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

Kazuhiko Hayakawa
Hiroshima University

M. Hashem Pesaran
USC and Trinity College, Cambridge

L. Vanessa Smith
University of York

May 16, 2014

Abstract

This paper proposes the transformed maximum likelihood estimator for short dynamic panel data models with interactive fixed effects, and provides an extension of Hsiao et al. (2002) that allows for a multifactor error structure. This is an important extension since it retains the advantages of the transformed likelihood approach, whilst at the same time allows for observed factors (fixed or random). Small sample results obtained from Monte Carlo simulations show that the transformed ML estimator performs well in finite samples and outperforms the GMM estimators proposed in the literature in almost all cases considered.

JEL Classifications: C12, C13, C23
Keywords: short T dynamic panels, transformed maximum likelihood, multi-factor error structure, interactive fixed effects

---

1   Introduction

There now exists an extensive literature on the estimation of linear dynamic panel data models where the time dimension (T) is short and fixed relative to the cross section dimension (N), which is large. Such panels are usually referred to as *micro panels*, and often arise in microeconometric applications. For example, many empirical applications based on survey data such as the British Household Panel Surveys (BHPS) and the Panel Study in Income Dynamics (PSID) are characterized by data covering relatively short time periods. Although it is now quite common to include dynamics in such studies, it is rare to find studies that allow for error cross section dependence as well. In most empirical applications time dummies are used to deal with cross section dependence, which is valid only if the time effect is homogeneous over the cross section units. Short T panels also arise in the cross country empirical growth literature where data is typically averaged over five to seven years to eliminate the business cycle effects. Both generalized method of moments (GMM) and likelihood approaches have been advanced to estimate such panel data models. See, for example, Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998), Hsiao et al. (2002) and Binder et al. (2005).[1] However, this literature assumes that the errors are cross sectionally independent, which might not hold in many applications where cross section units are subject to common unobserved effects, or possibly spatial or network spillover effects. Ignoring cross section dependence can have important consequences for conventional estimators of dynamic panels. Phillips and Sul (2007) study the impact of cross section dependence modelled as a factor structure on the inconsistency of the pooled least squares estimate of a short dynamic panel regression. Sarafidis and Robertson (2009) investigate the properties of a number of standard widely used generalized method of moments (GMM) estimators under cross section dependence and show that such estimators are inconsistent.

In applications where the spatial patterns are important and can be characterized by known spatial weight matrices, error cross section dependence is typically modelled as spatial autoregressions and estimated jointly with the other parameters of the dynamic panel data model. Lee and Yu (2010) provide a review. For small T, Elhorst (2005) and Su and Yang (2007) consider random effects as well as fixed effects specifications. In the latter case they apply the first-differencing operator to eliminate the fixed effects and then use the transformed likelihood approach of Hsiao et al. (2002) to deal with the initial value problem. The treatment of the initial values in spatial dynamic panel data models poses additional difficulties and requires further investigation. More recently Jacobs et al. (2009) discuss GMM estimation of dynamic fixed effect panel data models featuring spatially correlated errors and endogenous interaction.

However, in addition to the spatial effects it is also likely that the error cross section dependence could be a result of omitted unobserved common factor(s). This class of models has been the subject of intensive research over the past five years and robust estimation procedures have been advanced in the case of panels where N and T are both large.[2] In contrast, little work has been done so far on the estimation of short T dynamic panels where error cross section dependence is due to unobserved common factors. An early contribution by MaCurdy (1982) features panel models with an error structure that combines factor schemes with autoregressive-moving average models estimated by maximum likelihood and used to analyze the error process associated with the earnings of prime

---

[1] The analysis of Hsiao et al. (2002) is extended by Hayakawa and Pesaran (2012) to allow for a cross-sectionally heteroskedastic error term.

