taking averages: $\bar{y}_i = \bar{x}_i'(\beta+\gamma) + \bar{\varepsilon}_i$, so that with the *Between* estimator we would be estimating the sum $\beta+\gamma$,a biased estimator of the partial effects β we are interested on.

Even when it is consistent, the *Between* estimator will be inefficient, since it does not exploit the structure of autocorrelation and heteroscedasticity in the error term.

Relationship among estimators

The least-squares estimator can be written,

$$\begin{aligned}\hat{\beta}^{total} &= \left[S_{xx}^{total}\right]^{-1} S_{xy}^{total} = \left[S_{xx}^{within} + S_{xx}^{between}\right]^{-1}\left[S_{xy}^{within} + S_{xy}^{between}\right] = \\ &= \left[S_{xx}^{within} + S_{xx}^{between}\right]^{-1}\left[S_{xx}^{within}\hat{\beta}^{Within} + S_{xx}^{between}\hat{\beta}^{Between}\right]\end{aligned}$$

and if we define:

$$\begin{aligned}F^W &= \left[S_{xx}^{within} + S_{xx}^{between}\right]^{-1} S_{xy}^{within} \\ F^B &= I - F^W = \left[S_{xx}^{within} + S_{xx}^{between}\right]^{-1}\left[S_{xx}^{within} + S_{xx}^{between}\right] - \left[S_{xx}^{within} + S_{xx}^{between}\right]^{-1} S_{xy}^{within} = \\ &= \left[S_{xx}^{within} + S_{xx}^{between}\right]^{-1} S_{xy}^{between}\end{aligned}$$

then,

$$\hat{\beta}^{total} = F^W\hat{\beta}^W + F^B\hat{\beta}^B$$

so that the least-squares estimater can be written as a matrix linear convex combination of the *Within* and the *Between* estimators: $\hat{\beta}^{total} = F\hat{\beta}^{within} + (I-F)\hat{\beta}^{between}$. We will later see that it is not the only estimator admitting such a representation.

5 Estimating in first differences

An alternative transformation that eliminates individual effects is to take time differences in the model, obtaining:

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta\varepsilon_{it},\ i = 1, 2, ..., N,\ t = 2, 3, .., T$$

even though if the error term of the original model was a white noise, the error term in the first-differenced model will have a MA(1) structure, with first-order autocorrelation. So, we have changed the autocorrelation structure of the error term.

Estimating in *First differences* is useful no matter whether the *Random Effects* or the *Fixed Effects* models are appropriate. Estimating in *First differences* may be specially indicated in panels with a very short time dimension, for which individual sample means may be subject to important sampling error. However, a limitation of this approach is again the impossibility to estimate the coefficients in any time invariant explanatory variable.

Consistency of the *First-differences* estimator requires,

$$E\left[\left(\varepsilon_{it}-\varepsilon_{i,t-1}\right)/\left(x_{it}-x_{i,t-1}\right)\right]=0$$

a stronger condition than $E\left[\varepsilon_{it}/x_{it}\right]=0$, but weaker than the strong exogeneity condition that is need for consistency of the *Within* estimator.

We have,

$$Var(\Delta\varepsilon_{it})=\sigma^2 H$$

where H is a symmetric, $(T-1)$x$(T-1)$ matrix whose elements are equal to +2 along the main diagonal, equal to -1 in the two diagonals next to the main diagonal, and equal to -1 everywhere else.

The least squares estimator is:

$$\begin{aligned}
\hat{\beta} &= \left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)\left(\Delta x_{it}\right)'\right]^{-1}\left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)\left(\Delta y_{it}\right)\right] \\
Var(\hat{\beta}) &= \sigma_{\varepsilon}^{2}\left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)\left(\Delta x_{it}\right)'\right]^{-1}\left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)H\left(\Delta x_{it}\right)'\right]\left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)\left(\Delta x_{it}\right)'\right]^{-1}
\end{aligned}$$

This approach will provide consistent, although inefficient, estimates. Matrices in these expressions have $T-1$ rows. An alternative would be to use the Newey-West robust estimate of the variance-covariance matrix, since we know the exact order of autocorrelation in the error term.

Since the structure of the covariance matrix of the error term is known, we could also try to improve efficiency by using Generalized least squares:

$$\hat{\beta}=\hat{\beta}=\left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)H^{-1}\left(\Delta x_{it}\right)'\right]^{-1}\left[\sum_{i=1}^{N}\left(\Delta x_{it}\right)H^{-1}\left(\Delta y_{it}\right)\right]$$

In practice, it is usually the case that Generalized least squares estimates in levels and in first differences are noticeably different, which suggests the existence of unobservable individual effects that bias the estimation in levels.

This approach is not preferable to other estimation methods.

It is specifically appropriate for estimation of *Treatment effects* in two-period panels, with a specification like,

$$y_{it}=\alpha_i+x_{it}'\beta+\theta S_t+\varepsilon_{it},$$

with $t = 1, 2$, where $S_t = 0$ in $t = 1$, and $S_t = 1$ in $t = 2$. The first period is the before-treatment period, while the second period comes after the treatment has been applied. The treatment effect is:

$$E\left[\Delta y_{it} \mid (\Delta x_{it} = 0)\right] = \theta,$$

which it can therefore be estimated as the constant in the model in first differences.

The first-differences estimator is less efficient that the *Within* estimator for $T > 2$ if ε_{it} is $i., i.d.$. It coincides with the between estimator in panels with $T = 2$, since: $y_{i1} - \bar{y} = y_{i1} - \frac{y_1+y_2}{2} = \frac{y_{i1}-y_{i2}}{2}$ and $y_{i2} - \bar{y} = y_{i2} - \frac{y_1+y_2}{2} = -\frac{y_{i1}-y_{i2}}{2}$, and similarly for the x_{it} variables. Under the assumption that the ε_{it} are i.,i.d., then it can be shown that the GLS estimator of the *First-differences* model equals the *Within* estimator. However, the *First-Differenced* model estimates the first differenced equation by OLS and it is therefore less efficient than the *Within* estimator.

