2022 STATA ECONOMETRICS WINTER SCHOOL

Anabela Carneiro[1] João Cerejeira[2] Miguel Portela[2,3] Paulo Guimarães[1,4]

[1]FEP and CEF.UP – U.Porto
[2]NIPE – UMinho
[3]IZA, Bonn
[4]Banco de Portugal

Faculdade de Economia da Universidade do Porto, Portugal

January 25-26, 2022

Outline

1. Instrumental variables
2. Tests for endogeneity and validity of the instruments
3. Generalized method of moments (*GMM*)
4. Static panel data and first-differences
5. Time-effects
6. Endogenous variables
7. Dynamic panel data: Arellano & Bond type of estimation
8. Issues with the use of instruments in first-differences
9. System *GMM* estimation
10. Unit root tests

Instrumental variables

Take the example of a wage regression (*logwage*) with education as a regressor. Unobserved heterogeneity will affect both wages and education:

$$\begin{aligned} lwage_i &= \beta_0 + \beta_1 educ_i + u_i \\ &= \beta_0 + \beta_1 educ_i + (abil_i + v_i) \end{aligned} \qquad (1)$$

$$E(educ_i \times u_i) = E(educ_i \times abil_i) \neq 0 \qquad (2)$$

which implies that the *OLS* estimator is biased and inconsistent.
There are two ways to solve the problem: (i) search for a *proxy* for ability; (ii) search for an instrumental variable for education.
An instrument for *educ* is a variable that:

- explains part of the variation in education
- its not correlated with the variation in the unobserved component of equation (1) $u_i = abil_i + v_i$

Two-stage least squares (2SLS)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_j x_{ij} + u_i \tag{3}$$

First stage regression (reduced form)

$$educ_i = \pi_0 + \pi_1 fatheduc_i + \pi_2 motheduc_i + \pi_j x_{ij} + w_i \tag{4}$$

Get the estimates

$$\hat{educ}_i = \hat{\pi}_0 + \hat{\pi}_1 fatheduc_i + \hat{\pi}_2 motheduc_i + \hat{\pi}_j x_{ij} \tag{5}$$

and apply *OLS* (second stage regression)

$$lwage_i = \beta_0 + \beta_1 \hat{educ}_i + \beta_j x_{ij} + u_i \tag{6}$$

This is the consistent estimator *IV/2SLS*.

Endogeneity test

Using the residuals from the first regression

$$\hat{w}_i = educ_i - \hat{educ}_i \tag{7}$$

one test of exogeneity is of the type $H_0 : \delta = 0$ in the regression

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_j x_{ij} + \delta \hat{w}_i + u_i \tag{8}$$

which is estimated by *OLS*. If we reject the null hypothesis there is evidence favorable to the endogeneity of *educ*.

- additional instrumental variables can increase the efficiency of the estimator (lower variance)
- When we have more instruments than endogenous variables we have an overidentified model

Test for the validity of the instruments

In an overidentified model we can test the validity of the set of instruments used; *i.e.*, test if the instruments are not correlated with the error term in the original model.
One should get the residuals using the estimated parameters from the second stage regression

$$\hat{u}_i = lwage_i - \left(\hat{\beta}_0^{IV} + \hat{\beta}_1^{IV} \times educ_i + \hat{\beta}_j^{IV} x_{ij}\right) \tag{9}$$

Now we implement the regression of $\hat{u}_i$ on a constant, *fatheduc*, *motheduc* as well as on the variables x_{ij}, retaining the R^2. Under the null hypothesis of valid instruments we know that

$$nR^2 \sim \chi^2_q \tag{10}$$

where q stands for the degrees of freedom, which equals the number of instruments minus the endogenous variables $(2 - 1 = 1)$. The null hypothesis is rejected when nR^2 is bigger than the critical value associated with the test's significance level, α (for example, 5%).

Generalized method of moments (GMM)

The GMM is a flexible estimation procedure, which allows for heteroskedasticity.
Let *fatheduc*, *motheduc* and x_{ij} be instruments within z_{ij}. Now, the moment conditions for the population can be written as

$$E\left(z_i u_i\right) = 0 \tag{11}$$

which defines the *IV/GMM* estimator. The estimator minimizes the following criteria

$$\min_b \left(\frac{1}{N}\sum\nolimits_{i=1}^{N} e_i z_i'\right) W_N \left(\frac{1}{N}\sum\nolimits_{i=1}^{N} z_i e_i\right) \tag{12}$$

as a function of the parameters, $e_i = y_i - x_i' b$, where W_N is a weighting matrix that determines the efficiency properties of the *IV/GMM* estimator. We thus obtain the estimator $\hat{\beta}$, with residuals $\hat{u}_i$.

GMM (cont.)

When

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} z_i z_i' \right)^{-1} \tag{13}$$

we obtain the $2SLS$ estimator

$$\hat{\beta} = \left(X'Z \left(Z'Z \right)^{-1} Z'X \right)^{-1} X'Z \left(Z'Z \right)^{-1} Z'Y \tag{14}$$

This estimator is efficient when the error term, u_i, is homoskedastic, $E\left(u_i^2 | z_i\right) = \sigma_u^2$.

When we face conditional heteroskedasticity, for example $E\left(u_i^2 | z_i\right) = \sigma_i^2\left(z_i\right)$, then the two step GMM estimator asymptotically efficient is obtained by using the following weighting matrix

$$W_N = W_N\left(\hat{\beta}_1\right) = \left(\frac{1}{N} \sum_{i=1}^{N} z_i \hat{u}_{i1}^2 z_i' \right)^{-1} \tag{15}$$

where $\hat{u}_{i1}$ are residuals of the first stage estimator, and $\hat{\beta}_1$ is, for example, the $2SLS$ estimator.