[2] See, for example, Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011), Chudik et al. (2011), and Kapetanios et al. (2011).

age males. In subsequent work, for the case of a single factor, Holtz-Eakin et al. (1988) and Ahn et al. (2001), suggest a quasi-differencing approach to purge the factor structure and then use GMM to consistently estimate the model parameters.[3] Nauges and Thomas (2003) follow this approach in addition to prior first-differencing to eliminate the fixed effect, which they consider separately from the single common factor structure assumed for the errors. Ahn et al. (2013) extend this approach to the more general case of a multifactor error structure.

More recently, 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 errors. They show that the resulting estimator is asymptotically more efficient than the GMM estimator based on quasi-differencing as it exploits extra restrictions implied by the model. Elhorst (2010) considers a fixed effects dynamic panel with contemporaneous endogenous interaction effects under small T. For estimation purposes, he adopts both the maximum likelihood estimator of Hsiao et al. (2002) and the GMM estimator of Arellano and Bond (1991). Bai (2013) suggests a quasi-maximum likelihood (ML) approach applied to the original dynamic panel without differencing (simple or quasi), and uses the approach of Mundlak (1978) and Chamberlain (1982) to deal with the correlation between the factor loadings and the regressors, but continues to assume that all factor loadings (including the one associated with the intercepts) are uncorrelated with the errors.[4]

In this paper, following Hsiao et al. (2002), we propose an alternative quasi ML approach applied to the panel data model after first-differencing. In this way, we account for heterogeneity of the initial values and the common factors in an integrated framework. The proposed estimation procedure includes the transformed likelihood procedure of Hsiao et al. (2002) as a special case. It allows for both fixed and interactive effects (the latter based on a random coefficient specification), and can be used to test the validity of the fixed effects specification against the more general model with interactive effects. Our procedure differs from the one proposed by Bai (2013) since he proposes to apply the maximum likelihood (ML) procedure to the level model without time-invariant fixed effects, whilst we propose to apply the ML procedure to the first-differenced model where time-invariant fixed effects are removed. The application of the ML approach to dynamic panel data models without first-differencing requires the fixed effects in the processes generating the regressors to be uncorrelated with the errors. Otherwise, as shown in Hsiao et al. (2002), the initial values (y_{i0}) could be subject to an incidental parameter problem. More specifically, reliance on the Mundlak-Chamberlain device for the specification of y_{i0} employed by Bai (2013) will be valid only under random effects specification of the processes generating the regressors. However, this assumption is not required under the transformed likelihood approach, where the quasi ML approach is applied to first differences. The proposed method can also be readily extended to a panel VAR framework as in Binder et al. (2005). Monte Carlo simulations are carried out to investigate the finite sample performance of the transformed ML estimator including a comparison with several GMM estimators. We find that the transformed ML estimator performs well in almost all cases considered, while the GMM estimators perform (sometimes) substantially poorly.

The rest of this paper is organized as follows. Section 2 sets out the dynamic model (with and without regressors) and develops the transformed likelihood approach. Initially we consider the relatively simple case where in addition to fixed effects the model contains a single unobserved common factor with interactive effects. In subsection 2.3 we extend our analysis to models with multiple factors. In Section 3, a review of the GMM approach is provided. In Section 4, we describe the Monte Carlo experiments and compare bias, root mean square errors , size and power

---

[3]The quasi-differencing transformation was originally proposed by Chamberlain (1984). Holtz-Eakin et al. (1988) implement it in the context of a bivariate panel autoregression.

[4]See also Sarafidis and Wansbeek (2012) for a recent survey of panel data models with error cross section dependence when T is short.

of the proposed transformed ML estimator to a number of different GMM estimators.[5] Section 5 concludes.

2 The Likelihood Approach

2.1 AR(1) model

Consider the following first order autoregressive, AR(1), panel data model

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

where T is fixed and small relative to N which could be large, α_i for $i = 1, 2, ..., N$ are the fixed effects, f_t is an unobserved common factor for all i, u_{it} are the individual-specific (idiosyncratic) errors, λ_i for $i = 1, 2, ..., N$ are factor loadings distributed indepedently of u_{it} and f_t. No restrictions will be imposed on f_t except that $g_t = \Delta f_t \neq 0$ for at least some $t = 1, 2, ..., T$. Note that this requirement does not restrict the specification of the model since the excluded case of $f_t = C$ (a fixed constant for all t) is already covered by the explicit inclusion of fixed effects, α_i, in the model. We consider the problem of estimation of γ under the following assumptions:

Assumption 1 $|\gamma| < 1$ and the AR(1) model given in (1) has started from the infinite past.