6 The Random Effects estimator

Under this approach, we view all the factors that affect the dependent variable and have not been included as regressors, as being included in the random error term. The usual assumption for this model is that the unobserved α_i-terms are independently and identically distributed across individuals. The model is then,

$$y_{it} = \mu + x_{it}'\beta + (\alpha_i + \varepsilon_{it}),\ \varepsilon_{it} \sim i., i.d.(0, \sigma_\varepsilon^2),\ \alpha_i \sim i., i.d.(0, \sigma_\alpha^2)$$

with assumptions:

$$\begin{aligned}
E(\varepsilon_{it}/X) &= E(\alpha_i/X) = 0, \forall i \\
E(\varepsilon_{it}^2/X) &= \sigma_\varepsilon^2 \\
E(\alpha_i^2/X) &= \sigma_\alpha^2 \\
E(\varepsilon_{it}\alpha_j/X) &= 0\ \forall i, j, t \\
E(\varepsilon_{it}\varepsilon_{js}/X) &= 0\ \forall t \neq s, i \neq j \\
E(\alpha_i\alpha_j/X) &= 0\ \forall i \neq j
\end{aligned}$$

Even if ε_{it} is uncorrelated, there will be some serial correlation in the error terms $\alpha_i + \varepsilon_{it}$, coming from the α_i component. We assume that the components α_i and ε_{it} are independent from each other, as well as independent of the explanatory variables x_{is} for all time periods t, s. This leads to a particular form of time correlation, and the standard OLS covariance matrix is inappropriate, while the estimator itself is inefficient. For each individual i, all error terms can be stacked as the Tx1 column vector: $\alpha_i 1_T + \varepsilon_{it}$, with covariance matrix,

$$Var\left(\alpha_i 1_T + \varepsilon_{it}\right) = \Omega = \sigma_\alpha^2 1_T 1_T' + \sigma_\varepsilon^2 I_T = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & ... & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\varepsilon^2 + \sigma_\alpha^2 & ... & \sigma_\alpha^2 & \sigma_\alpha^2 \\ ... & ... & ... & ... & ... \\ \sigma_\alpha^2 & \sigma_\alpha^2 & ... & \sigma_\varepsilon^2 + \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & ... & \sigma_\alpha^2 & \sigma_\varepsilon^2 + \sigma_\alpha^2 \end{pmatrix}$$

To compute the GLS estimator, we transform the data by premultiplying each vector of variables by Ω^{-1}, where:

$$\Omega^{-1} = \sigma_\varepsilon^{-2}\left[I_T - \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} 1_T 1_T'\right] = \sigma_\varepsilon^{-2}\left[\left(I_T - \frac{1}{T} 1_T 1_T'\right) + \psi \frac{1}{T} 1_T 1_T'\right]$$

where: $\psi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$. Since $I_T - \frac{1}{T} 1_T 1_T'$ transforms the data in deviations from their individual means and $\frac{1}{T} 1_T 1_T'$ takes individual means, the GLS estimator for β can be written as,

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i) + \psi T \sum_{i=1}^{N}(\bar{x}_i - \bar{x})'(\bar{x}_i - \bar{x})\right)^{-1} \\ &\quad \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)'(y_{it} - \bar{y}_i) + \psi T \sum_{i=1}^{N}(\bar{x}_i - \bar{x})'(\bar{y}_i - \bar{y})\right) \end{aligned}$$

Two special cases deserve some discussion:

- when $T \to \infty$, the unobserved becomes observable, and it is unlikely that α_i can be constant, unless it is not random. The *Fixed Effects* estimator would then be efficient, and it would coincide with GLS,
- if $\sigma_\varepsilon^2/\sigma_\alpha^2 \to 0$, then the stochastic component is dominated by α_i, which are constant over time, so we are left again with the *Fixed Effects* estimator. In these two cases, the GLS estimator coincides with the *Fixed Effects* estimator.

6.1 Relationship to other estimators

As it was the case with the *Pooled* OLS estimator, we can show that the *Random Effects* GLS estimator is a vector convex linear combination of the *Between* and the *Fixed Effects* estimators.

From the general expression for the GLS estimator, it can be shown that,

$$\hat{\beta}_{GLS} = (I_k - \Delta)\hat{\beta}_B + \Delta\hat{\beta}_{FE}$$

where:

$$\begin{aligned}\Delta &= \left[S_{xx}^{within} + \psi S_{xx}^{between}\right]^{-1} S_{xx}^{within} = \\ &= \psi T \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)'(x_{it}-\bar{x}_i) + \psi T \sum_{i=1}^{N}(\bar{x}_{i.}-\bar{x})'(\bar{x}_{i.}-\bar{x})\right)^{-1} \left[\sum_{i=1}^{N}(x_{i.}-\bar{x}_i)'(x_{i.}-\bar{x}_i)\right]\end{aligned}$$

with ψ being the parameter that we defined above: $\psi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$.

The matrix Δ is proportional to the inverse of the covariance matrix of $\hat{\beta}_B$, so that the GLS estimator is a matrix-weighted average of the *Between* and the *Within* estimators, where the weight depends on the relative variances of the two estimators, the more accurate estimator receiving the heavier weight.

The *Between* estimator discards the time series information in the data set. The GLS estimator is the optimal combination of the *Between* and the *Within* estimators, and is therefore more efficient than either one of them. The POLS estimator is also a linear combination of the two estimators, as seen in previous sections, which differs from the previous one by the presence of the ψ parameter in the definition of the Δ weight. It is a special case of the previous linear combination, for $\psi = 1$. Hence, the *Pooled* OLS estimator is not the efficient linear combination of the *Between* and the *Fixed Effects* estimators. GLS will be more efficient than OLS, as usual.

It is easy to see that for $\psi = 0$ we get the *Fixed Effects* or *Within estimator*, since then, $\Delta = 0$. As we saw above, since $\psi \to 0$ when $T \to \infty$, it follows that the *Random Effects* and the *Fixed Effects* estimators are equivalent for large T. If $\psi = 1$, the GLS estimator reduces to the *Pooled* OLS estimator. The ψ parameter can be thought of as being the relevance given to variation across individuals in the panel. The *Fixed Effects* or *Within* estimator, with $\psi = 0$, ignores that variation. The *Pooled* least squares estimator, with $\psi = 1$, assigns to variation across individuals the same importance as to the variation over time among observations from a given individual, without taking into account that some of their variability comes from variation in α_i across individuals.

The GLS estimator will be unbiased if the explanatory variables are independent of all ε_{it} and all α_i. It will be consistent for N or T or both tending to infinity if in addition to *i*) $E[(x_{it} - \bar{x}_i)\varepsilon_{it}] = 0$ we also have *ii*) $E(\bar{x}_i\varepsilon_{it}) = 0$, and even most importantly, *iii*) $E(\bar{x}_i\alpha_i) = 0$. These conditions are also required for the *Between* estimator to be consistent (Verbeek).

Under weak regularity conditions, the *Random effects estimator*, $\hat{\beta}_{RE}$, also known as the Balestra-Nerlove estimator, is asymptotically Normal, with covariance matrix,

$$Var(\hat{\beta}_{RE}) = \sigma_\varepsilon^2 \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)'(x_{it}-\bar{x}_i) + \psi T \sum_{i=1}^{N}(\bar{x}_i-\bar{x})'(\bar{x}_i-\bar{x})\right)^{-1}$$

which shows that the *Random Effects* estimator is more efficient than the *Fixed Effects* estimator as long as $\psi > 0$. The gain in efficiency is due to the use of the between variation in the data $(\bar{x}_i - \bar{x})$ as it appears in the second term in the expression of the variance-covariance matrix. The covariance matrix above is obtained when estimating by OLS the transformed model (**??**).

We must remember that if we do not feel very confident on the analytical structure we are imposing on the variance-covariance matrix of the error term, we can always proceed by applying ordinary least-squares and a robust inference by using an appropriately corrected empirical covariance matrix, as explained in previous sections.