GMM (cont.)

- Different matrices W_N give origin to different consistent estimators, with different associated variance-covariance matrices
- The optimal weighting matrix is the inverse of the sample variance-covariance matrix of the sample moments
- This matrix, however, depends on the unknown parameters, which represents a problem

One solution is to use an estimation step procedure:

- (i) in the 1^{st} step we use the non-optimal matrix W_N, which does not depend on the parameters (for example, it can be the identity matrix)
- (ii) we obtain a first consistent estimate of our parameters of interest
- (iii) in the 2^{nd} step we build the optimal matrix W_N which uses these estimates and allow us to obtain an asymptotically efficient estimator

GMM (cont.)

Advantages of the *GMM* procedure:

- (1) we do not need to assume a probability distribution
- (2) we can deal with heteroskedasticity of unknown form
- (3) we can estimate parameters even when the model cannot be solved analytically using the first order conditions

The Sargan test of the overidentifying restrictions is given by

$$S\left(\hat{\beta}_2\right) = N\left(\frac{1}{N}\sum_{i=1}^{N} \hat{u}_{i2}^2 z_i'\right) W_N\left(\hat{\beta}_1\right)\left(\frac{1}{N}\sum_{i=1}^{N} z_i \hat{u}_{i2}^2\right) \qquad (16)$$

where $\hat{u}_{i2}$ are the residuals of the 2^{nd} step.
This test follows asymptotically a χ_q^2 distribution under the null hypothesis $E\left(z_i u_i\right) = 0$, where q is the number of instruments minus the parameters to be estimated.

Panel data (static)

As discussed before, panel data contain repeated observations in different moments for each of the observed units. The units can be individuals, firms, countries, among others.

These data allow to control for unobserved heterogeneity at the unit level

Generically, the model is defined as

$$\begin{aligned} y_{it} &= \beta_0 + x_{it1}\beta_1 + x_{it2}\beta_2 + \ldots + x_{itk}\beta_k + \eta_i + v_{it} \\ &= x'_{it}\beta + \eta_i + v_{it} \end{aligned} \tag{17}$$

where η_i represents the individual specific effects which are unknown/unobserved and constant over time, and $i = 1, \ldots, N$, $t = 1, \ldots, T$.

We assume: $E(v_{it}) = 0$; $E(v_{it}|x_{it}) = 0$.

Panel data: first-differences

In the model

$$\begin{aligned} y_{it} &= x'_{it}\beta + u_{it} & (18)\\ u_{it} &= \eta_i + v_{it} & (19) \end{aligned}$$

where individual unobserved effects are correlated with x_{it}. Taking first-differences we eliminate η_i:

$$y'_{it} - y_{it-1} = (x_{it} - x_{it-1})'\,\beta + (u_{it} - u_{it-1}) = (x_{it} - x_{it-1})'\,\beta + (v_{it} - v_{it-1}) \tag{20}$$

such that the OLS estimator is unbiased if $(v_{it} - v_{it-1})$ and $(x_{it} - x_{it-1})$ are not correlated. This is an assumption less restrictive compared with the strict exogeneity assumption of the *FE* estimator.

Panel data: first-differences (cont.)

Once more the *OLS* standard-errors are not correct as they do not take into account the correlation between $\left(v_{it} - v_{it-1}\right)$ and $\left(v_{it-1} - v_{it-2}\right)$

$$E\left[\left(v_{it} - v_{it-1}\right)\left(v_{it-1} - v_{it-2}\right)\right] = -\sigma_v^2 \tag{21}$$

$$E\left(v_i v_i'\right) = \sigma_v^2 \begin{bmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ 0 & \cdots & -1 & 2 \end{bmatrix} \tag{22}$$

assuming that v_{it} are not correlated over time.

Panel data: time effects

Usually the data only covers a few years, so the time effects (macroeconomic aggregate shocks) can be controlled for using specific dummies for periods. This way, the model to estimate is the following

$$y_{it} = x_{it}'\beta + \lambda_t + \eta_i + v_{it} \tag{23}$$

When a constant exists in the model, which is the case most of the time, the λ_t parameters measure the difference between the constant term in moment t and the coefficient of the excluded period. The interpretation is the same as in the *OLS*, *RE* and *FE*. For the model in first-differences we have

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})'\,\beta + (\lambda_t - \lambda_{t-1}) + (v_{it} - v_{it-1}) \tag{24}$$

and the estimated effects for time measure the differences between adjacent periods.

Endogenous variables

Consider the model in first-differences

$$y_{it} - y_{it-1} = \left(x_{it} - x_{it-1}\right)' \beta + \left(v_{it} - v_{it-1}\right) \tag{25}$$

where x_{it} is an endogenous variable if it is correlated with v_{it}.
It can also exist an effect from v_{it-1} to x_{it}, such that $E\left(x_{it} v_{it-1}\right) \neq 0$. In this case, the variable x_{it} is predetermined or *weakly exogenous*, as it is correlated with the error term in the previous period, but not contemporaneously.
In both cases we have $E\left[\left(x_{it} - x_{it-1}\right)\left(v_{it} - v_{it-1}\right)\right] \neq 0$, which implies that the OLS estimator is biased.
If v_{it} are not correlated over time, lagged values of x_{it} can be used as instruments for endogenous first-differences, and the model can be estimated by instrumental variables (IV).

