Transformed Maximum Likelihood Estimation of Short Dynamic Panel Data Models with Interactive Effects

Kazuhiko Hayakawa[1] M. Hashem Pesaran[2] L. Vanessa Smith[3]

[1]Hiroshima University [2]USC and Trinity College, Cambridge

[3]University of York

The 20th International Panel Data Conference
Hitotsubashi University, Tokyo, 9-10 July 2014

Outline of presentation

1. Introduction
2. Transformed ML approach
3. GMM approach
4. Monte Carlo
5. Conclusion

1. Introduction

Introduction

- There exists a wide literature on the estimation of linear dynamic panel data models where T is small and N is large.
- GMM and likelihood approaches have been advanced to estimate such panel data models (Anderson and Hsiao, 1981,1982; Arellano and Bond, 1991; Ahn and Schmidt, 1995; Arellano and Bover, 1995; Blundell and Bond, 1998; Hsiao, Pesaran and Tahmiscioglu, 2002; Binder, Hsiao and Pesaran, 2005, Hayakawa and Pesaran, 2014 etc.).
- This literature assumes that the errors are cross sectionally independent.
- Phillips and Sul (2007) and Sarafidis and Robertson (2009) show that the pooled least squares estimator and widely used IV and GMM estimators are inconsistent in the presence of cross section dependence.
- To deal with possible error cross section dependence some recent research has considered allowing for spatial effects in dynamic panel data models (Lee and Yu (2010)).

Introduction

- Error cross section dependence could be a result of omitted unobserved common factor(s).
- Several estimation procedures have been proposed for panels where N and T are both large.
- However, less work has been done so far on the estimation of short T dynamic panels.
- Holtz-Eakin, Newey and Rosen (1988) and Ahn, Lee and Schmidt (2001, 2013), suggest a quasi-difference approach to remove the factor structure, and then use GMM to consistently estimate the model parameters.
- Nauges and Thomas (2003) follow this approach, but take the first-differences to remove fixed effects prioir to estimation.
- Robertson and Sarafidis (2013) propose an instrumental variable estimation procedure that introduces new parameters to represent the unobserved covariances between the instruments and the factor component of the residual.

Introduction

- These are based on GMM approach.
- Instead, this paper uses the ML approach.
- We extend the transformed quasi-maximum likelihood (QML) approach proposed in Hsiao et al. (2002) to incorporate the factor structure(interactive fixed effects).
- Recently, Bai (2013) suggests a QML approach assuming that the loadings in the factor structure are random, and using the projection method of Mundlak (1978) and Chamberlain (1982) to deal with the correlation between the common effects and the regressors.
- Our procedure differs from that of Bai (2013) since he proposes to apply the ML procedure directly to the specification which includes the nuisance parameters, whilst we propose to apply the ML procedure to the transformed relations that are free from nuisance parameters.

2. Transformed Likelihood Approach

The Transformed Likelihood Approach

- Consider the following model

$$\begin{aligned} y_{it} &= \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}, \qquad (1) \\ \xi_{it} &= \lambda_i f_t + u_{it}, \qquad i = 1, 2, ..., N; t = 1, 2, ..., T. \end{aligned}$$

- α_i are fixed effects, f_t is an unobserved common effect where without loss of generality it is assumed that $g_t = \Delta f_t \neq 0$, for at least some $t = 1, 2, ..., T$, u_{it} are the individual-specific errors, and λ_i are random interactive effects.
- The regressor x_{it} is generated either by

$$x_{it} = \mu_i + ct + \vartheta_i f_t + \sum_{j=0}^{\infty} a_j \varepsilon_{i,t-j}, \quad \sum_{j=0}^{\infty} |a_j| < \infty, \qquad (2)$$

or

$$\Delta x_{it} = c + \vartheta_i g_t + \sum_{j=0}^{\infty} d_j \varepsilon_{i,t-j}, \quad \sum_{j=0}^{\infty} |d_j| < \infty, \qquad (3)$$

where μ_i are fixed effects , and κ_i are random interactive effects distributed independently of u_{it} and f_t.

The Transformed Likelihood Approach

- Make the following assumptions.

Assumption 1

The idiosyncratic shocks, u_{it}, $(i = 1, 2, ..., N,\ t = 1, 2, ..., T)$, are independently distributed both across i and t, with mean zero, and variance σ^2.

Assumption 2

The unobserved factor loadings, λ_i are i.i.d over i, and independent of the individual specific errors, u_{jt}, and the common factor, f_t, for all i, j and t with fixed means λ, and finite variance. In particular,

$$\lambda_i = \lambda + \eta_i,\ \eta_i \sim IID(0, \sigma_\eta^2).$$

Assumption 3

The error terms η_i and u_{it} are normally distributed.

The Transformed Likelihood Approach

Assumption 4

The dynamic process given by (4) has started from $y_{i,-S}$ with S finite such that $E(\Delta y_{i,-S+1} | \Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT}) = \tilde{b}$ for all i.

Assumption 5

The interactive effects ϑ_i in Δx_{it} have constant variance $var(\vartheta_i) = \sigma_{\vartheta}^2$, and are uncorrerated with λ_i and u_{it} for all i and t.

Assumption 6

The error terms ε_{it} in x_{it} are independently distributed over all i and t, with $E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2) = \sigma_{\varepsilon}^2$, and independent of $u_{it'}$ for all t' and t.