Assumption 2 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 3 The unobserved factor loadings, λ_i, are independently and identically distributed across i and of the individual specific errors, u_{jt}, and the common factor, f_t, for all i, j and t with fixed mean, λ, and a finite variance. In particular,

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

Assumption 4 The error terms η_i and u_{it} are normally distributed.

Remark 1 *Assumption 1 is made to simplify the exposition. In the next subsection we consider the case where the dynamic process has started from a finite past. In such a case it is also possible to allow for unit roots, namely the case where $\gamma = 1$.*

Remark 2 *For each i, the composite 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. In our approach f_t can be fixed or random.*

Remark 3 *Under Assumption 4, η_i and u_{it} are considered normally distributed for the application of the ML approach. The normality assumption is not required as $N \to \infty$, so long as the errors η_i and u_{it} have finite fourth-order moments.*

---

[5]In these comparisons we do not include Bai's recent estimator since the computer code for the implementation of this estimation method has not yet been released. Also, the Monte Carlo evidence provided in Bai (2013) is more illustrative in nature and does not cover cases where there are fixed effects in the processes generating the regressors that are correlated with the errors. Further, Bai (2013) does not provide any evidence on size and power of tests based on his proposed estimator. We intend to include Bai's estimation method in our comparative analysis once workable computer codes are released.

Remark 4 *No assumptions are made regarding the fixed effects, α_i. They could be correlated with λ_i and u_{it}, and need not be cross sectionally independent. For example, α_i could follow a spatial autoregressive specification where $cov(\alpha_i, \alpha_j) \neq 0$ for all i and j.*

Under Assumption 3 we can rewrite model (1) as

$$
\begin{aligned}
y_{it} &= \alpha_i + \gamma y_{i,t-1} + \lambda_i f_t + u_{it} \\
&= \alpha_i + \gamma y_{i,t-1} + \lambda f_t + \eta_i f_t + u_{it}.
\end{aligned}
$$

We eliminate the individual effects by first-differencing

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

Under Assumption 1, by recursive substitution, we have the following expression for $t = 1$

$$
\Delta y_{i1} = \lambda_i \tilde{g}_1 + v_{i1},
\tag{4}
$$

where $\tilde{g}_1 = \sum_{j=0}^{\infty} \gamma^j g_{1-j}$, $v_{i1} = \sum_{j=0}^{\infty} \gamma^j \Delta u_{i,1-j}$ with $E(v_{i1}) = 0$ and $var(v_{i1}) = \omega \sigma^2$. Although ω is given by $2/(1+\gamma)$ in this model, in general, we treat ω as a free parameter to be estimated.

To deal with the incidental parameter problem associated with λ_i, instead of quasi-differencing to eliminate λ_i, we use (2) and write (3) and (4) as

$$
\begin{aligned}
\Delta y_{i1} &= \lambda \tilde{g}_1 + \eta_i \tilde{g}_1 + v_{i1} \\
\Delta y_{it} &= \gamma \Delta y_{i,t-1} + \lambda g_t + \eta_i g_t + \Delta u_{it}, \qquad (t = 2, 3, ..., T).
\end{aligned}
$$

In matrix notation the above system of equations can be written as

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

where $\Delta \mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$, $\Delta \mathbf{W}_i = (0, \Delta y_{i1}, ..., \Delta y_{i,T-1})'$, $\mathbf{g} = (\tilde{g}_1, g_2, ..., g_T)'$, and $\boldsymbol{\xi}_i = \eta_i \mathbf{g} + \mathbf{r}_i$, with $\mathbf{r}_i = (v_{i1}, \Delta u_{i2}, ..., \Delta u_{iT})'$. From Hsiao et al. (2002) we have that

$$
E(\mathbf{r}_i \mathbf{r}_i') = \sigma^2
\begin{pmatrix}
\omega & -1 & & & & 0 \\
-1 & 2 & \ddots & & & \\
& & \ddots & & & \\
& & & \ddots & 2 & -1 \\
0 & & & & -1 & 2
\end{pmatrix}
= \sigma^2 \boldsymbol{\Omega}.
\tag{6}
$$

