Linear dynamic panel data models

Laura Magazzini

Sant'Anna School of Advanced Studies
laura.magazzini@santannapisa.it

Dynamic panel data models

Notation & Assumptions

- For a randomly drawn cross section observation, the basic linear multivariate dynamic panel data model can be written as $(t = 2, 3, ..., T)$

$$y_{it} = \rho y_{it-1} + \mathrm{x}'_{it}\beta + c_i + u_{it}$$

  or, alternatively as

$$y_{it} = \mathrm{w}'_{it}\gamma + c_i + u_{it}$$

  with $\mathrm{w}_{it} = (y_{it-1}, \mathrm{x}'_{it})'$ and $\gamma = (\rho, \beta')'$

  - u_{it} denotes the idiosyncratic error term
  - c_i captures individual heterogeneity

FE or RE?

- Is it possible to apply FE?
- Obviouly no: the model violates both the orthogonality condition and the strict exogeneity assumption
- The regressor y_{it-1} is correlated both with c_i and u_{it-1}:

$$y_{it-1} = \rho y_{it-2} + x'_{it-1}\beta + c_i + u_{it-1}$$

The idea of "internal" instruments

- Suppose that u_{it} is uncorrelated over time and consider the first difference of our model

$$y_{it} - y_{it-1} = \rho(y_{it-1} - y_{it-2}) + (\mathrm{x}_{it} - \mathrm{x}_{it-1})'\beta + (c_i - c_i) + (u_{it} - u_{it-1})$$

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \mathrm{x}_{it}'\beta + \Delta u_{it}$$

- OLS is inconsistent: Δy_{it-1} is correlated with Δu_{it}
  - ▷ You need to apply IV estimation
- Can you think of a valid instrument?

The idea of "internal" instruments

- Suppose that u_{it} is uncorrelated over time and consider the first difference of our model

$$y_{it} - y_{it-1} = \rho(y_{it-1} - y_{it-2}) + (\mathrm{x}_{it} - \mathrm{x}_{it-1})'\beta + (c_i - c_i) + (u_{it} - u_{it-1})$$

$$\Delta y_{it} = \rho\Delta y_{it-1} + \Delta \mathrm{x}_{it}'\beta + \Delta u_{it}$$

- Can you think of a valid instrument?
- *When u_{it} is uncorrelated over time, values of y_{it-2} are valid instruments in the equation estimated in first differences* (Anderson & Hsiao, 1981, 1982)
  - ▷ Is it relevant?
  - ▷ Is it exogenous?

Sequential exogeneity

$$E(u_{it}|\mathrm{w}_{it}, \mathrm{w}_{it-1}, ..., \mathrm{w}_{i2}, c_i) = 0$$

- $\mathrm{w}_{it} = (y_{it-1}, \mathrm{x}'_{it})'$ $(t = 1, 2, \ldots T)$ are **sequentially exogenous** conditional on unobserved heterogeneity c_i
- One implication of sequential exogeneity is

  $$E(\mathrm{w}_{is} u_{it}) = 0$$

  with $t = 2, 3, ..., T$ and $s \leq t$
- In words, the idiosyncratic error in each time period is assumed to be uncorrelated only with past and present values of the explanatory variables but is allowed to be correlated with future values of the explanatory variables

- Show that sequential exogeneity implies $E(\mathrm{w}_{is}u_{it}) = 0$ with $t = 2, 3, ..., T$ and $s \leq t$

$$\begin{aligned}
E(\mathrm{w}_{is}'u_{it}) &= E[E(\mathrm{w}_{is}'u_{it}|\mathrm{w}_{is})] \\
&= E[\mathrm{w}_{is}'E(u_{it}|\mathrm{w}_{is})] \\
&= E[\mathrm{w}_{is}'E(E(u_{it}|\mathrm{w}_{it}, \mathrm{w}_{it-1}, ..., \mathrm{w}_{i2}, c_i)|\mathrm{w}_{is})] \\
&= E[\mathrm{w}_{is}'E(0|\mathrm{w}_{is})] \\
&= 0
\end{aligned}$$

- Another useful implication of sequential exogeneity is

$$E(u_{it}u_{it-j}) = 0 \quad \text{for } j > 0$$

- In words, under sequential exogeneity, the idiosyncratic error terms are serially uncorrelated
- In fact, by the Law of Total Expectations

$$E(u_{it}u_{it-j}) = E[E(u_{it}u_{it-j}|\mathrm{w}_{it}, \mathrm{w}_{it-1}, \mathrm{w}_{i2}, c_i)]$$

- Since u_{it-j} is a linear combination of the variables in the conditioning set

$$E[E(u_{it}u_{it-j}|\mathrm{w}_{it}, \mathrm{w}_{it-1}, \mathrm{w}_{i2}, c_i)] = E[u_{it-j}E(u_{it}|\mathrm{w}_{it}, \mathrm{w}_{it-1}, \mathrm{w}_{i2}, c_i)] = 0$$

The first order AR panel data model

$$
\begin{aligned}
y_{it} &= \rho y_{it-1} + v_{it} \\
v_{it} &= c_i + u_{it} \\
|\rho| &< 1 \quad t = 2, ..., T
\end{aligned}
$$