6.2 Practical implementation of the Random Effects estimator

An easy way to compute the GLS estimator is obtained by applying OLS to transformed variables:

$$y_{it} - \upsilon\bar{y}_i = \mu(1-\upsilon) + (x_{it} - \upsilon\bar{x}_i)'\beta + \varepsilon_{it}$$

where $\upsilon = 1 - \psi^{1/2}$, so that a fixed proportion υ of the individual means is subtracted from the data to obtain the transformed model.

The error term in this transformed regression is still i., i.d. over individuals and over time. Again, $\upsilon = 1\,(\psi = 0)$ corresponds to the *Fixed Effects* or *Within* estimator, while $\upsilon = 0$ corresponds to the *Pooled* OLS estimator. As $T \to \infty, \upsilon \to 1$, and we get the *Fixed Effects* estimator.

We need estimates of the variances of the two error components α_i and $\bar{\varepsilon}_i$,to implement GLS. To that end, we use the variance of the *Fixed Effects* residuals, with denominator $NT - N - k$ as the estimate of σ_ε^2. The denominator relfects the fact that we are estimating N intercepts and k slope coefficients. The error variance for the *Between regression* is $\sigma_\alpha^2 + \sigma_\varepsilon^2/T$, which can be consistently estimated by,

$$\hat{\sigma}_B^2 = \frac{1}{N-k}\sum_{i=1}^{N}(\bar{y}_i - \hat{\mu}_B - \bar{x}_i'\hat{\beta}_B)^2$$

This leads to a consistent estimator for σ_α^2 :

$$\hat{\sigma}_\alpha^2 = \hat{\sigma}_B^2 - \frac{1}{T}\hat{\sigma}_\varepsilon^2$$

Again, the correction for degrees of freedom can be achieved by subtracting $k+1$ from the denominator of $\hat{\sigma}_B^2$.

As an alternative, Green (6 ed.) proposes the equality,

$$\sigma_{POLS}^2 = \sigma_\varepsilon^2 + \sigma_\alpha^2$$

to compute an estimate of σ_α^2 after estimating by POLS and *Fixed Effects*. The Residual sum of squares from the *Pooled OLS* estimator must be divided by $NT - k - 1$, since there is a single intercept.

6.3 Summary

- The *Between* estimator exploits the differences between individuals, and it is determined as OLS in a regression of individual averages. Consistency, for $N \to \infty$, requires two types of conditions: *i)* $E(\bar{x}_i \alpha_i) = 0$, and *ii)* $E(\bar{x}_i \bar{\varepsilon}_i) = 0$, which will usually require explanatory variables to be uncorrelated with the individual effects α_i, as well as strictly exogenous.

- The *Fixed Effects* (or *Within*) estimator exploits the differences within individuals, and it is determined as OLS in a regression using all observations in deviations from individual means. It is consistent for $T \to \infty$ or $N \to \infty$ provided $E[(x_{it} - \bar{x}_i)\varepsilon_{it}] = 0$. This requires explanatory variables to be strictly exogenous, but it does not impose any restrictions upon the relationship between x_{it} and α_i.

- The OLS estimator exploits both dimensions, although less than efficiently. It is determined as OLS in the original model, and it can be written as a convex linear combination of the two previous estimators. Consistency for $T \to \infty$ or $N \to \infty$ requires that $E[x_{it}(\varepsilon_{it} + \alpha_i)] = 0$. This requires explanatory variables to be uncorrelated with α_i, but it does not impose that they are strictly exogenous. It suffices with x_{it} and ε_{it} to be contemporaneously uncorrelated. It also requires explanatory variables to have no correlation with the unobservable individual effects α_i.

- The *Random effects* estimator combines the information in the *Between* and the *Within* estimators in an efficient way. It is consistent for $T \to \infty$ or $N \to \infty$ under the combined conditions that imply consistency for the *Between* and the *Within* estimators. It can be obtained as the efficient weighted average of the *Within* and the *Between* estimators, or as the OLS estimator in a regression with variables transformed as $y_{it} - \upsilon \bar{y}_i$, with $\upsilon = 1 - \psi^{1/2} = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}$.

- *Fixed Effects* estimation is a conditional analysis, measuring the effects of x_{it} on y_{it}, controlling for the individual effects α_i. Prediction is possible only for individuals in the particular sample being used, and even then it is only possible if the panel is long enough that α_i can be consistently estimated. *Random Effects* estimation is instead an example of marginal analysis or population averaged analysis, as the individual effects are integrated out as *i., i.d.* random variables. The *Random Effects* estimator can be applied outside the sample. If the true model is a Random Effects model, then whether to perform a conditional or marginal analysis will vary with the application. If analysis is for a random sample of countries, then one uses random effects, but if one is intrinsically interested in the particular countries in the sample, then one does *Fixed Effects* estimation even though this can entail a loss of efficiency. However, if some unobserved individual specific effects are correlated with regressors, then the *Random Effects* estimator does not make sense, being inconsistent, and we will need either the *Fixed Effects* estimator or the *First Differences* estimator.

6.4 Testing for random effects

The treatment applied to the individual effects can imply substantial differences in numerical estimates in the usual case in which the time dimension of the panel data is small. The *Fixed effects* approach $\{E(y_{it}/x_{it}) = x'_{it}\beta + \alpha_i\}$ is conditional upon the values for α_i. It considers the distribution of y_{it} given α_i, where the $\alpha'_i s$ can be estimated. This makes sense if the individuals in the sample are "one of a kind", and cannot be taken as random draws from some underlying population. That would be the case if the number of units is relatively small. Inferences are made with respect to the effects that happen to be included in the sample. The *Random effects* approach $\{E(y_{it}/x_{it}) = x'_{it}\beta\}$ is not conditional upon the individual $\alpha'_i s$ but "integrates them out". We are then not usually interested in the value of α_i for a given individual. Inferences are made with respect to the population characteristics.

Even if we are interested in a large number of individual units and the *Random effects* approach seems appropriate, we may prefer the *Fixed effects* estimator if x_{it} is clearly correlated with α_i, since that would lead to inconsistent least-squares estimators as used in the *Random effects* estimator. This problem disappears in the *Fixed effects* estimator because α_i is eliminated from the model.

6.4.1 Hausman test

Hausman (1978) suggested a test for the null hypothesis that x_{it} and α_i are uncorrelated. Two estimators are compared: one that it is consistent under both the null and alternative hypothesis, and a second estimator which is consistent only under the null hypothesis. A significant difference between both estimators is interpreted as the null hypothesis not being true. In our case, the *Fixed Effects* estimator is consistent with independence of the possible correlation between x_{it} and α_i , while the *Random Effects* estimator will be consistent and efficient only if the null hypothesis of lack of correlation is true. Usually, to compare the two estimators, we would have to compute the covariance between the two estimates.

