Advanced Dynamic Panel Data Methods

GMM Techniques for Dynamic Simultaneous Panel Data Models

Sebastian Kripfganz

University of Exeter Business School, Department of Economics, Exeter, UK
www.kripfganz.de @Kripfganz

Universidad de Salamanca

July 17–19, 2023

Instruments and Moment Conditions

- Recall that the simple IV or 2SLS estimator with instruments $\mathbf{z}_{it}$ builds on the first-stage regression

$$\Delta y_{i,t-1} = \mathbf{z}'_{it}\boldsymbol{\pi} + \nu_{it}$$

with homogeneous coefficients $\boldsymbol{\pi}$.

- If the process is not stationary, the first-stage coefficients are generally heterogeneous:

$$\Delta y_{i,t-1} = \mathbf{z}'_{it}\boldsymbol{\pi}_t + \nu_{it}$$

  - Imposing homogeneous first-stage coefficients still yields consistent estimates. However, exploiting the first-stage heterogeneity can lead to efficiency gains.

Instruments and Moment Conditions

- We can easily implement an estimator with first-stage time-series heterogeneity by interacting the K_z instruments $\mathbf{z}_{it}$ with time dummies $d_s = \mathcal{I}(s = t)$, $s = 3, 4, \dots, T$:

$$\Delta y_{i,t-1} = \sum_{s=3}^{T} (d_s \mathbf{z}_{it})' \boldsymbol{\pi}_s + \nu_{it}$$

  - This yields a total of $K_z(T-2)$ instruments.

Instruments and Moment Conditions

- Re-consider the first-differenced regression model in vector form:

$$\Delta \mathbf{y}_i = \lambda \Delta \mathbf{y}_{i,-1} + \Delta \varepsilon_i$$

with single instrument

$$\mathbf{z}_i = \mathbf{y}_{i,-2} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,T-2} \end{pmatrix}$$

  - The simple IV estimator is

$$\hat{\lambda}_{IV} = \left(\sum_{i=1}^{N} \mathbf{z}_i' \Delta \mathbf{y}_{i,-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{z}_i' \Delta \mathbf{y}_i = \lambda + \left(\sum_{i=1}^{N} \mathbf{z}_i' \Delta \mathbf{y}_{i,-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{z}_i' \Delta \varepsilon_i$$

Instruments and Moment Conditions

- Interacted with time dummies, the matrix of instruments becomes the diagonal matrix

$$\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 & \cdots & 0 \\ 0 & y_{i2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & y_{i,T-2} \end{pmatrix}$$

  - The corresponding 2SLS estimator is

$$\hat{\lambda}_{2SLS} = \left(\sum_{i=1}^{N} \Delta\mathbf{y}_{i,-1}' \mathbf{Z}_i \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \Delta\mathbf{y}_{i,-1} \right)^{-1} \times \sum_{i=1}^{N} \Delta\mathbf{y}_{i,-1}' \mathbf{Z}_i \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \Delta\mathbf{y}_i$$

Instruments and Moment Conditions

- By the law of large numbers, consistency of $\hat{\lambda}_{IV}$ requires

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i' \Delta \boldsymbol{\varepsilon}_i = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=3}^{T} y_{i,t-2}\Delta\varepsilon_{it} = E\left[\sum_{t=3}^{T} y_{i,t-2}\Delta\varepsilon_{it}\right] = 0$$

while for consistency of $\hat{\lambda}_{2SLS}$ we need

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i' \Delta \boldsymbol{\varepsilon}_i = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} y_{i1}\Delta\varepsilon_{i3} \\ y_{i2}\Delta\varepsilon_{i4} \\ \vdots \\ y_{i,T-2}\Delta\varepsilon_{iT} \end{pmatrix} = E\left[\begin{pmatrix} y_{i1}\Delta\varepsilon_{i3} \\ y_{i2}\Delta\varepsilon_{i4} \\ \vdots \\ y_{i,T-2}\Delta\varepsilon_{iT} \end{pmatrix}\right] = \mathbf{0}$$

  - The 2SLS estimator with expanded instruments exploits the moment conditions $E[y_{i,t-2}\Delta\varepsilon_{it}] = 0$ separately for each time period t, while for the simple IV estimator they only need to hold on average, $\frac{1}{T-2}\sum_{t=3}^{T} E[y_{i,t-2}\Delta\varepsilon_{it}] = 0$.

Instruments and Moment Conditions

- Arellano and Bond (1991) noticed that there are further moment conditions that can be exploited:

$$E[y_{i,t-s}\Delta\varepsilon_{it}] = 0$$

for $s \geq 2$ (not just $s = 2$). This yields heterogeneous first-stage regressions in which the available number of instruments is growing with t:

$$\Delta y_{i,t-1} = \sum_{s=2}^{t-1} \pi_{t,t-s} y_{i,t-s} + \nu_{it}$$

  - Thus, for $t = 3$ there is the single instrument y_{i1}, for $t = 4$ there are two instruments (y_{i1}, y_{i2}), and eventually for $t = T$ there are $T - 2$ instruments $(y_{i1}, y_{i2}, \ldots, y_{i,T-2})$.
  - This results in a total of $K_z = \frac{(T-1)(T-2)}{2}$ instruments.

Instruments and Moment Conditions

- The matrix of instruments becomes the block-diagonal matrix

$$\mathbf{Z}_i = \left(\begin{array}{c|cc|c|cccc} y_{i1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline 0 & y_{i1} & y_{i2} & & 0 & 0 & \cdots & 0 \\ \hline \vdots & & & \ddots & & & & \\ \hline 0 & 0 & 0 & & y_{i1} & y_{i2} & \cdots & y_{i,T-2} \end{array}\right)$$

- In matrix notation, the corresponding 2SLS estimator is

$$\hat{\lambda}_{2SLS} = \left(\Delta\mathbf{y}'_{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Delta\mathbf{y}_{-1}\right)^{-1}\Delta\mathbf{y}'_{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Delta\mathbf{y}$$

Instruments and Moment Conditions

- In matrix notation, the moment conditions are

$$E[\mathbf{Z}_i' \Delta \varepsilon_i] = \mathbf{0}$$

- For any deterministic $K_z \times K_{\tilde{z}}$ transformation matrix $\mathbf{R}$,

$$E[\mathbf{R}'\mathbf{Z}_i' \Delta \varepsilon_i] = \mathbf{0}$$

  still yields valid moment conditions.
- If $\mathbf{R}$ is a square matrix of full rank, the resulting 2SLS estimator is unaffected because

$$\mathbf{ZR}(\mathbf{R}'\mathbf{Z}'\mathbf{ZR})^{-1}\mathbf{R}'\mathbf{Z}' = \mathbf{ZRR}^{-1}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}'^{-1}\mathbf{R}'\mathbf{Z}' = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

Instruments and Moment Conditions

- With a suitable choice of $\mathbf{R}$, which adds up and interchanges various columns of $\mathbf{Z}_i$, we can thus re-organize the instruments equivalently:

$$\mathbf{RZ}_i = \left(\begin{array}{ccccc|cccc|c} y_{i1} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \\ y_{i2} & y_{i1} & 0 & \cdots & 0 & y_{i2} & 0 & \cdots & 0 & \\ y_{i3} & y_{i2} & y_{i1} & & 0 & y_{i3} & y_{i2} & & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & & \vdots & \vdots & \ddots & & \\ y_{i,T-2} & y_{i,T-3} & y_{i,T-4} & \cdots & y_{i1} & y_{i,T-2} & y_{i,T-3} & \cdots & y_{i2} & \end{array} \right)$$

  - The first column is the Anderson and Hsiao (1981) instrument $\mathbf{y}_{i,-2}$. The difference of the first column in the second block and the second column in the first block corresponds to the instrument $\Delta\mathbf{y}_{i,-2}$.

Instrument Reduction Techniques

- The model is strongly overidentified unless T is very small. The number of instruments increases quadratically in T.
- Asymptotically (as $N \to \infty$ with T fixed), more valid instruments in principle improve the efficiency of the estimator.
- As discussed by Roodman (2009) in some detail, too many instruments relative to N can cause biased coefficient and standard error estimates and weakened specification tests.
  - Intuitively, looking at the extreme case, when the number of instruments approaches the sample size, the instruments perfectly predict the regressor in the first stage.
  - Finite-sample biases are aggravated if the additional instruments are weak:

$$Cov(y_{i,t-s}, \Delta y_{i,t-1}) = -\frac{\lambda^{s-2}}{1+\lambda}\sigma_\varepsilon^2 \to 0$$

    as $s \to \infty$ (for $s \geq 2$).