- Standard panel data estimation methods such as within group (WG) and first differences (FD) allow for unrestricted correlation between unobserved heterogeneity and past, present and future values of the right hand side variables
- However, in order to achieve consistency they require that all regressors are strictly exogenous
- This assumption is clearly violated in the case of the lagged dependent variable since u_{it} cannot be orthogonal (and therefore not correlated) to $y_{it}, y_{it-1}, ..., y_{iT}$

Inconsistent of pooled OLS, WG, FD

- Under sequential exogeneity it can be shown that pooled OLS, WG and FD estimators are all inconsistent for fixed T
- There is a difference however since under appropriate stability conditions ($|\rho| < 1$ in the AR(1) model) the inconsistency in the WG estimator – but not in the pooled OLS or in the FD estimator – is of order T^{-1}
- This in turn implies that the asymptotic bias of the WG estimator – but not of the OLS and FD estimators – goes to zero as T goes to infinity

Inconsistency of pooled OLS

- For the pooled OLS estimator we can write

$$\hat{\rho}_{OLS} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1} y_{it} \right)$$

- To study the asymptotic properties of OLS, we substitute $y_{it} = \rho y_{it-1} + v_{it}$ to get:

$$\hat{\rho}_{OLS} = \rho + \left(\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1} v_{it} \right)$$

Inconsistency of pooled OLS

- The asymptotic bias of the simple pooled OLS estimator for ρ is given by

$$\text{plim}(\hat{\rho}_{OLS} - \rho) = (1-\rho)\frac{\frac{\sigma_c^2}{\sigma_u^2}}{\frac{\sigma_c^2}{\sigma_u^2} + \frac{1-\rho}{1+\rho}} > 0$$

  where $E(c_i^2) = \sigma_c^2$ and $E(u_{it}^2) = \sigma_u^2$
- It follows that the pooled OLS estimator is *biased upwards* and its asymptotic bias does not depend on T
- The consistency of the pooled OLS estimator cannot be rescued by allowing the time dimension to grow without bounds

Inconsistency of WG

- For the WG regression, we can write

$$\hat{\rho}_{WG} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \ddot{y}_{it-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \ddot{y}_{it-1} \ddot{y}_{it} \right)$$

with $\ddot{y}_{it} = y_{it} - \bar{y}_i$ and $\bar{y}_i = \sum_T y_{it}/T$

- As in the previous case, to analyse the asymptotic properties of $\hat{\rho}_{WG}$ we substitute $\ddot{y}_{it} = \rho \ddot{y}_{it-1} + \ddot{u}_{it}$:

$$\hat{\rho}_{WG} = \rho + \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \ddot{y}_{it-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \ddot{y}_{it-1} \ddot{u}_{it} \right)$$

Inconsistency of WG

- Nickell (1981) derives the analytical expression for the asymptotic bias

$$\text{plim}(\hat{\rho}_{WG} - \rho) = -\frac{\frac{1+\rho}{T-1}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)}{1 - \frac{2\rho}{(1-\rho)(T-1)}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)}$$

- Nickell also provides a simple approximation

$$\text{plim}(\hat{\rho}_{WG} - \rho) \approx -\frac{1+\rho}{T-1}$$

  for reasonably large values of T
- This makes it clear that when $\rho > 0$, the WG estimator is biased downwards
- Furthermore, even with $T = 10$, which is the order of magnitude of most sets of panel data, if $\rho = 0.5$, than the asymptotic bias is $-.167$
- If T also goes to infinity, the asymptotic bias converges to 0 so that the WG estimator is consistent for ρ if both $N \to \infty$ and $T \to \infty$

Inconsistency of FD

- For the FD estimator, we can write

$$\hat{\rho}_{FD} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta y_{it-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta y_{it-1} \Delta y_{it} \right)$$

with $\Delta y_{it} = y_{it} - y_{it-1}$

- To study the asymptotic properties of $\hat{\rho}_{FD}$, we substitute $\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$ to get:

$$\hat{\rho}_{FD} = \rho + \left(\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1} \Delta u_{it} \right)$$

Inconsistency of FD

- The asymptotic bias of the FD estimator is given by:

$$\text{plim}(\hat{\rho}_{FD} - \rho) = -\frac{1+\rho}{2} < 0$$

  and therefore the FD estimator is biased downwards and its asymptotic bias does not depend on T

How can ρ be consistently estimated?

- The simplest linear dynamic panel data model is the first order autoregressive panel data model:

$$y_{it} = \rho y_{it-1} + c_i + u_{it}$$

with $|\rho| < 1$ and $t = 2, ..., T$

- In order to consistently estimate ρ:

1. Use the First Difference transformation to remove the individual effect

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$$

2. Exploit the sequential exogeneity assumption to find instruments for Δy_{it-1} (and, therefore, *moment conditions*)
3. ρ can be consistently estimated by using the **GMM estimator**

Assumption: sequential exogeneity

$$E(u_{it}|y_{it-1}, y_{it-2}, ..., y_{i1}, c_i) = E(u_{it}|\mathrm{y}_i^{t-1}, c_i) = 0$$

- Recall that the seq.exo. assumption implies lack of autocorrelation in u_{it}
  - Lagged u_{is} are linear combinations of the variables in the conditioning set (e.g. $u_{it-1} = y_{it-1} - \rho y_{it-2} - c_i$)
- How can this assumption be exploited to find instruments for Δy_{it-1}?