The Transformed Likelihood Approach

- For each i, the compositive error ξ_{it} in (1) is heteroskedastic even though it is assumed that $var(u_{it}) = \sigma^2$ is homoskedastic, namely for each i we have $Var(\xi_{it} | \lambda_i) = \lambda_i^2 \sigma_f^2 + \sigma^2$.
- As shown by Hayakawa and Pesaran (2012), in a recent extension of Hsiao et al. (2002), it could be possible to allow for heteroskedasticity in u_{it}, but this will not be pursued here.
- The normality assumption(Assumption 3) is not essential as $N \to \infty$, so long as the errors η_i and u_{it} have fourth-order moments.
- Assumption 4 is used to derive the marginal model for Δy_{i1} below.
- Assumption 6 implies that the regressor x_{it} is strictly exogenous.

The Transformed Likelihood Approach

- We eliminate the individual effects by first-differencing and using Assumption 2 we have

$$\begin{aligned}\Delta y_{it} &= \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \lambda_i g_t + \Delta u_{it} \\ &= \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \lambda g_t + \eta_i g_t + \Delta u_{it} \quad \text{for } t = 2, 3, ..., T. \end{aligned} \tag{4}$$

- By recursive substitution, we have the following expression for $t = 1$,

$$\begin{aligned}\Delta y_{i1} &= \gamma^S \Delta y_{i,-S+1} + \beta \sum_{j=0}^{S-1} \gamma^j \Delta x_{i,1-j} + \lambda_i \sum_{j=0}^{S-1} \gamma^j g_{1-j} + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j} \\ &= \gamma^S \Delta y_{i,-S+1} + \beta \sum_{j=0}^{S-1} \gamma^j \Delta x_{i,1-j} + \lambda_i \tilde{g}_{1S} + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j},\end{aligned}$$

where $\tilde{g}_{1S} = \sum_{j=0}^{S-1} \gamma^j g_{1-j}$.

The Transformed Likelihood Approach

- This expression shows that Δy_{i1} contains many unknown quantities such as unknown parameters or unobserved past variables.
- However, it is possible to derive an expression for Δy_{i1} based on observed variables and a finite number of parameters, as follows.

Theorem

Consider model (4), where x_{it} follows either (2) or (3). Suppose that Assumptions 1 to 6 hold. Then Δy_{i1} can be expressed as:

$$\Delta y_{i1} = b + \boldsymbol{\pi}' \Delta \mathbf{x}_i + v_{i1},$$

where b is a constant, $\boldsymbol{\pi}$ is a T-dimensional vector of constants, $\Delta \mathbf{x}_i = (\Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT})'$, and v_{i1} is independently distributed across i, such that $E(v_{i1}) = 0$, and $E(v_{i1}^2) = \omega\sigma^2$, with $0 < \omega < K < \infty$.

The Transformed Likelihood Approach

- Using the above Theorem and (4), the transformed model can be rewritten as

$$\Delta \mathbf{y}_i = \Delta \mathbf{W}_i \boldsymbol{\delta} + \lambda \mathbf{g} + \boldsymbol{\xi}_i, \tag{5}$$

where $\boldsymbol{\delta} = (b, \boldsymbol{\pi}', \gamma, \beta)'$, $\boldsymbol{\xi}_i = \eta_i \mathbf{g} + \mathbf{r}_i$ with $\mathbf{g} = (\tilde{g}_1, g_2, ..., g_T)'$, and

$$\Delta \mathbf{y}_i = \begin{pmatrix} \Delta y_{i1} \\ \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, \quad \Delta \mathbf{W}_i = \begin{pmatrix} 1 & \Delta \mathbf{x}_i' & 0 & 0 \\ 0 & \mathbf{0} & \Delta y_{i1} & \Delta x_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \Delta y_{i,T-1} & \Delta x_{iT} \end{pmatrix}, \quad \mathbf{r}_i = \begin{pmatrix} v_{i1} \\ \Delta u_{i2} \\ \vdots \\ \Delta u_{iT} \end{pmatrix}.$$

- From Hsiao et al. (2002), we have

$$E(\mathbf{r}_i \mathbf{r}_i') = \sigma^2 \begin{pmatrix} \omega & -1 & & & 0 \\ -1 & 2 & \ddots & & \\ & & \ddots & & \\ & & \ddots & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix} = \sigma^2 \boldsymbol{\Omega},$$

The Transformed Likelihood Approach

- Using $Var(\xi_i) = \sigma^2\boldsymbol{\Omega}+\sigma_\eta^2\mathbf{gg}' = \sigma^2\left(\boldsymbol{\Omega}+\phi\mathbf{gg}'\right)$ where $\phi = \sigma_\eta^2/\sigma^2$, the log-likelihood function is given by

$$
\begin{aligned}
\frac{\ell(\psi)}{N} &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln\left|\boldsymbol{\Omega}+\phi\mathbf{gg}'\right| \\
&\quad -\frac{1}{2N\sigma^2}\sum_{i=1}^{N}\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\delta} - \lambda\mathbf{g}\right)'\left(\boldsymbol{\Omega}+\phi\mathbf{gg}'\right)^{-1}\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\delta} - \lambda\mathbf{g}\right) \\
&= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln\left|\boldsymbol{\Omega}\right| - \frac{1}{2}\ln(1+\phi\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}) \\
&\quad -\frac{1}{2\sigma^2}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{\phi\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{g} - \lambda^2\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g} + 2\lambda\mathbf{g}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)}\right]
\end{aligned}
\tag{6}
$$

where $\mathbf{v}_i = \mathbf{v}_i(\boldsymbol{\delta}) = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\delta}$, $\bar{\mathbf{v}} = N^{-1}\sum_{i=1}^{N}\mathbf{v}_i$,

$$
\mathbf{B}_N = \mathbf{B}_N(\boldsymbol{\delta}) = \left(N^{-1}\sum_{i=1}^{N}\mathbf{v}_i(\boldsymbol{\delta})\mathbf{v}_i'(\boldsymbol{\delta})\right).
$$