Instrument Reduction Techniques

- It is usually beneficial to give up some efficiency in favor of lower bias by reducing the number of instruments (Kiviet, 2020). Technically, this is done by choosing a rank-deficient transformation matrix $\mathbf{R}$, which removes and/or linearly combines some columns from $\mathbf{Z}_i$.
- One such instrument reduction approach – commonly referred to as "collapsing" – just keeps the first block of the (re-organized) instrument matrix:

$$\mathbf{R}_{col}\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & \cdots & 0 \\ y_{i2} & y_{i1} & 0 & \cdots & 0 \\ y_{i3} & y_{i2} & y_{i1} & & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ y_{i,T-2} & y_{i,T-3} & y_{i,T-4} & \cdots & y_{i1} \end{pmatrix}$$

Instrument Reduction Techniques

- Another approach – "curtailing" – limits the lag depth – i.e., the maximum number of instruments in each period's first-stage regression:

$$\Delta y_{i,t-1} = \sum_{s=2}^{1+p} \pi_{t,t-s} y_{i,t-s} + \nu_{it}$$

  - For example, if $p = 2$, this becomes

$$\mathbf{R}_{cur}\mathbf{Z}_i = \left(\begin{array}{c|cc|cc|c|cc} y_{i1} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \hline 0 & y_{i1} & y_{i2} & 0 & 0 & \cdots & 0 & 0 \\ \hline 0 & 0 & 0 & y_{i2} & y_{i3} & & 0 & 0 \\ \hline \vdots & & & & & \ddots & & \\ \hline 0 & 0 & 0 & 0 & 0 & & y_{i,T-3} & y_{i,T-2} \end{array}\right)$$

Instrument Reduction Techniques

- Both approaches can be combined:

$$\mathbf{R}_{cc}\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 \\ y_{i2} & y_{i1} \\ y_{i3} & y_{i2} \\ \vdots & \vdots \\ y_{i,T-2} & y_{i,T-3} \end{pmatrix}$$

  - The resulting 2SLS estimator differs from the one using both Anderson and Hsiao (1981) instruments $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$, which are equivalent to $\mathbf{z}_{it} = (y_{i,t-2}, y_{i,t-3})$, because the latter drops time period $t = 3$ from the estimation (due to the missing value for y_{i0}), while in the above instrument matrix $\mathbf{Z}_i$ (and its transformations) all missing values are replaced by zeros.
  - This leads to a slightly peculiar situation, where the first-stage coefficients are homogeneous for $t > 3$, but the second coefficient in the first stage for $t = 3$ is restricted to 0.

Instrument Reduction Techniques

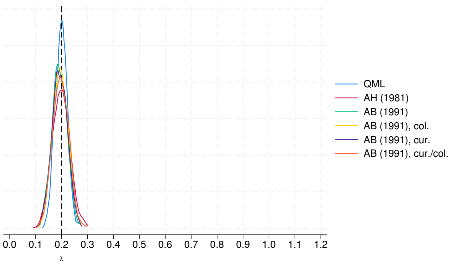
- A more logical approach might be to treat the first stage for the initial time period separately from the subsequent ones, which leads to the following transformed instrument matrix:

$$\mathbf{R}_{\tilde{cc}}\mathbf{Z}_i = \left(\begin{array}{c|cc} y_{i1} & 0 & 0 \\ \hline 0 & y_{i2} & y_{i1} \\ 0 & y_{i3} & y_{i2} \\ \vdots & \vdots & \\ 0 & y_{i,T-2} & y_{i,T-3} \end{array}\right)$$

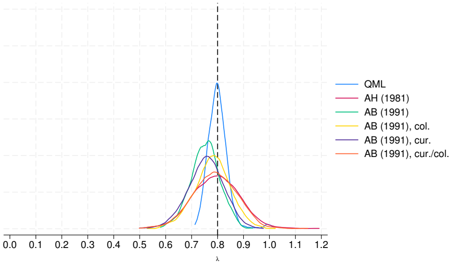
  - This idea is attributed to Jan Kiviet, but it has not received any attention in the empirical literature.
  - Using only the lower-right block of the above instrument matrix would now be equivalent to the estimator with both Anderson and Hsiao (1981) instruments.

Instrument Reduction Techniques

- Simulated distributions (kernel density estimates) based on 1,001 replications:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, and $\alpha_i \in \{-1,0,1\}$; $N = 300$, $T = 10$
  - Stationary initial observations: $y_{i1} = \frac{\alpha_i}{1-\lambda} + \nu_{i1}$, where $\nu_{i1} \sim \mathcal{N}\left(0, \frac{1}{1-\lambda^2}\right)$
  - All Arellano and Bond (1991) GMM estimators are efficient one-step estimators.



(a) $\lambda = 0.2$ (low persistence)



(b) $\lambda = 0.8$ (high persistence)

Weighting Matrix

- Recall that the first-differenced errors $\Delta\varepsilon_{it}$ exhibit first-order serial correlation if the untransformed idiosyncratic error component ε_{it} is serially uncorrelated:

$$
Var(\Delta\varepsilon_i) = \sigma_\varepsilon^2 \mathbf{D}\mathbf{D}' = \sigma_\varepsilon^2 \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & 0 \\ 0 & -1 & 2 & \ddots & \\ \vdots & & \ddots & \ddots & -1 \\ 0 & 0 & & -1 & 2 \end{pmatrix}
$$

  where **D** is the first-difference transformation matrix introduced earlier.
  - The 2SLS estimator is inefficient.

Weighting Matrix

- Let $\mathbf{Z}_{\Delta i}$ be the matrix of instruments (after application of any instrument reduction technique), such that $E[\mathbf{Z}'_{\Delta i}\Delta\varepsilon_i] = \mathbf{0}$.
- The asymptotic variance-covariance matrix of the moment functions is

$$Var(\mathbf{Z}'_{\Delta i}\Delta\varepsilon_i) = E[\mathbf{Z}'_{\Delta i}\Delta\varepsilon_i\Delta\varepsilon'_i\mathbf{Z}_{\Delta i}] = \sigma_\varepsilon^2 E[\mathbf{Z}'_{\Delta i}\mathbf{D}\mathbf{D}'\mathbf{Z}_{\Delta i}]$$

- An efficient GMM estimator uses the optimal weighting matrix

$$\mathbf{W} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_{\Delta i}\mathbf{D}\mathbf{D}'\mathbf{Z}_{\Delta i}\right)^{-1}$$

  - σ_ε^2 can be dropped because the estimator is invariant to multiplication of $\mathbf{W}$ by a constant scalar.
  - Since $\mathbf{D}\mathbf{D}'$ is a known matrix, no preliminary estimator is needed.

Weighting Matrix

- In practice, even if we retain the assumption of serially uncorrelated idiosyncratic errors ε_{it}, the homoskedasticity assumption usually needs to be relaxed.
  - The optimal weighting matrix then requires a preliminary consistent estimator:

$$\mathbf{W}(\hat{\lambda}_1) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_{\Delta i} \Delta\hat{\varepsilon}_i(\hat{\lambda}_1) \Delta\hat{\varepsilon}_i(\hat{\lambda}_1)' \mathbf{Z}_{\Delta i} \right)^{-1}$$

    where $\Delta\hat{\varepsilon}_i(\hat{\lambda}_1) = \Delta y_{it} - \hat{\lambda}_1 \Delta y_{i,t-1}$ are the first-differenced residuals, and $\hat{\lambda}_1$ is typically the inefficient but consistent one-step GMM estimator $\hat{\lambda}_1 = \hat{\lambda}_{GMM}(\mathbf{W})$ with the weighting matrix $\mathbf{W}$ that would be optimal under homoskedasticity.
  - As noted before, iterated GMM and continously-updating GMM are alternatives to the simple two-step procedure.

Standard Errors

- Since all moment functions are linear in the parameters, the (one-step, two-step, or iterated) GMM estimator can be obtained in closed form. For example, the two-step estimator is

$$\hat{\lambda}_{GMM}(\mathbf{W}(\hat{\lambda}_1)) = \left(\Delta\mathbf{y}'_{-1}\mathbf{Z}_\Delta\mathbf{W}(\hat{\lambda}_1)\mathbf{Z}'_\Delta\Delta\mathbf{y}_{-1}\right)^{-1}\Delta\mathbf{y}'_{-1}\mathbf{Z}_\Delta\mathbf{W}(\hat{\lambda}_1)\mathbf{Z}'_\Delta\Delta\mathbf{y}$$

- If the one-step estimator was used with non-optimal weighting matrix, robust standard errors should be computed with the conventional "sandwich" formula:

$$\widehat{Var}(\hat{\lambda}_{GMM}(\mathbf{W})) = (\Delta\mathbf{y}'_{-1}\mathbf{Z}_\Delta\mathbf{W}\mathbf{Z}'_\Delta\Delta\mathbf{y}_{-1})^{-1}\Delta\mathbf{y}'_{-1}\mathbf{Z}_\Delta\mathbf{W}\hat{\mathbf{V}}\mathbf{W}\mathbf{Z}'_\Delta\Delta\mathbf{y}_{-1} \times (\Delta\mathbf{y}'_{-1}\mathbf{Z}_\Delta\mathbf{W}\mathbf{Z}'_\Delta\Delta\mathbf{y}_{-1})^{-1}$$

where $\hat{\mathbf{V}} = \mathbf{W}(\hat{\lambda}_{GMM}(\mathbf{W}))^{-1}$ is a consistent estimate of $Var(\mathbf{Z}'_{\Delta i}\Delta\varepsilon_i)$.