Anderson & Hsiao (1981, 1982)

- One simple possibility suggested by Anderson and Hsiao (1981, 1982) is to use y_{it-2} as instrument for Δy_{it-1}
  - Relevance: Is y_{it-2} correlated with Δy_{it-1}?
  - Exogeneity: Is y_{it-2} uncorrelated with Δu_{it}? That is, $E[y_{it-2}\Delta u_{it}] = 0$?
- They also proposed an alternative where Δy_{it-2} is used instead as instrument
- Both these IV estimators thus exploit one moment condition in estimation, respectively

$$E[y_{it-2}\Delta u_{it}] = E[y_{it-2}(\Delta y_{it} - \rho \Delta y_{it-1})] = 0$$

and

$$E[\Delta y_{it-2}\Delta u_{it}] = E[\Delta y_{it-2}(\Delta y_{it} - \rho \Delta y_{it-1})] = 0$$

Arellano & Bond (1991)

- Arellano and Bond (1991) proposed to use the entire set of instruments (or at least a larger subset) in a GMM procedure
  - ▷ Under seq.exo. values of y lagged twice *or more* are valid instruments for Δy_{it-1} in the first differecing version of the equation of interest

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$$

  - ▷ So, at a generic time t, we can use $y_{i1}, ..., y_{it-2}$ as potential instruments for Δy_{it-1}
  - ▷ Number of moment conditions available is $0.5(T-1)(T-2)$
  - ▷ See also Holtz-Eakin, Newey, & Rosen (1988)

Arellano & Bond (1991)

- For example, when $T = 5$, six moment conditions are available $(0.5(T-1)(T-2) = 6)$
  - Moment conditions for $t = 3$

$$E[y_{i1}(\Delta y_{i3} - \rho \Delta y_{i2})] = 0$$

  - Moment conditions for $t = 4$

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

  - Moment conditions for $t = 5$

$$\begin{aligned} E[y_{i3}(\Delta y_{i5} - \rho \Delta y_{i4})] &= 0 \\ E[y_{i2}(\Delta y_{i5} - \rho \Delta y_{i4})] &= 0 \\ E[y_{i1}(\Delta y_{i5} - \rho \Delta y_{i4})] &= 0 \end{aligned}$$

Extensions

- This method extends to models with limited moving-average serial correlation in u_{it}
- For instance, if u_{it} is itself MA(1), then the first difference error term Δu_{it} is MA(2)
  - ▷ $u_{it} = \varepsilon_{it} + \alpha\varepsilon_{it-1}$ with $\varepsilon_{it} \sim D(0, \sigma^2_\varepsilon)$
  - ▷ $E(u_{it}u_{it-1}) \neq 0 \rightarrow E(\Delta u_{it}\Delta u_{it-1}) \neq 0$ & $E(\Delta u_{it}\Delta u_{it-2}) \neq 0$
  - ▷ $E(u_{it}u_{it-2}) = 0 \rightarrow E(\Delta u_{it}\Delta u_{it-3}) = 0$
- As a consequence, y_{it-2} would not be a valid instrumental variable in these first-differenced equations but y_{it-3} and longer lags remain available as instruments
- In this case ρ is identified provided $T \geq 4$

Asymptotic Distribution

- Under the assumption of sequential exogeneity, the first difference two-step GMM estimator is efficient and asymptotically normally distributed
- If we also assume homoskedasticity
  - ▷ Conditional homoskedasticity
  - ▷ Time series homoskedasticity

  the first difference one-step GMM estimator is consistent, efficient and asymptotically normal
- If homoskedasticity is not satisfied, the one-step estimator is no longer efficient (still, it remains consistent)
  - ▷ A heteroskedasticity robust estimator of the asymptotic variance matrix for the one-step estimator can be considered

Finite sample results

Monte Carlo set up by Arellano and Bond (1991)

- $y_{it} = \rho y_{it-1} + x_{it} + c_i + u_{it}$ with $y_{i0} = 0$
- $x_{it} = 0.8 x_{it-1} + \varepsilon_{it}$
- $c_i \sim iidN(0,1)$; $u_{it} \sim iidN(0,1)$; and $\varepsilon_{it} \sim iidN(0,1)$ and independent of c_i, u_{is} for all t, s
- Sample size: $N = 100$; $T = 7$
- Mean and standard deviations computed on the basis of 100 replications

Monte Carlo results

	OLS	WG	GMM1	GMM2
$\rho = 0.5$				
Mean	.7216	.3954	.4884	.4920
Std.Dev.	.0216	.0272	.0671	.0739
$\beta = 1$				
Mean	.7002	1.041	1.005	.9976
Std.Dev.	.0484	.0480	.0631	.0668

Suggestion for applied work

- The fact that the pooled OLS estimator is biased *upward* and the WG estimator is biased *downward* can be used to assess the performance of consistent estimators in applications

$$\hat{\rho}_{WG} \leq \rho \leq \hat{\rho}_{OLS}$$

- If the estimate produced by a consistent estimator lies outside the interval defined by the pooled OLS (upper bound) and the WG (loeer bound) estimates, the chosen consistent estimator is likely to have finite sample problems (as it is the case with GMM...)

