Lecture 3: Dynamic linear model with covariates

$$
\begin{aligned}
y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \eta_i + v_{it} \qquad\qquad |\alpha| < 1 \\
&= X_{it}\delta + u_{it}
\end{aligned}
$$

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

$$
X_{it} = (y_{i,t-1}, x_{it}) \qquad\qquad \delta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}
$$

We maintain the previous assumptions: error components, serially uncorrelated shocks, predetermined initial conditions.

Hence the moment conditions

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

remain valid for this more general model.

Different assumptions about the properties of x_{it} will imply different sets of moment conditions.

x_{it} may be correlated or uncorrelated with η_i.

x_{it} may be endogenous, predetermined or strictly exogenous wrt v_{it}.

More restrictive assumptions will imply the validity of additional moment conditions. This will increase efficiency if the stronger assumptions are valid, but will imply inconsistency if they are not.

Fortunately these are typically overidentifying assumptions, which can be tested.

Some Examples of assumptions on x_{it}

(1) Correlated individual effects, weak exogeneity

$E(x_{it}\eta_i) \neq 0.$ $\qquad E(x_{is}v_{it}) = 0 \quad \text{for } s \leq t.$

x_{it} is correlated with the individual effects, and predetermined wrt the (serially uncorrelated) shocks. There may be feedback from current shocks to future values of x_{is}.

Both $y_{i,t-1}$ and x_{it} are correlated with η_i.

We again transform the model to eliminate η_i, for example by first-differencing.

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

The assumption that x_{it} is predetermined implies the moment conditions

$$E(x_{i,t-s}\Delta v_{it}) = 0 \quad \text{for } t = 3, ..., T \text{ and } s \geq 1$$

i.e. $x_{i,t-1}$ (as well as $x_{i,t-2}$ and longer lags) is uncorrelated with $\Delta v_{it} = v_{it} - v_{i,t-1}$.

The complete set of (linear) moment conditions can be written as $E(Z_i' \Delta v_i) = 0$, where now

$$Z_i = \begin{pmatrix} y_{i1} & x_{i1} & x_{i2} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1} & x_{i2} & x_{i3} & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i,T-2} & x_{i1} & \dots & x_{i,T-1} \end{pmatrix}$$

GMM estimation then proceeds as before.

Letting ΔX denote the stacked $N(T-2) \times 2$ matrix of observations on $X_{it} = (y_{i,t-1}, x_{it})$, we obtain

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

Alternative choices of W_N produce one step and two step GMM estimators, as for the AR(1) model.

If instead x_{it} is endogenous wrt the (serially uncorrelated) v_{it} shocks $(E(x_{is}v_{it}) = 0 \quad \text{for } s < t)$, then only the subset

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

remain valid. In this case the treatment of x_{is} and y_{is} in the instrument matrix is symmetric.

If instead x_{it} is strictly exogenous wrt the v_{it} shocks ($E(x_{is}v_{it}) = 0$ for all s, t), then the larger set of moment conditions

$$E(x_{is}\Delta v_{it}) = 0 \quad \text{for } t = 3, ..., T \text{ and } s = 1, 2, ..., T$$

would be valid.

Implementation of these alternatives simply deletes columns or adds columns to the instrument matrix Z_i defined above.

(2) Uncorrelated individual effects, strict exogeneity,

$$E(x_{it}\eta_i) = 0. \qquad E(x_{is}v_{it}) = 0 \quad \text{for all } s, t.$$

x_{it} is uncorrelated with the individual effects, and strictly exogenous wrt the shocks.

This introduces some new issues, since now we have valid moment conditions for one or more of the equations in levels.

In particular, relative to the case in which x_{it} is strictly exogenous but correlated with η_i, and we use all the moment conditions that this implies for the equations in first-differences, as discussed above, this gives a further T non-redundant linear moment conditions.

Essentially we have T observations on x_{it}, and we assume that all of these are uncorrelated with η_i, so we obtain T additional moment conditions.

