(Short-T) Dynamic Panel Data Analysis in Stata

Sebastian Kripfganz

University of Exeter Business School, Department of Economics, Exeter, UK
www.ipdc2026.org / www.kripfganz.de

31st International Panel Data Conference
University of Exeter Business School, Research Methods Centre
July 8, 2026

Dynamic Models (Lagged Dependent Variable)

- This part of the workshop considers panel data settings with N subjects and T time periods, where $T \ll N$ ("short-T panels" or "micro panels").
- Linear model with state dependence (lagged dependent variable):

$$E[y_{it} | y_{i,t-1}, \mathbf{X}_i; \alpha_i] = \lambda y_{i,t-1} + \mathbf{x}_{it}'\beta + \alpha_i$$

with regression analogue

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}'\beta + \underbrace{\alpha_i + \varepsilon_{it}}_{\text{error components}}$$

where α_i is a time-invariant unobserved subject-specific effect and ε_{it} is an idiosyncratic error term.
- The lagged dependent variable $y_{i,t-1}$ can be motivated by habit formation and other reasons for partial-adjustment processes.

Dynamic Models (Lagged Dependent Variable)

- Reparameterization of the regression model in error correction form:

$$y_{it} - y_{i,t-1} = \underbrace{(\lambda - 1)}_{\substack{\text{(negative)}\\ \text{speed of}\\ \text{adjustment}}} \underbrace{\left(y_{i,t-1} - \mathbf{x}'_{i,t-1} \frac{\boldsymbol{\beta}}{1-\lambda} - \frac{\alpha_i}{1-\lambda} \right)}_{\substack{\text{deviation from long-run equilibrium}\\ E[y_{it}|\mathbf{X}_i;\alpha_i]=\mathbf{x}'_{it}\frac{\boldsymbol{\beta}}{1-\lambda}+\frac{\alpha_i}{1-\lambda}}} + (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + \varepsilon_{it}$$

- Objects of interest:
  - Short-run effects:

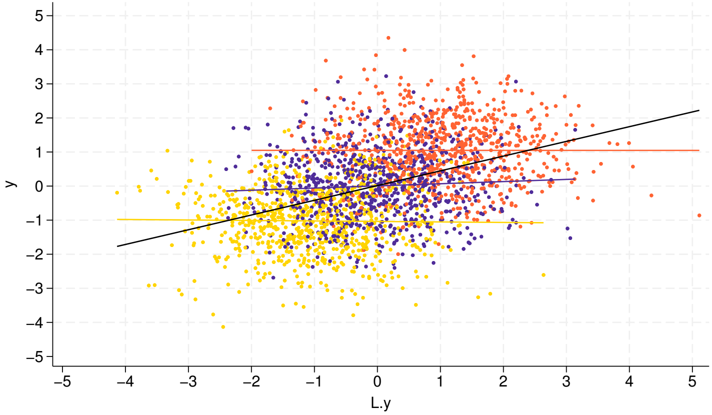
$$\frac{\partial E[y_{it}|y_{i,t-1}, \mathbf{X}_i; \alpha_i]}{\partial \mathbf{x}_{it}} = \boldsymbol{\beta}$$

  - Long-run effects:

$$\frac{\partial E[y_{it}|\mathbf{X}_i; \alpha_i]}{\partial \mathbf{x}_{it}} = \frac{\boldsymbol{\beta}}{1-\lambda}$$

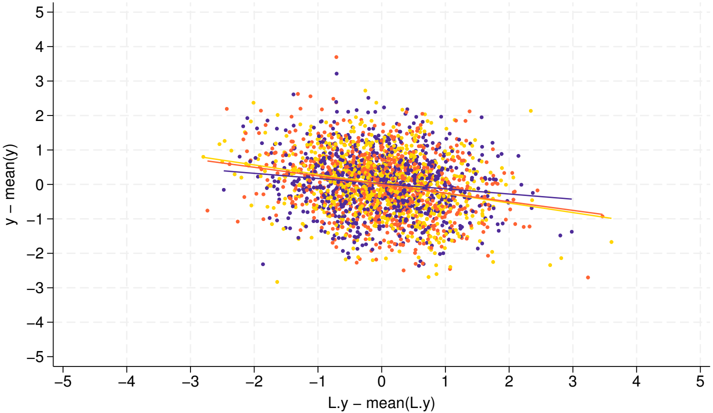
Dynamic Models (Lagged Dependent Variable)

- Simulated data, where $E[y_{it}|y_{i,t-1}] = E[y_{it}] = 0$ (i.e., $\lambda = 0$):
  - $y_{it} = \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0,1)$ and $\alpha_i \in \{-1, 0, 1\}$
  - $N = 600,\ T = 5$



Dynamic Models (Lagged Dependent Variable)

- Re-centering of y_{it} and $y_{i,t-1}$:
  - Deviations from subject-specific means, $y_{it} - \bar{y}_i$ and $y_{i,t-1} - \bar{y}_{i,-1}$, where $\bar{y}_i = \frac{1}{T-1} \sum_{s=2}^{T} y_{is}$ and $\bar{y}_{i,-1} = \frac{1}{T-1} \sum_{s=2}^{T} y_{i,s-1}$



Dynamic Models (Lagged Dependent Variable)

- The lagged dependent variable $y_{i,t-1}$ is correlated with the subject-specific error component α_i by construction of the model:

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

  - $y_{i,t-1}$ is not strictly exogenous, but only sequentially exogenous/weakly exogenous/predetermined – i.e., $E[\varepsilon_{it}|y_{i1}, y_{i2}, \dots, y_{i,t-1}, \mathbf{X}_i; \alpha_i] = 0$

Dynamic Models (Lagged Dependent Variable)