Ahn & Schmidt (1995)

A non-linear GMM estimator

- Ahn & Schmidt (1995) noticed that under AB assumptions other moment conditions are available
- As u_{it} is uncorrelated over time, and c_i is uncorrelated with u_{it}, it is also true that

$$E(v_{iT} \Delta v_{it}) = 0 \qquad t = 3, ..., T-1$$

  with $v_{it} = c_i + u_{it}$
- Moment conditions are quadratic, so that non-linear GMM estimation is required
- Efficiency gains with respect to AB91 when β close to 1 and high σ_c^2 / σ_u^2
  - Limited application so far as not available in most common econometric softwares (in STATA an unofficial command released in 2017)

The weak instruments problem

- Blundell and Bond (1998) show that the instruments used in the standard first difference GMM estimator become less informative in two important cases
- First, as the value of ρ increases towards unity, and second as the variance of the unobserved heterogeneity component, σ_c^2 increases with respect to the variance of the idiosyncratic error term, σ_u^2
- To see this we can focus on the case with $T = 3$:
  - ▷ We are left with only one moment condition $E(y_{i1}\Delta u_{i3}) = 0$
  - ▷ ρ is just identified

Weak instruments problem with $T = 3$

- In this case we use one equation in first differences ($i = 1, ..., N$)

$$\Delta y_{i3} = \rho \Delta y_{i2} + \Delta u_{i3}$$

  where we use y_{i1} as an instrument for Δy_{i2}
- The GMM estimator reduces to a simple IV estimator with the following reduced form equation

$$\Delta y_{i2} = \pi y_{i1} + r_i$$

- For sufficiently high values of ρ or σ_c^2/σ_u^2, the OLS estimate of the reduced form coefficient π can be made arbitrarily close to zero
- In this case y_{i1} is only weakly correlated with Δy_{i2} and thus a **weak instruments problem** arise

Weak instruments problem with $T = 3$

- It can be shown that

$$\text{plim}\hat{\pi}_{LS} = (\rho - 1)\frac{\frac{(1-\rho)^2}{(1-\rho^2)}}{\frac{\sigma_c^2}{\sigma_u^2} + \frac{(1-\rho)^2}{(1-\rho^2)}}$$

- For instance, with $\frac{\sigma_c^2}{\sigma_u^2} = 1$ and $\rho = 0.8$, $\text{plim}\hat{\pi}_{LS} = -0.02$
- In words, the instruments available for the equations in first-differences are likely to be weak when the individual series are persistent, that is when they have near unit root properties
- Blundell and Bond (1998) propose additional moment conditions that allow to solve this problem
    - ▷ See also Arellano and Bover (1995)

GMM "system" estimator

$$y_{it} = \rho y_{it-1} + c_i + u_{it}$$

- Arellano and Bond (1991) GMM-DIF estimator uses lags of y_{it} (lag two or older) as instruments for the equation in first differences
- Blundell and Bond (1998) suggest using lags of Δy_{it} (lag one or older) as instruments for the equation in levels

$$E(v_{it}\Delta y_{it-1}) = 0 \quad t = 3, 4, ..., T$$

- An additional assumption is needed for consistency: the **mean stationarity assumption**

The mean stationarity assumption

Roodman (2009)

$$y_{it} = \rho y_{it-1} + c_i + u_{it}$$

- Entities in this system can evolve much like GDP per worker in the Solow growth model, converging towards mean stationarity
- By itself, a positive fixed effect, for instance, provides a constant, repeated boost to y in each period, like investment does for the capital stock
- But, assuming $|\rho| < 1$, this increment is offset by reversion towards the mean
- The observed entities therefore converge to steady-state levels defined by

$$E[y_{it}|c_i] = E[y_{it+1}|c_i] \Rightarrow y_{it} = \rho y_{it} + c_i \Rightarrow y_{it} = \frac{c_i}{1-\rho}$$

- The fixed effects and the autocorrelation coefficient interact to determine the long-run mean of the series

The mean stationarity assumption

Roodman (2009)

- Now consider the momemt conditions proposed by Blundell and Bond (1998): $E(v_{it}\Delta y_{it-1}) = 0,\ t = 3, 4, ..., T$
- We can write $(t \geq 3)$

$$
\begin{aligned}
E(v_{it}\Delta y_{it-1}) &= E[(c_i + u_{it})(y_{it-1} - y_{it-2})] \\
&= E[(c_i + u_{it})(\rho y_{it-2} + c_i + u_{it-1} - y_{it-2})] \\
&= E[(c_i + u_{it})((\rho - 1)y_{it-2} + c_i + u_{it-1})] \\
&= E[c_i((\rho - 1)y_{it-2} + c_i)] = 0
\end{aligned}
$$

which is equivalent to $E[c_i((\rho - 1)y_{it} + c_i)] = 0$ for $t \geq 1$

The mean stationarity assumption

Roodman (2009)

$$E[c_i((\rho - 1)y_{it} + c_i)] = 0 \quad \text{for } t \geq 1$$

- By dividing this condition by $(1 - \rho)$ we get:

$$E\left[c_i\left(y_{it} - \frac{c_i}{1-\rho}\right)\right] = 0$$

- Deviations from long-run means must not be correlated with the fixed effects
- It is possible to show that if this condition holds in t then it also holds in all subsequent periods
- Effectively, this is a condition on the *initial observation*

GMM "system" estimator

- The conditions imposed by Blundell and Bond (1998) imply that additional $0.5(T-1)(T-2)$ moment conditions for the equation in levels are available
- However, all but $T-2$ moment restrictions are redundant since they can be expressed as linear combinations of the GMM-DIF moment restrictions