Standard Errors

- For the two-step GMM estimator with optimal weighting matrix, GMM standard errors computed with the conventional formula

$$\widehat{Var}(\hat{\lambda}_{GMM}(\mathbf{W}(\hat{\lambda}_1))) = \left(\Delta \mathbf{y}_{-1}' \mathbf{Z}_\Delta \mathbf{W}(\hat{\lambda}_1) \mathbf{Z}_\Delta' \Delta \mathbf{y}_{-1}\right)^{-1}$$

can be severely downward biased in small samples. This is due to the ignored sampling variation in the estimation of the preliminary estimator $\hat{\lambda}_1$.

  - This can be accounted for with the Windmeijer (2005) correction, which is now standard practice.
  - A further refinement was proposed by Hwang, Kang, and Lee (2022). Their "doubly-robust" standard errors additionally correct for a bias resulting from the estimator's overidentification.
  - Similar adjustments need to be made for the iterated GMM estimator (Hansen and Lee, 2021).

Additional Regressors

- Additional regressors $\mathbf{x}_{it}$ can be accommodated in a straightforward way, assuming that they are not time invariant:

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}' \beta + \varepsilon_{it}$$

- Maintaining the assumption of serially uncorrelated ε_{it}, valid instruments can be found by classifying the regressors $\mathbf{x}_{it} = (\mathbf{x}_{1,it}, \mathbf{x}_{2,it}, \mathbf{x}_{3,it})$ as strictly exogenous $(\mathbf{x}_{1,it})$, predetermined $(\mathbf{x}_{2,it})$, or endogenous $(\mathbf{x}_{3,it})$.
  - The matrix of instruments $\mathbf{Z}_\Delta = (\mathbf{Z}_{\Delta,y_{-1}}, \mathbf{Z}_{\Delta,x_1}, \mathbf{Z}_{\Delta,x_2}, \mathbf{Z}_{\Delta,x_3})$ can be partitioned into separate blocks for each variable. Each block has a similar structure to the one in the simple panel AR(1) model. Instrument reduction techniques should generally be applied unless T is very small relative to N.
  - Further variables validly excluded from the regression model could be added as instruments, if available.
  - All results about estimator efficiency and robust standard errors carry over.
  - Instruments obtained from regressors $\mathbf{x}_{it}$ can also provide additional identification strength for the coefficient λ of the lagged dependent variable.

Additional Regressors

- Strictly exogenous regressors $\mathbf{x}_{1,it}$ satisfy the moment conditions

$$E[\mathbf{x}_{1,i,t-s}\Delta\varepsilon_{it}] = \mathbf{0}$$

  for all s.
  - It is customary to restrict $s \geq 0$.
- Predetermined regressors $\mathbf{x}_{2,it}$ satisfy the moment conditions

$$E[\mathbf{x}_{2,i,t-s}\Delta\varepsilon_{it}] = \mathbf{0}$$

  for all $s \geq 1$.
  - Notice that variables are classified as predetermined with respect to ε_{it}. This implies that they are endogenous with respect to $\Delta\varepsilon_{it}$.
- Endogenous regressors $\mathbf{x}_{3,it}$ satisfy the moment conditions

$$E[\mathbf{x}_{3,i,t-s}\Delta\varepsilon_{it}] = \mathbf{0}$$

  for all $s \geq 2$.

Serial Correlation

- The instruments have been obtained under the assumption of serially uncorrelated idiosyncratic errors ε_{it} (which corresponds to first-order serial correlation in $\Delta\varepsilon_{it}$).
- If we suspect first-order serial correlation in ε_{it} (which corresponds to second-order serial correlation in $\Delta\varepsilon_{it}$), the instrument $y_{i,t-2}$ becomes invalid, but all further lags $y_{i,t-3}, y_{i,t-4}, \ldots$ remain valid (and similarly for regressors $\mathbf{x}_{it}$).
- A feasible strategy for dealing with serial correlation would thus be to adjust the starting lag for the instruments accordingly.
- However, keep in mind that deeper lags tend to be weaker instruments.
  - A more promising strategy is often to view serially correlated errors as evidence of model misspecification, and to adjust the model appropriately with the aim to obtain a dynamically complete model. This can be done by adding higher-order lags of the dependent variable as further regressors or by adding distributed lags of $\mathbf{x}_{it}$.

Overidentification Test

- It is standard practice to test for the validity of the $K_z - 1 - K_x$ overidentifying restrictions with the Sargan (1958) test for the one-step GMM estimator,

$$J = N\ \mathbf{m}(\hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}))'\ \mathbf{W}\ \mathbf{m}(\hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W})) \xrightarrow{d} \chi^2(K_z - 1 - K_x)$$

or the Hansen (1982) test for the two-step GMM estimator,

$$J = N\ \mathbf{m}(\hat{\boldsymbol{\theta}}_{GMM}\mathbf{W}(\hat{\boldsymbol{\theta}}_1))'\ \mathbf{W}(\hat{\boldsymbol{\theta}}_1)\ \mathbf{m}(\hat{\boldsymbol{\theta}}_{GMM}\mathbf{W}(\hat{\boldsymbol{\theta}}_1)) \xrightarrow{d} \chi^2(K_z - 1 - K_x)$$

where $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$

  - This is not a test of the validity of all K_z instruments. It requires the maintained assumption that any $1 + K_x$ instruments (or linear combinations of instruments) are valid.
  - The Sargan (1958) test based on the one-step GMM estimator is asymptotically invalid if the weighting matrix is not optimal.

Overidentification Test

- A rejection of the overidentifying restrictions is often not informative about the type of model misspecification. The leading cases worth investigating are:
  - The classification of regressors into strictly exogenous, predetermined, and endogenous variables might be incorrect.
  - The model might not be dynamically complete, possibly due to omitted (higher-order) lags of the dependent variable or regressors, which can cause serial correlation of the idiosyncratic error term.
    - Blundell and Bond (2000) provide a theoretical justification for an error term with MA(1) structure resulting from measurement error.
    - Separate tests for serial correlation can aid in identifying the source of the problem.
  - Other relevant explanatory variables might be omitted from the model.
- Importantly, as emphasized by Roodman (2009), overidentification tests tend to substantially underreject the null hypothesis of no model misspecification when there are (too) many instruments, thus lulling applied researchers into a false sense of security.

Incremental Overidentification Tests

- To assess the correct regressor classification, incremental overidentification tests in the spirit of Eichenbaum, Hansen, and Singleton (1988) can be used.
  - Assuming that the moment conditions $E[\mathbf{x}_{2,i,t-s}\Delta\varepsilon_{it}] = 0$ are always valid for $s > 2$, predetermined regressors satisfy the additional moment restrictions $E[\mathbf{x}_{2,i,t-1}\Delta\varepsilon_{it}] = 0$.
  - Similarly, assuming that it can be taken for granted that $E[\mathbf{x}_{1,i,t-s}\Delta\varepsilon_{it}] = 0$ for $s > 1$, strictly exogenous regressors satisfy the additional moment restrictions $[\mathbf{x}_{1,it}\Delta\varepsilon_{it}] = 0$.
  - This suggests to contrast two estimators with and without the additional moment functions. A rejection then indicates that the variables under consideration should be classified as endogenous rather than predetermined.

Incremental Overidentification Tests

- The test statistic in its basic form is

$$J_2 - J_1 \xrightarrow{d} \chi^2(df_2 - df_1)$$

where $J_1 \xrightarrow{d} \chi^2(df_1)$ is the overidentification test statistic from the maintained model, and $J_2 \xrightarrow{d} \chi^2(df_2)$ is the overidentification test statistic from the extended model with the additional moment restrictions.

  - In finite samples, the incremental overidentification test statistic can become negative because the moment functions are weighted with separately estimated weighting matrices.