The Transformed Likelihood Approach

- Note that the likelihood (6) is a function of a fixed number of unknown parameters, $\boldsymbol{\psi} = (\boldsymbol{\delta}', \omega, \sigma^2, \phi, \lambda, \mathbf{g}')'$ for a fixed T.
- In the interactive case where $\phi \neq 0$, $\mathbf{g}$ is not identified separately from ϕ.
- Without loss of generality, we can set $\mathbf{q} = \sqrt{\phi}\mathbf{g}$, and write the log-likelihood function as

$$
\begin{aligned}
N^{-1}\ell(\boldsymbol{\varphi}) \quad = \quad & -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}) \\
& -\frac{1}{2\sigma^2}N^{-1}\left[\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \right. \\
& \left. -\frac{\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \varrho^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + 2\varrho\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}}\right], \qquad (7)
\end{aligned}
$$

where $\boldsymbol{\varphi} = (\boldsymbol{\delta}', \omega, \sigma^2, \varrho, \mathbf{q}')'$, and $\varrho = \lambda/\sqrt{\phi}$.

The Transformed Likelihood Approach

- Taking partial derivatives with respect to ϱ and σ^2 and solving out for these, we have

$$\hat{\varrho} = \frac{\mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}}{\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}, \qquad \hat{\sigma}^2 = T^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\mathbf{\Omega}^{-1}\mathbf{v}_i - \frac{\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} + \frac{\left(\mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}\right)^2}{\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}}{1+\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}\right]$$

- Substituting these into (7), we have the following concentrated log-likelihood function:

$$\begin{aligned} N^{-1}\ell(\boldsymbol{\theta}) \quad \propto \quad & -\frac{T}{2}\ln\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\mathbf{\Omega}^{-1}\mathbf{v}_i - \frac{\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} + \frac{\left(\mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}\right)^2}{\left(\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)}}{1+\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}\right] \\ & -\frac{1}{2}\ln|\mathbf{\Omega}| - \frac{1}{2}\ln\left(1+\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right), \end{aligned} \tag{8}$$

where $\boldsymbol{\theta} = (\boldsymbol{\delta}', \omega, \mathbf{q}')'$.

The Transformed Likelihood Approach

- For a fixed T and as $N \to \infty$, using standard results from the asymptotic theory of ML estimation, we have

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to^{d} N(\mathbf{0}, \mathbf{H}^{-1}(\boldsymbol{\theta})),$$

where

$$\mathbf{H}(\boldsymbol{\theta}) = p \lim_{N \to \infty} \left[-\frac{1}{N} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right].$$

- A consistent estimator of $Var(\widehat{\boldsymbol{\theta}})$ can be obtained

$$\widehat{Var}(\widehat{\boldsymbol{\theta}}) = \left[-\frac{\partial^2 \ell\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \tag{9}$$

where the second partial derivatives are evalauted at the MLE, $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\delta}}', \widehat{\omega}, \widehat{\mathbf{q}}')'$.

Extension to the multifactor case

- Consider the extension of model (1) to the multifactor case

$$\begin{aligned} y_{it} &= \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}, \\ \xi_{it} &= \mathbf{f}_t' \boldsymbol{\lambda}_i + u_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T), \end{aligned} \quad (10)$$

  where $\mathbf{f}_t$ and $\boldsymbol{\lambda}_i$ are $m \times 1$ vectors of unobserved common effects and random interactive effects, respectively, the latter distributed independently of u_{it} and $\mathbf{f}_t$.

- Without loss of generality it is assumed that $\mathbf{g}_t = \Delta \mathbf{f}_t \neq \mathbf{0}$ for at least some $t = 1, 2, ..., T$. The remaining parameters are the same as before.

- It is assumed that the number of factors m is known and that $m < T$.

Extension to the multifactor case

- To accommodate multiple factors the following modified versions of Assumptions 2 and 3 are needed:

Assumption 7

The unobserved factor loadings, $\boldsymbol{\lambda}_i$, are independently and identically distributed across i and of the individual specific errors, u_{jt}, and the common factor, $\mathbf{f}_t$, for all i, j and t, with fixed means, $\boldsymbol{\lambda}$, and a finite variance. In particular,

$$\boldsymbol{\lambda}_i = \boldsymbol{\lambda} + \boldsymbol{\eta}_i,\ \boldsymbol{\eta}_i \sim IID(\mathbf{0}, \Omega_\eta), \tag{11}$$

where Ω_η is an $m \times m$ symmetric positive definite matrix, $\|\boldsymbol{\lambda}\| < K$ and $\|\Omega_\eta\| < K$ for some positive constant $K < \infty$.