There are many equivalent ways to write these. One possibility that is quite elegant (but less useful in practice if we have to deal with unbalanced panel data) is

$$E(x_{is}u_{iT}) = 0 \quad \text{for } s = 1, 2, ..., T$$

where $u_{iT} = \eta_i + v_{iT}$ is the error term for the untransformed equation in the final time period.

The complete set of (linear) moment conditions can then be written as $E(Z_i^{+\prime} u_i^+) = 0$ where

$$Z_i^+ = \begin{pmatrix} y_{i1} & x_{i1} & \dots & x_{iT} & 0 & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & y_{i1} & y_{i2} & x_{i1} & \dots & x_{iT} & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & y_{i1} & \dots & y_{i,T-2} & x_{i1} & \dots & x_{iT} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & x_{i1} & \dots & x_{iT} \end{pmatrix}$$

or

$$Z_i^+ = \begin{pmatrix} Z_i^D & 0 & \dots & 0 \\ 0 & x_{i1} & \dots & x_{iT} \end{pmatrix}$$

where Z_i^D is the corresponding instrument set (i.e. assuming strict exogeneity) for the first-differenced equations, and

$$u_i^+ = \begin{pmatrix} \Delta v_{i3} \\ \vdots \\ \Delta v_{iT} \\ \eta_i + v_{iT} \end{pmatrix}$$

In effect, we augment the system of first-differenced equations by adding the levels equation for the final period, and augment the instrument matrix by adding the T valid instruments for this equation.

[With unbalanced panel data, where not all time periods are observed for all individuals, the observation for the final period may be missing for some individuals. An equivalent but less elegant alternative augments the system of first-differenced equations by adding all the available levels equations.]

As before, GMM estimators are based on the sample analogue of these moment conditions, which here gives

$$\widehat{\delta}_{GMM} = (X^{+\prime} Z^{+} W_N Z^{+\prime} X^{+})^{-1} X^{+\prime} Z^{+} W_N Z^{+\prime} y^{+}$$

where y_i^+ and X_i^+ are defined analogously to u_i^+, and y, X and Z are stacked across the N individuals as before.

One further difference compared to the basic first-differenced GMM estimators is that there is no compelling choice for the one step weight matrix when we add one (or more) levels equations to the system. Assuming that $v_{it} \sim iid(0, \sigma_v^2)$ does not give a form for $E(u_i^+ u_i^{+\prime})$ that is proportional to a known matrix (since the variance of u_{iT} depends on σ_η^2 as well as σ_v^2).

For slightly obscure reasons, the GAUSS and OX implementations now use

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

where

$$H^+ = \begin{pmatrix} H & 0 \\ 0 & I \end{pmatrix}$$

and H is the $(T-2) \times (T-2)$ matrix with 2 on the main diagonal, -1 on the first off-diagonal, and 0 elsewhere.

The STATA implementation allows this and other choices.

A corollary is that we may expect the optimal two step GMM estimator to provide a more substantial improvement in efficiency here, compared to the one step GMM estimator considered. In contrast to the situation for the basic first-differenced estimators, there is no reason to suppose that the one step weight matrix we use for this augmented estimator will be close to the optimal weight matrix.

The optimal (two step) GMM estimator is obtained using $W_N = \widehat{V}_N^{-1}$, where

$$\widehat{V}_N = \frac{1}{N}\sum_{i=1}^{N} Z_i^{+\prime}\widehat{u}_i^{+}\widehat{u}_i^{+\prime}Z_i$$

and

$$\widehat{u}_i^{+} = \widehat{y}_i^{+} - X_i^{+}\widehat{\delta}$$

for some consistent initial estimator $\widehat{\delta}$.

If instead x_{it} is only predetermined wrt v_{it} (but uncorrelated with η_i), we still have an additional T moment conditions for the equations in levels, which can again be written as

$$E(x_{is}u_{iT}) = 0 \quad \text{for } s = 1, 2, ..., T$$

These are used to obtain $E(Z_i^{+\prime}u_i^+) = 0$ where Z_i^+ again has the form

$$Z_i^+ = \begin{pmatrix} Z_i^D & 0 & \dots & 0 \\ \\ 0 & x_{i1} & \dots & x_{iT} \end{pmatrix}$$

but now Z_i^D contains the smaller set of instruments that are valid for the first-differenced equations when x_{it} is only predetermined.