Incremental Overidentification Tests

- As an alternative that is guaranteed to be nonnegative, the relevant partition of the estimated weighting matrix from the extended model can be used to weight the moment functions from the maintained model (Newey, 1985).
- It is important to keep in mind that (incremental) overidentification tests are only meaningful if the baseline specification is correct.
  - There must be sufficiently many valid moment conditions to (just-)identify all parameters, which is an untestable assumption.
  - For the incremental test, the maintained model must be correctly specified.

Incremental Overidentification Tests

- An alternative to the incremental overidentification test is a generalized Hausman (1978) test, directly contrasting the two estimators with and without the moment functions under investigation:

$$H = (\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1)'(Var(\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1))^{-1}(\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1) \xrightarrow{d} \chi^2(\min(df_2 - df_1, 1 + K_x))$$

  - A robust estimate of $Var(\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1)$ in the spirit of White (1982) should be used that does not rely on one of the estimators to be fully efficient.
  - When the number of additional overidentifying restrictions, $df_2 - df_1$, is not larger than the number of contrasted coefficients, $1 + K_x$, then the generalized Hausman test is asymptotically equivalent to incremental overidentification tests.

Underidentification Tests

- The coefficients are underidentified if there is insufficient correlation of the instruments with the regressors, as an extreme case of weak identification.
- Windmeijer (2021) highlights that underidentification tests are overidentification tests in an auxiliary regression of any endogenous variable (in the wider sense) on the remaining regressors:

$$\Delta y_{i,t-1} = \Delta \mathbf{x}'_{it} \psi + \xi_{it}$$

  with the same instruments $\mathbf{Z}_{\Delta i}$ as before.
    - The null hypothesis is that the model is underidentified – i.e., a rejection of the test is desirable.
    - Sanderson and Windmeijer (2016) propose closely related weak-identification tests.

Transformed Level GMM Estimator

- Re-consider the simple panel AR(1) model. With the first-difference transformation matrix $\mathbf{D}$, such that $\Delta\boldsymbol{\varepsilon}_i = \mathbf{D}\boldsymbol{\varepsilon}_i$ and $\mathbf{D}\boldsymbol{\iota}_{T-1} = \mathbf{0}$, the moment conditions can be rewritten in terms of the untransformed level errors, as noted by Arellano and Bover (1995):

$$E[\mathbf{Z}'_{\Delta i}\Delta\boldsymbol{\varepsilon}_i] = E[\mathbf{Z}'_{\Delta i}\mathbf{D}(\alpha_i\boldsymbol{\iota}_{T-1} + \boldsymbol{\varepsilon}_i)] = E[\mathbf{Z}'_{Di}(\alpha_i\boldsymbol{\iota}_{T-1} + \boldsymbol{\varepsilon}_i)] = \mathbf{0}$$

where $\mathbf{Z}_{Di} = \mathbf{D}'\mathbf{Z}_{\Delta i}$.

  - This does not mean that differencing of the model is equivalent to differencing of the instruments, because $\mathbf{D} \neq \mathbf{D}'$
  - For example, if $T = 4$, the (unreduced) instrument matrix becomes

$$\mathbf{D}'\mathbf{Z}_{\Delta i} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{i1} & 0 & 0 \\ 0 & y_{i1} & y_{i2} \end{pmatrix} = \begin{pmatrix} -y_{i1} & 0 & 0 \\ y_{i1} & -y_{i1} & -y_{i2} \\ 0 & y_{i1} & y_{i2} \end{pmatrix} = \mathbf{Z}_{Di}$$

Transformed Level GMM Estimator

- The "difference GMM" estimator can then be written in terms of level variables, here for the one-step estimator:

$$\hat{\lambda}_{\Delta GMM}(\mathbf{W}) = \left(\mathbf{y}'_{-1}\mathbf{Z}_D\mathbf{W}\mathbf{Z}'_D\mathbf{y}_{-1}\right)^{-1}\mathbf{y}'_{-1}\mathbf{Z}_D\mathbf{W}\mathbf{Z}'_D\mathbf{y}$$

with weighting matrix

$$\mathbf{W} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_{Di}\mathbf{Z}_{Di}\right)^{-1} = \frac{1}{N}(\mathbf{Z}'_D\mathbf{Z}_D)^{-1}$$

such that the one-step GMM estimator equals the 2SLS estimator for the level model.

Forward-Orthogonal Deviations

- Instead of first differencing, Arellano and Bover (1995) suggest to transform the model into deviations from their forward mean:

$$\overrightarrow{\Delta} y_{it} = \lambda \overrightarrow{\Delta} y_{i,t-1} + \overrightarrow{\Delta} \varepsilon_{it}$$

where

$$\overrightarrow{\Delta} y_{it} = \sqrt{\frac{T-t+1}{T-t}} \left(y_{it} - \frac{1}{T-t+1} \sum_{s=t}^{T} y_{is} \right)$$

$$\overrightarrow{\Delta} y_{i,t-1} = \sqrt{\frac{T-t+1}{T-t}} \left(y_{i,t-1} - \frac{1}{T-t+1} \sum_{s=t}^{T} y_{i,s-1} \right)$$

$$\overrightarrow{\Delta} \varepsilon_{it} = \sqrt{\frac{T-t+1}{T-t}} \left(\varepsilon_{it} - \frac{1}{T-t+1} \sum_{s=t}^{T} \varepsilon_{is} \right)$$

Forward-Orthogonal Deviations

- In compact notation, the model in forward-orthogonal deviations is

$$\underbrace{\overrightarrow{\Delta}\mathbf{y}_i}_{\begin{pmatrix} \overrightarrow{\Delta} y_{i2} \\ \overrightarrow{\Delta} y_{i3} \\ \vdots \\ \overrightarrow{\Delta} y_{i,T-2} \end{pmatrix}} = \lambda \underbrace{\overrightarrow{\Delta}\mathbf{y}_{i,-1}}_{\begin{pmatrix} \overrightarrow{\Delta} y_{i1} \\ \overrightarrow{\Delta} y_{i2} \\ \vdots \\ \overrightarrow{\Delta} y_{i,T-2} \end{pmatrix}} + \underbrace{\overrightarrow{\Delta}\boldsymbol{\varepsilon}_i}_{\begin{pmatrix} \overrightarrow{\Delta} \varepsilon_{i2} \\ \overrightarrow{\Delta} \varepsilon_{i3} \\ \vdots \\ \overrightarrow{\Delta} \varepsilon_{i,T-1} \end{pmatrix}}$$

  - The transformation is again orthogonal to any time-invariant variable, especially the unit-specific error component α_i.
  - The scaling factor $\sqrt{\frac{T-t+1}{T-t}}$ ensures that $Var(\varepsilon_{it}) = Var(\overrightarrow{\Delta}\varepsilon_{it}) = \sigma_\varepsilon$ under homoskedasticity.

Forward-Orthogonal Deviations

- While the first-differenced model is defined for periods $t = 3, 4, \ldots, T$, the deviations from forward means can be computed for periods $t = 2, 3, \ldots, T - 1$.
- In the forward-orthogonally transformed model, already the first lag of the dependent variable qualifies as an instrument:

$$E[y_{i,t-s} \overrightarrow{\Delta} \varepsilon_{it}] = 0$$

for $s \geq 1$ (instead of $s \geq 2$ as in the first-differenced model).

Forward-Orthogonal Deviations

- Recall that any transformation of the moment conditions

$$E[\mathbf{R}\mathbf{Z}'_{\Delta i}\mathbf{D}\varepsilon_i] = \mathbf{0}$$

with nonsingular transformation matrix $\mathbf{R}$ yields identical estimates.

  - As long as $\mathbf{Z}_{\Delta i}$ is of the unreduced block-diagonal structure with all available instruments, Arellano and Bover (1995) show that the instrument transformation matrix $\mathbf{R}$ can be chosen in a particular block-diagonal way such that

$$\mathbf{R}\mathbf{Z}'_{\Delta i}\mathbf{D}\varepsilon_i = \mathbf{Z}'_{\Delta i}\mathbf{K}\varepsilon_i$$

    for any upper-trapezoidal model transformation matrix $\mathbf{K}$ – i.e., all elements outside of the upper triangle of matrix $\mathbf{K}$ equal 0 – that satisfies $\mathbf{K}\boldsymbol{\iota}_{T-1} = \mathbf{0}$
  - This holds for forward-orthogonal deviations, where $\mathbf{K} = \mathbf{F}$ such that $\mathbf{F}\varepsilon_i = \overrightarrow{\Delta}\varepsilon_i$.