The essential result in Hausman (1978) is that the covariance between an efficient estimator and its difference with respect to an inefficient estimator is zero. Hence, since the *Random Effects* estimator is efficient under the null, then if the null hypothesis is true, we will have:

$$Cov(\hat{\beta}_{RE}, \hat{\beta}_{FE} - \hat{\beta}_{RE}) = 0$$

so that,

$$Cov(\hat{\beta}_{RE}, \hat{\beta}_{FE}) = -Var(\hat{\beta}_{RE})$$

and therefore,

$$Var(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})$$

and the test statistic is computed as:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \left[Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE}) \right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

where the two variance-covariance matrices must be substituted by their respective estimates. Under the null hypothesis, the statistic follows a chi-squared distribution with k degrees of freedom, where k is the number of elements in β. A word of caution: the matrix in square brackets may not be positive definite in small samples. We should in that case conclude that the covariance matrices are not different, thereby not rejecting the *Random effects* model, since if the two estimators were different, then the statistic should be positive and relatively large. Even if the statistic turned out to be negative, we would still be able to implement the test for a subset of elements in β. Another strategy would be to move to asymptoticalley equivalent versions of the test statistic. One of them is,

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{B})' \left[Var(\hat{\beta}_{FE}) + Var(\hat{\beta}_{B}) \right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{B})$$

Hausman test can be applied to any other pair of estimators with properties similar to the ones we have used here, as the estimator in *First differences* versus the *Pooled* OLS estimator, since, in the absence of *Random Effects*, the POLS estimator is efficient.

6.4.2 Alternative tests for the comparison between the Fixed Effects and the Random Effects models

When we introduced the Between estimator, we mentioned Mundlak (1978) assumption that the conditional expectation $E(\alpha_i/X_i)$ can be assumed to be a function of the group means: $E(\alpha_i/X_i) = \bar{x}_i' \, \gamma$. That led to the model:

$$y_{it} = \mu + x_{it}'\beta + \bar{x}_i'\gamma + \varepsilon_{it}$$

Mundlak's assumption preserves the specification of the *Random Effects* model while modelling the correlation between individual effects and the observed time varying explanatory variables. This specification is also a compromise between the *Fixed Effects* model and the *Random Effects* model, the difference between them coming from the vector of coefficients γ. Hence, a significance test for this vector of coefficients is an alternative to the Hausman specification testy described above, so long as the assumption on $E(\alpha_i/X_i)$ is approximately correct.

An asymptotically equivalent way to implement the specification test is to perform the Wald test of $\gamma = 0$ in the auxiliary OLS regression,

$$y_{it} - \upsilon\bar{y}_i = (1-\upsilon)\,\mu + (x_{it} - \upsilon\bar{x}_i)'\,\beta_1 + (x_{it} - \bar{x}_i)'\,\gamma + u_{it}$$

where υ is the same parameter used in the alternative implementation of the Random Effects estimator, which is a special case for $\gamma = 0$. If instead, the Fixed Effects estimator is appropriate, then the error term ($u_{it} \equiv (1-\upsilon)\alpha_i + (\varepsilon_{it} - \upsilon\bar{\varepsilon}_i)$) will be correlated with the regressors, and additional functions of the regressors such as $(x_{it} - \bar{x}_i)$ may have significant coefficients in the previous equation.

Breusch and Pagan (1980) proposed a Lagrange multiplier type of test for significance of random effects, $H_0 : \sigma_\alpha^2 = 0$, versus the alternative that it is positive, based on OLS residuals. We therefore, test for lack of autocorrelation in the sum $\varepsilon_{it} + \alpha_i$. The Lagrange multiplier statistic,

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left[\sum_{t=1}^T \hat{\varepsilon}_{it}\right]^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}^2} - 1 \right]^2 = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left[T\bar{\hat{\varepsilon}}_i\right]^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}^2} - 1 \right]^2$$

follows a chi-square distribution with one degree of freedom. The residuals in this expression come from the restricted model, estimated with OLS.

Example: Green (6ed., examples 9.5 and 9.6) applies this test for the logged wages model that excludes the time invariant characteristics, and also computes estimates of the variance component parameters.

6.5 Goodness of fit in panel data models

Goodness of fit under panel data has peculiar features, since we want to weight differently the ability of a model to explain the Between and the Within variation in the data. On the other hand, the R^2 is appropriate only under OLS estimation. It is standard to use a R^2 defined as the square of the correlation between the actual and fitted values, which is always in $[0, 1]$, and collapses to the usual R^2 under OLS estimation. Since Total variation can be decomposed into Between and Within variation:

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - \bar{y})^2 = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - \bar{y}_i)^2 + \frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i - \bar{y})^2$$

The *Fixed Effects* estimator is constructed to explain the Within variations, and it maximizes the Within R^2:

$$R^2_{within} = \left[corr\left\{\hat{y}_{it}^{FE} - \hat{y}_i^{FE}, y_{it} - \bar{y}\right\}\right]^2 = \left[corr\left\{(x_{it} - \bar{x}_i)\hat{\beta}^{FE}, y_{it} - \bar{y}\right\}\right]^2$$

The *Between* estimator maximizes the Between R^2:

$$R^2_{between} = \left[corr\left\{\hat{y}_i^B, \bar{y}\right\}\right]^2 = \left[corr\left\{\bar{x}_i'\hat{\beta}^B, \bar{y}\right\}\right]^2$$

The OLS estimator maximizes the Overall goodness of fit:

$$R^2_{overall} = \left[corr\left\{\hat{y}_{it}, y_{it}\right\}\right]^2$$

where $\hat{y}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{y}_{it}$ and $\hat{y}_i = \frac{1}{TN}\sum\sum\hat{y}_{it}$ where the intercept terms are omitted. If we take into account the variation explained by the N estimated intercepts $\hat{\alpha}_i$, then the fixed effects estimator captures perfectly the between variation. This however, does not mean that it fits the data well, since it is only that the dummy variables capture the data perfectly, and that should not be incorporated into a goodness of fit measure.

The point is that it is possible to define Within, Between and Overall R^2 measures for any arbitrary estimator, using fitted values $\hat{y}_{it}$ and averages $\hat{y}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{y}_{it}$ and $\hat{y}_i = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{y}_{it}$, omitting intercept terms. As we have mentioned, for the Fixed Effects estimator, this would ignore the variation captured by the $\hat{\alpha}_i$ individual intercept estimates.

For the *Random Effects* estimator, the Within, the Between and the Overall R^2 will necessarily be smaller than for the *Fixed Effects*, *Between* and *OLS* estimators, respectively. This again, shows that goodness of fit measures by themselves are not adequate to choose between alternative (potentially non-nested) specifications of the model.

Example: Verbeek (p. 358), logged wages. RATS program.

6.6 Instrumental variables estimators of the Random Effects model

As we have seen, the use of the *Fixed Effects* estimator to solve the problem of correlation between explanatory variables and individual effects may be undesirable, if we are interested in the effect of time invariant variables on the dependent variable.