Using (6) and recalling that η_i and u_{it} are independently distributed we have

$$
Var(\boldsymbol{\xi}_i) = \sigma^2 \boldsymbol{\Omega} + \sigma_\eta^2 \mathbf{g}\mathbf{g}' = \sigma^2 \left(\boldsymbol{\Omega} + \phi \mathbf{g}\mathbf{g}' \right),
$$

where

$$
\phi = \frac{\sigma_\eta^2}{\sigma^2}.
$$

Hence, the log-likelihood function of the transformed model (5) is given by

$$
\begin{aligned}
\ell(\boldsymbol{\psi}) = {} & -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln(\sigma^2) - \frac{N}{2} \ln \left| \boldsymbol{\Omega} + \phi \mathbf{g}\mathbf{g}' \right| \\
& -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (\Delta \mathbf{y}_i - \Delta \mathbf{W}_i \gamma - \lambda \mathbf{g})' \left(\boldsymbol{\Omega} + \phi \mathbf{g}\mathbf{g}' \right)^{-1} (\Delta \mathbf{y}_i - \Delta \mathbf{W}_i \gamma - \lambda \mathbf{g}).
\end{aligned}
\tag{7}
$$

The log-likelihood in (7) is a function of a fixed number of unknown parameters, $\boldsymbol{\psi} = (\gamma, \omega, \sigma^2, \phi, \lambda, \mathbf{g}')'$. After some algebra (see Section A.2 of the Appendix) it can be written as

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

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

$$
\mathbf{B}_N = \mathbf{B}_N(\gamma) = N^{-1}\sum_{i=1}^{N}\mathbf{v}_i(\gamma)\mathbf{v}_i'(\gamma). \quad (9)
$$

It is clear that if $\phi = 0$, the log-likelihood simplifies to the case of panels with (pure) time effects and λ is not separately identified from the elements of $\mathbf{g}$. In such a case λ is typically set to unity and T time dummies are introduced to estimate $\mathbf{g}$. In the interactive case where $\phi \neq 0$, $\mathbf{g}$ is not identified separately from ϕ and without loss of generality we can set $\mathbf{q} = \sqrt{\phi}\mathbf{g}$ and write the log-likelihood function in (8) as

$$
\begin{aligned}
N^{-1}\ell(\boldsymbol{\theta}) =\ & -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln(1 + \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}) \\
& -\frac{1}{2\sigma^2}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{1 + \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}}\right],
\end{aligned} \quad (10)
$$

where $\boldsymbol{\theta} = (\gamma, \omega, \sigma^2, \kappa, \mathbf{q}')'$, and $\kappa = \lambda/\sqrt{\phi}$.

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

$$
\hat{\kappa} = \left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^{-1}\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}, \quad (11)
$$

and

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

But using (9) the above expression can also be written as

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

or equivalently as

$$
\hat{\sigma}^2 = T^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega} + \mathbf{q}\mathbf{q}'\right)^{-1}\mathbf{v}_i - \frac{\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)^2}{(1 + \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q})(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q})}\right]. \quad (12)
$$

In practice, $\bar{\mathbf{v}}$ is likely to be small for sufficiently large N, which ensures a positive estimate for σ^2, although this is not guaranteed if N is small.