Forward-Orthogonal Deviations

- The $(T-2) \times (T-1)$ forward-orthogonal transformation matrix is

$$\mathbf{F} = \begin{pmatrix} \sqrt{\frac{T-1}{T-2}} & 0 & \cdots & 0 \\ 0 & \sqrt{\frac{T-2}{T-3}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{T-2}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} \\ 0 & \frac{T-3}{T-2} & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

  where the diagonal matrix at the front contains the variance-equating scaling factors.
  - Due to the scaling factor, $\mathbf{F}\mathbf{F}' = \mathbf{I}_{T-2}$
  - $\mathbf{F} = \mathbf{S}\mathbf{D}$ is a nonsingular transformation of the first-difference transformation matrix $\mathbf{D}$.

Forward-Orthogonal Deviations

- To illustrate the model transformation equivalence, consider the case for $T = 4$:

$$\underbrace{\left(\begin{array}{cc|c} \multicolumn{2}{c|}{\begin{pmatrix} \sqrt{\frac{3}{2}} & 0 \\ 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{2} \end{pmatrix}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline 0 & 0 & \sqrt{2}\left(-\frac{1}{2}\right) \end{array}\right)}_{\mathbf{R}} \underbrace{\begin{pmatrix} y_{i1} & 0 \\ 0 & y_{i1} \\ 0 & y_{i2} \end{pmatrix}}_{\mathbf{Z}'_{\Delta i}} \underbrace{\begin{pmatrix} \Delta\varepsilon_{i3} \\ \Delta\varepsilon_{i4} \end{pmatrix}}_{\Delta\boldsymbol{\varepsilon}_i}$$

$$= \underbrace{\begin{pmatrix} y_{i1} & 0 \\ 0 & y_{i1} \\ 0 & y_{i2} \end{pmatrix}}_{\mathbf{Z}'_{\Delta i}} \underbrace{\begin{pmatrix} \sqrt{\frac{3}{2}} & 0 \\ 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{2} \end{pmatrix}}_{\mathbf{S}} \underbrace{\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{pmatrix}}_{\boldsymbol{\varepsilon}_i}$$

$$= \underbrace{\begin{pmatrix} y_{i1} & 0 \\ 0 & y_{i1} \\ 0 & y_{i2} \end{pmatrix}}_{\mathbf{Z}'_{\Delta i}} \underbrace{\begin{pmatrix} \sqrt{\frac{3}{2}} & 0 \\ 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{pmatrix}}_{\boldsymbol{\varepsilon}_i} = \underbrace{\begin{pmatrix} y_{i1} & 0 \\ 0 & y_{i1} \\ 0 & y_{i2} \end{pmatrix}}_{\mathbf{Z}'_{\Delta i}} \underbrace{\begin{pmatrix} \overrightarrow{\Delta}\varepsilon_{i2} \\ \overrightarrow{\Delta}\varepsilon_{i3} \end{pmatrix}}_{\overrightarrow{\Delta}\boldsymbol{\varepsilon}_i}$$

Forward-Orthogonal Deviations

- First differencing and forward-orthogonal deviations yield identical estimators $\hat{\lambda}_{\Delta GMM}(\mathbf{W}) = \hat{\lambda}_{\overrightarrow{\Delta} GMM}(\mathbf{W})$ if all available instruments are used without application of any instrument reduction techniques.
- The estimators differ in unbalanced panel data sets with gaps.
  - A missing observation in the time series for y_{it} in period t implies a loss of two observations in first differences, because neither Δy_{it} nor $\Delta y_{i,t-1}$ can be computed. However, while we similarly cannot compute $\overrightarrow{\Delta} y_{it}$, any other $\overrightarrow{\Delta} y_{is}$ for $s \neq t$ can still be computed by skipping the missing observation in the calculation of the forward means. In such a scenario, forward-orthogonal deviations thus retain more information and lead to more efficient estimation.

Forward-Orthogonal Deviations

- Similar to the use of first-differenced instruments $\Delta y_{i,t-s}$, $s \geq 2$, for the first-differenced model, Hayakawa (2009) and Hayakawa, Qi, and Breitung (2019) proposed backward-orthogonally transformed instruments $\overleftarrow{\Delta} y_{i,t-s}$, $s \geq 1$, for the forward-orthogonally transformed model, where

$$\overleftarrow{\Delta} y_{i,t-s} = \sqrt{\frac{t-s}{t-s-1}} \left(y_{i,t-s} - \frac{1}{t-s} \sum_{l=1}^{t-s} y_{il} \right)$$

  - This becomes relevant if the initial observations only satisfy

$$E\left[\left(y_{i1} - \frac{\alpha}{1-\lambda} \right) \overrightarrow{\Delta} \varepsilon_{it} \right] = 0$$

    instead of the joint assumption $E[y_{i1}\varepsilon_{it}] = 0$ and $E[\alpha_i \varepsilon_{it}] = 0$ for $t \geq 2$.
  - There are no instruments available anymore for $t = 2$ in this case.

Forward-Orthogonal Deviations

- Instruments for additional regressors $\mathbf{x}_{it}$ can be found in a similar way:
  - Strictly exogenous regressors $\mathbf{x}_{1,it}$ satisfy the moment conditions

$$E[\mathbf{x}_{1,i,t-s} \overrightarrow{\Delta} \varepsilon_{it}] = \mathbf{0}$$

    for all s.
  - Predetermined regressors $\mathbf{x}_{2,it}$ satisfy the moment conditions

$$E[\mathbf{x}_{2,i,t-s} \overrightarrow{\Delta} \varepsilon_{it}] = \mathbf{0}$$

    for all $s \geq 0$.
  - Endogenous regressors $\mathbf{x}_{3,it}$ satisfy the moment conditions

$$E[\mathbf{x}_{3,i,t-s} \overrightarrow{\Delta} \varepsilon_{it}] = \mathbf{0}$$

    for all $s \geq 1$.

Forward-Orthogonal Deviations

- Based on the moment conditions $E[\mathbf{Z}'_{\overrightarrow{\Delta}i} \overrightarrow{\Delta} \varepsilon_i] = \mathbf{0}$, where $\mathbf{Z}_{\overrightarrow{\Delta}i} = \mathbf{Z}_{\Delta i}$ when no instrument reduction is applied, the optimal one-step weighting matrix under homoskedasticity is

$$\mathbf{W} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_{\overrightarrow{\Delta}i} \mathbf{F} \mathbf{F}' \mathbf{Z}_{\overrightarrow{\Delta}i} \right)^{-1} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_{\overrightarrow{\Delta}i} \mathbf{Z}_{\overrightarrow{\Delta}i} \right)^{-1}$$

- An optimal weighting matrix for two-step estimation is obtained in the usual way as

$$\mathbf{W}(\hat{\lambda}_1) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_{\overrightarrow{\Delta}i} \overrightarrow{\Delta} \hat{\varepsilon}_i(\hat{\lambda}_1) \overrightarrow{\Delta} \hat{\varepsilon}_i(\hat{\lambda}_1)' \mathbf{Z}_{\overrightarrow{\Delta}i} \right)^{-1}$$

where $\overrightarrow{\Delta} \hat{\varepsilon}_i(\hat{\lambda}_1) = \overrightarrow{\Delta} y_{it} - \hat{\lambda}_1 \overrightarrow{\Delta} y_{i,t-1}$ are the forward-orthogonally transformed residuals, and $\hat{\lambda}_1$ is a consistent preliminary estimator.

Within-Groups Transformation

- In general, the within-groups transformation $\mathbf{M}_\iota$ that yields deviations from within-group means, such that $\mathbf{M}_\iota \varepsilon_i = \bar{\Delta}\varepsilon_i$, is not suitable for models with dynamic feedback (as in the case of a lagged dependent variable). Because $\bar{\Delta}\varepsilon_{it} = \varepsilon_{it} - \frac{1}{T-1}\sum_{s=2}^{T}\varepsilon_{is}$ is not just a function of future but also all past errors, most instruments become invalid.
- We can still use it to formulate valid moment conditions for the subset of strictly exogenous variables $\mathbf{x}_{1it}$:

$$E[\mathbf{Z}'_{\bar{\Delta},x_1,i}\bar{\Delta}\varepsilon_i] = \mathbf{0}$$

Combination of Model Transformations

- It can be reasonable to combine different model transformations.
  - Because it is less intuitive (although perfectly valid) to use future observations – i.e., leads – as instruments for strictly exogenous regressors $\mathbf{x}_{1it}$, we could instead jointly use the moment conditions

$$E[\mathbf{x}_{1it} \bar{\Delta} \varepsilon_{it}] = 0$$
$$E[\mathbf{x}_{1i,t-s} \overrightarrow{\Delta} \varepsilon_{it}] = 0$$

    for $s \geq 0$. The former – akin to the moment conditions for the traditional FE estimator – are only valid for strictly exogenous regressors, while the latter are also valid for predetermined regressors.
  - An incremental overidentification test of the strict-exogeneity assumption can then be used to contrast estimators with and without the instruments for the within-groups transformed model.