- The fixed-effects (FE) estimator does not successfully eliminate the bias/inconsistency (Nickell, 1981) because $\bar{\Delta} y_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$ is correlated with $\bar{\Delta}\varepsilon_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$ (when T is small) in the regression

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

- The first-difference estimator is biased/inconsistent as well because $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ is correlated with the transformed error term $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$ (for any T) in the regression

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

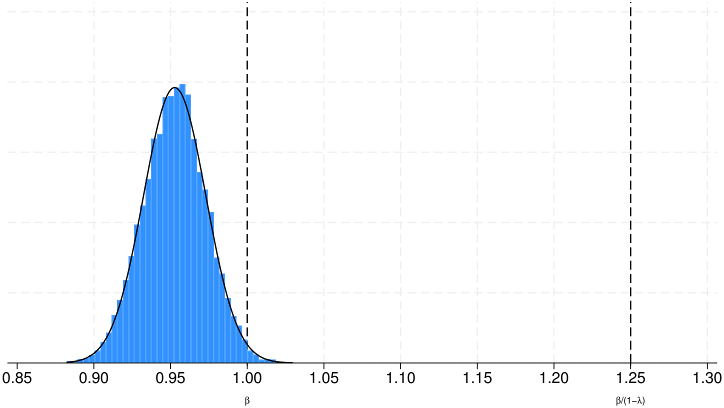
- Similar problems occur also without a lagged dependent variable when $\mathbf{x}_{it}$ violates the strict-exogeneity assumption.

Omitted Dynamics

- Given the difficulties of estimating a (short-T) dynamic panel data model, applied researchers often decide to stick to a static panel data model instead. However, estimating
$$y_{it} = \mathbf{x}_{it}'\mathbf{b} + a_i + e_{it}$$
with the FE estimator $\hat{\mathbf{b}}_{FE}$ generally yields biased estimates of the short-run effects $\frac{\partial E[y_{it}|y_{i,t-1},\mathbf{X}_i;\alpha_i]}{\partial \mathbf{x}_{it}} = \beta$ when the true data-generating process is dynamic.
  - Even worse, many researchers using a static model specification implicitly aim to estimate long-run "equilibrium" effects $\frac{\partial E[y_{it}|\mathbf{X}_i;\alpha_i]}{\partial \mathbf{x}_{it}} = \frac{\beta}{1-\lambda}$.

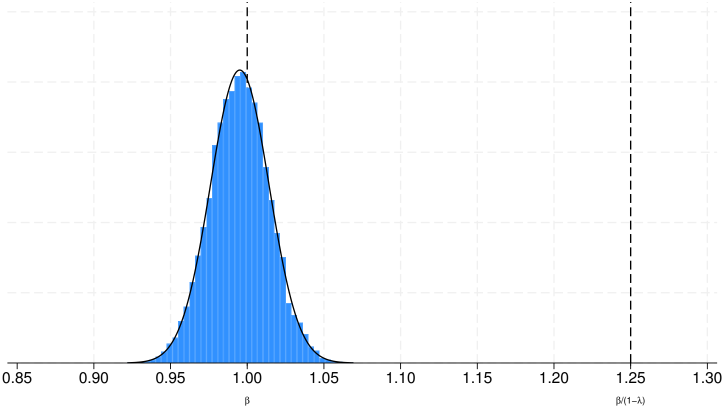
Omitted Dynamics

- Static FE estimator based on 10,000 replications:
  - $y_{it} = \lambda y_{i,t-1} + \beta x_{it} + \alpha_i + \varepsilon_{it}$, where $x_{it} \sim \mathcal{N}(0,1)$, $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \in \{-1,0,1\}$, $\lambda = 0.2$, and $\beta = 1$; $N = 600$, $T = 5$



Omitted Dynamics

- Static FE estimator based on 10,000 replications:
  - $y_{it} = \lambda y_{i,t-1} + \beta x_{it} + \alpha_i + \varepsilon_{it}$, where $x_{it} \sim \mathcal{N}(0,1)$, $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \in \{-1,0,1\}$, $\lambda = 0.2$, and $\beta = 1$; $N = 60$, $T = 50$



Estimation Methods

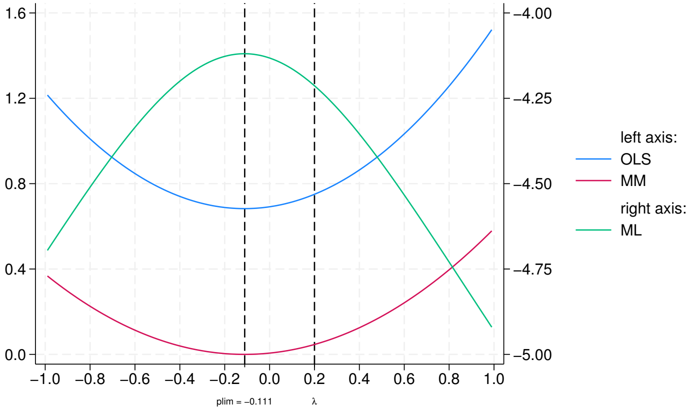
- Let us initially consider a simple autoregressive panel model – panel AR(1) – with error components structure:

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

- There are three leading estimation principles:
  1. Ordinary least-squares (OLS) estimation minimizes the sum of squared residuals (after adding dummy variables for the fixed effects or transforming all variables into deviations from within-group averages).
  2. Method of moments (MM) estimation is based on the fact that some population moments of the data shall be zero – e.g., the first-order condition $E[m_{FE,i}(\lambda)] = 0$ from the FE estimator – and then minimizes the sum of the squared sample moments.
  3. Maximum likelihood (ML) estimation finds parameter values that maximize the joint density of the residuals given the data. The first-order condition is the same as the one from the OLS estimation.