The *Fixed Effects* estimator can be written:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)'x_{it}\right)^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)'y_{it}\right]$$

which can be interpreted as an instrumental variable estimator in model:

$$y_{it} = \mu + x'_{it}\beta + \alpha_i + \varepsilon_{it}$$

where each explanatory variable is instrumented by its value in deviations from the individual specific mean. Since $\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(x_{it} - \bar{x}_i)'\alpha_i\right] = \sum_{i=1}^{N}\left[\sum_{t=1}^{T}(x_{it} - \bar{x}_i)'\right]\alpha_i = 0$, then all that it is needed for consistency is $E\left[(x_{it} - \bar{x}_i)'\varepsilon_{it}\right] = 0$, which is implied by the strict exogeneity of the x_{it} variables. If a particular element in x_{it} happens to be uncorrelated with α_i, it can be used as its own instrument without taking differences with respect to the individual mean. That is the case of time invariant effects, whose effect on the dependent variable can therefore be estimated under this approach.

6.6.1 The Hausman and Taylor estimator

A more general approach was introduced by Hausman and Taylor (1981), in the line of the Random Effects estimate, as follows: The random effects approach to the linear model:

$$y_{it} = x'_{it}\beta + z'_i\alpha + \varepsilon_{it}$$

is based on the assumption that the unobserved individual specific effects z_i are uncorrelated with the included variables x_{it}. This is a major shortcoming, since it is a very strong assumption to make. However, the Random Effects treatment allows for observed time-invariant characteristics, to appear explicitely in the estimated model, while the Fixed Effects estimator does not, since they are absorbed into the fixed effects. Hausman and Taylor's (1981) estimator suggests a way to overcome the first limitation while accommodating the second advantage, and using only the information in the model.

These authors consider the model

$$y_{it} = x'_{1it}\beta_1 + x'_{2it}\beta_2 + z'_{1i}\alpha_1 + z'_{2i}\alpha_2 + (\varepsilon_{it} + u_i)$$

where x_1 is a k_1-vector, x_2 is a k_2-vector, z_1 is a l_1-vector, z_2 is a l_2-vector, and all individual effects in z_i are assumed to be observed. Unobserved effects would be contained into the individual-specific random term u_i. Variables with the 2-index are correlated with u_i, while those carrying the 1-index are assumed to be uncorrelated with u_i. Hence, OLS and GLS estimates will be biased and inconsistent. Assumptions on random terms are:

$$\begin{aligned}
E(u_i/x_{1it}, z_{1i}) &= 0,\ \textit{although}\ E(u_i/x_{2it}, z_{2i}) \neq 0 \\
Var\,(u_i/x_{1it}, x_{2it}, z_{1i}, z_{2i}) &= \sigma_u^2; \\
Cov\,(\varepsilon_{it}, u_i/x_{1it}, x_{2it}, z_{1i}, z_{2i}) &= 0; \\
Var\,(\varepsilon_{it} + u_i/x_{1it}, x_{2it}, z_{1i}, z_{2i}) &= \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2; \\
Corr\,(\varepsilon_{it} + u_i, \varepsilon_{is} + u_i/x_{1it}, x_{2it}, z_{1i}, z_{2i}) &= \rho = \sigma_u^2/\sigma^2
\end{aligned}$$

The group mean deviations $x_{1it} - \bar{x}_{1i}, x_{2it} - \bar{x}_{2i}$ can be used as $k_1 + k_2$ instrumental variables. Since z_1 is uncorrelated with the disturbances, it can be used as a set of l_1 instrumental variables for themselves. So, we need another l_2 instrumental variables. Hausman and Taylor show that the individual (group) means for x_1 can be used as such, so the identification condition[4] is $k_1 \geq l_2$.

Feasible GLS is better than OLS, and it is also an improvement on the simple instrumental variable estimator, which is consistent, but inefficient.

Taking deviations from group means:

[4] To estimate the original model, Hausman and Taylor suggest using $x_{1it}, z_{1i}, x_{2it} - \bar{x}_2$ and $\bar{x}_{1i}$ as instruments. We can use time averages of those time-varying regressors that are uncorrelated with α_i as instruments for the time-invariant regressors. The identification condition is then that we have enough of those instruments: $k_1 \geq l_2$.

$$y_{it}-\bar{y}_i=(x_{1it}-\bar{x}_{1i})'\beta_1+(x_{2it}-\bar{x}_{2i})'\beta_2+(\varepsilon_{it}-\bar{\varepsilon}_i)$$

which can be consistently estimated by LS, in spite of the correlation between x_2 and u. This is, of course, the Fixed Effects, Least-Squares Dummy Variable (LSDV) estimator. However, it would not identify the values of coefficients for time invariant variables. It is also inefficient, since x_{1t} is needlessly instrumented.

We can describe four steps to compute the Hausman and Taylor instrumental variable estimator, the first three of which provide us with the ψ-parameter neded to transform the data and compute the estimator in a last step.

- Step 1: Obtain the LSDV (fixed-effects) estimator of $\beta=(\beta_1,\beta_2)$ based on x_1 and x_2. The residual variance from this step is a consistent estimator of σ^2_ε.

- Step 2: Form the within groups residuals e_{it} from LSDV regression in Step 1. Stack the group (individual) means, conveniently repeated, in a full sample length data vector, $e^*_{it}=\bar{e}_i, i=1,2,...,N,\ t=1,2,...,T$. The residuals are computed excluding the estimate of the constant term. These are used as the dependent variable in an instrumental variable regression on z_1 and z_2 with instrumental variables z_1 and x_1 (assuming $k_1\geq l_2$). Time invariant variables are repeated T times in the data matrices in this regression. This provides a consistent estimator of α_1,α_2.

- Step 3: The residual variance from step 2 is a consistent estimator of $\sigma^{*2}=\sigma^2_\alpha+\sigma^2_\varepsilon/T$. From this estimator and the estimator of σ^2_ε from step 1, we deduce an estimator: $\sigma^2_\alpha=\sigma^{*2}-\sigma^2_\varepsilon/T$, and compute the weight for the GLS estimator: $\psi=\sqrt{\frac{\sigma^2_\varepsilon}{T\sigma^2_\alpha+\sigma^2_\varepsilon}}$

- Step 4: A weighted instrumental variable estimator. Consider the full set of explanatory variables: $w'_{it}=(x'_{1it},x'_{2it},z'_{1i},z'_{2i})$, for which we have nT observations. We perform the usual GLS transformation as for the random effects model: $w^{*\prime}_{it}=w'_{it}-(1-\hat{\psi})\bar{w}'_i,\ y^*_{it}=y_{it}-(1-\hat{\psi})\bar{y}_i$, and collect these transformed data in a matrix W^* and a column vector y^*. For the time-invariant variables, the group mean is equal to the original variable, and the transformation just multiplies the original data by $1-\hat{\psi}$. The instrumental variables are: $v'_{it}=\left[(x'_{1it}-\bar{x}_{1i})',(x'_{2it}-\bar{x}_{2i})',z'_{1i},\bar{x}'_{1i}\right]$. These are stacked as rows in an nTx$(k_1+k_2+l_1+l_2)$ matrix V. For the third and fourth sets of instruments, the time invariant variables and group means are repeated for each time period for that individual or group. The instrumental variable estimator would be:

$$\left(\hat{\beta}',\hat{\alpha}'\right)'_{IV}=\left[(W^{*\prime}V)(V'V)^{-1}(V'W^*)\right]^{-1}\left[(W^{*\prime}V)(V'V)^{-1}(V'y^*)\right]$$

For the sake of comparison, the $FGLS$ random-effects[5] estimator would be: $\left(\hat{\beta}', \hat{\alpha}'\right)'_{RE} = (W^{*\prime}W^{*})^{-1} W^{*\prime}y^{*}$.