Combination of Model Transformations

- Let $\mathbf{Z}_{\bar{\Delta}i}$ contain the instruments for the within-groups transformed model and $\mathbf{Z}_{\vec{\Delta}i}$ the instruments for the forward-orthogonally transformed model. The combined moment conditions are

$$E\left[\begin{pmatrix} \mathbf{Z}'_{\bar{\Delta}i}\bar{\Delta}\varepsilon_i \\ \mathbf{Z}'_{\vec{\Delta}i}\vec{\Delta}\varepsilon_i \end{pmatrix}\right] = E\left[\begin{pmatrix} \mathbf{Z}'_{\bar{\Delta}i}\mathbf{M}_\iota\varepsilon_i \\ \mathbf{Z}'_{\vec{\Delta}i}\mathbf{F}\varepsilon_i \end{pmatrix}\right] = E\left[\left(\mathbf{M}_\iota\mathbf{Z}_{\bar{\Delta}i} \mid \mathbf{F}'\mathbf{Z}_{\vec{\Delta}i} \right)'\varepsilon_i\right] = \mathbf{0}$$

  - Because all model transformations can be recast as instrument transformations, the resulting estimator is a conventional GMM estimator for the untransformed level model with instruments $\mathbf{Z}_i = (\mathbf{M}_\iota\mathbf{Z}_{\bar{\Delta}i}, \mathbf{F}'\mathbf{Z}_{\vec{\Delta}i})$

Nonlinear Moment Conditions

- Absence of serial correlation in ε_{it} is a necessary condition for the validity of many of the instruments.
- Ahn and Schmidt (1995) suggest to explicitly exploit this assumption in the form of the additional $T-3$ quadratic moment conditions:

$$E[(\alpha_i + \varepsilon_{iT})\Delta\varepsilon_{it}] = 0$$

  for $t = 3, 4, \dots, T-1$, provided that $T \geq 4$
  - These additional moment conditions improve efficiency and help with potential identification problems when $\lambda \to 1$, without requiring additional assumptions.
  - For the purpose of avoiding first-stage overfitting, a "collapsed" version can be implemented as

$$E\left[(\alpha_i + \varepsilon_{iT})\sum_{s=3}^{T-1}\Delta\varepsilon_{is}\right] = 0$$

Nonlinear Moment Conditions

- Under the weaker initial-observations assumption mentioned earlier (which only guarantees validity of $\Delta y_{i,t-s}$ instead of $y_{i,t-s}$, $s \geq 2$ as valid instruments for the first-differenced model), these nonlinear moment conditions need to be replaced by

$$E[\Delta y_{i,t-2}\Delta\varepsilon_{i,t-1} + (\Delta\varepsilon_{i,t-1})^2 + \Delta y_{i,t-1}\Delta\varepsilon_{it}] = 0$$

  for $t = 4, 5, \ldots, T$, as noted by Chudik and Pesaran (2022).
  - In this case, it is further required that there are no endogenous regressors $\mathbf{x}_{3it}$ in the model.

Nonlinear Moment Conditions

- Under homoskedasticity (and the stronger initial-observations assumption), Ahn and Schmidt (1995) propose to replace the nonlinear moment conditions by the $T-2$ nonlinear moment conditions

$$E[(\alpha_i + \bar{\varepsilon}_i)\Delta\varepsilon_{it}] = 0$$

for $t = 2, 3, \ldots, T$, and the additional $T-3$ linear moment conditions

$$E[y_{i,t-2}\Delta\varepsilon_{i,t-1} - y_{i,t-1}\Delta\varepsilon_{it}] = 0$$

for $t = 4, 5, \ldots, T$

  - Thus, homoskedasticity implies an extra $T-2$ overidentifying restrictions, which can be tested with an incremental overidentification test or a generalized Hausman test.

Nonlinear Moment Conditions

- The linear moment functions – e.g., $\mathbf{m}_{\Delta i}(\boldsymbol{\theta}) = \mathbf{Z}'_{\Delta i}\mathbf{D}\boldsymbol{\varepsilon}_i$ for the first-differenced model, or $\mathbf{m}_{\vec{\Delta}}(\boldsymbol{\theta}) = \mathbf{Z}'_{\vec{\Delta} i}\mathbf{F}\boldsymbol{\varepsilon}_i$ for the forward-orthogonally transformed model – can be stacked with the nonlinear moment functions $\mathbf{m}_{nl,i}(\boldsymbol{\theta})$:

$$\mathbf{m}_i(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{m}_{\Delta i}(\boldsymbol{\theta}) \\ \mathbf{m}_{nl,i}(\boldsymbol{\theta}) \end{pmatrix}$$

- With nonlinear moment functions, no closed-form solution exists. The GMM estimator is obtained by numerically minimizing the objective function:

$$\hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}(\hat{\boldsymbol{\theta}}_1)) = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}(\hat{\boldsymbol{\theta}})' \; \mathbf{W}(\hat{\boldsymbol{\theta}}_1) \; \mathbf{m}(\hat{\boldsymbol{\theta}})$$

  - An optimal weighting matrix for one-step GMM estimation does not exist in this case. A two-step, iterated, or continuously-updating GMM estimator is required for efficient estimation.

Instruments for Level Model

- Instead of exploiting nonlinear moment conditions to address the problem of weak identification when $\lambda \rightarrow 1$, further linear moment conditions can be found for the untransformed level model by imposing a stronger initial-observations condition.
- In addition to serially uncorrelated ε_{it} and $E[y_{i1}\varepsilon_{it}] = 0$, Blundell and Bond (1998) consider the assumption $E[\Delta y_{i2}\alpha_i] = 0$
    - In the simple panel AR(1) model, the latter assumption can be rewritten as

$$E\left[\left(y_{i2} - \frac{\alpha_i}{1-\lambda}\right)\alpha_i\right] = 0$$

      That is, a unit's (initial) deviation from their (unit-specific) long-run equilibrium $\frac{\alpha_i}{1-\lambda}$ – "steady state" – should be unrelated to the long-run equilibrium itself.
    - As Roodman (2009) notes, this creates a tension because this assumption is more likely to be violated when λ is close to 1 – i.e., when any deviations persist for long times – which is precisely the situation for which the new assumption is intended to provide additional identification strength.

Instruments for Level Model

- A sufficient but not necessary condition for this initial-observations assumption to hold is mean stationarity of the process for y_{it} (jointly with the processes for any $\mathbf{x}_{it}$ regressors).
- The recursive structure of the model then implies

$$E[\Delta y_{i,t-s}(\alpha_i + \varepsilon_{it})] = 0$$

for $s \geq 1$ and all $t \geq 3$ (not just the initial observations).

  - Thus, lagged first differences of the dependent variable qualify as instruments for the untransformed level model.

Instruments for Level Model

- It turns out that beyond $\Delta y_{i,t-1}$ all deeper lags $\Delta y_{i,t-s}$, $s \geq 2$, become redundant when the new level moment conditions are combined with the moment conditions $E[y_{i,t-s}\Delta\varepsilon_{it}]$, $s \geq 2$, for the model in first differences.
  - For example, if $T = 4$, the matrix with all (transformed) instruments for the level model becomes

$$(\mathbf{D}'\mathbf{Z}_{\Delta i}, \mathbf{Z}_{li}) = \left(\begin{array}{ccc|ccc} -y_{i1} & 0 & 0 & 0 & 0 & 0 \\ y_{i1} & -y_{i1} & -y_{i2} & \Delta y_{i2} & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \Delta y_{i2} & \Delta y_{i3} \end{array}\right)$$

    but column 5 equals a linear combination of columns 2 to 4 – column 3 minus column 2 plus column 4.
  - This redundancy result only holds if no instrument reduction techniques are applied, but it is customary to only include the first lagged difference, $\Delta y_{i,t-1}$, as an instrument for the level model in any case.

Instruments for Level Model

- The matrix of $T-2$ non-redundant additional instruments for the level model therefore becomes

$$\mathbf{Z}_{li} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & \Delta y_{i3} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Delta y_{i,T-1} \end{pmatrix}$$

  - To reduce the number of instruments, matrix $\mathbf{Z}_{li}$ can be collapsed into the column vector $\mathbf{Z}_{li} = (0, \Delta y_{i2}, \Delta y_{i3}, \ldots, \Delta y_{i,T-1})'$, imposing homogeneity on the first-stage coefficients.