Assumption 8

The error terms $\boldsymbol{\eta}_i$ and u_{it} are normally distributed.

Extension to the multifactor case

- Under Assumptions 1, 4, 5, 6, 7 and 8, and following similar derivations as in the single factor case we have

$$
\begin{aligned}
N\bar{\ell}(\boldsymbol{\theta}) \quad \propto \quad & -\frac{1}{2}\ln|\Omega| - \frac{1}{2}\ln\left|\mathbf{I}_m + \mathbf{Q}'\Omega^{-1}\mathbf{Q}\right| \\
& -\frac{T}{2}\ln\left\{\begin{array}{c} N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\Omega+\mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i \\ -\bar{\mathbf{v}}'\Omega^{-1}\mathbf{Q}\mathbf{A}^{-1}\left(\mathbf{Q}'\Omega^{-1}\mathbf{Q}\right)^{-1}\mathbf{Q}'\Omega^{-1}\bar{\mathbf{v}} \end{array}\right\},
\end{aligned} \tag{12}
$$

  where $\boldsymbol{\theta} = \left(\gamma, \beta, \omega, vec(\mathbf{Q})'\right)'$, $\mathbf{Q} = \sigma^{-1}\mathbf{G}\Omega_\eta^{1/2}$ with $\mathbf{G} = (\widetilde{\mathbf{g}}_1, \mathbf{g}_2, ..., \mathbf{g}_T)'$ and $\widetilde{\mathbf{g}}_1 = \sum_{j=0}^{\infty}\gamma^j\mathbf{g}_{1-j}$, and $\mathbf{A} = \mathbf{I}_m + \mathbf{Q}'\Omega^{-1}\mathbf{Q}$.

- The restrictions implied by $\mathbf{Q} = \sigma^{-1}\mathbf{G}\Omega_\eta^{1/2}$ are not binding, in the sense that the log-likelihood function is invariant to the choice of the normalization and they are used to identify the multifactor structure $\boldsymbol{\lambda}'\mathbf{g}_t$.

- Since $\boldsymbol{\lambda}$ and $\mathbf{g}_t$ are not separately identified their inner product can be equivalently written as $\boldsymbol{\delta}'\mathbf{q}_t$ where $\boldsymbol{\delta} = \sigma\Omega_\eta^{-1/2}\boldsymbol{\lambda}$, and $\mathbf{q}_t$ is the t^{th} row of $\mathbf{Q}$.

- It is also easily verified that (12) reduces to (8) when $m = 1$.

3. The GMM Approach

Ahn, Lee and Schmidt (2001,2013)

- Consider the following model:

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{w}_{it}'\boldsymbol{\delta} + \boldsymbol{\lambda}_i'\mathbf{f}_t + \varepsilon_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T), \\ &= \mathbf{w}_{it}'\boldsymbol{\delta} + \tilde{\boldsymbol{\lambda}}_i'\tilde{\mathbf{f}}_t + \varepsilon_{it}, \end{aligned} \tag{13}$$

where $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}_{it}')'$, $\mathbf{x}_{it}$ is $k \times 1$, $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$, $\tilde{\boldsymbol{\lambda}}_i = (\alpha_i, \lambda_{1i}, ..., \lambda_{mi})'$, and $\tilde{\mathbf{f}}_t = (1, f_{1t}, ..., f_{mt})'$ are $(\tilde{m} \times 1)$ vectors with $\tilde{m} = m + 1$.

- In matrix notation:

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \tilde{\mathbf{F}}\tilde{\boldsymbol{\lambda}}_i + \boldsymbol{\varepsilon}_i, \tag{14}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ and $\tilde{\mathbf{F}} = (\tilde{\mathbf{f}}_1, ..., \tilde{\mathbf{f}}_T)'$ is a $T \times \tilde{m}$ matrix.

- To separately identify $\tilde{\mathbf{F}}$ from $\tilde{\boldsymbol{\lambda}}_i$ they impose $\tilde{m}^2$ restrictions on the factors as follows:

$$\tilde{\mathbf{F}} = (\boldsymbol{\Psi}', \mathbf{I}_{\tilde{m}})'$$

where $\boldsymbol{\Psi}$ is a $(T - \tilde{m}) \times \tilde{m}$ matrix of unrestricted parameters.

Ahn, Lee and Schmidt (2001,2013)

- Under the above restriction, the model to be estimated becomes

$$\dot{\mathbf{y}}_i = \dot{\mathbf{W}}_i\delta + \mathbf{\Psi}\mathbf{y}_i - \mathbf{\Psi}\ddot{\mathbf{W}}_i\delta + \mathbf{v}_i,$$

  where $\dot{\mathbf{y}}_i = (y_{i1}, ..., y_{i,T-\tilde{m}})'$, $\ddot{\mathbf{y}}_i = (y_{i,T-\tilde{m}+1}, ..., y_{iT})'$,
  $\dot{\mathbf{W}}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{i,T-\tilde{m}})'$, $\ddot{\mathbf{W}}_i = (\mathbf{w}_{i,T-\tilde{m}+1}, ..., \mathbf{w}_{iT})'$, $\mathbf{\Psi}' = (\psi_1, ..., \psi_{T-\tilde{m}})$,
  $\dot{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i1}, ..., \varepsilon_{i,T-\tilde{m}})'$, $\ddot{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i,T-\tilde{m}+1}, ..., \varepsilon_{iT})'$ and
  $\mathbf{v}_i = \underline{\boldsymbol{\varepsilon}}_i - \mathbf{\Psi}\boldsymbol{\varepsilon}_i = (v_{i1}, ..., v_{i,T-\tilde{m}})'$.