The instrumental variable is consistent if the data is not weighted, that is, if W, rather than W^{*}, is used in estimation. But that would be inefficient, in the same way as OLS is inefficient in estimation of the simpler random effects model.

7 Dynamic linear models

7.1 Linear autoregressive models

Consider an autoregressive panel data model with a vector of exogenous explanatory variables:

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + x'_{it}\beta + \varepsilon_{it},\ \varepsilon_{it} \sim i., i.d.(0, \sigma^2_{\varepsilon})$$

Here the problem is that the lagged dependent variable will depend upon α_i irrespective of how we treat the individual effect α_i. To see this, assume, for simplicity, that there are not exogenous explanatory variables:

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \varepsilon_{it},\ \varepsilon_{it} \sim i., i.d.(0, \sigma^2_{\varepsilon})$$

Denoting $\bar{y}_{i,-1} = \sum_{t=2}^{T} y_{i,t-1}/(T-1)$, different from $\bar{y}_i = (1/T)\sum_{t=1}^{T} y_{i,t}$, the Fixed Effects estimator is:

$$\hat{\gamma}_{FE} = \frac{\sum\sum(y_{it} - \bar{y}_i)(y_{it-1} - \bar{y}_{i,-1})}{\sum\sum(y_{it-1} - \bar{y}_{i,-1})^2} = \gamma + \frac{\frac{1}{N(T-1)}\sum\sum(\varepsilon_{it} - \bar{\varepsilon}_i)(y_{it-1} - \bar{y}_{i,-1})}{\frac{1}{N(T-1)}\sum\sum(y_{it-1} - \bar{y}_{i,-1})^2}$$

which will be biased and inconsistent for $N \to \infty$ and fixed T. This is because the last term in the right-hand side does not have expectation zero due to the correlation between $\bar{y}_{i,-1}$ and $\bar{\varepsilon}_i$, and it does not converge to zero. In fact Nickell (1981), Hsiao (2003) show that:

$$p\lim \frac{1}{NT}\sum\sum(\varepsilon_{it} - \bar{\varepsilon}_i)(y_{it-1} - \bar{y}_{i,-1}) = -\frac{\sigma^2_{\varepsilon}}{T^2}\frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2} \neq 0$$

Notice that the inconsistency is not produced by any assumption we can make on the α_i, since it gets eliminated in the transformation, but rather, by the fact that the *Within* transformed lagged dependent variable is correlated with the *Within* transformed error.[6] Therefore, at a difference of what happens

[5] This denotes the Feasible GLS estimator of the *Random Effects* model, the noe we described in the *Implementation* section.

[6] $Cov(y_{i,t-1}, c_i + \varepsilon_i) = \sigma^2_c + \gamma Cov(y_{i,t-2}, c_i + \varepsilon_i)$, and the Covariance would converge, for T large, to $\frac{\sigma^2_c}{1-\delta}$.

in a static model, the *Fixed Effects* estimator does not solve the inconsistency in a dynamic model.

On the other hand, if $T \to \infty$, then the expression above converges to zero, and the *Fixed Effects* estimator is consistent if both $T \to \infty$ and $N \to \infty$. But in finite samples, this lack of consistency can be a serious problem. For instance, if $\gamma = 0.5$, then we have, as $N \to \infty$:

$$\begin{aligned} p\lim \hat{\gamma}_{FE} &= -0.25 \; if \; T = 2 \\ p\lim \hat{\gamma}_{FE} &= -0.04 \; if \; T = 3 \\ p\lim \hat{\gamma}_{FE} &= 0.33 \; if \; T = 10 \end{aligned}$$

To avoid the inconsistency, we make a different transformation to eliminate the individual effects α_i, by taking *First differences*:

$$y_{it} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}), \; t = 2, 3, ..., T \; \varepsilon_{it} \sim i., i.d.(0, \sigma_\varepsilon^2)$$

Once again, least squares would be inconsistent in this model because of the correlation between $y_{i,t-1}$ and $\varepsilon_{i,t-1}$, even when $T \to \infty$. But the transformation suggests an instrumental variable approach (Anderson and Hsiao (1981)) so long as ε_{it} does not exhibit autocorrelation, since $y_{i,t-2}$ is clearly correlated with the explanatory variable, but not with the error term,

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (y_{it} - y_{it-1}) y_{i,t-2}}{\sum \sum_{t=3} (y_{it-1} - y_{i,t-2}) y_{i,t-2}}$$

A standard argument shows that consistency of this instrumental variable estimator depends on $p\lim \frac{1}{N(T-2)} \sum_{i=1}^{N} \sum_{t=3}^{T} (\varepsilon_{it} - \varepsilon_{i,t-1}) y_{it-2} = 0$ for either N, T or both going to ∞. Anderson and Hsiao suggested an alternative instrumental variable estimator, using $y_{i,t-2} - y_{i,t-3}$ as instrumental variable:

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} (y_{it} - y_{it-1})(y_{i,t-2} - y_{i,t-3})}{\sum_{i=1}^{N} \sum_{t=3} (y_{it-1} - y_{i,t-2})(y_{i,t-2} - y_{i,t-3})}$$

which will be consistent if $p\lim \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=3}^{T} (\varepsilon_{it} - \varepsilon_{i,t-1})(y_{it-2} - y_{i,t-3})$ $= 0$ for either N, T or both going to ∞. As in the previous estimator, this condition will hold whenever ε_{it} lacks serial correlation. If there are exogenous regressors in the model, then not only their contemporaneous and lagged values, but also their future values, are valid instruments as well. If they are predetermined, their contemporaneous and lagged values will be valid instruments. The number of instruments increases with time, and it can easily get very large. However, the latter set of instruments requires an additional lag, and hence, we lose an additional sample period.