Endogenous variables (cont.)

If x_{it} is endogenous, $E(x_{it}v_{it}) \neq 0$ and $E(x_{it-1}v_{it-1}) \neq 0$. So, we have as valid instruments the variables x_{is}, with $s = 1, ..., t-2$, because $E(x_{it-2}v_{it}) = 0$ and $E(x_{it-2}v_{it-1}) = 0$.
If x_{it} is predetermined, $E(x_{it}v_{it-1}) \neq 0$, but $E(x_{it-1}v_{it-1}) = 0$. So, a set of valid instruments is defined by x_{is}, with $s = 1, ..., t-1$.

Dynamic panel data

Sometimes economic theory leads to a dynamic specification of the model. This is the case, for example, of tobacco consumption by an individual that can be defined as a function of tobacco consumption in the previous period.
A dynamic panel data model is specified as

$$y_{it} = \alpha y_{it-1} + x_{it}' \beta + \eta_i + v_{it} \tag{26}$$

The short-run effect of the regressors is defined as $\beta \left(\partial y_{it} / \partial x_{it} \right)$, while the long-run effect (*steady-state*) is given by $\beta / \left(1 - \alpha\right)$.

$$\begin{aligned} \frac{\partial y_{it+1}}{\partial x_{it}} &= \alpha \frac{\partial y_{it}}{\partial x_{it}} = \alpha \beta \\ \frac{\partial y_{it+2}}{\partial x_{it}} &= \alpha \frac{\partial y_{it+1}}{\partial x_{it}} = \alpha \left(\alpha \beta \right) = \alpha^2 \beta \end{aligned}$$

The long-run effect is given by:
$\beta + \alpha\beta + \alpha^2\beta + \ldots = \beta \left(1 - \alpha^{\infty}\right) / \left(1 - \alpha\right) = \beta / \left(1 - \alpha\right)$.

Consider the following model without regressors apart from the lagged dependent variable:

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it} \tag{27}$$

This way, y_{it-1} $(= \alpha y_{it-2} + \eta_i + v_{it-1})$ is correlated, by construction, with η_i, which implies that the *OLS* estimator is biased.
The fixed effects estimator is negatively biased (the bias becomes small when T increases).
The *FE* estimator for α is given by

$$\hat{\alpha}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)(y_{it-1} - \bar{y}_{i,-1})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it-1} - \bar{y}_{i,-1})^2} \tag{28}$$

where $\bar{y}_i = (1/T) \sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i,-1} = (1/T) \sum_{t=1}^{T} y_{it-1}$. Substituting equation (27) in equation (28) we obtain

$$\hat{\alpha}_{FE} = \alpha + \frac{(1/(NT)) \sum_{i=1}^{N} \sum_{t=1}^{T} (v_{it} - \bar{v}_i)(y_{it-1} - \bar{y}_{i,-1})}{(1/(NT)) \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it-1} - \bar{y}_{i,-1})^2} \tag{29}$$

However, this estimator is biased and inconsistent for $N \longrightarrow \infty$ and T fixed, as the last term in equation (29) does not have an expected value of zero, and it does not converge to zero if N goes to infinity. In the literature it has been shown that

$$\begin{aligned} p \lim_{N \longrightarrow \infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (v_{it} - \bar{v}_i)(y_{it-1} - \bar{y}_{i,-1}) &= \\ -\frac{\sigma_v^2}{T^2} \frac{(T-1) - T\alpha + \alpha^T}{(1-\alpha)^2} &\neq 0 \end{aligned} \tag{30}$$

This way, when T is fixed we have an inconsistent estimator, result that does not depend of the assumptions on η_is, as these are eliminated in the estimation. The problem occurs as a result of the transformation of the lagged dependent variable; it is correlated with the transformed residual. If $T \longrightarrow \infty$, the 2^o component in equation (29) converges to zero, such that the FE estimator becomes a consistent estimator for α if $T \longrightarrow \infty$ and $N \longrightarrow \infty$.

Dynamic panel data (cont.)

An alternative that also eliminates η_i is the model in first-differences (originally proposed by Anderson and Hsiao, 1981)

$$y_{it} - y_{it-1} = \alpha \left(y_{it-1} - y_{it-2} \right) + \left(v_{it} - v_{it-1} \right) \tag{31}$$

where y_{it-1} is correlated with v_{it-1} (y is predetermined), such that the *OLS* estimator applied to this equation is highly negatively biased.
However, a set of valid instruments for $\left(y_{it-1} - y_{it-2} \right)$ includes lagged levels of the variable y, y_{it-2}, y_{it-3}, ..., y_{i1}, as, for example, $E\left[y_{it-2} \left(v_{it} - v_{it-1} \right) \right] = 0$.
An *IV* estimator that combines in an optimal way this information is the *GMM* (*Generalized Method of Moments; solution traditionally associated with* Arellano and Bond, 1991) estimator.
The consistency of the estimator depends on the absence of serial correlation in v_{it}.

Dynamic panel data (cont.)

Let Δv_i be the vector of error terms for individual i in the equation in first-differences:

$$\Delta v_i = \begin{bmatrix} v_{i3} - v_{i2} \\ v_{i4} - v_{i3} \\ \vdots \\ v_{iT} - v_{iT-1} \end{bmatrix} = \begin{bmatrix} \Delta y_{i3} - \alpha \Delta y_{i2} \\ \Delta y_{i4} - \alpha \Delta y_{i3} \\ \vdots \\ \Delta y_{iT} - \alpha \Delta y_{iT-1} \end{bmatrix} \tag{32}$$

and Z_i the matrix of valid instruments for individual i

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & 0 \\ 0 & y_{i1} \; y_{i2} & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & y_{i1} \; y_{i2} \; \ldots \; y_{iT-2} \end{bmatrix} \tag{33}$$

Each line in the matrix contains the instruments that are valid in a given period.