- Then, if $\mathbf{x}_{it}$ is strictly exogenous, the following $(T-\tilde{m})(T-\tilde{m}+1)/2 + kT(T-\tilde{m})$ moment conditions hold

$$E\left[\mathbf{z}_{it}v_{it}\right] = \mathbf{0}, \qquad (t = 1, ..., T-\tilde{m})$$

  where $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-1}, \mathbf{x}'_{i1}, ..., \mathbf{x}'_{iT})'$.

- In Monte Carlo study below, we use $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-1}, x_{it}, ..., x_{iT})'$ to avoid a large finite sample bias.

Nauges and Thomas (2003)

- Nauges and Thomas (2003) consider the single factor dynamic panel model given by

$$\begin{aligned} y_{it} &= \mathbf{w}_{it}'\boldsymbol{\delta} + u_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T), \\ u_{it} &= a_i + \lambda_i f_t + \varepsilon_{it}, \end{aligned} \tag{15}$$

  with $|\gamma| < 1$ and the initial values y_{i0} are observed and stochastic.

- The parameters a_i and λ_i may be correlated, while

$$E(y_{i0}\varepsilon_{it}) = 0,\ E(a_i\varepsilon_{it}) = 0,\ E(\lambda_i\varepsilon_{it}) = 0,\ E(\varepsilon_{it}\varepsilon_{is}) = 0, \tag{16}$$

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

- First difference to eliminate a_i so that the (15) becomes

$$\Delta y_{it} = \boldsymbol{\delta}'\Delta\mathbf{w}_{it} + \lambda_i g_t + \Delta\varepsilon_{it}. \tag{17}$$

  where $g_t = \Delta f_t$.

Nauges and Thomas (2003)

- By using a quasi-differencing transformation suggested by Holtz-Eakin et al. (1988), they obtain the following model:

$$(\Delta y_{it} - r_t \Delta y_{i,t-1}) = \boldsymbol{\delta}'(\Delta \mathbf{w}_{it} - r_t \Delta \mathbf{w}_{i,t-1}) + (\Delta \varepsilon_{it} - r_t \Delta \varepsilon_{i,t-1})$$
$$(i = 1, 2, ..., N; \quad t = 3, 4, ..., T)$$

  where $r_t = g_t / g_{t-1} = (f_t - f_{t-1})/(f_{t-1} - f_{t-2})$.

- Given assumptions (16), if x_{it} is strictly exogenous, the following $(T-2)(T-1)/2 + kT\ (T-2)$ moment conditions hold:

$$E[\mathbf{z}_{it}(\Delta \varepsilon_{it} - r_t \Delta \varepsilon_{i,t-1})] = \mathbf{0}, \qquad (t = 3, 4, ..., T),$$

  where $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-3}; \mathbf{x}'_{i1}, ..., \mathbf{x}'_{iT})'$.

- In Monte Carlo studies below, we use $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-m-2}, x_{i1}, ..., x_{it})'$ to avoid a large finite sample bias.

4. Monte Carlo simulation

ARX(1) with single factor

- The y_{it} are generated as

$$\begin{aligned} y_{it} &= \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}, \\ \xi_{it} &= \lambda_i f_t + u_{it},\ u_{it} \sim iid\mathcal{N}(0, \sigma^2). \end{aligned}$$

  for for $i = 1, 2, ..., N$; $t = -S+1, -S+2, .., 0, 1, ..., T$.

- The regressors, x_{it}, are generated as

$$x_{it} = \mu_i + \vartheta_i f_t + \breve{x}_{it},\ \ ,\breve{x}_{it} = \rho_x \breve{x}_{i,t-1} + \sqrt{1-\rho_x^2}\varepsilon_{it}, \tag{18}$$

  with $\breve{x}_{i,-S} = 0$, for $t = -S+1, ..., 0, 1, ..., T$, $\rho_x = 0.8$, $\mu_i \sim iid\mathcal{N}(0, 1)$, $\varepsilon_{it} \sim iid\mathcal{N}(0, 1)$.

- We discard the first $S = 50$ observations, using the observations $t = 0$ through T for estimation.

ARX(1) with single factor

- We generate the factor loadings independently as

$$\vartheta_i \sim iid\mathcal{N}(0.5, \sigma_\vartheta^2), \;\; \lambda_i \sim iid\mathcal{N}(0.5, \sigma_\lambda^2), \tag{19}$$

- For the unobserved common factor, f_t, we consider two cases:

$$\begin{aligned} \text{trend} \qquad f_t &= \begin{cases} 0 & t = -S+1, ..., -1, 0 \\ t & t = 1, 2, ..., T \end{cases}, \\ \text{AR(1)} \qquad f_t &= \rho_f f_{t-1} + \sqrt{1-\rho_f^2}\varepsilon_{ft}, \text{ for } t = -S+1, ..., -1, 0, 1, .., T. \end{aligned}$$

- $\varepsilon_{ft} \sim iid\mathcal{N}(0, 1)$, and $\rho_f = 0.9$ with $f_{-S} = 0$.
- We scale the resultant f_t values such that $T^{-1} \sum_{t=1}^{T} f_t^{\,2} = 1$.
- Each f_t is generated once and the same $f_t' s$ are used in all replications of a given experiment.