The instrumental variable estimator is,

$$\hat{\theta}_{IV} = \left[\left(\sum_{i=1}^{n} \Delta X_i' Z_i \right) \left(\sum_{i=1}^{n} Z_i' Z_i \right)^{-1} \left(\sum_{i=1}^{n} Z_i' \Delta X_i \right) \right]^{-1} \left[\left(\sum_{i=1}^{n} \Delta X_i' Z_i \right) \left(\sum_{i=1}^{n} Z_i' Z_i \right)^{-1} \left(\sum_{i=1}^{n} Z_i' \Delta y_i \right) \right]$$

where the X matrix includes the lagged endogenous variable in addition to possible predetermined or exogenous variables, and Z is the matrix of chosen instruments. The variance-covariance matrix is,

$$Var\left(\hat{\theta}_{IV}\right)=\sigma^2_{\Delta\varepsilon}\left[\left(\sum_{i=1}^{n}\Delta X_i' Z_i\right)\left(\sum_{i=1}^{n}Z_i' Z_i\right)^{-1}\left(\sum_{i=1}^{n}Z_i'\Delta X_i\right)\right]^{-1}$$

where an estimate for $\sigma^2_{\Delta\varepsilon}$ could be obtained from the residual sum of squares of the diferenced model: $\hat{\sigma}^2_{\Delta\varepsilon}=RSS/[N(T-2)]$. But this will be an underestimate, since it ignores the fact that the difference operator introduces first order serial correlation. In fact, the previous footnote suggests that the previous calculation will be an approximate estimate of $2\sigma^2_{\varepsilon}$. But there is also the additional problem that the observations are autocorrelated. Hence, the standard IV variance-covariance matrix is inappropriate, and we must use,

$$Var(\hat{\theta}_{IV})=A\left[\left(\sum_{i=1}^{n}\Delta X_i' Z_i\right)\left(\sum_{i=1}^{n}Z_i' Z_i\right)^{-1}\hat{\sigma}^2_{\varepsilon}\left(\sum_{i=1}^{n}Z_i' G Z_i\right)\left(\sum_{i=1}^{n}Z_i' Z_i\right)^{-1}\left(\sum_{i=1}^{n}Z_i'\Delta X_i\right)\right]A$$

with[7] G being a TxT matrix: $G=\begin{pmatrix} 2 & -1 & 0 & ... \\ -1 & 2 & ... & 0 \\ 0 & ... & ... & -1 \\ ... & 0 & -1 & 2 \end{pmatrix}$,where,

$$A=\left[\left(\sum_{i=1}^{n}\Delta X_i' Z_i\right)\left(\sum_{i=1}^{n}Z_i' Z_i\right)^{-1}\left(\sum_{i=1}^{n}Z_i'\Delta X_i\right)\right]^{-1}$$

While one could discuss whether it is preferable to use levels or differences as instruments, the Generalized Method of Moments provides a unified approach to instrumental variable estimation.

7.2 General Method of Moments (GMM) estimation

Rather than arguing about which instrumental variable estimator we should use, a GMM argument would lead us to using both instruments, while eliminating the disadvantage of reduced sample sizes.

The two previous instrumental variable estimators use the moment conditions: $E\left[(\varepsilon_{it}-\varepsilon_{i,t-1})(y_{it-2}-y_{i,t-3})\right]=0$ and $E\left[(\varepsilon_{it}-\varepsilon_{i,t-1})y_{it-2}\right]=0$. Arellano and Bond (1991) suggest that the list of instruments can be extended by exploiting additional moment conditions and letting their number vary with t , thereby increasing efficiency. For instance, when $T=4$, we have, for $t=2$ the moment condition:[8]

[7] $Var(\varepsilon_{i2}-\varepsilon_{i1})=2\sigma^2_{\varepsilon}, Cov(\varepsilon_{i2}-\varepsilon_{i1},\varepsilon_{i3}-\varepsilon_{i2})=-\sigma^2_{\varepsilon}$

[8] Assuming there is an initial y_0 observation. Otherwise, we would have one moment condition less at each point in time,

$$E[(\varepsilon_{i2} - \varepsilon_{i1})y_{i0}] = 0$$

while for $t = 3$, we have:

$$\begin{aligned} E[(\varepsilon_{i3} - \varepsilon_{i2})y_{i1}] &= 0 \\ E[(\varepsilon_{i3} - \varepsilon_{i2})y_{i0}] &= 0 \end{aligned}$$

and, for $t = 4$:

$$\begin{aligned} E[(\varepsilon_{i4} - \varepsilon_{i3})y_{i0}] &= 0 \\ E[(\varepsilon_{i4} - \varepsilon_{i3})y_{i1}] &= 0 \\ E[(\varepsilon_{i4} - \varepsilon_{i3})y_{i2}] &= 0 \end{aligned}$$

So, in general, we have a matrix of instruments:

$$Z_i = \begin{pmatrix} [y_{i,0}] & 0 & ... & 0 \\ 0 & [y_{i0}, y_{i1}] & ... & 0 \\ ... & & ... & 0 \\ 0 & ... & ... & [y_{i0}, y_{i1}, ..., y_{i,T-2}] \end{pmatrix}$$

and the vector of transformed error terms:

$$\Delta\varepsilon_i = \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ ... \\ \varepsilon_{i,T} - \varepsilon_{i,T-1} \end{pmatrix}$$

and a set of $1 + 2 + 3 + ... + (T-1) = \frac{(T-1)T}{2}$ moment conditions:[9]

$$E[Z_i'\Delta\varepsilon_i] = E[Z_i'(\Delta y_i - \gamma\Delta y_{i,-1})] = 0$$

Since the number of moment conditions will usually exceed the number of unknown parameters, as it is the case in this example, we will minimize the quadratic form:

$$\min_{\gamma} \left[\frac{1}{N}\sum_{i=1}^{N} Z_i'(\Delta y_i - \gamma\Delta y_{i,-1})\right]' W_N \left[\frac{1}{N}\sum_{i=1}^{N} Z_i'(\Delta y_i - \gamma\Delta y_{i,-1})\right]$$

where W_N is a symmetric, positive definite weighting matrix which will depend on the sample size, N. Differentiating with respect to γ and solving:

[9] With $T = 4$ time observations, we will have 6 instruments or orthogonality conditions if there is an initial condition y_{i0}, and 3 such conditions if there is not known initial condition y_{i0}.

$$\gamma_{GMM} = \left[\left(\sum_{i=1}^{N} \Delta y'_{i,-1} Z_i \right) W_N \left(\sum_{i=1}^{N} Z'_i \Delta y_{i,-1} \right) \right]^{-1} \left[\left(\sum_{i=1}^{N} \Delta y'_{i,-1} Z_i \right) W_N \left(\sum_{i=1}^{N} Z'_i \Delta y_i \right) \right]$$

This estimator is consistent for any choice of positive definite weighting matrix W_N so long as orthogonality (moment) conditions are true. GMM theory shows that the optimal choice of weighting matrix, in order to minimize the variance-covariance matrix of the resulting estimator, is the inverse of the covariance matrix of the sample moments:

$$\underset{N \to \infty}{p \lim} W_N = \left[Var(Z'_i \Delta \varepsilon_i) \right]^{-1} = \left[E(Z'_i \Delta \varepsilon_i \Delta \varepsilon_i Z_i) \right]^{-1}$$

If no restrictions are imposed upon the covariance matrix, then it can be estimated by the sample average of a function of the residuals $\hat{\varepsilon}$ from a consistent initial estimate. Usually, this is obtained with the identity matrix as the initial weighting matrix:

$$\hat{W}_N^{opt} = \left(\frac{1}{N} \sum_{i=1}^{N} Z'_i \Delta \hat{\varepsilon}_i \Delta \hat{\varepsilon}_i Z_i \right)^{-1}$$

where $\hat{\varepsilon}_i$ denote the residuals from an initial GMM estimate obtained with an identity as weighting matrix: $W_N = I$.