Dynamic panel data (cont.)

There is a column for each moment and available lag for that period; the number of instruments is quadratic in T.

In order to restrict the number of instruments, we can restrict the lag in the definition of the set of instruments.

An alternative that allows for the reduction of instruments is to apply "collapse" to the matrix of instruments

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots \\ y_{i2} & y_{i1} & 0 & \cdots \\ y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{34}$$

Dynamic panel data (cont.)

So

$$E\left(Z_i'\Delta v_i\right) = 0 \tag{35}$$

such that it exists a total $\left(T-1\right)\left(T-2\right)/2$ of valid conditions for the moments.

The *GMM* estimator uses these conditions to estimate the parameters of the model in a consistent and efficient way in two stages. In a first stage the estimator minimizes

$$J_N = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta v_i\right)' W_N \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta v_i\right) \tag{36}$$

where W_N is a symmetric weighting matrix defined positive; N indicates that the matrix depends on the sample size.

Choosing $W_N = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'Z_i\right)^{-1}$ leads to the estimator $2SLS$ (Two stage least squares).

$2SLS$ is a good estimator under the homoskedasticity hypothesis; when this assumption is not valid we should use the *GMM* estimator.

Dynamic panel data (cont.)

The *GMM* estimator of one step, as it is implemented in Stata, uses the following matrix

$$W_{N1} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' A_N Z_i\right)^{-1} \tag{37}$$

$$A_N = \begin{bmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \tag{38}$$

being efficient when the error term is homoskedastic and not serially correlated. Such assumption is generally too restrictive. However, the results for the first step are inconsistent, being easy to obtain robust (to heteroskedasticity and serial correlation) standard-errors.

Dynamic panel data (cont.)

The *GMM* estimator in two stages is efficient under more general conditions, such as in the presence of heteroskedasticity. The efficient weighting matrix is computed as

$$W_N\left(\hat{\alpha}_1\right) = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta\hat{v}_i\Delta\hat{v}_i' Z_i\right)^{-1} \tag{39}$$

$$\Delta\hat{v}_i = \Delta y_i - \hat{\alpha}_1 \Delta y_{it-1} \tag{40}$$

where $\hat{\alpha}_1$ is a consistent estimator (for example, obtained from the *GMM* estimator in one step, were we can have $W_N = I$).
There is a problem for small samples (small number of individuals), as the estimated standard-errors for the *GMM* estimator in two steps tend to be too small (there is a correction available).

Dynamic panel data (cont.)

Consider the previous model

$$y_{it} = \alpha y_{it-1} + x_{it}'\beta + \eta_i + v_{it} \tag{41}$$

Depending on the assumptions about x_{it} we can have additional sets of instruments.
If x_{it} are strictly exogenous, meaning they are not correlated with all residuals v_{is}, we can have

$$E\{x_{is}\Delta v_{it}\} = 0 \text{ for each } s \text{ and } t \tag{42}$$

such that $x_{i1}, \ldots, x_{iT}$ can be added to the list of instruments for the equation in first differences in each period.

Exogeneity/endogeneity of the additional regressors and the set of instruments

Consider the dynamic model with one explanatory variable

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + v_{it} \tag{43}$$

and the model in first-differences

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \beta \Delta x_{it} + \Delta v_{it} \tag{44}$$

Consider the case when $T = 4$. When the variable x is strictly exogenous with respect to v, the instruments are

$$Z_i = \begin{bmatrix} y_{i1}, x_{i1}, \ldots, x_{i4} & 0 \\ 0 & y_{i1}, y_{i2}, x_{i1}, \ldots, x_{i4} \end{bmatrix} \tag{45}$$

Dynamic models with exogenous variables

If the variables x_{it} are not strictly exogenous, but predetermined, in which case the contemporaneous and lagged values of $x_{it}s$ are not correlated with the residuals, we only have $E\{x_{it}v_{is}\} = 0$ for $s \geq t$.
In this case, only $x_{it-1}, ..., x_{i1}$ are valid instruments for the equation in first-differences in period t. The moment conditions are defined as

$$E\{x_{it-j}\Delta v_{it}\} = 0 \text{ for } j = 1, ..., t-1 \text{ (for each } t\text{)} \tag{46}$$

Exogeneity/endogeneity of the additional regressors and the set of instruments

When the variable x is predetermined with relation to v, the instruments become

$$Z_i = \begin{bmatrix} y_{i1}, x_{i1}, x_{i2} & 0 \\ 0 & y_{i1}, y_{i2}, x_{i1}, x_{i2}, x_{i3} \end{bmatrix} \tag{47}$$

When the variable x is endogenous with relation to v, the instruments are

$$Z_i = \begin{bmatrix} y_{i1}, x_{i1} & 0 \\ 0 & y_{i1}, y_{i2}, x_{i1}, x_{i2} \end{bmatrix} \tag{48}$$

Sargan test for the overidentification restrictions

The null hypothesis of this test is that the instruments are valid; *i.e.*, the instruments are not correlated with the residuals of the equation in first-differences. The Sargan test is computed as

$$S = NJ_N\left(\hat{\alpha}_2\right) = N\left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta\hat{v}_{i2}\right)' W_N\left(\hat{\alpha}_1\right)\left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta\hat{v}_{i2}\right) \quad (49)$$

Under the null hypothesis the statistic of the test follows a χ^2_q distribution, where q equals to the number of instruments minus the number of parameters in the model.