Equivalence of Estimation Methods

- FE objective functions for the pooled OLS, MM, and ML methods:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.2$



Estimation Methods

- Given that $y_{i,t-1}$ is predetermined with respect to ε_{it} and correlated with α_i, the following approaches are commonly applied to obtain consistent estimates:
  - (Quasi-)maximum likelihood (QML) estimation, making further assumptions about the initial observations.
  - Bias-corrected (BC) estimation, utilizing the known bias expression for the FE estimator by either correcting the estimator itself or by adjusting the first-order condition.
  - Instrumental variables (IV) and generalized method of moments (GMM) estimation, exploiting knowledge about the model dynamics to construct valid instruments.
- In the following, we illustrate the methods with some toy examples using the Arellano and Bond (1991) data set.

```
. webuse abdata
```

Quasi-Maximum Likelihood Estimation

- The inconsistency/bias of the uncorrected maximum likelihood estimator results from conditioning on the initial observations of the dependent variable – i.e., treating those initial observations as exogenous even though they are correlated with the transformed error term (in deviations from within-group means or in first differences).
- To solve the problem, we can maximize the unconditional (log-)likelihood function by incorporating the marginal density of the initial observations.
- In Stata, this is done by the `xtdpdqml` command (Kripfganz, 2016).
- A fixed-effects (Hsiao, Pesaran, and Tahmiscioglu, 2002) and random-effects (Bhargava and Sargan, 1983) version are available.
- Occasionally, the default initial values for the variance parameters are incompatible with the data. Alternative initial values must be specified in such cases.

```
. xtdpdqml n w k yr1978-yr1984, fe vce(robust)
. xtdpdqml n w k yr1978-yr1984, re vce(robust) initval(.1 .2 .2 .3)
```

Quasi-Maximum Likelihood Estimation

- A generalized Hausman (1978) test for FE versus RE can be implemented, although this is not directly built into the `xtdpdqml` package but requires a manual implementation:

```
. xtdpdqml n w k yr1978-yr1984, mlparams
. estimates store fe
. xtdpdqml n w k yr1978-yr1984, re initval(.1 .2 .2 .3) mlparams
. estimates store re
. suest fe re, vce(cluster id)
. test ([fe__model]LD.n = [re__model]L.n) ([fe__model]D.w = [re__model]w)
> ([fe__model]D.k = [re__model]k)
```

Dynamic Panel Bias

- Nickell (1981) obtained an expression for the downward bias/inconsistency of the FE estimator:

$$\operatorname*{plim}_{N\to\infty} (\hat{\lambda}_{FE} - \lambda) = - \frac{\frac{1+\lambda}{T-2}\left(1 - \frac{1-\lambda^{T-1}}{(1-\lambda)(T-1)}\right)}{1 - \frac{2\lambda}{(1-\lambda)(T-2)}\left(1 - \frac{1-\lambda^{T-1}}{(1-\lambda)(T-1)}\right)}$$

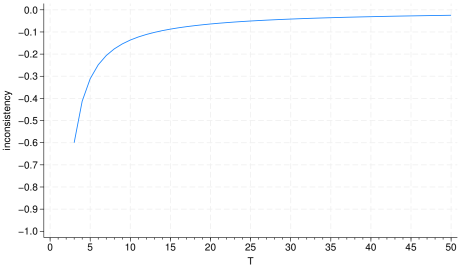
  - A simple approximation is

$$\operatorname*{plim}_{N\to\infty} (\hat{\lambda}_{FE} - \lambda) \approx - \frac{1+\lambda}{T-1}$$

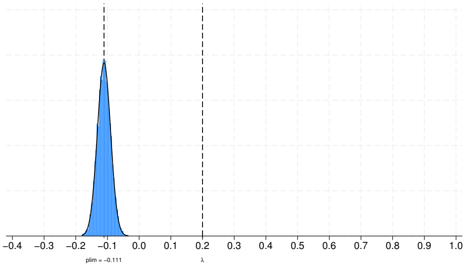
  - Note that $\text{plim}_{N\to\infty}(\hat{\lambda}_{FE} - \lambda) \to 0$ as $T \to \infty$.

Dynamic Panel Bias

- Inconsistency/bias of the dynamic FE estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.2$ (low persistence)



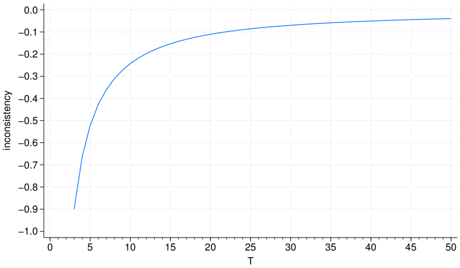
(a) inconsistency for varying T



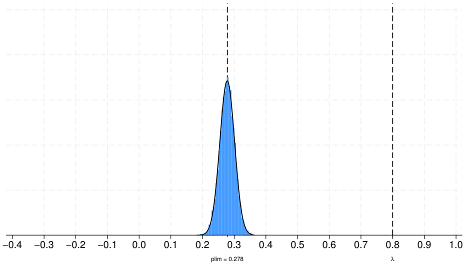
(b) bias for $N = 600$, $T = 5$; $\varepsilon_{it} \sim \mathcal{N}(0,1)$; 10,000 replications

Dynamic Panel Bias

- Inconsistency/bias of the dynamic FE estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.8$ (high persistence)