Instruments for Level Model

- Non-redundant instruments for additional regressors $\mathbf{x}_{it}$ can be added accordingly under the additional assumption $E[\Delta \mathbf{x}_{it}\alpha_i] = \mathbf{0}$ (Blundell, Bond, and Windmeijer, 2001):
  - Strictly exogenous regressors $\mathbf{x}_{1,it}$ satisfy the $K_{x_1}(T-1)$ non-redundant moment conditions
  $$E[\Delta \mathbf{x}_{1,it}(\alpha_i + \varepsilon_{it})] = \mathbf{0}$$
  - Likewise, predetermined regressors $\mathbf{x}_{2,it}$ satisfy the $K_{x_2}(T-1)$ moment conditions
  $$E[\Delta \mathbf{x}_{2,it}(\alpha_i + \varepsilon_{it})] = \mathbf{0}$$
  - Endogenous regressors $\mathbf{x}_{3,it}$ satisfy the $K_{x_3}(T-2)$ moment conditions
  $$E[\Delta \mathbf{x}_{3,i,t-1}(\alpha_i + \varepsilon_{it})] = \mathbf{0}$$

Instruments for Level Model

- If instead of $E[\Delta \mathbf{x}_{it}\alpha_i] = \mathbf{0}$ the regressors $\mathbf{x}_{it}$ satisfy the stronger "random-effects" assumption $E[\mathbf{x}_{it}\alpha_i] = \mathbf{0}$, the following additional K_x non-redundant moment conditions arise:

$$E[\mathbf{x}_{i1}(\alpha_i + \varepsilon_{i2})] = \mathbf{0}$$

  - Consequently, this assumption can be tested with an incremental overidentification test or a generalized Hausman test.
  - In practice, one might also replace $E[\Delta \mathbf{x}_{1,it}(\alpha_i + \varepsilon_{it})] = \mathbf{0}$ by $E[\mathbf{x}_{1,it}(\alpha_i + \varepsilon_{it})] = \mathbf{0}$ for all t (and similarly for $\mathbf{x}_{2,it}$ and $\mathbf{x}_{3,i,t-1}$). However, the estimator is invariant to this alteration.

System GMM as Level GMM

- Combining the instruments for the level model with those for the transformed model (either in first differences or forward-orthogonal deviations) yields a so-called "system GMM" estimator.
- Regarding the origin of the name "system GMM", notice that the stacked moment functions can be written as

$$\begin{pmatrix} \mathbf{m}_{\Delta i}(\hat{\theta}) \\ \mathbf{m}_{li}(\hat{\theta}) \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_{\Delta i}\Delta\varepsilon_i \\ \mathbf{Z}'_{li}\varepsilon_i \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{\Delta i} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{li} \end{pmatrix}' \begin{pmatrix} \Delta\varepsilon_i \\ \alpha_i \boldsymbol{\iota}_{T-1} + \varepsilon_i \end{pmatrix}$$

which constitutes a system of the transformed and untransformed model.

System GMM as Level GMM

- While the system approach is intuitive for selecting valid instruments, it is often convenient to write the estimator in terms of the level model only:

$$E\left[\begin{pmatrix} \mathbf{Z}'_{\Delta i}\Delta\boldsymbol{\varepsilon}_i \\ \mathbf{Z}'_{li}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \end{pmatrix}\right] = E\left[\begin{pmatrix} \mathbf{Z}'_{\Delta i}\mathbf{D}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \\ \mathbf{Z}'_{li}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \end{pmatrix}\right]$$

$$= E\left[\begin{pmatrix} \mathbf{D}'\mathbf{Z}_{\Delta i} & | & \mathbf{Z}_{li} \end{pmatrix}'(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i)\right] = \mathbf{0}$$

  - We can even combine multiple model transformations, if applicable:

$$E\left[\begin{pmatrix} \mathbf{Z}'_{\bar{\Delta} i}\bar{\Delta}\boldsymbol{\varepsilon}_i \\ \mathbf{Z}'_{\overrightarrow{\Delta} i}\overrightarrow{\Delta}\boldsymbol{\varepsilon}_i \\ \mathbf{Z}'_{li}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \end{pmatrix}\right] = E\left[\begin{pmatrix} \mathbf{Z}'_{\bar{\Delta} i}\mathbf{M}_{\iota}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \\ \mathbf{Z}'_{\overrightarrow{\Delta} i}\mathbf{F}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \\ \mathbf{Z}'_{li}(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i) \end{pmatrix}\right]$$

$$= E\left[\begin{pmatrix} \mathbf{M}_{\iota}\mathbf{Z}_{\bar{\Delta} i} & | & \mathbf{F}'\mathbf{Z}_{\overrightarrow{\Delta} i} & | & \mathbf{Z}_{li} \end{pmatrix}'(\alpha_i\boldsymbol{\iota}_{T-1}+\boldsymbol{\varepsilon}_i)\right] = \mathbf{0}$$

Time Effects

- Recall that common time-specific effects can be accounted for by including a set of time dummies $d_s = \mathcal{I}(s = t)$ as regressors:

$$\mathbf{y}_i = \lambda \mathbf{y}_{i,-1} + \mathbf{X}_i \beta + \alpha_i \boldsymbol{\iota}_{T-1} + \boldsymbol{\varepsilon}_i$$

where

$$\mathbf{X}_i = \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & \mathbf{x}'_{i2} \\ 1 & 1 & & 0 & \mathbf{x}'_{i3} \\ \vdots & & \ddots & & \vdots \\ 1 & 0 & & 1 & \mathbf{x}'_{iT} \end{array} \right) \quad \text{or} \quad \mathbf{X}_i = \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & \mathbf{x}'_{i2} \\ 0 & 1 & & 0 & \mathbf{x}'_{i3} \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & \mathbf{x}'_{iT} \end{array} \right)$$

depending on whether or not an intercept – i.e., a vector of ones – is included instead of the time dummy d_2.

Time Effects

- The time dummies can be treated as uncorrelated with both error components α_i and ε_{it}. Consequently, they can be instrumented by themselves. Lagged time dummies are redundant due to their deterministic nature.
- With balanced panel data, once time dummies are instrumented for the untransformed model, they become redundant as instruments for the transformed model.
  - For example, if $T = 4$, the respective (transformed) instrument matrices (without intercept) would be

$$(\mathbf{D}'\mathbf{Z}_{\Delta,d,i}, \mathbf{Z}_{l,d,i}) = \left(\begin{array}{cc|ccc} -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

    where column 1 equals column 4 minus column 3, and column 2 equals column 5 minus column 4.
  - An equivalent redundancy result holds if the first dummy is replaced by an intercept.

Time-Invariant Regressors

- Since any of the considered model transformations is orthogonal to any variable that is constant over time, $\mathbf{K}\boldsymbol{\iota}_{T-1} = \mathbf{0}$ for $\mathbf{K} \in \{\mathbf{D}, \mathbf{F}, \mathbf{M}\}$, the effects of time-invariant regressors $\mathbf{c}_i$ can only be identified in the untransformed model:

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{c}_i'\boldsymbol{\gamma} + \alpha_i + \varepsilon_{it}$$

- If the coefficients $\boldsymbol{\gamma}$ are not of particular interest, the time-invariant regressors can simply be subsumed under the unit-specific error component: $\tilde{\alpha}_i = \mathbf{c}_i'\boldsymbol{\gamma} + \alpha_i$
- If the coefficients $\boldsymbol{\gamma}$ are the objects of interest (or if including them helps to make it more plausible that time-varying regressors $\mathbf{x}_{it}$ are uncorrelated with α_i), appropriate instruments for the level model are needed.

Time-Invariant Regressors

- In empirical research, instead of explicitly specifying strong instruments for time-invariant regressors, identification of γ is occasionally implicitly assumed through the overidentifying restrictions from the other instruments under the Blundell and Bond (1998) initial-observations assumption.
- However, if $E[\Delta y_{i,t-1}\alpha_i] = 0$ holds, it is difficult to justify that at the same time $E[\Delta y_{i,t-1}\mathbf{c}_i] \neq \mathbf{0}$, and similarly for (lagged) first differences of $\mathbf{x}_{it}$.
  - Unless such an approach can be theoretically justified (by making peculiar assumptions on the data-generating process), any estimates of γ obtained this way are driven by spurious finite-sample correlation between the instruments and the time-invariant regressors (Kripfganz and Schwarz, 2019).

Time-Invariant Regressors

- It might be appropriate to assume that $E[\mathbf{c}_i \alpha_i] = \mathbf{0}$ (and $E[\mathbf{c}_i \varepsilon_{it}] = 0$), in which case the variables $\mathbf{c}_i$ can serve as their own instruments.
- Alternatively, a Hausman and Taylor (1981) strategy could be employed. If it holds for (a subset of) time-varying regressors $\mathbf{x}_{it}$ that $E[\mathbf{x}_{it} \alpha_i] = 0$ and $E[\mathbf{x}_{it} \mathbf{c}_i']$ is of full rank, then $\mathbf{x}_{it}$ can serve as instruments for $\mathbf{c}_i$.
- Any other omitted variables satisfying valid exclusion restrictions could potentially serve as instruments as well.

