Dynamic linear model

$$y_{it} = \alpha y_{i,t-1} + x_{it}\beta + \eta_i + v_{it} \qquad\qquad |\alpha| < 1$$

for $i = 1, ..., N$ and $t = 2, ..., T$.

NB. The first observation is y_{i1}, so that the first available equation is

$$y_{i2} = \alpha y_{i1} + x_{i2}\beta + \eta_i + v_{i2}$$

and we have $T-1$ equations in levels. Some authors assume y_{i0} is observed, and thus have T equations in levels.

Two important properties of the lagged dependent variable.

$E[y_{i,t-1}\eta_i] > 0$ since η_i is part of the process that generates $y_{i,t-1}$ according to our specification.

$E[y_{i,t-1}v_{i,t-1}] > 0$ for the same reason.

Thus $y_{i,t-1}$ is correlated with the individual effects, and is not strictly exogenous.

Most of the estimation issues are present in the simpler dynamic model

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \qquad\qquad |\alpha| < 1$$

that we focus on initially.

Useful to establish the properties of pooled OLS and Within Groups estimators in this setting.

Assuming $E[y_{i,t-1}v_{it}] = 0$, then $p\lim \widehat{\alpha}_{OLS} > \alpha$ as a result of the positive correlation between $y_{i,t-1}$ and η_i.

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \qquad\qquad |\alpha| < 1$$

Consider

$$\widetilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T-1}(y_{i1} + ... + y_{i,T-1})$$

and

$$\widetilde{v}_{it} = v_{it} - \frac{1}{T-1}(v_{i2} + ... + v_{iT})$$

Notice that all correlations of order $\frac{1}{T-1}$ are negative. E.g. $corr(y_{i,t-1}, \frac{-1}{T-1}v_{i,t-1})$ and $corr(v_{it}, \frac{-1}{T-1}y_{it})$.

This suggests that $E[\widetilde{y}_{i,t-1}\widetilde{v}_{it}] < 0$ and is of order $\frac{1}{T-1}$ (i.e. $E[\widetilde{y}_{i,t-1}\widetilde{v}_{it}] \rightarrow 0$ as $(T-1) \rightarrow \infty$).

These properties can be shown more formally (e.g. Nickell, *Econometrica* 1981).

Thus $p\lim_{N\to\infty} \widehat{\alpha}_{WG} < \alpha$ for fixed T. The Within estimator is inconsistent as the cross-section dimension of the panel (only) becomes large.

Also $p\lim_{T\to\infty} \widehat{\alpha}_{WG} = \alpha$. The Within estimator is consistent as the time dimension of the panel becomes large.

The inconsistency of the Within Groups estimator is of order $\frac{1}{T-1}$.

However calculations of this inconsistency, and Monte Carlo experiments, suggest the bias of the Within estimator remains non-negligible in cases with $T = 10$ or $T = 15$.

Also note that the inconsistency does not disappear as $\alpha \to 0$. So, unless T is large, the Within estimator does not provide reliable evidence on whether a lagged dependent variable should be included in the model.

However, in practice it is useful to know that OLS levels is likely to be biased upwards, and (in short panels) Within is likely to be biased downwards. Supposedly consistent estimators that give $\widehat{\alpha} >> \widehat{\alpha}_{OLS}$ or $\widehat{\alpha} << \widehat{\alpha}_{WG}$ should be viewed with suspicion.

Instrumental variables

A popular class of estimators that are consistent as $N \to \infty$ with T fixed first transform the model to eliminate the individual effects, and then apply instrumental variables.

The Within transformation is not useful in this context, since it introduces the shocks from all time periods into the transformed error term.

The first-differencing transformation is more promising.

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it}$$

for $i = 1, ..., N$ and $t = 3, ..., T$.

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it}$$

First-differenced OLS is not consistent (as $N \to \infty$ or $T \to \infty$, or both).

Since $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and $\Delta v_{it} = v_{it} - v_{i,t-1}$, we have

$$E[\Delta y_{i,t-1} \Delta v_{it}] < 0$$

However if we are willing to assume that $E[y_{i,t-1} v_{it}] = 0$, then $y_{i,t-2}$ or $\Delta y_{i,t-2}$ are valid instrumental variables for $\Delta y_{i,t-1}$ in the first-differenced equations.