Note: the results of the 2^{nd} step are only used in the Sargan test.

In face of heteroskedasticity we use the Hansen test to evaluate the validity of the instruments.

The test becomes fragile when the number of instruments is high.

Test for the serial correlation in the residuals

If the error term v_{it} is serially correlated, the *GMM* estimator of the dynamic model is inconsistent. So, it is important to test for serial correlation in the residuals. Stata implements automatically the test for serial correlation for the residuals in first differences. We should observe 1^{st} order serial correlation, but there should be no 2^{nd} order correlation. The tests are for the hypothesis

$$H_0 \quad : \quad E\left(\Delta v_{it} \Delta v_{it-1}\right) \neq 0 \tag{50}$$
$$H_0 \quad : \quad E\left(\Delta v_{it} \Delta v_{it-2}\right) = 0 \tag{51}$$

We can also test for correlation across additional lags.
Under the null hypothesis of no serial correlation in the residuals, the test statistic follows a reduced normal distribution.
Additionally, we assume that the residuals are not correlated across individuals.
Our model should include time dummies in order to avoid correlation between individuals, and N should be large.

System GMM (*weak instruments*)

Remember that the instruments should satisfy the following conditions:

1. Uncorrelated with the error term in the model
2. Correlated with the endogenous explanatory variable

The Sargan test allows for the test of the correlation between the instruments and the error term. If the test rejects the null hypothesis of no correlation, the *IV* estimator is biased and inconsistent.

However, even if the instruments are not correlated with the error term, it can occur a bias associated with too small samples if the instruments are *weekly correlated* with the endogenous regressor.

System GMM (cont.)

For the dynamic panel data model in first differences

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta v_{it} \tag{52}$$

the lagged levels of $y_{it-2}, ..., y_{i1}$ when used as instruments for Δy_{it-1} become less informative when α increases. For the extreme case of a unit root, $y_{it} = y_{it-1} + v_{it}$, α we cannot identify the *GMM* model in first-differences. The resulting bias due to weak instruments tends to go in the same direction of the bias in the *FE* model (*i.e.*, negative; the parameter is under-estimated).

If y is close to a random walk, the *GMM* estimator in first-differences has a week performance because lag levels of the variable have little information about future changes, and, in that sense, lags of y are weak instruments for the transformed variables.

System GMM (cont.)

We can use additional moment conditions when the initial conditions satisfy

$$E\left(\eta_i \Delta y_{i2}\right) = 0 \tag{53}$$

which happens when when we have a *mean stationary* process

$$\begin{aligned} y_{i1} &= \frac{\eta_i}{1-\alpha} + v_{i1} && (54) \\ E\left(v_{i1}\right) &= E\left(\eta_i v_{i1}\right) = 0 && (55) \end{aligned}$$

The $T-2$ additional moment conditions (additional to the conditions for the model in first differences) are in this case

$$\begin{aligned} E\left(u_{it}\Delta y_{it-1}\right) &= E\left[\left(\eta_i + v_{it}\right)\Delta y_{it-1}\right] \\ &= E\left[\left(y_{it} - \alpha y_{it-1}\right)\Delta y_{it-1}\right] = 0 && (56) \end{aligned}$$

which relates to the absence of correlation between the composite error term of the equation in levels (27) and Δy_{it-1}.

System GMM (cont.)

For Blundell and Bond (1998) the process should be convergent, with $|\alpha| < 1$.
If $y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$, conditional on η_i, we expect that y_{it} converges to $\eta_i / (1 - \alpha)$.
The deviations from the initial observations, y_{i1}, of the long-run convergence values should not be correlated with the fixed effects: $E(\eta_i (y_{i1} - \eta_i / (1 - \alpha))) = 0$.
If this condition is met in the first period, it will also be met in the following ones.
This assumption about initial conditions implies that units with bigger growth (units with bigger fixed effects) are not systematically close or away of their stationary state, when compared with the units with smaller growth.

System GMM (cont.)

- The *GMM* estimator that combines the moment conditions of the model in first-differences and the model in levels is called the *System* estimator (Arellano and Bover, (1995); Blundell and Bond, 1998), and it has been shown that it works much better (lower bias and more precision), particularly when α is big; *i.e.*, when a time series is persistence. This result is due to the fact that Δy_{it-1} is a good instrument for y_{it-1}, explains well y_{it-1}, independently of the value of α. The Sargan test should be used in order to evaluate the validity of the additional moment conditions.
- The central assumption of this procedure is that the first-differences of instruments are uncorrelated with the fixed effects.
- The introduction of more instruments increases the efficiency of the estimator.

Reduced form for the model in differences

When $T = 3$ there exists only one instrument, y_{i1}, for the endogenous variable Δy_{i2}. The reduced form is

$$\Delta y_{i2} = \pi_d y_{i1} + r_i \tag{57}$$

given that

$$\Delta y_{i2} = (\alpha - 1)\, y_{i1} + \eta_i + v_{i2} \tag{58}$$

the *plim* of the *OLS* estimator for π is (assuming covariance stationarity)

$$p\lim \hat{\pi}_d = (\alpha - 1)\, \frac{k}{\sigma_\eta^2 / \sigma_v^2 + k}; \;\; k = \frac{1-\alpha}{1+\alpha} \tag{59}$$

such that *plim* $\hat{\pi}_d \longrightarrow 0$ if $\alpha \longrightarrow 1$ and/or $\frac{\sigma_\eta^2}{\sigma_v^2} \longrightarrow \infty$.