For example, with $T = 5$

$(^*)$ restrictions are redundant

- Moment conditions when $t = 3$

$$E[\Delta y_{i2}(y_{i3} - \rho y_{i2})] = 0$$

- Moment conditions when $t = 4$

$$\begin{aligned} E[\Delta y_{i3}(y_{i4} - \rho y_{i3})] &= 0 \\ E[\Delta y_{i2}(y_{i4} - \rho y_{i3})] &= 0 \; (^*) \end{aligned}$$

- Moment conditions when $t = 5$

$$\begin{aligned} E[\Delta y_{i4}(y_{i5} - \rho y_{i4})] &= 0 \\ E[\Delta y_{i3}(y_{i5} - \rho y_{i4})] &= 0 \; (^*) \\ E[\Delta y_{i2}(y_{i5} - \rho y_{i4})] &= 0 \; (^*) \end{aligned}$$

GMM "system" estimator

- Summarizing, the enlarged set of moment conditions is

DIF. $E[y_{i,t-j}(\Delta y_{it} - \rho \Delta y_{it-1})] = 0$ for $t = 3, 4, ..., T$; $j \geq 2$

LEV. $E[\Delta y_{it-1}(y_{it} - \rho y_{it-1})] = 0$ for $t = 3, 4, ..., T$

- The GMM method can be applied to the "SYS" of equation ("DIF" & "LEV") for the estimation of ρ

The weak instrument problem is solved

- To show that these additional moment restrictions remain informative when ρ increases towards unity and/or when the variance of the unobserved heterogeneity component, σ_c^2 increases with respect to the variance of the idiosyncratic error term, σ_u^2, we can again focus on the case with $T = 3$
- We are left with only one moment condition $E(v_{i3}\Delta y_{i2}) = 0$, so that ρ is just identified
- Here, we can use one equation in level

$$y_{i3} = \rho y_{i2} + c_i + u_{i3}$$

  where we use Δy_{i2} as an instrument for y_{i2}
- The corresponding GMM estimator is an IV estimator with the following reduced form equation

$$y_{i2} = \pi \Delta y_{i2} + r_i$$

The weak instrument problem is solved

- It is possible to show that, in this case

$$\text{plim}\hat{\pi}_{LS} = \frac{1}{2}\frac{1-\rho}{1-\rho^2}$$

- This moment condition stays informative for high values of ρ, in contrast to the moment condition available for the first differenced method
- For instance, $\text{plim}\hat{\pi}_{LS} = 0.28$ for $\rho = 0.8$
- More recently, Bun & Windmeijer (2010) point to the presence of a weak instrument problem also for the BB98 estimator

Finite sample results

Monte Carlo set up by Blundell and Bond (1998)

- $y_{it} = \rho y_{it-1} + c_i + u_{it}$
- The mean stationarity assumption is satisfied: $y_{i0} = \frac{c_i}{1-\rho} + \varepsilon_{i1}$
- $c_i \sim iidN(0,1)$; $u_{it} \sim iidN(0,1)$
- $\varepsilon_{i1} \sim iidN(0, \sigma_\varepsilon^2)$ and independent of c_i, u_{is} for all t, s
- Sample size: $N = 100$; $T = 4, 11$
- Mean and standard deviations computed on the basis of 1000 replications

Monte Carlo results

$T = 4$

	GMM-DIF.			GMM-SYS		
ρ	Mean	Std.Dev	RMSE	Mean	Std.Dev	RMSE
0.0	-0.0044	0.1227	0.1227	0.0100	0.0990	0.0994
0.3	0.2865	0.1849	0.1853	0.3132	0.1215	0.1221
0.5	0.4641	0.2674	0.2693	0.5100	0.1330	0.1330
0.8	0.4844	0.8224	0.8805	0.8101	0.1618	0.1620
0.9	0.2264	0.8264	1.0659	0.9405	0.1564	0.1615

Monte Carlo results

$T = 11$

	GMM-DIF			GMM-SYS		
ρ	Mean	Std.Dev	RMSE	Mean	Std.Dev	RMSE
0.0	-0.0138	0.0463	0.0483	-0.0183	0.0431	0.0468
0.3	0.2762	0.0541	0.0591	0.2728	0.0487	0.0558
0.5	0.4629	0.0623	0.0725	0.4689	0.0535	0.0568
0.8	0.6812	0.1036	0.1576	0.7925	0.0651	0.0655
0.9	0.6455	0.1581	0.2996	0.9259	0.0453	0.0522

Comments

- In these simulations, when $T = 4$ we observe both the poor performance of the first-differenced GMM estimator at high values of ρ and the dramatic improvement in precision that results from exploiting the additional moment conditions
- The poor performance of the FD-GMM estimator improves with the number of time periods available
- However, even with $T = 11$, for high values of ρ there remain sizable gains in bias, precision and RMSE

Multivariate dynamic panel data models

- The multivariate dynamic panel data model is ($|\rho| < 1$; $t = 2, 3, ..., T$)

$$y_{it} = \rho y_{it-1} + \mathrm{x}'_{it}\beta + c_i + u_{it}$$

- Sequential exogeneity assumption

$$E(u_{it}|\mathrm{x}_i^t, \mathrm{y}_i^{t-1}, c_i) = 0$$

with $\mathrm{x}_i^t = (x_{i1}, x_{i2}, ..., x_{it})$ and $\mathrm{y}_i^{t-1} = (y_{i1}, y_{i2}, ..., y_{it-1})$