ARX(1) with single factor

- Fixed effects, α_i, are generated so that it is correlated with the regressors x_{it} and the errors u_{it}

$$\alpha_i = T^{-1} \sum_{t=1}^{T} x_{it} + \lambda_i \bar{f} + \bar{u}_i + v_i,$$

  where $\bar{f} = T^{-1} \sum_{t=1}^{T} f_t$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$ and $v_i \sim iid\mathcal{N}(0,1)$.
- We set $\beta = 1$ and determine σ^2, σ^2_λ, and σ^2_ϑ such that $R^2_y - \gamma^2 = 0.1$.(the details are omitted)
- $\gamma = (0.4, 0.8)$,
- $T = (6, 10)$
- $N = (150, 200, 500)$.
- The significance level is 5% and all experiments are replicated 1,000 times.

ARX(1) with single factor

- Findings are as follows:
  1. ML estimator has small biases and RMSEs.
  2. All GMM estimators perform very poorly. Biases and RMSEs are quite large.
  3. Sizes of ML estimator are close to the nominal one.
  4. Size distortion of GMM is substantial.

ARX(1) with single factor

$T = 6$, $\beta = 1$, $f_t \sim$AR(1) (Bias and RMSE($\times$100))

$\gamma = 0.4$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.19	4.29	-0.05	7.41	-0.05	2.30	0.01	4.08
ALS(1step)	0.81	16.60	-3.11	17.23	-1.71	8.16	-0.86	8.51
NT(1step)	-3.70	32.16	4.70	14.58	15.80	20.97	8.52	10.84
$\gamma = 0.8$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.06	2.38	-0.07	4.33	-0.01	1.32	0.01	2.42
ALS(1step)	-1.35	5.39	3.03	8.29	-2.32	3.16	4.29	5.56
NT(1step)	-1.37	14.76	0.51	6.36	9.39	11.85	0.79	3.65

ARX(1) with single factor

$T = 6, \quad \beta = 1, \quad f_t \sim trend \quad$ (Bias and RMSE($\times$100))

$\gamma = 0.4$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.07	5.82	-0.28	8.96	-0.07	2.95	0.01	4.77
ALS(1step)	9.07	36.37	-13.00	40.27	3.71	37.09	-6.38	39.69
NT(1step)	54.52	55.27	-4.96	16.36	59.81	59.82	-6.11	10.67
$\gamma = 0.8$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.09	3.10	-0.14	4.88	-0.01	1.65	0.01	2.64
ALS(1step)	4.65	10.21	-4.82	14.47	3.66	9.12	-2.99	12.24
NT(1step)	18.93	19.50	1.40	6.85	19.90	19.90	1.44	3.99

ARX(1) with single factor

$T = 10$, $\beta = 1$, $f_t \sim$AR(1) (Bias and RMSE($\times$100))

$\gamma = 0.4$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.03	2.58	0.12	5.53	-0.04	1.42	0.07	3.06
ALS(1step)	0.62	6.01	-6.28	11.10	0.44	3.36	-5.74	7.82
NT(1step)	21.94	26.73	7.16	11.51	33.75	34.16	6.39	8.10
$\gamma = 0.8$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.04	1.24	0.08	3.03	-0.02	0.67	0.03	1.67
ALS(1step)	-0.10	1.75	1.32	4.59	-0.09	1.00	1.32	2.82
NT(1step)	11.26	15.39	3.56	5.66	19.17	19.26	3.36	4.12

ARX(1) with single factor

$T = 10, \quad \beta = 1, \quad f_t \sim trend \quad$ (Bias and RMSE($\times$100))

$\gamma = 0.4$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	0.09	3.21	-0.10	6.01	0.01	1.76	0.06	3.35
ALS(1step)	15.28	29.71	-22.34	36.05	16.93	33.49	-23.36	39.17
NT(1step)	53.85	54.45	-0.06	11.72	59.90	59.90	-1.21	6.59
$\gamma = 0.8$								
	$N = 150$				$N = 500$			
	γ		β		γ		β	
Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	-0.03	1.69	0.02	3.25	0.01	0.89	0.04	1.77
ALS(1step)	3.77	7.11	-3.49	11.63	1.34	3.97	0.69	6.82
NT(1step)	19.51	19.64	3.53	5.91	19.90	19.90	4.06	4.80

ARX(1) with single factor

$T = 6, \quad \gamma = 0.4, \quad \beta = 1, \quad f_t \sim$AR(1) (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H$_1$)	Size	Power(H$_1$)	Power(H$_1$)	Size	Power(H$_1$)
Estimators \ γ	0.30	0.40	0.50	0.30	0.40	0.50
ML	68.2	5.8	65.6	99.0	5.9	99.0
ALS1(1step)	37.2	15.2	26.3	62.4	15.6	37.2
NT1(1step)	29.5	35.2	41.4	16.7	46.1	73.0
	β					
	Power(H$_1$)	Size	Power(H$_1$)	Power(H$_1$)	Size	Power(H$_1$)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	26.3	3.9	23.3	68.0	4.7	68.4
ALS1(1step)	21.7	11.5	17.5	34.8	7.5	30.2
NT1(1step)	13.3	14.4	34.2	6.6	32.1	84.6