Substituting (11) and (12) into (10), we have

$$
\begin{aligned}
N^{-1}\ell(\gamma, \omega, \mathbf{q}) \propto\ & -\frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln\left(1 + \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right) \\
& -\frac{T}{2}\ln\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega} + \mathbf{q}\mathbf{q}'\right)^{-1}\mathbf{v}_i - \frac{\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)^2}{(1 + \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q})(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q})}\right].
\end{aligned} \quad (13)
$$

The transformed ML estimator is obtained by maximizing the above concentrated log-likelihood function. Having obtained the ML estimators of γ, ω and $\mathbf{q}$, (which we denote by $\hat{\gamma}$, $\hat{\omega}$ and $\hat{\mathbf{q}}$), the MLE of σ^2 and κ can then be computed using (11) and (12). To compute the variance-covariance matrix of $\hat{\boldsymbol{\theta}} = (\hat{\gamma}, \hat{\omega}, \hat{\sigma}^2, \hat{\kappa}, \hat{\mathbf{q}}')'$ we need to make use of the unconcentrated log-likelihood function given by (10) and compute its second derivatives, either analytically or numerically. For a fixed T and as $N \to \infty$, using standard results from the asymptotic theory of ML estimation we have

$$AsyVar(\sqrt{N}\hat{\boldsymbol{\theta}}) = \mathbf{H}^{-1}(\boldsymbol{\theta}),$$

where (using the unconcentrated log-likelihood function given by (10))

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

A consistent estimator of $AsyVar(\hat{\boldsymbol{\theta}})$ can be obtained as

$$\widehat{Var}(\hat{\boldsymbol{\theta}}) = \left[-\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1}, \tag{14}$$

where the second partial derivatives are evaluated at the MLE, $\hat{\boldsymbol{\theta}} = (\hat{\gamma}, \hat{\omega}, \hat{\sigma}^2, \hat{\kappa}, \hat{\mathbf{q}}')'$. The first and second derivatives of the log-likelihood function are provided in Section A.3 of the Appendix.

2.2 ARX(1) model

Consider next the case where an exogenous variable is included in model (1) and consider the augmented AR(1) model (which we denote by ARX(1))

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \lambda_i f_t + u_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T). \tag{15}$$

For simplicity we assume that x_{it} is a scalar. Extension to the case of multiple regressors is straightforward at the expense of notational complexity. Taking the first-difference of (15) and using Assumption 3 we have

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \lambda g_t + \eta_i g_t + \Delta u_{it}, \qquad (t = 2, 3, ..., T). \tag{16}$$

We assume that the regressor x_{it} is generated either by

$$x_{it} = \mu_i + ct + \vartheta_i f_t + \sum_{j=0}^{\infty} a_j \varepsilon_{i,t-j}, \qquad \sum_{j=0}^{\infty} |a_j| < \infty, \tag{17}$$

or

$$\Delta x_{it} = c + \vartheta_i g_t + \sum_{j=0}^{\infty} d_j \varepsilon_{i,t-j}, \qquad \sum_{j=0}^{\infty} |d_j| < \infty, \tag{18}$$

where μ_i are fixed effects (which could be correlated with u_{it} and/or ε_{it}), and ϑ_i are random interactive effects distributed independently of u_{it} and f_t.

Assumption 5 *The dynamic process given by (16) 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 6 *The interactive effects ϑ_i in Δx_{it} have constant variance $var(\vartheta_i) = \sigma_\vartheta^2$ and are uncorrelated with λ_i and u_{it} for all i and t.*

Assumption 7 *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.*

Remark 5 *Assumption 5 imposes the restriction that the expected changes in the initial values are the same across all individuals, but does not necessarily require that $|\gamma| < 1$ or that all individuals should start from the same position.*

Remark 6 *Assumption 7 requires that x_{it} is strictly exogenous. This can be relaxed by considering vector autoregressions as in Holtz-Eakin et al. (1988). See also Binder et al. (2005).*

Remark 7 *While the time variant individual effects, λ_i, are treated as random they could be correlated with the regressor(s) x_i, such that $\lambda_i = \pi' x_i + \epsilon_i$, so long as the Mudlank-Chamberlain device is used to control for this correlation. However, the $\lambda_i's$ cannot be correlated across i.*

By recursive substitution we have

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

where $\tilde{g}_{1S} = \sum_{j=0}^{S-1} \gamma^j g_{1-j}$. 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 1 *Consider model (16) where x_{it} follows either (17) or (18). Suppose that Assumptions 2, 3, 5, 6, and 7 hold. Then Δy_{i1} can be expressed as:*

$$
\Delta y_{i1} = b + \boldsymbol{\pi}' \Delta \mathbf{x}_i + v_{i1}, \tag{19}
$$

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 $0 < \omega < K < \infty$.