  - ▷ As in the first order autoregressive panel data model, note that sequantial exogeneity implies lack of serial correlation in u_{it}

Predetermined versus endogenous variables

- The sequential exogeneity assumption, $E(u_{it}|\mathrm{x}_i^t, \mathrm{y}_i^{t-1}, c_i) = 0$, implies that the idiosyncratic error in each time period is assumed to be uncorrelated only with past and present values of the explanatory variables (predetermined variables)
- This allows for feedback effects but not for simultaneity
- To allow for simultaneity, the idiosyncratic error term in each time period has to be assumed to be uncorrelated only with past values of the explanatory variables (endogenous variables)
- In this case we write:

$$E(u_{it}|\mathrm{x}_i^{t-1}, \mathrm{y}_i^{t-1}, c_i) = 0$$

Consistency

- The sequential exogeneity assumption also implies that the following $0.5(T-1)(T-2)+0.5(T+1)(T-2)K$ linear moment restrictions hold

$$\begin{aligned} E[\mathrm{y}_i^{t-2}(\Delta y_{it} - \rho\Delta y_{it-1} - \Delta \mathrm{x}_{it}\beta)] &= 0 \\ E[\mathrm{x}_{ki}^{t-1}(\Delta y_{it} - \rho\Delta y_{it-1} - \Delta \mathrm{x}_{it}\beta)] &= 0 \end{aligned}$$

- The estimating equation is in first differences

$$\Delta y_{it} = \rho\Delta y_{it-1} + \Delta \mathrm{x}_{it}\beta + \Delta u_{it}$$

- Under the sequential exogeneity assumption, values of y lagged twice or more and values of x lagged once or more are valid instruments for Δy_{it-1} and $\Delta \mathrm{x}_{it}$

Example: $T = 5$

- $1 + 2K$ moment conditions for $t = 3$

$$
\begin{aligned}
E[y_{i1}(\Delta y_{i3} - \rho \Delta y_{i2} - \Delta \mathrm{x}_{i3}\beta)] &= 0 \\
E[\mathrm{x}_{i1}(\Delta y_{i3} - \rho \Delta y_{i2} - \Delta \mathrm{x}_{i3}\beta)] &= 0 \\
E[\mathrm{x}_{i2}(\Delta y_{i3} - \rho \Delta y_{i2} - \Delta \mathrm{x}_{i3}\beta)] &= 0
\end{aligned}
$$

- $2 + 3K$ moment conditions for $t = 4$

$$
\begin{aligned}
E[y_{i1}(\Delta y_{i4} - \rho \Delta y_{i3} - \Delta \mathrm{x}_{i4}\beta)] &= 0 \\
E[y_{i2}(\Delta y_{i4} - \rho \Delta y_{i3} - \Delta \mathrm{x}_{i4}\beta)] &= 0 \\
E[\mathrm{x}_{i1}(\Delta y_{i4} - \rho \Delta y_{i3} - \Delta \mathrm{x}_{i4}\beta)] &= 0 \\
E[\mathrm{x}_{i2}(\Delta y_{i4} - \rho \Delta y_{i3} - \Delta \mathrm{x}_{i4}\beta)] &= 0 \\
E[\mathrm{x}_{i3}(\Delta y_{i4} - \rho \Delta y_{i3} - \Delta \mathrm{x}_{i4}\beta)] &= 0
\end{aligned}
$$

- Finally, $3 + 4K$ moment conditions for $t = 5$

$$\begin{aligned}
E[y_{i1}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0 \\
E[y_{i2}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0 \\
E[y_{i3}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0 \\
E[\mathrm{x}_{i1}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0 \\
E[\mathrm{x}_{i2}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0 \\
E[\mathrm{x}_{i3}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0 \\
E[\mathrm{x}_{i4}(\Delta y_{i5} - \rho \Delta y_{i4} - \Delta \mathrm{x}_{i5}\beta)] &= 0
\end{aligned}$$

- The GMM method allows you to obtain consistent estimates of the parameters

GMM-SYS estimator

- Assumptions
  - Sequential exogeneity
  - Mean stationarity also for x
- These assumptions lead to additional non-redundant moment conditions which can be exploited in the context of the GMM-SYS estimator

$$
\begin{aligned}
E[\Delta y_{it-1}(y_{it} - \rho y_{it-1} - x_{it}\beta)] &= 0 \\
E[\Delta x_{it}(y_{it} - \rho y_{it-1} - x_{it}\beta)] &= 0
\end{aligned}
$$

Asymptotically...