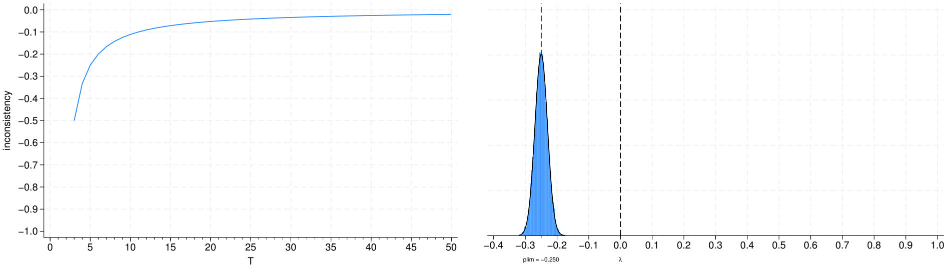
(a) inconsistency for varying T



(b) bias for $N = 600$, $T = 5$; $\varepsilon_{it} \sim \mathcal{N}(0, 1)$; 10,000 replications

Dynamic Panel Bias

- Inconsistency/bias of the dynamic FE estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0$ (no persistence)



(a) inconsistency for varying T

(b) bias for $N = 600$, $T = 5$; $\varepsilon_{it} \sim \mathcal{N}(0,1)$; 10,000 replications

Additive Bias Correction

- Based on the closed-form expression

$$b_T(\lambda) = \operatorname*{plim}_{N\to\infty} \hat{\lambda}_{FE} - \lambda$$

an additive bias correction can be applied to the FE estimator:

$$\hat{\lambda}_{aBCFE} = \hat{\lambda}_{FE} - b_T(\hat{\lambda}_0)$$

with a bias estimate obtained from an initial consistent estimator $\hat{\lambda}_0$.

  - Kiviet (1995) operationalized this idea based on a refined bias approximation, also allowing for additional strictly exogenous regressors $\mathbf{x}_{it}$.
  - A downside of this approach is its reliance on an initial consistent estimator, and the necessity to obtain standard errors with bootstrap methods.
  - In Stata, this is implemented in the command `xtlsdvc` (Bruno, 2005).

Iterative Bias Correction

- To avoid reliance on an initial consistent estimator, we can consider the iterations

$$\hat{\lambda}_j = \hat{\lambda}_{FE} - b_T(\hat{\lambda}_{j-1})$$

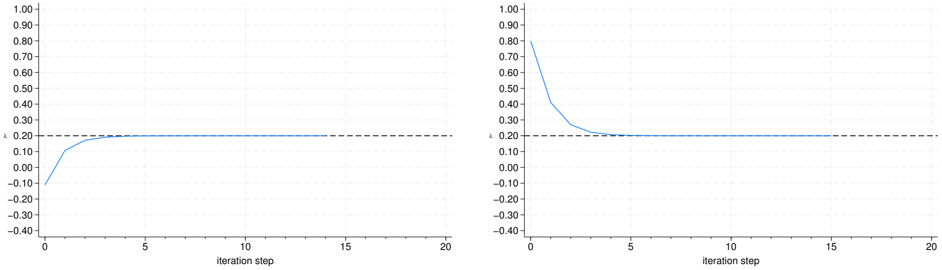
  starting with an initial guess $\hat{\lambda}_0 \in (-1, 1)$, which could be $\hat{\lambda}_{FE}$ or any other reasonable value.
- The iterative bias-corrected estimator is the fixed point of this updating rule obtained upon convergence:

$$\hat{\lambda}_{iBCFE} = \hat{\lambda}_{FE} - b_T(\hat{\lambda}_{iBCFE})$$

  - The estimator proposed by Bun and Carree (2005) is based on this idea, again extended to accommodate strictly exogenous regressors $\mathbf{x}_{it}$.

Iterative Bias Correction

- Numerical illustration (proof of concept) for the iterative bias correction, where $\hat{\lambda}_{FE}$ is replaced by its probability limit instead of using actual data.
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.2$; $T = 5$
  - $\hat{\lambda}_j = (\text{plim}_{N\to\infty} \hat{\lambda}_{FE}) - b_T(\hat{\lambda}_{j-1})$
  - Convergence criterion: $|\hat{\lambda}_j - \hat{\lambda}_{j-1}| < 10^{-7}$



(a) $\hat{\lambda}_0 = \text{plim}_{N\to\infty} \hat{\lambda}_{FE}$ (b) $\hat{\lambda}_0 = 0.8$

Bias-Corrected Method of Moments Estimation

- The bias of the FE estimator manifests itself in a violation of the first-order condition:

$$E[m_{FE,i}(\lambda)] = d_T(\lambda, \sigma_\varepsilon^2) = -\frac{\sigma_\varepsilon^2}{(1-\lambda)(T-1)} \left(1 - \frac{1-\lambda^{T-1}}{(1-\lambda)(T-1)}\right)$$

- An alternative approach is to directly correct the moment functions:

$$m_{BCFE,i}(\lambda, \sigma_\varepsilon^2) = m_{FE,i}(\lambda) - d_T(\lambda, \sigma_\varepsilon^2)$$

such that the moment condition $E[m_{BCFE,i}(\lambda)] = 0$ holds again.

Bias-Corrected Method of Moments Estimation

- The bias-corrected MM estimator of Breitung, Kripfganz, and Hayakawa (2022) then solves

$$\hat{\lambda}_{BCMM} = \arg\min_{\hat{\lambda}} \left(\frac{1}{N} \sum_{i=1}^{N} m_{BCFE,i}(\hat{\lambda}, \hat{\sigma}_{\varepsilon}^2(\hat{\lambda})) \right)^2$$

  - The estimator is equivalent to the iterative bias-corrected estimator of Bun and Carree (2005) and the adjusted profile likelihood estimator of Dhaene and Jochmans (2016).
  - The bias formula can be adjusted for higher-order autoregressive models:

$$y_{it} = \sum_{j=1}^{p} \lambda_j y_{i,t-j} + \alpha_i + \varepsilon_{it}$$

  - Strictly exogenous regressors $\mathbf{x}_{it}$ can be incorporated as well.
  - Standard errors are easy to compute using well established asymptotic results for method of moments estimators.