Time-Invariant Regressors

- As another alternative, if $E[\mathbf{c}_i \alpha_i] \neq \mathbf{0}$, it might be reasonable to assume that $E[\mathbf{c}_i \tilde{\alpha}_i] = \mathbf{0}$ after including within-group averages $\bar{\mathbf{x}}_i$ as additional regressors in the spirit of the "correlated random-effects" approach proposed by Mundlak (1978):

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}'_{it}\beta + \mathbf{c}'_i\gamma + \underbrace{\bar{\mathbf{x}}'_i\phi + \tilde{\alpha}_i}_{\alpha_i} + \varepsilon_{it}$$

  - This requires that $\mathbf{x}_{it}$ are strictly exogenous. If they are predetermined, $\bar{\mathbf{x}}_i$ could be replaced by the initial observations $\mathbf{x}_{i1}$ (and possibly also y_{i1}), as noted by Kripfganz and Schwarz (2019).

Time-Invariant Regressors

- The identifying assumptions $E[\mathbf{c}_i\alpha_i] = \mathbf{0}$, $E[\mathbf{c}_i\tilde{\alpha}_i] = \mathbf{0}$, or $E[\mathbf{x}_{it}\alpha_i] = 0$ cannot be tested (in their entirety).
  - If there are more relevant instruments – e.g., because $K_x > K_c$ under the Hausman and Taylor (1981) approach – it is possible to test at least the overidentifying restrictions in the usual way.
  - If the coefficients γ are just-identified – e.g., under the assumption $E[\mathbf{c}_i\alpha_i] = \mathbf{0}$ – an incremental overidentification test comparing estimators with and without the instruments $\mathbf{c}_i$ is not helpful, because the coefficients γ are not identified without those instruments; see discussion above.
- If the coefficients γ are overidentified, incorrect exogeneity assumptions about the added instruments can cause inconsistency not just of the estimator for γ but also for λ and β.

Time-Invariant Regressors

- As an alternative to estimating all coefficients in a single stage, Kripfganz and Schwarz (2019) propose a two-stage procedure:
  1. Estimate the coefficients λ and β with any consistent estimator (BC-MM, QML, GMM) from
  $$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}'\beta + \tilde{\alpha}_i + u_{it}$$
  where $\tilde{\alpha}_i = \mathbf{c}_i'\gamma + \alpha_i$
  2. Estimate the coefficients γ from
  $$y_{it} - \hat{\lambda} y_{i,t-1} - \mathbf{x}_{it}'\hat{\beta} = \mathbf{c}_i'\gamma + \alpha_i + \zeta_{it}(\hat{\lambda}, \hat{\beta})$$
  Because the errors $\zeta_{it}(\hat{\lambda}, \hat{\beta}) = \varepsilon_{it} - (\hat{\lambda} - \lambda) y_{i,t-1} - \mathbf{x}_{it}'(\hat{\beta} - \beta)$ are a function of the first-stage estimation uncertainty, standard errors need to be corrected accordingly.
  - The two-stage approach is generally less efficient than a single-stage GMM estimator, but the first stage is robust to misspecification at the second stage.

Weighting Matrix

- Under the classical error components structure with serially uncorrelated ε_{it} and homoskedasticity of both ε_{it} and α_i, an optimal weighting matrix would be a function of the unknown variance ratio $\tau = \sigma_\alpha^2/\sigma_u^2$:

$$\mathbf{W} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' (\tau \boldsymbol{\iota}_{T-1} \boldsymbol{\iota}_{T-1}' + \mathbf{I}_{T-1}) \mathbf{Z}_i \right)^{-1}$$

  where $\mathbf{Z}_i$ is the matrix of all (transformed) instruments – e.g., $\mathbf{Z}_i = (\mathbf{D}'\mathbf{Z}_{\Delta i}, \mathbf{Z}_{li})$.
  - Efficient one-step GMM estimation is infeasible, unless all moment conditions refer to the transformed model (because $\mathbf{D}_i \boldsymbol{\iota}_{T-1} = \mathbf{0}$) or τ is known.
- An optimal weighting matrix $\mathbf{W}(\hat{\theta}_1) = (\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i{}' \hat{\varepsilon}_i(\hat{\theta}_1) \hat{\varepsilon}_i(\hat{\theta}_1)' \mathbf{Z}_i)^{-1}$ requires a preliminary consistent estimator $\hat{\theta}_1$.

Weighting Matrix

- The leading candidate for an initial weighting matrix is

$$\mathbf{W} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1} = \left(\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_{\Delta i}{}'\mathbf{D}_i\mathbf{D}_i'\mathbf{Z}_{\Delta i} & \mathbf{Z}_{\Delta i}{}'\mathbf{D}_i\mathbf{Z}_{li} \\ \mathbf{Z}_{li}{}'\mathbf{D}_i'\mathbf{Z}_{\Delta i} & \mathbf{Z}_{li}{}'\mathbf{Z}_{li}\end{pmatrix}\right)^{-1}$$

  which leads to 2SLS estimation and is optimal when $\sigma_\alpha^2 = 0$ (Windmeijer, 2000).

  - Alternatively, Blundell, Bond and Windmeijer (2001) suggested

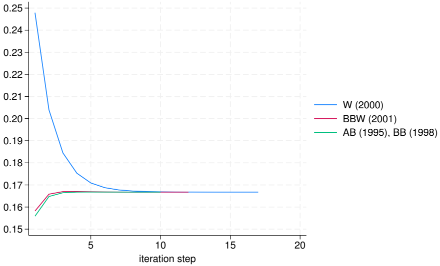
$$\mathbf{W} = \left(\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_{\Delta i}{}'\mathbf{D}_i\mathbf{D}_i'\mathbf{Z}_{\Delta i} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{li}{}'\mathbf{Z}_{li}\end{pmatrix}\right)^{-1}$$

    while Arellano and Bover (1995) and Blundell and Bond (1998) proposed

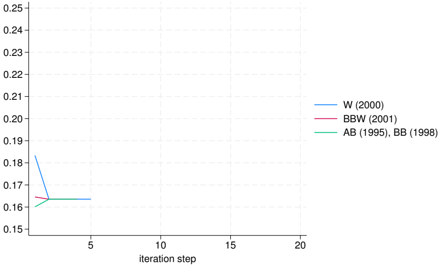
$$\mathbf{W} = \left(\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_{\Delta i}{}'\mathbf{Z}_{\Delta i} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{li}{}'\mathbf{Z}_{li}\end{pmatrix}\right)^{-1}$$

Weighting Matrix

- The iterated GMM estimator avoids the lack of finite-sample robustness to the (arbitrary) choice of the initial weighting matrix.
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \in \{-1,0,1\}$, and $\lambda = 0.2$; $T = 5$



(a) $N = 60$



(b) $N = 600$

Specification Test

- As mentioned earlier, depending on the particular application, the validity of the assumption $E[\Delta y_{i2} \alpha_i] = 0$ might be contested. If there is no clear guidance from economic theory, a statistical test is desirable.
- As noted by Blundell and Bond (1998), the $T - 3$ nonlinear moment conditions obtained earlier become redundant once those additional $T - 2$ instruments for the level model are introduced.
  - Thus, there is 1 overidentifying restriction due to the Blundell and Bond (1998) initial-observations assumption, which can be tested with an incremental overidentification test or a generalized Hausman test
  - To assess this assumption, it is (unfortunately) common practice to contrast the "system GMM" estimator with a "difference GMM" estimator (without nonlinear moment functions). A test based on this comparison has lower power to detect a violation of the initial-observations assumption. Instead, nonlinear moment functions should be included in the baseline estimator (Magazzini and Calzolari, 2020).

Interim Conclusion

- The recursive nature of the model provides a potentially large number of internal instruments.
  - Their validity relies on the absence of (higher-order) serial correlation. Testing is essential.
- Too many (weak) instruments hamper the reliability of the estimator. Unless T is very small (relative to N), instrument reduction techniques should be employed.
  - Collapsing seems preferable, possibly combined with curtailing when T becomes relatively large.
- To address the concern of weak instruments when the process is very persistent, additional nonlinear moment conditions can be useful.
- The popular "system GMM" estimator adds further comparatively strong instruments. However, it is often not straightforward to justify the required initial-observations assumption.
- Correctly classifying all of the regressors as strictly exogenous, predetermined, or endogenous is another crucial task.