- When moment conditions are satisfied (in the population), GMM estimates are

i) consistent

ii) asymptotically normal

iii) can reach "efficiency"
(one-step with spherical disturbances, otherwise two-step)

iv) Variance can be estimated
(different formula with "optimal" weighting matrix, but robust version can also be applied)

Testing the assumptions

Autocorrelation test by Arellano & Bond

- The consistency of the GMM-FD (& SYS) estimator relies on the absence of first (and higher) order serial correlation in the disturbances

$$E(u_{it}u_{it-1}) = 0 \quad t = 3, 4, ..., T$$

- Since the model is estimated in first differences it is useful to rewrite this condition as

$$E(\Delta u_{it}\Delta u_{it-2}) = 0 \quad t = 4, 5, ..., T$$

  - ▷ Lack of autocorrelation in u_{it} implies that Δu_{it} is autocorrelated of order 1, whereas higher order autocorrelation is zero

Testing the assumption

- This assumption can be tested by using the standardized average residual autocovariance

$$m_2 = \frac{\sum_{i=1}^{N} \sum_{t=5}^{T} \Delta \hat{u}_{it} \Delta \hat{u}_{it-2}}{(*)}$$

- The test statistic is asymptotically N(0,1) under the null of no serial correlation in u_{it}
- $(*)$: for details refer to Arellano and Bond (1991)

Sargan test

- More generally, the Sargan test of overidentifying restrictions can also be computed when A_N is chosen optimally
- The test statistic s is asymptotically χ^2 with as many degrees of freedom as overidentifying restrictions, under the null hypothesis of the validity of the instruments
  - ▷ # overidentifying restrictions = # moment conditions − # estimated parameters

Testing the lack of correlation in u_{it}

Arellano & Bond (1991) set up: $H_0 : E(u_{it} u_{is}) = 0$ for any t, s

LoS	m_2^{1s}	m_2^{1sr}	m_2^{2s}	Sargan
10%	12	15	14	7
5%	5	5	6	4
1%	1	1	1	0

- LoS: level of significance
- Experiments with $\rho = 0.5$
- The table reports the share of rejections (H_0 true; size of the test)

Power of the test statistics

Arellano & Bond (1991) set up: $H_0 : E(u_{it} u_{is}) = 0$ for any t, s

- Experiments with different degree of serial correlation: $corr(u_{it}, u_{it-1}) = \rho_u \neq 0$

LoS	m_2^{1s}	m_2^{1sr}	m_2^{2s}	Sargan
$\rho_u = 0.2$				
10%	54	53	53	33
5%	45	46	46	20
1%	24	25	25	2
$\rho_u = 0.3$				
10%	95	96	96	71
5%	91	92	92	47
1%	78	77	78	30

Robustness of the test statistics

Arellano & Bond (1991) set up: $H_0 : E(u_{it} u_{is}) = 0$ for any t, s

- Experiments with heteroskedasticity: $\text{var}(u_{it}) = x_{it}^2$

LoS	m_2^{1s}	m_2^{1sr}	m_2^{2s}	Sargan
10%	23	9	10	8
5%	15	2	2	0
1%	3	0	0	0

Testing mean stationarity

- The validity of the additional “LEV” orthogonality conditions (spanning from the mean stationarity assumption) can be tested by using the so-called Sargan difference test of overidentified retstrictions
- The model is estimated twice, without and with the additional moment conditions
- The difference in the value of the Sargan statistic is found to have a χ^2 distribution under the null (the validity of these additional moment conditions), where the degrees of freedom are given by the number of additional restrictions
- Magazzini & Calzolari (2021) propose a Lagrange Multiplier (LM) test that has more power in detecting violation of the mean stationarity assumption
    - ▷ In the “pure dynamic” model, asy. eq. to comparing SYS with non-linear GMM (Ahn & Schmidt, 1995)

Let us start with the problems...

Estimation of second-sted standard errors

- Let us go back to Arellano & Bond (1991) Monte Carlo set up
- The authors also report the estimated s.e.

	GMM1	1sASE	R1sASE	GMM2	2sASE
$\rho = 0.5$					
Mean	.4884	**.0683**	**.0677**	.4920	**.0604**
Std.Dev.	**.0671**	.0096	.0120	**.0739**	.0106
$\beta = 1$					
Mean	1.005	**.0612**	**.0607**	.9976	**.0548**
Std.Dev.	**.0631**	.0031	.0055	**.0668**	.0052

1sASE: one-step asymptotic standard errors

R1sASE: robust one-step asymptotic standard errors

2sASE: two-step asymptotic standard errors

Comments - 2sASE

- In the Monte Carlo simulations the estimator of the asymptotic standard errors of GMM2 in the last column (2sASE) shows a sizable downward bias relative to the finite-sample standard deviation reported in the second column
- Why?
- The efficient GMM estimator uses A_N^{2s} as the weighting matrix

$$A_N^{2s} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathrm{Z}'_{Di} \Delta \hat{\mathrm{u}}_i \Delta \hat{\mathrm{u}}'_i \mathrm{Z}_{Di} \right)^{-1}$$

- A_N^{2s} is a function of estimated *fourth* moments
- Generally, it takes a substantially larger sample size to estimate fourth moments reliably than to estimate first and second moments
- Also, A_N^{2s} is a function of the parameter and this should be taken into account when computing s.e.