Bias-corrected fixed-effects estimator

- The bias-corrected MM estimator is implemented in the `xtdpdbc` Stata package (Kripfganz and Breitung, 2022):

```
. xtdpdbc n w k, fe vce(robust)
. estimates store fe
```

- The command supports higher-order autoregressive models and provides a simple option to add time dummies:

```
. xtdpdbc n w k, fe vce(robust) lags(2) teffects
```

Bias-corrected random-effects estimator

- A random-effects (RE) version of the bias-corrected method of moments estimator can be used when the regressors $\mathbf{x}_{it}$ are assumed to be uncorrelated with the unobserved effects α_i:

```
. xtdpdbc n w k, re vce(robust)
```

- Because the RE estimator simply adds additional moment conditions, we can then test for RE versus FE with a Hansen (1982) overidentification test or a generalized version of the Hausman (1978) test:

```
. estat overid
. estat hausman fe
```

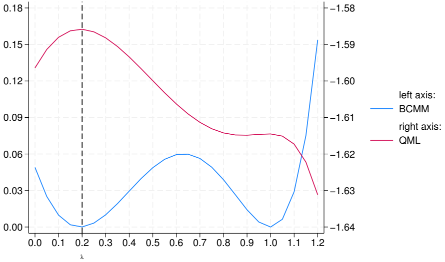
Bias-corrected random-effects estimator

- With time effects, we need to be careful with the degrees of freedom for the specification tests. With unbalanced panel data, extra moment conditions are added for the time effects under the RE assumption. However, they are asymptotically redundant, which needs to be accounted for when comparing the FE and RE estimates.

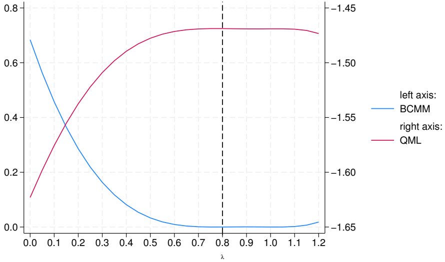
```
. xtdpdbc n w k, fe vce(robust) teffects
. estimates store fe
. xtdpdbc n w k, re vce(robust) teffects
. estat overid
. estat hausman fe
. estat hausman fe, df(2)
. estat hausman fe (L.n w k), df(2)
```

Multiple local optima and flat objective functions

- Objective functions of the bias-corrected MM estimator and the QML estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$; $T = 5$



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



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

Multiple local optima and flat objective functions

- Due to the nonlinearity of the bias-corrected moment function, there are multiple solutions and the numerical algorithm occasionally converges to an incorrect solution. At the correct solution, all eigenvalues of the gradient should be negative. The command `xtdpdbc` automatically re-initializes the optimization with alternative initial parameter values – random draws from the uniform distribution for λ – until a correct solution is found.

```
. xtdpdbc n w k, fe vce(robust) from(0.99 0 0 0, copy)
```

Multiple local optima and flat objective functions

- When the objective function is almost flat in a neighborhood of the correct solution, the maximum eigenvalue is close to zero as well. In such cases, the numerical algorithm might converge to a solution with marginally positive eigenvalue in finite samples. We can increase the tolerance for the maximum eigenvalue by which it is allowed to exceed 0 for declaring convergence, although the results need to be treated with caution.

```
. xtdpdbc n w k, re vce(robust) lags(2) teffects
. xtdpdbc n w k, re vce(robust) lags(2) teffects eigtolerance(0.1)
```

Convergence problems

- Especially when there are many exogenous regressors, the computational burden can be reduced by concentrating out their coefficients β from the objective function by using the available closed-form solutions for them, and only optimize with respect to the autoregressive parameter λ. The resulting estimates will be the same.

```
. xtdpdbc n w k, fe vce(robust) concentration
```

- In some cases, formal convergence might not be declared when the objective function is virtually flat. It might help to turn off the convergence criterion for the scaled gradient, although the estimation method might just not be reliable in such situations.

```
. xtdpdbc n L(0/2).(w k), fe vce(robust) lags(3)
. xtdpdbc n L(0/2).(w k), fe vce(robust) lags(3) nonrtolerance
```

Generalized Method of Moments Estimation

- In many applications, one wants to allow for the regressors $\mathbf{x}_{it}$ to be predetermined or endogenous – i.e., correlated with past (and contemporaneous) shocks (e.g., dynamic or simultaneous feedback from the outcome variable). In this case, GMM estimators with a suitable choice of instruments are predominantly used.
- Stata's `xtdpd` command suite (with the specialized commands `xtabond` and `xtdpdsys`) can be used but offers limited flexibility. The `gmm` command is more flexible but less user-friendly and still has limitations.
- The `xtabond2` command (Roodman, 2009b) is very popular, but users need to be careful when relying on its default settings.
- In the following, I will focus on the `xtdpdgmm` package (Kripfganz, 2019) which provides the largest flexibility.

Instruments in first-differenced model

- Recall the first-differenced model:

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

- Under the assumption that ε_{it} is serially uncorrelated, lags of the dependent variable $y_{i,t-2}, y_{i,t-3}, \ldots$ are valid instruments for $\Delta y_{i,t-1}$ (Anderson and Hsiao, 1981; Arellano and Bond, 1991). Equivalently, the following moment conditions can be exploited:

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

- Instead of using standard instruments, a much larger number of "GMM-type" instruments can be obtained by interacting them with a full set of time dummies.
- Intuitively, this creates separate instruments for each time period, while standard instruments impose homogeneity of the first-stage coefficients across time periods.