(3) Uncorrelated individual effects, endogenity exogeneity

If instead x_{it} is endogenous wrt v_{it} (but uncorrelated with η_i), we only have an additional $T - 1$ useful moment conditions for the equations in levels. These can be written as

$$E(x_{is}u_{iT}) = 0 \quad \text{for } s = 1, 2, ..., T - 1$$

In this case $E(x_{iT}v_{iT}) \neq 0$ prevents us from exploiting the orthogonality between x_{iT} and η_i.

Now that we are familiar with exploiting moment conditions that relate to the equations in levels, we can notice an intermediate possibility between the assumptions that x_{it} is correlated or uncorrelated with η_i.

This arises when x_{it} is correlated with η_i, but some known function of x_{it} is uncorrelated with η_i, and thus can be used to form valid instruments for the equations in levels.

A leading example is where the covariance between x_{it} and η_i is constant over time, so that the first-differences of x_{it} are uncorrelated with η_i.

$$E(x_{it}\eta_i) = \omega_i \neq 0 \quad \text{for } t = 1, ..., T$$

$$\Rightarrow E[(x_{it} - x_{i,t-1})\eta_i] = \omega_i - \omega_i = 0$$

In this case - suggested by Arellano and Bover (*Journal of Econometrics*, 1995) - suitably dated first-differences Δx_{is} can be used as instruments for the levels equations.

'Suitably dated' depends on the correlation between Δx_{is} and v_{it}, which in turn depends on whether x_{it} is itself assumed to be endogenous, predetermined or strictly exogenous wrt v_{it}.

Details will not be given here, but a similar case will be considered shortly.

When we have variables (e.g. x_{it} itself, or Δx_{it}) that are uncorrelated with the individual effects, so that the levels equations can be used in estimation, it becomes possible to identify coefficients on time-invariant explanatory variables, i.e. γ in the more general model

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

where the w_i are observed explanatory variables that do not vary over time.

If all valid moment conditions require the model to be transformed to eliminate the unobserved η_i, the transformation also eliminates the observed w_i and we do not identify γ.

Coefficients on time-invariant explanatory variables can always be estimated under the assumption that $E[w_i(\eta_i + v_{it})] = 0$, in which case the levels w_i can themselves be used as instruments for the equations in levels.

Whether these estimates are consistent depends on whether the assumption $E[w_i(\eta_i + v_{it})] = 0$ is valid.

The possibility of using (transformations of) time-varying explanatory variables as instruments for the equations in levels allows us to estimate γ consistently without invoking this assumption about w_i itself.

This extends an idea introduced into the static panel data literature by Hausman and Taylor (*Econometrica*, 1981). Arellano and Bover (1995) provide further details.

The observation that Δx_{it} may be uncorrelated with η_i even though x_{it} itself is correlated with η_i raises the possibility that Δy_{it} may also be uncorrelated with η_i.

This turns out to require a further restriction on the process generating the initial conditions (y_{i1}).

To consider this further, and to explore when this initial conditions restriction is particularly informative, we return to the simpler AR(1) specification considered previously.

'System GMM'

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

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

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

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

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

Recall that these assumptions imply $m = (T-2)(T-1)/2$ linear moment conditions

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

and a further $T-3$ quadratic moment conditions, which we now write as

$$E(\Delta u_{is} u_{iT}) = 0 \quad \text{for } s = 3, 4, ..., (T-1)$$

Additional (initial conditions) assumption

$$E(\Delta y_{i2}\eta_i) = 0$$

This has two implications.

i) Now we have $E(\Delta y_{is}\eta_i) = 0$ for $s = 2, ..., T$, since the AR(1) specification implies

$$\Delta y_{it} = \alpha^{t-2}\Delta y_{i2} + \sum_{s=0}^{t-3} \alpha^s \Delta v_{i,t-s} \quad \text{for } t = 3, ..., T$$

This then implies an additional $T-2$ non-redundant linear moment conditions for the equations in levels, which can be written (for example) as

$$E(\Delta y_{is}u_{iT}) = 0 \quad \text{for } s = 2, 3, ..., (T-1)$$

ii) Given these additional linear moment restrictions, the quadratic moment restrictions are now redundant.