Reduced form for the model in differences (cont.)

For $T = 3$ the IV estimator for α is defined as

$$\hat{\alpha}_d = \frac{y_1' \Delta y_3}{y_1' \Delta y_2} \tag{60}$$

with asymptotic variance

$$asyvar\left(\hat{\alpha}_d\right) = 2\left(1+\alpha\right)\left(\frac{\left(1-\alpha\right)+\left(1+\alpha\right)\frac{\sigma_\eta^2}{\sigma_v^2}}{\left(1-\alpha\right)^2}\right) \tag{61}$$

which goes to infinity if $\alpha \longrightarrow 1$ and/or $\frac{\sigma_\eta^2}{\sigma_v^2} \longrightarrow \infty$.

Reduced form for the model in levels

Once more, when $T = 3$ the reduced form for the endogenous y_{i2} is

$$y_{i2} = \pi_l \Delta y_{i2} + s_{i2} \tag{62}$$

$$p \lim \hat{\pi}_l = \frac{1}{2} \tag{63}$$

and the parameter of interest does not go to 0.

The *IV* estimator, using only the level moment condition is given by

$$\hat{\alpha}_l = \frac{\Delta y_2' y_3}{\Delta y_2' y_2} \tag{64}$$

$$asyvar\left(\hat{\alpha}\right) = 2\left(1+\alpha\right)\left(1+\frac{\sigma_\eta^2}{\sigma_v^2}\right) \tag{65}$$

which does not go to infinity if $\alpha \longrightarrow 1$ (but goes to infinity if $\frac{\sigma_\eta^2}{\sigma_v^2} \longrightarrow \infty$).

In the $AR(1)$ model, let u_i^+ be the vector of error terms for individual i for the first-differences and level equations

$$u_i^+ = \begin{bmatrix} v_{i3} - v_{i2} \\ \vdots \\ v_{iT} - v_{iT-1} \\ \eta_i + v_{i3} \\ \vdots \\ \eta_i + v_{iT} \end{bmatrix} = \begin{bmatrix} \Delta y_{i3} - \alpha \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} - \alpha \Delta y_{iT-1} \\ y_{i3} - \alpha y_{i2} \\ \vdots \\ y_{iT} - \alpha y_{iT-1} \end{bmatrix} \tag{66}$$

The system GMM estimator (cont.)

and let Z_i^s be a matrix of instruments for individual i

$$Z_i^s = \begin{bmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i1} \; y_{i2} & 0 & 0 & 0 & 0 & 0 \\ & & \ddots & & & & \\ 0 & 0 & 0 & y_{i1} \; y_{i2} \; \ldots \; y_{iT-2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta y_{i2} & 0 & 0 \\ & & & & & \ddots & \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta y_{iT-1} \end{bmatrix} \tag{67}$$

The system GMM estimator (cont.)

So

$$E\left(Z_i^{s\prime} u_i^{+}\right) = 0 \tag{68}$$

are the moment conditions for the system, with a total $(T-2)+(T-1)(T-2)/2$ moment conditions, used to estimate α by (linear) *GMM*.

For the level equations, we do not transform the regressors to eliminate the fixed-effect in order to make them exogenous with respect to this term.

For the variables close to a random walk, past changes can be better for the prediction of current levels.

The validity of these instruments depends once more on the no serial correlation in v_{it}.

The system GMM estimator (cont.)

Contrary to the *GMM* estimator for first-differences, there is no efficient weighting matrix for the one step estimator when v_{it} are homoskedastic and without serial correlation. For Stata ($\mathrm{XTABOND2}$) the weighting matrix is given by $W_N = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{s\prime} H_N Z_i^s\right)^{-1}$, which is efficient if $\sigma_\eta^2 = 0$.

$$H_N = \begin{bmatrix} 2 & -1 & -1 & \cdots & 1 & 0 & 0 & \cdots \\ -1 & 2 & -1 & \cdots & -1 & 1 & 0 & \cdots \\ -1 & -1 & 2 & \cdots & 0 & -1 & 1 & \cdots \\ . & . & . & \cdots & . & . & . & \cdots \\ 1 & -1 & 0 & \cdots & 1 & 0 & 0 & \cdots \\ 0 & 1 & -1 & \cdots & 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots & 0 & 0 & 1 & \cdots \\ . & . & . & \cdots & . & . & . & \cdots \end{bmatrix} \tag{69}$$

The system GMM estimator: guidelines

1. Apply the estimators to panels with small T and large N
2. Include time dummies
3. Use orthogonal deviations in panels with gaps
4. Try to put all regressors in the instruments' matrix
5. Report and pay attention to the number of instruments used in the estimations
6. Verify the validity of the assumptions underlying the *GMM* estimator
7. In the results mention all the choices made in the specification of the model

Some Monte Carlo results

Consider the model

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + v_{it} \tag{70}$$

$$x_{it} = \rho x_{it-1} + \tau \eta_i + \theta v_{it} + e_{it} \tag{71}$$

with $T = 8$, $N = 500$, $\beta = 1$, $\tau = 0.25$, $\theta = -0.1$, $\sigma^2_\eta = 1$, $\sigma^2_v = 1$, $\sigma^2_e = 0.16$, *Normal*, 10000 replications.

Some Monte Carlo results (cont.)