(Too) many GMM-type instruments

- There are $\frac{(T-1)(T-2)}{2}$ of these GMM-type instruments for the lagged dependent variable. Their number thus increases quadratically in T, which can quickly lead to first-stage overfitting. Consequences include biased coefficient and standard error estimates, and weakened specification tests (Roodman, 2009a).
- Remedies include "collapsing" into standard IVs and/or "curtailing" of the maximum lag order for the IVs (Kiviet, 2020).

Instruments for other regressors

- For the regressors $\mathbf{x}_{it}$, standard or GMM-type instruments are constructed similarly depending on their classification:
  - Strictly exogenous regressors satisfy $E[\mathbf{x}_{i,t-s}\Delta\varepsilon_{it}] = 0$ for all s (but in practice usually restricted to $s \geq 0$), leading to a large number of potential instruments.
  - Predetermined regressors satisfy $E[\mathbf{x}_{i,t-s}\Delta\varepsilon_{it}] = 0$ for $s \geq 1$. In other words, valid instruments are $\mathbf{x}_{i,t-1}, \mathbf{x}_{i,t-2}, \ldots$.
  - Endogenous regressors satisfy $E[\mathbf{x}_{i,t-s}\Delta\varepsilon_{it}] = 0$ for $s \geq 2$, implying one instrument less per variable (and time period) than for predetermined regressors. Valid instruments are $\mathbf{x}_{i,t-2}, \mathbf{x}_{i,t-3}, \ldots$.

Two-step, iterated, and continuously-updating GMM estimation

- Due to the large number of instruments, the model is (highly) overidentified. The two-stage least squares (2SLS) estimator is generally inefficient. An efficient "two-step GMM" estimator can be obtained as follows:
  1. Estimate the model by 2SLS or GMM with an inefficient weighting matrix.
  2. Re-estimate the model by GMM with an "optimal" weighting matrix constructed from the residuals of the initial step.
- An asymptotically equivalent "iterated GMM" estimator iterates this procedure until convergence of the weighting matrix (Hansen, Heaton, and Yaron, 1996). This yields finite-sample robustness to the choice of the initial weighting matrix.
- Alternatively, the "continuously-updating GMM" estimator obtains the optimal weighting matrix simultaneously with the coefficient estimates (Hansen, Heaton, and Yaron, 1996). This can be computationally more demanding.
- Standard errors need to be corrected (in finite samples) to account for the first-step estimation error in the optimal weighting matrix (Windmeijer, 2005; Hwang, Kang, and Lee, 2022; Hansen and Lee, 2021).

Difference GMM estimation in Stata

- The so-called "difference GMM" estimator can be implemented in Stata with the `xtdpdgmm` command, here exemplarily with collapsing and curtailing (maximum lag order 4):

```
. xtdpdgmm n L.n w k, model(diff) collapse gmmiv(L.n, lagrange(1 4))
> gmmiv(w, lagrange(2 4)) gmmiv(k, lagrange(0 4)) teffects nolevel
> twostep vce(robust)
```

- An equivalent syntax using the convenience tool `xtdpdgmmfe` is:

```
. xtdpdgmmfe n w k, lags(1) endogenous(w) exogenous(k) serial(0)
> curtail(4) collapse teffects nonl twostep vce(robust)
```

- Iterated and continuously-updating GMM estimators are easily obtained by replacing option `twostep` with either `igmm` or `cugmm`.

Overidentification tests

- While the joint validity of all instruments is untestable, we can test the overidentifying restrictions implied by the additional instruments after assuming that as many instruments are valid as there are regressors (Hansen, 1982).
- A rejection might be indicative of various model misspecifications:
  - The classification of regressors into strictly exogenous, predetermined, or endogenous variables might be incorrect.
  - The dynamic structure of the model might be incomplete, causing the idiosyncratic errors ε_{it} to be serially correlated.
  - Relevant explanatory variables might be omitted.

```
. estat overid
```

Overidentification tests

- Overidentification tests with too many instruments might incorrectly signal validity of the overidentifying restrictions when in fact the model is misspecified.
- Instead of (or in addition to) testing all overidentifying restrictions jointly, we can test subsets with incremental overidentification tests (Eichenbaum, Hansen, and Singleton, 1988).
- These incremental tests are especially useful to assess the regressor classification (Kiviet, 2020).

```
. xtdpdgmm n L.n w k, model(diff) collapse gmmiv(L.n, lagrange(1 4))
> gmmiv(w, lagrange(2 4)) gmmiv(k, lagrange(1 4)) gmmiv(k, lagrange(0 0))
> teffects nolevel twostep vce(robust) overid
. estat overid, difference
. estimates store dgmm
```

Overidentification tests

- Effectively, the incremental overidentification tests compare two models with and without the suspicious instruments; a significant difference indicates invalidity, similar (and asymptotically equivalent) to a Hausman-type specification test.

```
. xtdpdgmm n L.n w k, model(diff) collapse gmmiv(L.n, lagrange(1 4))
> gmmiv(w, lagrange(2 4)) gmmiv(k, lagrange(1 4)) teffects nolevel
> twostep vce(robust)
. estat overid
. estat overid dgmm
. estat hausman dgmm
```

Serial-correlation tests

- A key assumption is absence of serial correlation in the idiosyncratic error component ε_{it}.
- Widely applied tests check for serial correlation (of order 2 or higher) in the first-differenced errors $\Delta\varepsilon_{it}$ (Arellano and Bond, 1991; Yamagata, 2008). Those tests have low power against some alternatives, especially if there is high autocorrelation near to a random walk.

```
. estat serial
```