For example

$$\begin{aligned}\Delta u_{i,t-1}u_{iT} &= (\Delta y_{i,t-1} - \alpha\Delta y_{i,t-2})u_{iT} \\ &= \Delta y_{i,t-1}u_{iT} - \alpha\Delta y_{i,t-2}u_{iT}\end{aligned}$$

So $E(\Delta y_{i,t-1}u_{iT}) = 0$ and $E(\Delta y_{i,t-2}u_{iT}) = 0$ jointly imply $E(\Delta u_{i,t-1}u_{iT}) = 0$.

Conveniently, the complete set of moment conditions implied by our standard assumptions and the initial conditions restriction $E(\Delta y_{i2}\eta_i) = 0$ can be written as

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

and

$$E(\Delta y_{is}u_{iT}) = 0 \ \ \text{for } s = 2, 3, ..., (T-1)$$

and can thus be exploited using a linear GMM estimator.

However this is not just a matter of convenience. When this additional initial conditions assumption is valid, exploiting the additional moment conditions for the equations in levels can in some cases provide a dramatic improvement in efficiency, and reduction in finite sample bias, compared to the basic first-differenced GMM estimator. This is particularly important as $\alpha \rightarrow 1$, or as the y_{it} series becomes more persistent.

In this case the correlation between $\Delta y_{i,t-1}$ and lagged levels $y_{i,t-s}$ for $s \geq 2$ becomes weaker, and the first-differenced GMM has poor finite sample properties associated with weak instruments - imprecise parameter estimates, and serious finite sample bias.

In this context, exploiting the quadratic moment conditions could make a substantial improvement (Ahn and Schmidt, 1995).

It turns out that exploiting the additional linear moment conditions provides much more dramatic gains, provided that the additional initial conditions restriction is valid (Blundell and Bond, 1998).

Monte Carlo evidence (survey paper, Table 2).

Why is this a restriction on the initial conditions?

The AR(1) specification determines y_{i2} given y_{i1}, so to guarantee that Δy_{i2} is uncorrelated with η_i we require a restriction on the behaviour of y_{i1}.

What kind of restriction?

This is a form of stationarity restriction on the y_{it} series.

The representation

$$\Delta y_{it} = \alpha^{t-2}\Delta y_{i2} + \sum_{s=0}^{t-3} \alpha^s \Delta v_{i,t-s} \quad \text{for } t = 3, ..., T$$

suggests (using backward recursion for earlier periods) that if the same model has generated the y_{it} series for long enough prior to our sample period, the observations on Δy_{it} would indeed be uncorrelated with η_i.

$$
\begin{aligned}
\Delta y_{it} &= \alpha \Delta y_{t-1} + \Delta v_{it} \\
&= \alpha(\alpha \Delta y_{t-2} + \Delta v_{it-1}) + \Delta v_{it} \\
&= \alpha^2 \Delta y_{t-2} + \alpha \Delta v_{it-1} + \Delta v_{it}, etc.
\end{aligned}
$$

‘Long enough’ means long enough for any influence of the true start-up of the process to have become negligibly small (which in turn depends on the true value of α). Example: this may be violated for new entrants and young firms.

More formally, write

$$\begin{aligned} y_{i2} &= \alpha y_{i1} + \eta_i + v_{i2} \\ (y_{i2} - y_{i1}) &= (\alpha - 1) y_{i1} + \eta_i + v_{i2} \end{aligned}$$

Define (wlog)

$$e_{i1} = y_{i1} - \left(\frac{\eta_i}{1-\alpha}\right)$$

$$y_{i1} = \left(\frac{\eta_i}{1-\alpha}\right) + e_{i1}$$

Then

$$\begin{aligned}\Delta y_{i2} &= (\alpha - 1)\left(\frac{\eta_i}{1-\alpha}\right) + (\alpha - 1)e_{i1} + \eta_i + v_{i2} \\ &= (\alpha - 1)e_{i1} + v_{i2}\end{aligned}$$

The standard error components assumption implies $E(v_{i2}\eta_i) = 0$.