The general GMM approach does not need that the ε_{it} be *i., i.d.* over individuals, and the optimal weighting matrix is estimated without imposing such constraint. However, the moment conditions are valid only under lack of autocorrelation. And if autocorrelation is present, there is no point in computing a robust estimate of the variance-covariance matrix of estimates, since they will be inconsistent.

Under weak regularity conditions, the GMM estimator for γ is asymptotically Normal for $N \to \infty$ and fixed T, with covariance matrix,

$$\underset{N \to \infty}{p \lim} \left[\left(\sum_{i=1}^{N} \Delta y'_{i,-1} Z_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} Z'_i \Delta \hat{\varepsilon}_i \Delta \hat{\varepsilon}_i Z_i \right)^{-1} \left(\sum_{i=1}^{N} Z'_i \Delta y_{i,-1} \right) \right]^{-1}$$

With i., i.d. errors, the middle term reduces to,

$$\sigma_\varepsilon^2 W_N^{opt} = \sigma_\varepsilon^2 \left(\frac{1}{N} \sum_{i=1}^{N} Z'_i G Z_i \right)^{-1}$$

with[10] G being a TxT matrix: $G = \begin{pmatrix} 2 & -1 & 0 & ... \\ -1 & 2 & ... & 0 \\ 0 & ... & ... & -1 \\ ... & 0 & -1 & 2 \end{pmatrix}$ so long as there is no autocorrelation in the error term. Alvarez and Arellano (2003) show

[10] $Var(\varepsilon_{i2} - \varepsilon_{i1}) = 2\sigma_\varepsilon^2, Cov(\varepsilon_{i2} - \varepsilon_{i1}, \varepsilon_{i3} - \varepsilon_{i2}) = -\sigma_\varepsilon^2$

that the GMM estimator is also consistent when both, N and T tend to infinity despite the fact that the number of moment conditions tends to infinity with the sample size.

For large T, however, the GMM estimator will be close to the Fixed Effects estimator, which provides a more attractive alternative.

7.3 Dynamic models with exogenous variables

In the case of the more general model:

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + x'_{it}\beta + \varepsilon_{it},\ \varepsilon_{it} \sim i., i.d.(0, \sigma^2_\varepsilon)$$

we will have different instruments as a function of the assumptions we make on the x_{it} variables. If they are assumed to be strictly exogenous, in the sense of being uncorrelated with all error terms at all time periods, we will also have: $E(x_{it}\Delta\varepsilon_{is}) = 0\ \forall s, t$, so that $x_{i1}, x_{i2}, ..., x_{iT}$ can be added as instruments to the model in first differences. But that would make the number of rows in Z_i too large. Almost the same amount of information can be obtained if we use the first differenced x_{it} as their own instruments. Then, we would be imposing moment conditions:

$$E(\Delta x_{it}\Delta\varepsilon_{it}) = 0, \forall t$$

and the matrix of instruments can be written:

$$Z_i = \begin{pmatrix} [y_{i,0}, \Delta x'_{i2}] & 0 & ... & 0 \\ 0 & [y_{i0}, y_{i1}, \Delta x'_{i3}] & ... & 0 \\ ... & & ... & 0 \\ 0 & ... & ... & [y_{i0}, y_{i1}, ..., y_{i,T-2}, \Delta x'_{iT}] \end{pmatrix}$$

If the x_{it} variables are not strictly exogenous, but only predetermined: $E(x_{it}\varepsilon_{is}) = 0, \forall s \geq t$. Then, $E\left[(x_{it} - x_{i,t-1})(\varepsilon_{it} - \varepsilon_{i,t-1})\right] \neq 0$, and only $x_{i,t-1}, ..., x_{i1}$ are valid instruments for the first-differenced equation in period t . The moment conditions imposed would then be:

$$E(x_{i,t-j}\Delta\varepsilon_{it}) = 0,\ for\ j = 1, 2, ..., t-1,\ for\ each\ t$$

Usually, one should expect to have a mixture of some exogenous and some predetermined variables to be used as instruments. Arellano and Bond (1995) explain how this approach can be integrated into the instrumental variable estimator of Hausman and Taylor (1981). They also discuss how information in levels of original variables can also be used in estimation.

Example: Verbeek

Verbeek refers to the estimation of a demand for labour equation based on data from 2800 large Belgium firms over 1986-1994. Using a theoretical model of union bargaining as reference, the authors estimate a static version:

$$\log L_{it} = \beta_0 + \beta_1 \log w_{it} + \beta_2 \log K_{it} + \beta_3 \log Y_{it} + \beta_4 \log w_{jt} + u_{it}$$

where w_{jt} denotes the industry average real wage, acting as an indicator of the reference negotiation wage level for unions, K_{it} is the stock of capital and Y_{it} is output, as well as adynamic version of the demand equation,

$$\log L_{it} = \beta_0 + \beta_1 \log w_{it} + \beta_2 \log K_{it} + \beta_3 \log Y_{it} + \beta_4 \log w_{jt} + \gamma \log L_{i,t-1} + \alpha_i + \varepsilon_{it}$$

where it is assumed that the error term has two components, the first one being unobservable firm-specific time-invariant heterogeneity. If we first-difference the equation, then $\Delta \log L_{i,t-1}$ will be correlated with $\Delta\varepsilon_{it}$. In addition, it is very likely that wages and employment are jointly bargained, wages then becoming an endogenous explanatory variable in the previous equation. Therefore,

$$E(\Delta \log w_{it} \Delta\varepsilon_{it}) \neq 0,$$

and we need to use an instrumental variables approach. Valid instruments for $\Delta \log w_{it}$ are $\log w_{i,t-2}, \log w_{i,t-3}, ...$ while $\log L_{i,t-2}, \log L_{i,t-3}, ...$ could be valid instruments for $\Delta \log L_{i,t-1}$. Hence, the number of instruments increases with t .

Estimation Labour demand equation [Konings and Roodhooft (1997)]
Dependent variable: $\log L_{it}$

	Static model	*Dynamic model*
$\log L_{i,t-1}$		0.60(0.045)
$\log Y_{it}$	0.021(0.009)	0.008(0.005)
$\log w_{it}$	$-1.78(0.60)$	$-0.66(0.19)$
$\log w_{jt}$	0.16(0.07)	0.054(0.033)
$\log K_{it}$	0.08(0.011)	0.078(0.006)
Test for overidentifying restrictions	$29.7(df = 15, p = 0.013)$	$51.66(df = 29, p = 0.006)$
Number of observations		

The p-values for both models are clos to 1%. The estimated short-run wage elasticity of labour demand is -0.66%, but the long-run elasticity is -1.64%, higher than it had been estimated with macro data.[11]

[11] Although there were several difficulties with the way the data had been constructed. See original article in *De Economist.*