- The new Stata command `xtdpdserial` (Kripfganz, 2024) implements various variants of serial-correlation tests using a unified framework. It works as a postestimation command and can also be used after `xtdpdbc`.

```
. xtdpdserial, statistics(dc(2 2) d)
```

Serial-correlation tests

- A recently proposed portmanteau test has superior power when T is very small (Jochmans, 2020), but suffers from a curse of dimensionality when T is growing (similar to the too-many-instruments problem). This can bite even for moderately small T when N is not very large.
- New tests based on longer/seasonal differences (with collapsing) outperform the existing tests under most scenarios, and they retain power when T increases or the shocks are very persistent (Kripfganz, Hosseinkouchack, and Demetrescu, 2026).

```
. xtdpdserial, statistics(pm sdc)
. xtdpdserial
```

Dynamically complete model

- Besides incorrect regressor classification as exogenous, predetermined, or endogenous, a common cause of serial correlation – and consquently invalid instruments – is dynamic model misspecification. To make the model dynamically complete, it might help to add (higher-order) lags as regressors.

```
. xtdpdgmm n L(1/2).n L(0/1).(w k), model(diff) collapse
> gmmiv(L.n, lagrange(1 4)) gmmiv(w k, lagrange(2 4)) teffects nolevel
> twostep vce(robust)
. estat overid
. xtdpdserial
```

- In fact, it is recommended to start with the least-restrictive model and to gradually add stronger assumptions if they are supported by the data (and economic theory).

Identification failure

- Difference GMM estimators can suffer from weak instruments if the first-differenced regressors are poorly correlated with their past levels, resulting in biases and large standard errors.
- In the extreme case, identification might break down. While not routinely applied in the empirical practice yet, underidentification tests can shed light on this issue (Sanderson and Windmeijer, 2016; Windmeijer, 2024). In Stata, they are implemented in the `underid` command (Schaffer and Windmeijer, 2020).

```
. underid, jgmm2s sw
```

Instruments for the untransformed level model

- A potential cause of weak identification can be a high persistence of the dependent variable (or a regressor). Additional instruments might re-establish identification strength.
- Such instruments can be found in the original model in levels:

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

- Under the additional assumption of joint mean stationarity of y_{it} and $\mathbf{x}_{it}$, (lagged) first differences $\Delta y_{i,t-1}$ and $\Delta \mathbf{x}_{i,t(-1)}$ qualify as valid instruments, maintaining the assumption that ε_{it} is serially uncorrelated (Blundell and Bond, 1998).
- Mean stationarity implies that $\Delta y_{i,t-1}$ and $\Delta \mathbf{x}_{i,t(-1)}$ are uncorrelated with α_i, and therefore

$$E[\Delta y_{i,t-1}(\alpha_i + \varepsilon_{it})] = 0, \quad E[\Delta \mathbf{x}_{i,t(-1)}(\alpha_i + \varepsilon_{it})] = 0$$

Instruments for the untransformed level model

- Higher-order lags for these instruments are usually not included because they are redundant when combined with all GMM-type instruments for the first-differenced model.
- For strictly exogenous or predetermined regressors, contemporaneous differences $\Delta \mathbf{x}_{it}$ qualify as instruments. For endogenous regressors, lagged differences $\Delta \mathbf{x}_{i,t-1}$ need to be used.
- The mean stationarity assumption is sometimes called an initial condition because it implies that the initial deviations y_{i0} from its subject-specific long-run equilibrium (steady state) should be unrelated to the long-run equilibrium itself.
- This requirement is insufficiently recognized in the empirical practice. In many (macroeconomic) applications, it is hard to justify.

System GMM estimation

- The combination of instruments for the first-differenced and the level model creates a "system" of estimating equations. The resulting estimator is thus called "system GMM" estimator. To be asymptotically efficient, it again needs to be implemented as a two-step, iterated, or continuously-updating GMM estimator.

```
. xtdpdgmm n L(1/2).n L(0/1).(w k), model(diff) collapse
> gmmiv(L.n, lagrange(1 4)) gmmiv(w k, lagrange(2 4))
> gmmiv(L.n L.w L.k, difference model(level) lagrange(0 0))
> teffects twostep vce(robust) overid
```

System GMM estimation

- In addition to the usual specification tests, the incremental overidentification test comparing the system GMM to the difference GMM estimator is especially useful to assess the validity of the mean stationarity assumption.

```
. estat overid
. estat overid, difference
. xtdpdserial
```

Deviations from within-subject means

- This system can be extended by adding instruments for other model transformations.
- To maximize the instrument strengths for strictly exogenous regressors, instruments $\mathbf{x}_{it}$ can be added for the model in deviations from within-group means (akin to a fixed-effects estimation):

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

where

$$\bar{\Delta} y_{it} = y_{it} - \frac{1}{T}\sum_{s=1}^{T} y_{is} \quad \text{etc.}$$

- With `xtdpdgmm`, this can be done by using the suboption `model(mdev)`.