Two-stage least squares (2SLS) estimators of this type were suggested by Anderson and Hsiao (*Journal of the American Statistical Association*, 1981). E.g.

$$\widehat{\alpha}_{AH} = (\Delta y'_{-1} Z (Z'Z)^{-1} Z' \Delta y_{-1})^{-1} \Delta y'_{-1} Z (Z'Z)^{-1} Z' \Delta y$$

where Δy is the stacked $N(T-2) \times 1$ vector of observations on Δy_{it}, Δy_{-1} is the stacked $N(T-2) \times 1$ vector of observations on $\Delta y_{i,t-1}$, and Z is the stacked $N(T-2) \times 1$ vector of observations on $y_{i,t-2}$.

One further time series observation is lost if $\Delta y_{i,t-2}$ rather than $y_{i,t-2}$ is used as the instrument.

The assumption that $y_{i,t-1}$ is predetermined follows naturally from assuming that the v_{it} are serially uncorrelated shocks, provided the initial conditions (y_{i1}) are also uncorrelated with subsequent v_{it} shocks.

Note that a minimum of $T = 3$ time series observations are required to identify α using this approach.

The Anderson-Hsiao 2SLS estimators are consistent as $N \to \infty$ for fixed T. But they are not efficient, except in the special case with $T = 3$.

With $T > 3$, further valid instruments become available for the first-differenced equations in the later time periods. Efficiency can be improved by exploiting these additional instruments.

The transformed error term Δv_{it} has a particular moving average form of serial correlation, under the maintained assumption that v_{it} is serially uncorrelated. More generally, Δv_{it} may be heteroskedastic. These features can be exploited to improve efficiency when $T > 3$ (i.e. when the model is overidentified).

Generalised method of moments (GMM)

Holtz-Eakin, Newey and Rosen (*Econometrica*, 1988) and Arellano and Bond (*Review of Economic Studies*, 1991) applied the generalised method of moments approach developed by Hansen (*Econometrica*, 1982) to exploit this additional information in the dynamic panel data problem.

We first briefly review some properties of GMM estimators.

GMM formulates a set of orthogonality restrictions (moment conditions) related to an econometric model, and finds parameter estimates that come as close as possible to achieving these orthogonality properties in the sample.

Example

$$y_i - f(x_i, \beta) = u_i \qquad i = 1, ..., N \qquad \beta \text{ is } q \times 1$$

$$g(x_i) = z_i \quad s.t. \;\; E(z_i' u_i) = 0 \qquad z_i \text{ is } 1 \times r$$

The model specifies r (population) orthogonality restrictions: $E(z_i' u_i) = 0$

Sample analogue:

$$b_N(\beta) = \frac{1}{N} \sum_{i=1}^{N} z_i' u_i(\beta) = \frac{1}{N} \sum_{i=1}^{N} z_i' [y_i - f(x_i, \beta)]$$

GMM estimators choose $\widehat{\beta}_{GMM}$ to minimise the distance of $b_N(\beta)$ from zero.

$r = q$ (just identified): $\widehat{\beta}_{GMM}$ is the unique solution to $b_N(\beta) = 0$.

$r > q$ (overidentified): $\widehat{\beta}_{GMM}$ minimises a weighted quadratic distance, i.e.

$$\widehat{\beta}_{GMM} = \arg\min_{\beta} \ J_N(\beta) = b_N(\beta)' W_N b_N(\beta)$$

for some weight matrix W_N.

Note that this gives a family of GMM estimators, based on the same moment conditions $E(z_i' u_i) = 0$, for different choices of the weight matrix W_N.

Properties

Under conditions given in Hansen (1982)

i) Strong consistency: $\widehat{\beta}_{GMM} \xrightarrow{a.s.} \beta$ as $N \to \infty$.

ii) Asymptotic normality: $\sqrt{N}(\widehat{\beta}_{GMM} - \beta) \xrightarrow{D} N(0, avar(\widehat{\beta}_{GMM}))$ where

$$avar(\widehat{\beta}_{GMM}) = (D'WD)^{-1}D'WS_W WD(D'WD)^{-1}$$

and

$$\begin{aligned} D &= p\lim_{N\to\infty}\left(\frac{\partial b_N(\beta)}{\partial \beta'}\right) \\ W &= p\lim_{N\to\infty} W_N \\ S_W &= \frac{1}{N}\sum_{i=1}^{N} E(z_i' u_i u_i' z_i) \end{aligned}$$

S_W is the average covariance matrix for the moment conditions.