A sufficient conditon for $E(\Delta y_{i2}\eta_i) = 0$ is thus the restriction

$$E(e_{i1}\eta_{i1}) = 0$$

Interpretation

$\left(\frac{\eta_i}{1-\alpha}\right)$ is the level that our model specifies the y_{it} series will converge towards for individual i , if the process continues for long enough.

$e_{i1} = y_{i1} - \left(\frac{\eta_i}{1-\alpha}\right)$ is the deviation from this convergent level at the start of our sample period.

We require that these initial deviations are uncorrelated with η_i, or equivalently are uncorrelated with the convergent level itself.

The initial observations y_{i1} can deviate randomly, but not systematically, from these convergent levels.

This imposes a stationarity restriction on the mean of the y_{it} series, but does not impose any restriction on the variance. Sometimes known as 'mean stationarity'.

Whether this is a mild or strong restriction will depend on the context, and particularly on the nature of the initial observations in our sample.

As noted earlier, the restriction will hold automatically if the same process has generated the series for long enough in the past.

Thus if we believe the AR(1) specification, and there is nothing special about our first observation period, it is reasonable to expect this restriction to hold.

But if our first observation corresponds to the true start-up of the process, it may be an unreasonable restriction.

Computation of the extended (or 'system') GMM estimator is similar to the case where levels (or first-differences) of some x_{it} variable can be used to obtain instruments for the equations in levels.

We add one (or more) equations in levels to the set of first-differenced equations, for example

$$y_i^+ = \alpha y_{i(-1)}^+ + u_i^+$$

$$\begin{pmatrix} \Delta y_{i3} \\ \vdots \\ \Delta y_{iT} \\ y_{iT} \end{pmatrix} = \alpha \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{i,T-1} \\ y_{i,T-1} \end{pmatrix} + \begin{pmatrix} \Delta v_{i3} \\ \vdots \\ \Delta v_{iT} \\ \eta_i + v_{iT} \end{pmatrix}$$

and write the complete set of moment conditions as $E(Z_i^{+\prime}u_i^{+}) = 0$, where

$$Z_i^{+} = \begin{pmatrix} Z_i^D & 0 & \dots & 0 \\ & & & \\ 0 & \Delta y_{i2} & \dots & \Delta y_{i,T-1} \end{pmatrix}$$

Defining $b_N(\alpha) = \frac{1}{N}\sum_{i=1}^{N} Z_i^{+\prime}u_i^{+}(\alpha)$ and choosing α to minimise $J_N(\alpha) = b_N(\alpha)' W_N b_N(\alpha)$ gives

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

Models with serially correlated errors

MA errors

Example

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

$$v_{it} = \varepsilon_{it} + \phi \varepsilon_{i,t-1}$$

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

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

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

Now

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

for $t = 3, ..., T$.

The first equation for which we have a valid moment condition is now

$$\Delta y_{i4} = \alpha \Delta y_{i3} + \Delta \varepsilon_{i4} + \phi \Delta \varepsilon_{i3}$$

since $E[y_{i1}(\Delta \varepsilon_{i4} + \phi \Delta \varepsilon_{i3})] = 0$.

Hence we now require $T \geq 4$ to identify α.

More generally we have a set of linear moment conditions

$$E[y_{i,t-s}(\Delta\varepsilon_{it} + \phi\Delta\varepsilon_{i,t-1})] = 0 \quad \text{for } t = 4, ..., T \text{ and } s \geq 3$$

that can be used to obtain consistent GMM estimators of α in the presence of an MA(1) error process.

This extends naturally to higher order MA(q) error processes, provided the minimum number of time series observations needed to identify α are available.

NB. Models with MA errors arise naturally in dynamic models with (serially uncorrelated) errors in variables.