Forward-orthogonal deviations

- Instead of first differencing, the model can be transformed into deviations from their forward mean (Arellano and Bover, 1995):

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

where

$$\overrightarrow{\Delta} y_{it} = y_{it} - \frac{1}{T-t+1} \sum_{s=t}^{T} y_{is} \quad \text{etc.}$$

- For the lagged dependent variable, valid instruments are $y_{i,t-1}, y_{i,t-2}, \ldots$, and similarly for the regressors $\mathbf{x}_{it}$.

```
. xtdpdgmm n L(1/2).n L(0/1).(w k), model(fodev) collapse
> gmmiv(L.n, lagrange(0 3)) gmmiv(w k, lagrange(1 3))
> gmmiv(L.n L.w L.k, difference model(level) lagrange(0 0))
> teffects twostep vce(robust)
```

Nonlinear moment conditions

- Throughout, we maintained the assumption of serially uncorrelated idiosyncratic errors ε_{it}. This allows exploiting nonlinear moment conditions (Ahn and Schmidt, 1995):

$$E[(\alpha_i + \varepsilon_{iT})\Delta\varepsilon_{it}] = 0 \quad \text{for } t < T$$

- These moment conditions are redundant for the system GMM estimator (without collapsing and curtailing), but they can improve identification and efficiency (at the only cost of slightly higher computational complexity) for the difference GMM estimator when the mean stationary assumption does not hold.

```
. xtdpdgmm n L(1/2).n L(0/1).(w k), model(diff) collapse
> gmmiv(L.n, lagrange(1 4)) gmmiv(w k, lagrange(2 4)) teffects nl(noserial)
> twostep vce(robust)
. estat overid
. xtdpdserial
. estimates store noserial
```

Nonlinear moment conditions

- Alternative nonlinear moment conditions can further improve efficiency under a homoskedasticity assumption (Ahn and Schmidt, 1995).

```
. xtdpdgmm n L(1/2).n L(0/1).(w k), model(diff) collapse
> gmmiv(L.n, lagrange(1 4)) gmmiv(w k, lagrange(2 4)) teffects nl(iid)
> twostep vce(robust)
. estat overid noserial
```

- Yet another set of nonlinear moment conditions becomes available if none of the regressors $\mathbf{x}_{it}$ are classified as endogenous (Chudik and Pesaran, 2022), maintaining the assumption of no serial correlation. The respective `xtdpdgmm` option is `nl(predetermined)`.

Summary

- Conventional fixed-effects (and random-effects) estimators are biased in dynamic panel data models when T is small.
- Under strict exogeneity of the regressors (other than the lagged dependent variable), we can exploit our knowledge about the time series dependence to construct consistent maximum-likelihood or bias-corrected estimators.
  - Stata: `xtdpdqml` and `xtdpdbc`
- GMM estimators offer more flexibility by allowing for endogeneity of the regressors, but might suffer from weak identification.
  - Stata: `xtdpdgmm`
- Specification tests can aid the estimator choice but should not be used for specification "hacking". Also be aware of pre-testing implications for statistical inference.
  - Stata: `xtdpdserial`

References

- Ahn, S. C., and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68* (1): 5–27.
- Anderson, T. W., and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76* (375), 598–606.
- Arellano, M., and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58* (2), 277–297.
- Arellano, M., and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68* (1), 29–51.
- Bhargava, A., and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51* (6), 1635–1659.
- Blundell, R., and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87* (1), 115–143.
- Breitung, J., S. Kripfganz, and K. Hayakawa (2022). Bias-corrected method of moments estimators for dynamic panel data models. *Econometrics and Statistics 24*, 116–132.
- Bruno, G. S. F. (2005). Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *Stata Journal 5* (4), 473–500.
- Bun, M. J. G., and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23* (2), 200–210.
- Bun, M. J. G., and J. F. Kiviet (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics 132* (2), 409–444.
- Chudik, A., and M. H. Pesaran (2022). An augmented Anderson-Hsiao estimator for dynamic short-T panels. *Econometric Reviews 41* (4), 416–447.
- Dhaene, G., and K. Jochmans (2016). Likelihood inference in an autoregression with fixed effects. *Econometric Theory 32* (5), 1178–1215.
- Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics 103* (1), 51–78.

References

- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50* (4), 1029–1054.
- Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics 14* (3), 262–280.
- Hansen, B. E., and S. Lee (2021). Inference for iterated GMM under misspecification. *Econometrica 89* (3), 1419–1447.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46* (6), 1251–1271.
- Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109* (1), 107–150.
- Hwang, J., B. Kang, and S. Lee (2022). A doubly corrected robust variance estimator for linear GMM. *Journal of Econometrics 229* (2), 276–298.
- Jochmans, K. (2020). Testing for correlation in error-component models. *Journal of Applied Econometrics 35* (7), 860–878.
- Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68* (1), 53–78.
- Kiviet, J. F. (2020). Microeconometric dynamic panel data methods: Model specification and selection issues. *Econometrics and Statistics 13*, 16–45.
- Kripfganz, S. (2016). Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal 16* (4), 1013–1038.
- Kripfganz, S. (2019). Generalized method of moments estimation of linear dynamic panel data models. *Proceedings of the 2019 London Stata Conference*.
- Kripfganz, S. (2024). Robust testing for serial correlation in linear panel-data models. *Proceedings of the 2024 London Stata Conference*.
- Kripfganz, S., and J. Breitung (2022). Bias-corrected estimation of linear dynamic panel data models. *Proceedings of the 2022 London Stata Conference*.
- Kripfganz, S., M. Hosseinkouchack, and M. Demetrescu (2026). Serial correlation testing in error component models with moderately small T. *Department of Economics Discussion Papers 26/03*, University of Exeter.
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49* (6), 1417–1426.

References

- Roodman, D. (2009a). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71* (1), 135–158.
- Roodman, D. (2009b). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal 9* (1), 86–136.
- Sanderson, E., and F. Windmeijer (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics 190* (2), 212–221.
- Schaffer, M. E., and F. Windmeijer (2020). UNDERID: Stata module producing postestimation tests of under- and over-identification after linear IV estimation. *Statistical Software Components S458805*, Boston College Department of Economics.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126* (1), 25–51.
- Windmeijer, F. (2024). Testing underidentification in linear models, with applications to dynamic panel and asset pricing models. *Journal of Econometrics 240* (2), 105104.
- Yamagata, T. (2008). A joint serial correlation test for linear panel data models. *Journal of Econometrics 146* (1), 135–145.