The optimal GMM estimator, which minimises $avar(\widehat{\beta}_{GMM})$ for a given set of moment conditions, chooses W_N s.t. $W = S_W^{-1}$, giving

$$avar(\widehat{\beta}_{GMM}) = (D'WD)^{-1}$$

Typically implemented in two steps, choosing $W_N = \widehat{S}_W^{-1}$, where $\widehat{S}_W$ is a consistent estimate of S_W, obtained from some consistent initial estimate of β (e.g. a 'one step' GMM estimator, using some known weight matrix W_N).

Note:

OLS is a (just-identified) GMM estimator for $f(x_i, \beta) = x_i\beta$ linear and $z_i = x_i$.

2SLS is a (possibly overidentified) GMM estimator for $f(x_i, \beta) = x_i\beta$ and $W_N = (z'z)^{-1}$. This weight matrix is optimal if $u_i \sim iid(0, \sigma^2)$, but not more generally.

The AR(1) panel data model

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \qquad\qquad |\alpha| < 1$$

for $i = 1, ..., N$ and $t = 2, ..., T$.

Assumption (error components)

$$E(\eta_i) = E(v_{it}) = E(\eta_i v_{it}) = 0$$

Assumption (serially uncorrelated shocks)

$$E(v_{is} v_{it}) = 0 \quad \text{for } s \neq t$$

Assumption (predetermined initial conditions)

$$E(y_{i1} v_{it}) = 0 \quad \text{for } t = 2, ..., T$$

These assumptions specify a finite number of linear moment conditions, which can be exploited using a linear GMM estimator.

$$
\begin{array}{rcll}
 & & \text{First-differenced equations} & \text{Valid instruments} \\
(y_{i3} - y_{i2}) & = & \alpha(y_{i2} - y_{i1}) + (v_{i3} - v_{i2}) & y_{i1} \\
(y_{i4} - y_{i3}) & = & \alpha(y_{i3} - y_{i2}) + (v_{i4} - v_{i3}) & y_{i1}, y_{i2} \\
 & & \vdots & \\
(y_{iT} - y_{i,T-1}) & = & \alpha(y_{i,T-1} - y_{i,T-2}) + (v_{iT} - v_{i,T-1}) & y_{i1}, y_{i2}, ..., y_{i,T-2}
\end{array}
$$

Clearly $E(y_{i1}\Delta v_{i3}) = 0$ follows from assuming predetermined initial conditions.

$E(y_{i1}\Delta v_{i4}) = 0$ follows similarly.

$$y_{i2} = \alpha y_{i1} + \eta_i + v_{i2}$$

$E(\eta_i \Delta v_{i4}) = 0$ follows from the error components specification.

$E(v_{i2}\Delta v_{i4}) = 0$ follows from serially uncorrelated shocks.

Hence we obtain $E(y_{i2}\Delta v_{i4}) = 0$.

Similar arguments establish the $m = (T-2)(T-1)/2$ moment conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0 \quad \text{for } t = 3, ..., T \text{ and } s \geq 2$$

Giving the set of valid instruments proposed in the previous table.

These can also be written as $E(Z_i'\Delta v_i) = 0$ where

$$Z_i = \begin{pmatrix} y_{i1} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & y_{i1} & y_{i2} & \dots & y_{i,T-2} \end{pmatrix} \text{ and } \Delta v_i = \begin{pmatrix} \Delta v_{i3} \\ \Delta v_{i4} \\ \vdots \\ \Delta v_{iT} \end{pmatrix}$$

$$(T-2) \times m \qquad\qquad (T-2) \times 1$$

Sample analogue

$$b_N(\alpha) = \frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta v_i(\alpha)$$

For $T = 3$, we have 1 moment condition $E(y_{i1}\Delta v_{i3}) = 0$ and 1 parameter. α is just identified, the weight matrix is irrelevant, and the optimal GMM estimator coincides with the Anderson-Hsiao 2SLS estimator.

For $T > 3$, we have $m > 1$ moment conditions. α is overidentified.