If the AR(1) specification is correct for true values of a series y_{it}^*

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

but we observe the noisy measure

$$y_{it} = y_{it}^* + m_{it} \leftrightarrow y_{it}^* = y_{it} - m_{it}$$

then the model for the observed series y_{it} has the form

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

which has an MA(1) error component if the measurement errors m_{it} are serially uncorrelated.

AR errors

Example

$$y_{it} = \beta x_{it} + \eta_i + v_{it}$$

$$v_{it} = \rho v_{i,t-1} + \varepsilon_{it}$$

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

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

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

Suppose that x_{it} is correlated with η_i, and predetermined or endogenous (but not strictly exogenous) wrt ε_{it}.

Then there are no valid instruments for the first-differenced equations

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta v_{it}$$

Lagged values of both y_{it} and x_{it} are correlated with past ε_{it} shocks, and the autoregressive error term $\Delta v_{it} = \rho \Delta v_{i,t-1} + \Delta \varepsilon_{it}$ is also related to these past shocks.

Consistent estimation of β requires that we transform the 'static' model to obtain a dynamic representation with serially uncorrelated shocks.

$$\begin{aligned} y_{it} &= \beta x_{it} + \eta_i + v_{it} \\ \rho y_{i,t-1} &= \rho\beta x_{i,t-1} + \rho\eta_i + \rho v_{i,t-1} \\ y_{it} - \rho y_{i,t-1} &= \beta x_{it} - \rho\beta x_{i,t-1} + (1-\rho)\eta_i + v_{it} - \rho v_{i,t-1} \\ y_{it} &= \rho y_{i,t-1} + \beta x_{it} - \rho\beta x_{i,t-1} + (1-\rho)\eta_i + \varepsilon_{it} \end{aligned}$$

This is a dynamic model with serially uncorrelated shocks, that we know how to estimate. The explanatory variables (x_{it} and $x_{i,t-1}$) are correlated with the individual effects $(1-\rho)\eta_i$, and predetermined or endogenous wrt the serially uncorrelated shocks ε_{it}.

Taking first-differences gives the equations

$$\Delta y_{it} = \pi_1 \Delta y_{i,t-1} + \pi_2 \Delta x_{it} + \pi_3 \Delta x_{i,t-1} + \Delta \varepsilon_{it}$$

for $t = 3, ..., T$, for which we have the moment conditions

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

$$E(x_{i,t-s} \Delta \varepsilon_{it}) = 0 \quad \text{for } s \geq 2 \text{ or for } s \geq 1$$

Standard methods can thus be used to estimate the unrestricted parameters (π_1, π_2, π_3).

The non-linear 'common factor' restriction $\pi_3 = -\pi_1 \pi_2$ can then be tested and imposed if required, for example using minimum distance methods.

Alternatively β and ρ could be estimated directly from

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \beta \Delta x_{it} - \beta \rho \Delta x_{i,t-1} + \Delta \varepsilon_{it}$$

using the same moment conditions to implement a non-linear GMM estimator.

Specification tests

The basic specification test for GMM estimators is the Sargan (1958)/Hansen (1982) test of overidentifying restrictions (or J test).

Recall that these estimators were motivated by setting sample analogues of population orthogonality restrictions as close to zero as possible.

If the model is just identified, we achieve this exactly.

But if the model is overidentified, we can test whether the overidentifying restrictions are set close enough to zero to be consistent with their validity, when evaluated at the optimal GMM parameter estimates. If they are small enough, we do not reject the validity of the moment conditions used. Otherwise we reject. Very loosely, this is like testing for correlation between the model residuals and (a subset of) the instruments used.