ARX(1) with single factor

$T = 6, \quad \gamma = 0.8, \quad \beta = 1, \quad f_t \sim$AR(1) (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ γ	0.70	0.80	0.90	0.70	0.80	0.90
ML	99.8	5.1	97.2	100.0	6.4	100.0
ALS1(1step)	92.9	14.6	62.6	99.9	21.2	94.4
NT1(1step)	9.2	5.0	6.3	0.3	6.2	30.9
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	61.9	5.1	62.0	98.8	4.6	98.5
ALS1(1step)	24.0	12.5	58.1	37.0	21.4	97.9
NT1(1step)	32.7	5.1	40.2	73.1	4.8	83.4

ARX(1) with single factor

$T = 6,\quad \gamma = 0.4,\quad \beta = 1,\quad f_t \sim \textit{trend}$ (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ γ	0.30	0.40	0.50	0.30	0.40	0.50
ML	47.1	6.3	43.2	91.7	4.4	92.1
ALS1(1step)	59.9	44.1	38.6	90.6	71.1	46.5
NT1(1step)	78.8	88.6	94.0	100.0	100.0	100.0
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	21.7	6.9	20.3	53.5	4.6	56.7
ALS1(1step)	37.0	36.5	41.9	39.0	52.7	74.1
NT1(1step)	22.1	10.2	10.4	52.3	15.3	11.3

ARX(1) with single factor

$T = 6, \quad \gamma = 0.8, \quad \beta = 1, \quad f_t \sim \textit{trend} \quad$ (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H$_1$)	Size	Power(H$_1$)	Power(H$_1$)	Size	Power(H$_1$)
Estimators \ γ	0.70	0.80	0.90	0.70	0.80	0.90
ML	92.1	6.1	89.0	100.0	5.7	99.9
ALS1(1step)	82.9	49.1	69.8	98.2	48.0	87.1
NT1(1step)	0.5	0.3	0.7	0.0	0.0	40.8
	β					
	Power(H$_1$)	Size	Power(H$_1$)	Power(H$_1$)	Size	Power(H$_1$)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	57.9	5.7	54.8	97.1	5.2	96.5
ALS1(1step)	50.8	38.8	36.0	51.4	52.0	78.4
NT1(1step)	26.3	5.0	41.8	63.5	7.5	86.5

ARX(1) with single factor

$T = 10, \quad \gamma = 0.4, \quad \beta = 1, \quad f_t \sim$AR(1) (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ γ	0.30	0.40	0.50	0.30	0.40	0.50
ML	97.1	5.2	96.4	100.0	4.5	100.0
ALS1(1step)	45.0	6.8	54.5	87.4	7.1	90.0
NT1(1step)	36.5	64.6	82.1	98.2	99.9	100.0
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	42.4	4.2	43.7	90.1	5.3	91.2
ALS1(1step)	48.2	13.7	7.2	86.4	23.4	16.2
NT1(1step)	9.6	19.1	58.3	15.9	32.2	93.3

ARX(1) with single factor

$T = 10, \quad \gamma = 0.8, \quad \beta = 1, \quad f_t \sim \text{AR}(1) \quad$ (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H$_1$)	Size	Power(H$_1$)	Power(H$_1$)	Size	Power(H$_1$)
Estimators \ γ	0.70	0.80	0.90	0.70	0.80	0.90
ML	100.0	4.5	100.0	100.0	5.0	100.0
ALS1(1step)	99.1	4.0	99.1	99.0	5.0	99.0
NT1(1step)	2.8	7.5	53.0	0.0	93.5	99.7
	β					
	Power(H$_1$)	Size	Power(H$_1$)	Power(H$_1$)	Size	Power(H$_1$)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	91.0	5.0	91.6	100.0	5.1	100.0
ALS1(1step)	49.9	6.1	72.7	93.5	8.4	98.8
NT1(1step)	30.3	13.2	86.5	75.2	28.5	99.9

ARX(1) with single factor

$T = 10, \quad \gamma = 0.4, \quad \beta = 1, \quad f_t \sim \textit{trend} \quad$ (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ γ	0.30	0.40	0.50	0.30	0.40	0.50
ML	85.3	4.4	88.4	100.0	4.9	100.0
ALS1(1step)	47.1	36.1	53.6	79.1	47.5	55.5
NT1(1step)	96.4	99.4	99.8	100.0	100.0	100.0
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	38.2	5.0	37.3	83.9	5.3	86.7
ALS1(1step)	53.2	37.9	32.4	58.2	42.1	47.3
NT1(1step)	21.3	9.5	21.7	49.4	8.0	35.2

ARX(1) with single factor

$T = 10, \quad \gamma = 0.8, \quad \beta = 1, \quad f_t \sim \textit{trend} \quad$ (Size and power(%))

	$N = 150$			$N = 500$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ γ	0.70	0.80	0.90	0.70	0.80	0.90
ML	100.0	5.6	100.0	100.0	5.5	100.0
ALS1(1step)	90.3	28.1	92.3	97.7	16.0	97.7
NT1(1step)	0.1	0.1	49.6	0.0	23.8	100.0
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
Estimators \ β	0.90	1.00	1.10	0.90	1.00	1.10
ML	86.6	5.8	87.5	100.0	5.0	100.0
ALS1(1step)	44.8	22.1	51.6	65.9	14.9	94.7
NT1(1step)	28.1	12.7	82.4	62.0	34.8	99.9

ARX(1) with two factors

- y_{it} for the ARX(1) model is generated as

$$\begin{aligned} y_{it} &= \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}, \quad , \\ \xi_{it} &= \boldsymbol{\lambda}_i' \mathbf{f}_t + u_{it},\ u_{it} \sim iid\mathcal{N}(0, \sigma^2). \end{aligned}$$

  for $i = 1, 2, ..., N$; $t = -S+1, -S+2, .., 0, 1, ..., T$.