GMM estimators minimise a weighted quadratic distance

$$\begin{aligned}
\widehat{\alpha}_{GMM} &= \arg\min_{\alpha} J_N(\alpha) = b_N(\alpha)' W_N b_N(\alpha) \\
&= \arg\min_{\alpha} \left(\frac{1}{N} \sum_{i=1}^{N} \Delta v_i' Z_i \right) W_N \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta v_i \right) \\
&= (\Delta y_{-1}' Z W_N Z' \Delta y_{-1})^{-1} \Delta y_{-1}' Z W_N Z' \Delta y
\end{aligned}$$

where Δy and Δy_{-1} are the stacked $N(T-2) \times 1$ vectors of observations on Δy_{it} and $\Delta y_{i,t-1}$ as before, and $Z = (Z_1, ..., Z_N)'$ is the stacked $N(T-2) \times m$ matrix of observations on the instruments.

This simple generalised instrumental variables expression for the GMM estimators follows from the linearity of the moment conditions in α

$$\begin{aligned} E(y_{i,t-s}\Delta v_{it}) &= E(y_{i,t-s}(\Delta y_{it} - \alpha \Delta y_{i,t-1})) \\ &= E(y_{i,t-s}\Delta y_{it}) - \alpha E(y_{i,t-s}\Delta y_{i,t-1}) = 0 \end{aligned}$$

Compare

$$\widehat{\alpha}_{GMM} = (\Delta y'_{-1} Z W_N Z' \Delta y_{-1})^{-1} \Delta y'_{-1} Z W_N Z' \Delta y$$

and

$$\widehat{\alpha}_{AH} = (\Delta y'_{-1} Z (Z'Z)^{-1} Z' \Delta y_{-1})^{-1} \Delta y'_{-1} Z (Z'Z)^{-1} Z' \Delta y$$

For $T > 3$, two sources of greater efficiency:

$\widehat{\alpha}_{GMM}$ exploits more moment conditions ($m > 1$). The Anderson-Hsiao instrument matrix is a linear combination of the GMM instrument matrix.

2SLS weight matrix ($W_N = (Z'Z)^{-1}$) is not optimal for the first-differenced specification.

General results for GMM estimators indicate that $\widehat{\alpha}_{GMM}$ is strongly consistent (as $N \to \infty$ for fixed T) and asymptotically normal.

For an arbitrary W_N

$$avar(\widehat{\alpha}_{GMM}) = N(\Delta y'_{-1} Z W_N Z' \Delta y_{-1})^{-1} \Delta y'_{-1} Z W_N \widehat{V}_N W_N Z' \Delta y_{-1} (\Delta y'_{-1} Z W_N Z' \Delta y_{-1})^{-1}$$

where

$$\widehat{V}_N = \frac{1}{N} \sum_{i=1}^{N} \left(Z_i \widehat{\Delta v}_i \widehat{\Delta v}'_i Z_i \right)$$

and $\widehat{\Delta v}_{it} = \Delta y_{it} - \widehat{\alpha}\ \Delta y_{i,t-1}$ are consistent estimates of the first-differenced residuals, based on some consistent estimator $\widehat{\alpha}$.

The optimal (two step) GMM estimator thus sets $W_N = \widehat{V}_N^{-1}$, giving

$$avar(\widehat{\alpha}_{GMM}) = N(\Delta y'_{-1} Z W_N Z' \Delta y_{-1})^{-1}$$

One step weight matrix

For the special case in which $v_{it} \sim iid(0, \sigma_v^2)$, we can obtain a one step GMM estimator that is asymptotically equivalent.

NB. homoskedasticity is not required to derive the moment conditions we are using.

Role here is merely to suggest a good choice for the one step weight matrix.

For the first-differenced equations, this choice is not 2SLS, due to the serial correlation in Δv_{it} introduced by the transformation.

$$\Delta v_{it} = v_{it} - v_{i,t-1} \qquad \Delta v_{i,t-1} = v_{i,t-1} - v_{i,t-2}$$

$$E(\Delta v_{it}^2) = 2\sigma_v^2$$

$$E(\Delta v_{it}\Delta v_{i,t-1}) = -\sigma_v^2$$

$$E(\Delta v_{it}\Delta v_{i,t-s}) = 0 \quad \text{for } s \geq 2$$

$$E(\Delta v_i \Delta v_i') = \sigma_v^2 \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0\dots & \dots & 2 \end{pmatrix} = \sigma_v^2 H$$

In the *iid* case, the weight matrix

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' H Z_i \right)^{-1}$$

gives a one step GMM estimator for the first-differenced equations that is asymptotically equivalent to the (generally) optimal two step GMM estimator.