For the model

$$y_i = X_i\delta + u_i \qquad i = 1, ..., N$$

$$E(Z_i' u_i^*) = 0$$

where u_i^* denotes some transformation of the u_i (e.g. first-differencing), the test of overidentifying restrictions is

$$
\begin{aligned}
S &= NJ_N(\widehat{\delta}_{GMM}) \\
&= N(\frac{1}{N}\sum_{i=1}^{N}\widehat{\widehat{u}}_i^{*\prime} Z_i)(\frac{1}{N}\sum_{i=1}^{N} Z_i^{\prime}\widehat{u}_i^{*}\widehat{u}_i^{*\prime} Z_i)^{-1}(\frac{1}{N}\sum_{i=1}^{N} Z_i^{\prime}\widehat{\widehat{u}}_i^{*})
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{u}_i^{*} &= y_i^{*} - X_i^{*}\widehat{\delta} \\
\widehat{\widehat{u}}_i^{*} &= y_i^{*} - X_i^{*}\widehat{\delta}_{GMM}
\end{aligned}
$$

$\widehat{\delta}$ is an initial consistent estimator, s.t. $(\frac{1}{N}\sum_{i=1}^{N} Z_i^{\prime}\widehat{u}_i^{*}\widehat{u}_i^{*\prime} Z_i)^{-1} = W_N$ is the weight matrix used to calculate the optimal two step estimator.

$\widehat{\delta}_{GMM}$ is the optimal two step estimator, so that $\widehat{\widehat{u}}_i^{*}$ are the two step residuals.

If we have p columns in Z_i and k columns in X_i (or k rows in δ),

$$S \overset{a}{\sim} \chi^2(p - k)$$

under the null hypothesis that $E(Z_i' u_i^*) = 0$.

Values of the Sargan/Hansen test statistic that are too large, relative to critical values of the appropriate χ^2 distribution, reject the validity of the set of moment conditions used.

Tests of serial correlation

Difference Sargan test

This test procedure can be used to test nested hypotheses, where (only) a subset of the moment conditions that are valid under the null hypothesis remain valid under a less restrictive alternative hypothesis.

Example

$H_0 : v_{it} \sim MA(0)$

$H_1 : v_{it} \sim MA(1)$

For the basic AR(1) model, under the null we have the set of moment conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0 \text{ for } t = 3, ..., T \text{ and } s \geq 2 \leftrightarrow E(Z_i'\Delta v_i) = 0$$

While under the alternative we have the smaller set of moment conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0 \text{ for } t = 4, ..., T \text{ and } s \geq 3 \leftrightarrow E(Z_i^{A\prime}\Delta v_i) = 0$$

Importantly, Z_i^A is a strict subset of Z_i.

We estimate both under the null and under the alternative, and construct the Sargan statistics S under H_0 and S^A under H_1.

If the null is valid, they should both not reject, and the difference between them should not be too large.

If the null if invalid but the alternative is valid, then S should reject while S^A should not reject. The difference between them should be larger.

Less obviously, if both the null and the alternative are invalid, both S and S^A should reject. Again in this case the difference between them should be larger than in the case where the null is valid.

The difference $S - S^A$ is therefore informative about the validity of the null hypothesis.

Formally

$$DS = S - S^A \overset{a}{\sim} \chi^2(p - p^A)$$

where p is again the number of columns in Z_i and p^A is the number of columns in Z_i^A.

Large values of the DS statistic again reject the null hypothesis.

NB. Rejecting the null does not imply acceptance of the specific alternative that was considered. If we reject the null of no serial correlation against the MA(1) alternative, the next step should be to test the null of MA(1) shocks against a more general alternative (e.g. MA(2)).

Hausman test

The Hausman test has the same sequential reasoning as the Difference Sargan test, but focuses on the difference in the estimated parameter vectors under the null and under the alternative, rather than on the difference in the corresponding Sargan statistics.

Let $\widehat{\delta}_{GMM}$ denote the optimal two step GMM estimator under H_0 (i.e. using instruments Z_i) and let $\widehat{\delta}^{A}_{GMM}$ denote the optimal two step GMM estimator under H_1 (i.e. using instruments Z_i^A).

If the null is valid, both estimators are consistent, and the difference between them should not be too large.

If the null is invalid, then $\widehat{\delta}_{GMM}$ (at least) is inconsistent, and in this case the difference $\widehat{\delta}^A_{GMM} - \widehat{\delta}_{GMM}$ should be larger.