- The regressors, x_{it}, are generated as

$$x_{it} = \mu_i + \boldsymbol{\vartheta}_i' \mathbf{f}_t + \breve{x}_{it},\ \ , \breve{x}_{it} = \rho_x \breve{x}_{i,t-1} + \sqrt{1-\rho_x^2}\varepsilon_{it}, \tag{20}$$

  with $\breve{x}_{i,-S} = 0$ for $t = -S+1, ..., 0, 1, ..., T$, where $\boldsymbol{\vartheta}_i = (\vartheta_{1i}, \vartheta_{2i})'$, $\mu_i \sim iid\mathcal{N}(0, 1)$, $\varepsilon_{it} \sim iid\mathcal{N}(0, 1)$.

- $f_{\ell t}$, $\ell = 1, 2$, are generated as in the AR(1) case, and $\rho_x = 0.8$.
- The factor loadings $\boldsymbol{\vartheta}_i = (\vartheta_{1i}, \vartheta_{2i})'$ and $\boldsymbol{\lambda}_i = (\lambda_{1i}, \lambda_{2i})'$ are generated independently as

$$\vartheta_{\ell i} \sim iid\mathcal{N}(0.5, \sigma_{\ell\vartheta}^2),\ \lambda_{\ell i} \sim iid\mathcal{N}(0.5, \sigma_{\ell\lambda}^2),\ \ell = 1, 2, \tag{21}$$

ARX(1) with two factors

- The fixed effects, α_i, are generated so that it is correlated with the regressors, as well as with the errors

$$\alpha_i = \bar{x}_i + \lambda_{1i}\bar{f}_1 + \lambda_{2i}\bar{f}_2 + \bar{u}_i + v_i,$$

  where $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$.
- The remaining parameters are set as $\beta = 1$ and $R_y^2 - \gamma^2 = 0.1$.
- GMM estimators are not reported since it did not perform well in the single factor case.

- The findings are as follows:
  1. The results are similar to the single factor case.
  2. The ML estimator performs well: small bias and RMSE without size distortions.

ARX(1) with two factors

Two factors case (Bias and RMSE($\times$100))

	$T = 6$				$T = 10$			
	$\gamma = 0.4$		$\beta = 1.0$		$\gamma = 0.4$		$\beta = 1.0$	
N	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
150	-0.16	3.82	0.36	5.86	0.00	2.05	0.01	4.14
500	-0.12	2.01	-0.02	3.29	0.00	1.08	0.08	2.21
	$T = 6$				$T = 10$			
	$\gamma = 0.8$		$\beta = 1.0$		$\gamma = 0.8$		$\beta = 1.0$	
N	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
150	0.01	1.46	0.07	2.32	0.01	0.69	0.02	1.57
500	-0.03	0.77	-0.04	1.31	-0.02	0.39	0.01	0.86

ARX(1) with two factors

Two factors case (Size and power(%))

	$T = 6, \gamma = 0.4, \beta = 1.0$			$T = 10, \gamma = 0.4, \beta = 1.0$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
$N \setminus \gamma$	0.30	0.40	0.50	0.30	0.40	0.50
150	78.5	4.9	73.7	99.8	6.0	99.7
500	100.0	4.4	99.8	100.0	5.4	100.0
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
$N \setminus \gamma$	0.90	1.00	1.10	0.90	1.00	1.10
150	36.6	4.6	40.3	68.7	5.3	68.4
500	86.5	5.5	85.6	99.1	5.3	99.2

ARX(1) with two factors

Two factors case (Size and power(%))

	$T=6, \gamma=0.8, \beta=1.0$			$T=10, \gamma=0.8, \beta=1.0$		
	γ					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
$N \setminus \gamma$	0.70	0.80	0.90	0.70	0.80	0.90
150	100.0	4.8	100.0	100.0	4.0	100.0
500	100.0	4.4	100.0	100.0	4.9	100.0
	β					
	Power(H_1)	Size	Power(H_1)	Power(H_1)	Size	Power(H_1)
$N \setminus \gamma$	0.90	1.00	1.10	0.90	1.00	1.10
150	98.9	4.6	98.8	100.0	5.1	100.0
500	100.0	4.7	100.0	100.0	4.6	100.0

5. Conclusion

Conclusion

- This paper studies short dynamic panel data models with interactive fixed effects.
- Extended the transformed ML approach by Hsiao et al. (2002) to include interactive fixed effects.
- Monte Carlo simulations are carried out to investigate the finite sample performance.
- Simulation results revealed that the transformed ML estimator works well in finite sample and performs (sometimes substantially) better than existing GMM estimators.