While asymptotic results for the two step estimator only require an initial estimator that is consistent, small sample properties tend to be better when the estimate of the optimal weight matrix

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} \left(Z_i \widehat{\Delta v_i} \widehat{\Delta v}_i' Z_i \right) \right)^{-1}$$

uses residuals $\widehat{\Delta v_i}$ based on an initial estimator that is as efficient as possible.

Two step inference

Making explicit the dependence of the estimated optimal weight matrix on the initial consistent estimator

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

indicates a small sample problem with the usual asymptotic variance estimator

$$avar(\widehat{\alpha}_{GMM}) = N(\Delta y_{-1}' Z W_N(\widehat{\alpha}) Z' \Delta y_{-1})^{-1}$$

This neglects variation introduced by using an estimate $\widehat{\alpha}$ to construct the optimal weight matrix.

In very large samples, this variation is negligible, and the usual expression for the asymptotic variance is correct.

But in (reasonably large) finite samples, this additional variation makes inference based on $avar(\widehat{\alpha}_{GMM})$ unreliable.

In fact, $avar(\widehat{\alpha}_{GMM})$ provides a good estimate of the variance of an infeasible GMM estimator, that uses the true value α rather than the initial estimate $\widehat{\alpha}$ to construct the optimal weight matrix.

Windmeijer (*Journal of Econometrics*, 2005) proposes a finite sample correction that provides more accurate estimates of the variance of two step GMM estimators.

This finite sample correction is now implemented in programs such as Stata and PC-Give/Ox.

t-tests based on these corrected standard errors are found to be as reliable as those based on the one step GMM estimator (where no parameters are estimated in the construction of the weight matrix).

(As usual with t and Wald tests) these are quite accurate in cases where the estimated parameter(s) are not subject to serious finite sample bias problems (to be discussed). Bond and Windmeijer (*Econometric Reviews*, 2005) provide Monte Carlo evidence, and consider alternative tests of linear restrictions in the GMM context.

Quadratic moment conditions

The ‘standard assumptions’ stated earlier imply a further $T-3$ moment conditions that are quadratic in α.

These can be written as

$$E(\Delta v_{i,t-1}u_{it}) = 0 \ \text{ for } t = 4, ..., T$$

where $u_{it} = \eta_i + v_{it}$ is the error term for the untransformed equations in levels.

Exploiting these additional moment conditions requires numerical optimisation procedures, and is less common in practice.

The resulting optimal non-linear GMM estimator is efficient for this model, whereas the linear ‘Arellano-Bond’ GMM estimator is not.

See Ahn and Schmidt (*Journal of Econometrics*, 1995) for further discussion.

Homoskedasticity over time

Under the additional homoskedasticity assumption

$$E(v_{it}^2) = \sigma_i^2$$

there are a further $T - 3$ linear moment conditions

$$E(y_{i,t-2}\Delta v_{i,t-1} - y_{i,t-1}\Delta v_{it}) = 0 \quad \text{for } t = 4, ..., T$$

suggested by Ahn and Schmidt (1995).

These are simple to implement. They will improve efficiency if the additional homoskedasticity assumption is valid, but may introduce inconsistency if not.

Alternative transformations

We introduced these GMM estimators using the first-differencing transformation to eliminate the time-invariant individual effects from $u_{it} = \eta_i + v_{it}$.

The two key properties of first-differencing are:

* eliminates time-invariant variables

* does not introduce lagged shocks earlier than $v_{i,t-1}$ into the transformed error term

Any transformation that shares these properties would do equally well.

Arellano and Bover (*Journal of Econometrics*, 1995) show that the optimal GMM estimator is invariant (numerically!) to the particular transformation used within this class.

An alternative transformation of some interest is the forward Helmert's or 'orthogonal deviations' transformation

$$v_{it}^{O} = \left(\frac{T-t+1}{T-t+2}\right)^{\frac{1}{2}} \left[v_{i,t-1} - \frac{1}{T-t+1}(v_{it} + v_{i,t+1} + \ldots + v_{iT}) \right]$$

This estimates the mean for individual i using future observations on the series only, and takes (weighted) deviations from this mean.