Formally

$$h = (\widehat{\delta}^A_{GMM} - \widehat{\delta}_{GMM})'[avar(\widehat{\delta}^A_{GMM}) - avar(\widehat{\delta}_{GMM})]^{-1}(\widehat{\delta}^A_{GMM} - \widehat{\delta}_{GMM}) \overset{a}{\sim} \chi^2(r)$$

where $r = rank[avar(\widehat{\delta}^A_{GMM}) - avar(\widehat{\delta}_{GMM})]$.

When this covariance matrix is of full rank, we have $r = k$, the number of parameters in δ.

When this is not of full rank, we can either use a generalised inverse to evaluate h, and adjust the degrees of freedom accordingly; or more simply, apply the test using a subset of the parameters in δ for which this complication does not arise.

Direct test for serial correlation

Arellano and Bond (1991) proposed a direct test for serial correlation in the residuals of the first-differenced specification.

Since $\Delta v_{it} = v_{it} - v_{i,t-1}$ is MA(1) under the null hypothesis that $E(v_{is}v_{it}) = 0$ for $s \neq t$, we expect to find (negative) first-order serial correlation in the first-differenced residuals.

However it is informative to test for the absence of second-order serial correlation in the first-differenced residuals.

Let $\widehat{\Delta v}$ be the stacked $N(T-2) \times 1$ vector of first-differenced residuals $\widehat{\Delta v}_{it}$

Let $\widehat{\Delta v}_{-2}$ be the $N(T-4) \times 1$ vector of observations on the second lags of these first-differenced residuals $\widehat{\Delta v}_{i,t-2}$

Let $\widehat{\Delta v}_{*}$ be the $N(T-4) \times 1$ vector of observations on $\widehat{\Delta v}_{it}$ for the same periods in which $\widehat{\Delta v}_{i,t-2}$ is observed

Then

$$m_2 = \frac{\widehat{\Delta v}'_{-2}\widehat{\Delta v}_{*}}{se} \overset{a}{\sim} N(0,1)$$

under $H_0 : E(\Delta v_{it}\Delta v_{i,t-2}) = 0$.

The expression for the standard error of this autocovariance (se) can be found in Arellano and Bond (1991).

Values outside the range ± 1.96 thus reject the null at the 95% level.

Note that the sign of the test statistic is also informative about the sign of any correlation that is detected.

Tests of this type can also be calculated for estimates of the residuals for the first-differenced equations constructed using estimators that have been based on other transformations.

E.g. suppose that $\widehat{\delta}_{GMM}$ is a 'system GMM' estimator.

We can still construct

$$\widehat{\Delta v}_i = \Delta y_i - \Delta X_i \widehat{\delta}_{GMM}$$

and test for the absence of second-order serial correlation in these first-differenced residuals.

The corresponding test for the absence of first-order serial correlation in the first-differenced residuals (m_1) can - and should - be used to check that significant negative first-order serial correlation is detected, if some of the moment conditions used depend on the assumption that the levels of the v_{it} shocks are serially uncorrelated.

Tests of other assumptions

The Difference Sargan and Hausman test procedures can also be used to test assumptions about the status of x_{it} explanatory variables, and to test the initial conditions restriction required for lagged values of Δy_{is} to be valid instruments in the levels equations.

For example, suppose we want to test the null hypothesis $E(\Delta y_{i2}\eta_i) = 0$ in the simple AR(1) model against the alternative $E(\Delta y_{i2}\eta_i) \neq 0$.

We compute the 'system GMM' estimator under the null and the basic first-differenced GMM estimator under the alternative. The Hausman test considers whether these estimators are similar enough for the null hypothesis not to be rejected; while the Difference Sargan test considers whether the difference between the corresponding Sargan statistics is small enough for the null hypothesis not to be rejected.

Although we should be aware that these tests are likely to have low power in cases where the moment conditions available for the equations in first-differences provide only weak identification (i.e. if the true value of α is close to 1 in the simple AR(1) model; or if some of the x_{it} series are close to being random walks in more general specifications).