		OLS		*WG*		*DIF*		*SYS*	
		Mean	*StD*	*Mean*	*StD*	*Mean*	*StD*	*Mean*	*StD*
$\rho = 0.5$									
	ρ	.762	.012	.265	.018	.494	.034	.501	.024
$\alpha = 0.5$	α	.820	.007	.311	.017	.480	.040	.511	.027
	β	.775	.034	.490	.045	.930	.136	.997	.124
$\alpha = 0.95$	α	.990	.001	.662	.016	.548	.177	.979	.011
	β	.581	.035	.388	.044	.226	.356	.983	.101

Some Monte Carlo results (cont.)

		OLS		*WG*		*DIF*		*SYS*	
		Mean	*StD*	*Mean*	*StD*	*Mean*	*StD*	*Mean*	*StD*
$\rho = 0.95$									
	ρ	.997	.001	.591	.017	.676	.222	.958	.031
$\alpha = 0.5$	α	.650	.009	.396	.015	.480	.033	.518	.021
	β	.830	.022	.796	.040	.800	.290	1.075	.059
$\alpha = 0.95$	α	.962	.001	.882	.009	.927	.025	.957	.003
	β	.902	.017	.745	.040	.615	.400	1.019	.031

Dynamic panel data models and variance correction

1. The hypothesis testing for the efficient *GMM* estimator in two steps over-reject in small samples due to a negative bias in the estimation of the standard-error
2. We need to correct asymptotic variance for small samples due to the presence of estimated parameters in the weighting matrix (Windmeijer, 2005, solution for the 2^{nd} step standard-errors)

Windmeijer's (2005) procedure is valid if:

1. The *GMM* estimator does not have a significant bias in small samples
2. The *GMM* estimator has a symmetric distribution

Dynamic panel data models and variance correction (cont.)

Consider the model

$$\begin{aligned} y_{it} &= \beta_0 x_{it} + u_{it} && (72) \\ u_{it} &= \eta_i + v_{it} && (73) \end{aligned}$$

$E(x_{it}\eta_i) \neq 0$ and x_{it} is a predetermined variable,

$$\begin{aligned} E(x_{it}v_{it+s}) &= 0;\;\; s = 0, ..., T - t && (74) \\ E(x_{it}v_{it-r}) &\neq 0;\;\; r = 0, ..., t - 1 && (75) \end{aligned}$$

$T(T-1)/2$ moment conditions for the first-differences model

$$\begin{aligned} \Delta y_{it} &= \beta_0 \Delta x_{it} + \Delta u_{it} && (76) \\ \Delta u_{it} &= \Delta v_{it} && (77) \\ E\left(x_i^{t-1}\Delta u_{it}\right) &= 0; x_i^{t-1} = \left(x_{i1}, ..., x_{it-1}\right) && (78) \end{aligned}$$

Dynamic panel data models and variance correction (cont.)

In small samples the asymptotic standard-errors associated with the 2-step *GMM* estimator are too small. This occurs because of the presence of the estimated parameters $\hat{\beta}_1$ in the weighting matrix.
A correction for the variance-covariance matrix is defined as

$$\begin{aligned} V\hat{a}r_c\left(\hat{\beta}_2\right) &= N\left(\Delta x' Z W_N^{-1}\left(\hat{\beta}_1\right) Z' \Delta x\right)^{-1} + D_{\hat{\beta}_2, W_N\left(\hat{\beta}_1\right)} V\hat{a}r\left(\hat{\beta}_1\right) \\ &\quad + 2N\left(\Delta x' Z W_N^{-1}\left(\hat{\beta}_1\right) Z' \Delta x\right)^{-1} D_{\hat{\beta}_2, W_N\left(\hat{\beta}_1\right)} \end{aligned} \tag{79}$$

where

$$\begin{aligned} D_{\hat{\beta}_2, W_N\left(\hat{\beta}_1\right)} &= -\left(\Delta x' Z W_N^{-1}\left(\hat{\beta}_1\right) Z' \Delta x\right)^{-1} \Delta x' Z W_N^{-1}\left(\hat{\beta}_1\right) \frac{\partial W_N(\beta)}{\partial \beta}\Big|_{\hat{\beta}_1} \\ &\quad \times W_N^{-1}\left(\hat{\beta}_1\right) Z' \Delta \hat{u}_2 \end{aligned} \tag{80}$$

Unit root tests for panel data

Testing for unit roots in panels is important as we want to know if the *GMM* estimator in first-differences is identifiable, or if we need another estimator.

We can define two types of test: (i) those that are based on consistent estimates of α both under the null hypothesis of a unit root, and under the alternative; (ii) those that are based on estimators that are only consistent (or that have a bias that can be characterized) under the null hypothesis.

As the *OLS* estimator is consistent under the null hypothesis of a random walk, a simple test is based on a t test to the parameter estimated by *OLS*.

Under the null hypothesis $H_0 : \alpha = 1$, the *OLS* estimator in the model

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1} + u_{it} & (81)\\
u_{it} &= (1-\alpha)\eta_i + v_{it} & (82)
\end{aligned}
$$

is unbiased and consistent, and a t test based on the *OLS* results is given by

$$
t_{OLS} = \frac{\hat{\alpha}_{OLS} - 1}{\sqrt{Var(\hat{\alpha}_{OLS})}} \quad (83)
$$

where

$$
Var(\hat{\alpha}_{OLS}) = \left(y_{-1}' y_{-1}\right)^{-1} \sum_{i=1}^{N} y_{i,-1}' e_i e_i' y_{i,-1} \left(y_{-1}' y_{-1}\right)^{-1} \quad (84)
$$

with $e_i = y_i - y_{i,-1}\hat{\alpha}_{OLS}$, $y_i = (y_{i2}, ..., y_{iT})'$, $y_{i,-1} = (y_{i1}, ..., y_{iT-1})'$, and $y_{-1} = (y_1', y_2', ..., y_N')'$. Under the null, t_{OLS} follows asymptotically a reduced normal distribution. Under the alternative, $\alpha < 1$, the *OLS* estimator is positively biased.