Variance of the 2-step GMM estimator

Finite sample correction

- Montecarlo studies (including AB, 1991) have shown that estimated asymptotic standard errors of the efficient two-step GMM estimator are severely downward biased in small samples
- The weight matrix used in the calculation of the efficient two-step estimator is based on initial consistent parameter estimates
- Windmeijer (2005) shows that the extra variation due to the presence of these estimated parameters in the weight matrix accounts for much of the difference between the finite sample and the asymptotic variance of the two step estimator
- This difference can be estimated, resulting in a finite sample correction of the variance

Monte Carlo experiments

Windmejier (2005)

- $y_{it} = \beta x_{it} + c_i + u_{it}$
- $x_{it} = 0.5x_{it-1} + c_i + 0.5u_{it-1} + \varepsilon_{it}$
  - x_{it} are correlated with c_i and predetermined with respect to u_{it}
- $c_i \sim iidN(0,1)$ and $\varepsilon_{it} \sim iidN(0,1)$
- $u_{it} = \delta_i \tau_u \omega_{it}$
  with $\omega_{it} \sim \chi_1^2 - 1$; $\delta_i \sim U(0.5, 1.5)$; $\tau_t = 0.5 + 0.1(t-1)$
  - u_{it} are skewed and heteroskedastic over time and units
- $N = 100$
- 10000 replications

- Note that

Monte Carlo results

$\beta = 1$; results based on GMM-DIF

	Mean	Std.Dev.	Std.Err.	Corrected Std.Err.
$T = 4$				
$\hat{\beta}$, 1*st* step	0.9800	0.1538	0.1471	
$\hat{\beta}$, 2*nd* step	0.9868	0.1423	0.1244	0.1391
$T = 8$				
$\hat{\beta}$, 1*st* step	0.9784	0.0832	0.0809	
$\hat{\beta}$, 2*nd* step	0.9810	0.0721	0.0477	0.0715

Comments

- Non-corrected two-step standard errors are severely downward biased, especially when $T = 8$
- Both one-step and corrected two-step standard errors are very close to standard deviations
- Additional simulations also show that Wald tests based on non-corrected two-step estimated variances are oversized
- This is no more the case when corrected two-step variances are used instead
- Finally, the size performance of the Sargan/Hansen test is not affected by the first-step estimator used to construct the weight matrix

"Too many" instruments?

- The FD-GMM approach generates moment conditions prolifically, with the instrument count quadratic in T

T	3	4	5	6	7	8	9	10	15
with y_{it-1} only									
$0.5(T-1)(T-2)$	1	3	6	10	15	21	28	36	91
with y_{it-1} & two predetermined x									
$0.5(T-1)(T-2)$ $+(T+1)(T-2)$	5	13	24	38	55	75	98	124	299

- Of course, the count is higher, when SYS-GMM is considered

"Too many" instruments?

- When a two-step GMM is considered, we will need to estimate the optimal weighting matrix (var-cov matrix of moment conditions): since the number of elements in the estimated weighting matrix is quadratic in the instrument count, it is *quartic* in T
- A finite sample may lack adequate information to estimate such a large matrix
- In applications it is common (and wise) practice to use only less distant lags as instruments
    - $t-2, t-3$ or $t-2, t-3, t-4$ for y
    - $t-1$ and $t-2$ for x ($t-2$, $t-3$ if the variables in x are endogenous)
    - See also, "collapsing" the instruments (Roodman, 2009)

Other approaches to estimation

- Quasi maximum likelihood approach of the first differenced model (Hsiao, Pesaran & Tahmiscioglu, 2002)
- Methods based on correcting the fixed effect estimator
  - ▷ LSDVC - bias-corrected least squares dummy variable approach, started with Kiviet (1995)
  - ▷ Indirect inference (Gouriéroux, C., Phillips, P.C.B. and Yug, J., 2010)

Selected references

- Ahn, S.C. and Schmidt, P. (1995), Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.
- Anderson, T.W. and C. Hsiao (1981), Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association*, 76, 589-606
- Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Montecarlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297
- Arellano, M., and Bover, O. (1995), Another Look at the Instrumental Variable Estimation of Error-Component 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, 115-143

Selected references

- Bun, M.J.C., and Windmeijer, F. (2010), The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models, *Econometrics Journal* **13**, 95-126
- Holtz-Eakin, D., Newey, W., and Rosen, H.S. (1988), Estimating Vector Autoregressions with Panel Data, *Econometrica* **56**(6), 1371-1395
- Judson, R. and A. Owen (1999), Estimating Dynamic Panel Data Models: a Guide for Macroeconomists, *Economic Letters*, 65, 9-15
- Nickell, S. (1981), Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1417-1426
- Roodman, D. (2009), A Note on the Theme of Too Many Instruments, *Oxford Bulletin of Economics and Statistics*, 71, 135-157
- Windmeijer, F. (2005), A Finite Sample Correction for the Variance of Linear Two-step GMM Estimators, *Journal of Econometrics*, 126, 26-51

Further readings

Bias corrected LSDV approach

- Bun, M. (2003), Bias Correction in the Dynamic Panel Data Model with a Nonscalar Disturbance Covariance Matrix, *Econometric Reviews*, 22, 29-58
- Bun, M. and M. Carree (2006), Bias Corrected Estimation in Dynamic Panel Data Models with Heteroskedasticity, *Economic Letters*, 92, 220-227
- Bun, M. and J.F Kiviet (2006), The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models, *Journal of Econometrics*, 127, 409-444
- Kiviet, J. F. (1995), On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models, *Journal of Econometrics*, 68, 53-78

Further readings

Other estimation procedures

- Gouriéroux, C., Phillips, P.C.B. and Yug, J. (2010), Indirect Inference for Dynamic Panel Models, *Journal of Econometrics* **157**, 68-77
- Hsiao, C., Pesaran, M.H. and Tahmiscioglu A.K.: 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics* **109**(1), 107–150.