One property is that if the v_{it} are iid, then so are the v_{it}^{O}.

Hence the asymptotically efficient one step estimator for the *iid* special case is simply 2SLS.

Transformed model

$$y_{it}^{O} = \alpha y_{i,t-1}^{O} + v_{it}^{O} \quad \text{for } t = 3, ..., T$$

OLS on this transformed model coincides with Within Groups.

2SLS using all available linear moment conditions coincides with the one step first-differenced GMM estimator discussed above.

These results are useful for thinking about finite sample bias issues.

Overfitting

One source of finite sample bias is the use of ‘too many’ instruments relative to the sample size (N).

If we use all the available lagged instruments, the number of instruments grows rapidly with the time dimension of the panel.

Hence if T is moderately large relative to N, there is a danger of ‘overfitting’.

For 2SLS estimators, overfitting results in a finite sample bias in the direction of the OLS estimator.

By analogy with 2SLS on the orthogonal deviations model, this gives a (downward) finite sample bias, in the direction of the Within estimator.

We can investigate this by comparing GMM and Within estimators. If they are too close, we can reduce the number of lagged instruments used, to reduce this source of finite sample bias.

More generally, it is good practice to check the sensitivity of empirical results to using more or fewer lagged observations in the set of instruments for each equation.

This also indicates the properties of the first-differenced GMM estimator in the case where $T \to \infty$.

Overfitting leads first-differenced GMM to converge on Within Groups.

But recall that the Within estimator is consistent as $T \to \infty$.

So first-differenced GMM is also consistent (although not very useful) in the case where $T \to \infty$.

This is shown formally by Alvarez and Arellano (*Econometrica*, 2003).

The result is also much less useful in the context of models with endogenous explanatory variables, where the Within estimator is not consistent as $T \to \infty$, and hence not such a benign estimator to be converging on.

Weak instruments

Instrumental variables (and GMM) estimators have poor small sample properties in cases where the instruments, although valid, are only weakly correlated with the endogenous explanatory variables.

This is relevant for the first-differenced GMM estimator in the AR(1) model in the case where $\alpha \rightarrow 1$.

By analogy with random walks (innovations uncorrelated with past levels), the correlation between $\Delta y_{i,t-1}$ and the lagged levels $y_{i,t-s}$ for $s \geq 2$ becomes weaker as $\alpha \rightarrow 1$.

In the model we have focused on

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}$$

α remains formally identified as $\alpha \rightarrow 1$, and the first-differenced GMM estimator remains consistent as $N \rightarrow \infty$, provided $\sigma_\eta^2 = var(\eta_i) \neq 0$.

But Monte Carlo evidence suggests that first-differenced GMM estimators become very imprecise, and subject to serious finite sample biases, for values of α around 0.8 and above, unless the available samples are huge.

The finite sample bias is again found to be downward, in the direction of the Within estimator, consistent with findings for 2SLS estimators in simple cases where the weak instruments problem has been studied analytically.

Blundell and Bond (*Journal of Econometrics*, 1998) provide Monte Carlo evidence, and develop an extended GMM estimator that is more useful for estimating panel data models using very persistent series.

We will return to this after briefly considering how the basic GMM estimator studied so far can be adapted for models with additional explanatory variables.

Note also that if we combine

$$\begin{aligned} y_{it} &= \eta_i + \varepsilon_{it} \\ \varepsilon_{it} &= \alpha \varepsilon_{i,t-1} + v_{it} \end{aligned}$$

we obtain the alternative specification

$$y_{it} = \alpha y_{i,t-1} + (1-\alpha)\eta_i + v_{it}$$

In this specification, the process for y_{it} approaches a pure random walk as $\alpha \to 1$ (rather than a random walk with individual-specific drifts). Consequently lagged levels are completely uninformative instruments for $\Delta y_{i,t-1}$ in the case where $\alpha = 1$, and α is not identified using the moment conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0 \quad \text{for } t = 3, ..., T \text{ and } s \geq 2$$

Although for this model we can note that the OLS levels estimator is consistent when $\alpha = 1$.