This is a test of the type *Dickey − Fuller*. This is a modified version, based on the OLS estimator for α in the transformed model

$$y_{it} - y_{i1} = \alpha \left(y_{it-1} - y_{i1} \right) + w_{it},\ t = 3, \ldots, T \tag{85}$$

where $w_{it} = v_{it} - (1 - \alpha) \left(y_{i1} - \eta_i \right)$.
In this model the *OLS* estimator is unbiased when $\alpha = 1$, and a valid t test under the null of a unit root. When $\alpha < 1$ the *OLS* estimator is positively biased. The asymptotic bias is given by

$$p \lim_{N \longrightarrow \infty} \hat{\alpha}_{BM} = \frac{\alpha + 1}{2} \tag{86}$$

Unit root tests for panel data (cont.)

OLS in first-differences

Given the model

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta v_{it} \tag{87}$$

where

$$\sqrt{N}\left(\hat{\alpha}_{FD}\right) \longrightarrow N\left(0, \sigma^2_{FD}\right) \tag{88}$$

under the null hypothesis of a unit root. $\sigma^2_{FD} = \left(T-2\right)^{-1}$ when v_{it} are homoskedastic. The bias is the same when we have heteroskedasticity over time. Under a covariance stationarity hypothesis we have

$$p\lim \hat{\alpha} + 1 = \frac{\alpha + 1}{2} \tag{89}$$

Unit root tests for panel data (cont.)

Harris and Tzavalis (FE model)

The test for the unit root hypothesis is based on a correction for the bias of the fixed-effects estimator under the null. Under the hypothesis that $v_{it} \sim iid\ N\left(0, \sigma_v^2\right)$ and that y_{i1} are observable constants, y_{i1} is not correlated with the sequence $\left\{v_{it}\right\}$. Under the null hypothesis of a unit root in the model (81) we have

$$\sqrt{N}\left(\hat{\alpha}_{FE} - 1 - B\right) \longrightarrow N\left(0, C\right) \tag{90}$$

where $\hat{\alpha}_{FE}$ is the fixed-effects estimator of α, and B and C are defined as

$$B = -\frac{3}{T} \tag{91}$$

$$C = \frac{3\left(17\left(T-1\right)^2 - 20\left(T-1\right) + 17\right)}{5T^3\left(T-2\right)} \tag{92}$$

The t test for $H_0 : \alpha = 1$ is given by $t_{HT} = \left(\hat{\alpha}_{FE} - 1 - B\right) / \sqrt{C/N}$, which, under the null, follows asymptotically a reduced normal distribution.

Unit root tests for panel data (cont.)

Consider the following model

$$y_{it} = \alpha_i + \gamma_i y_{it-1} + \varepsilon_{it} \tag{93}$$

which can be written as

$$\Delta y_{it} = \alpha_i + \pi_i y_{it-1} + \varepsilon_{it} \tag{94}$$

where $\pi_i = \gamma_i - 1$.
The null hypothesis that all series have a unit root is defined by

$$H_0 : \pi_i = 0, \ \forall i \tag{95}$$

where one choice for the alternative hypothesis (stationarity of all series) can be defined as

$$H_1 : \pi_1 = \pi < 0, \text{ for each } i \tag{96}$$

Levin, Lin and Chu (2002) and Harris and Tzavalis (1999) are examples of this procedure.

Unit root tests for panel data (cont.)

Alternatively

$$H_1 : \pi_i < 0, \text{ for at least one } i \tag{97}$$

where one example is Im, Pesaran and Shin (2003).

- In all tests the null hypothesis defines that the series of all individual units have a unit root
- The null hypothesis can be rejected if any of the coefficients π_i is smaller than zero
- The rejection of the null hypothesis does not imply that all series are stationary
- One test for the stationarity of all series is interesting, as it will be rejected if one of the series is not stationary

References

- Arellano, M. (2003), Panel Data Econometrics, Oxford University Press: New York.
- Arellano, M. and S. Bond. (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", The Review of Economic Studies, 58: 277-97.
- Arellano, M. and O. Bover (1995), "Another look at the instrumental variable estimation of error-components models", Journal of Econometrics, 68: 29-51.
- Blundell, R., and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models", Journal of Econometrics, 87: 115-43.
- Bond, S. (2002), "Dynamic panel data models: A guide to micro data methods and practice", Portuguese Economic Journal, 1: 141-162.
- Greene, W.H. (2011), Econometric Analysis, 7th ed., Prentice-Hall: New Jersey.
- Roodman, D. (2009), "How to do xtabond2: An introduction to difference and system GMM in Stata", The Stata Journal, 9: 86-136.
- Windmeijer, F. (2005), "A finite sample correction for the variance of linear efficient two-step GMM estimators", Journal of Econometrics, 126: 25-51.
- Verbeek, M. (2012), A Guide to Modern Econometrics, 4th ed., John Wiley & Sons, Ltd.: Chichester, England.
- Wooldridge, J. (2008), Econometric Analysis of Cross Section and Panel Data, 2nd ed., The MIT Press: Cambridge, Massachusetts.