Proof. *See Section A.1 of the Appendix.* ■

Remark 8 *This theorem establishes the conditions under which the Mundlak-Chamberlain specification for the initial observations, Δy_{i1}, is valid. The key condition is the restrictions on the processes generating x_{it} or Δx_{it}. In our application, since we apply first-differencing before ML estimation we can allow for inclusion of fixed effects in the x_{it} process, but we must rule out the presence of fixed effects in the processes generating Δx_{it}. See Assumption 6.*

Using Theorem 1 and (16) the transformed model can be rewritten as

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

where $\boldsymbol{\varphi} = (b, \boldsymbol{\pi}', \gamma, \beta)'$, $\boldsymbol{\xi}_i = \eta_i \mathbf{g} + \mathbf{r}_i$, $\Delta \mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$, $\mathbf{r}_i = (v_{i1}, \Delta u_{i2}, ..., \Delta u_{iT})'$ and

$$
\Delta \mathbf{W}_i = \begin{pmatrix} 1 & \Delta \mathbf{x}_i' & 0 & 0 \\ 0 & \mathbf{0} & \Delta y_{i1} & \Delta x_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \Delta y_{i,T-1} & \Delta x_{iT} \end{pmatrix}. \tag{20}
$$

The rest of the analysis follows identically to the AR(1) case where the final expression for the log-likelihood function, $\bar{\ell}\left(\boldsymbol{\varphi}, \omega, \mathbf{q}\right)$, is given by (13), with the difference that $\mathbf{v}_i$ is now given by

$$\mathbf{v}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\varphi}, \tag{21}$$

where $\boldsymbol{\varphi} = (b, \boldsymbol{\pi}', \gamma, \beta)'$ and $\Delta\mathbf{W}_i$ is defined by (20).

2.3   Extension of the transformed maximum likelihood to the multifactor case

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

$$
\begin{aligned}
y_{it} &= \alpha_i + \gamma y_{i,t-1} + \xi_{it}, \\
\xi_{it} &= \mathbf{f}'_t\boldsymbol{\lambda}_i + u_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T),
\end{aligned}
\tag{22}
$$

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 specified as in Section 2.1. It is assumed that the number of factors m is known and that $m < T$. To accommodate multiple factors the following modified versions of Assumptions 3 and 4 are needed:

Assumption 8   *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,*

$$\boldsymbol{\lambda}_i = \boldsymbol{\lambda} + \boldsymbol{\eta}_i, \; \boldsymbol{\eta}_i \sim IID(\mathbf{0}, \boldsymbol{\Omega}_\eta), \tag{23}$$

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

Assumption 9   *The error terms $\boldsymbol{\eta}_i$ and u_{it} are normally distributed.*

Under Assumptions 8 and 9 and following similar derivations as in the single factor case we have

$$
\begin{aligned}
N\bar{\ell}\left(\boldsymbol{\theta}\right) \; \propto \; & -\frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln\left|\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\right| \\
& -\frac{T}{2}\ln\left\{ \begin{array}{c} N^{-1}\sum_{i=1}^N \mathbf{v}'_i\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i \\ -\bar{\mathbf{v}}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}(\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}} \end{array} \right\},
\end{aligned}
\tag{24}
$$

where $\boldsymbol{\theta} = \left(\gamma, \omega, vec(\mathbf{Q})'\right)'$, $\mathbf{Q} = \sigma^{-1}\mathbf{G}\boldsymbol{\Omega}_\eta^{1/2}$ with $\mathbf{G} = (\tilde{\mathbf{g}}_1, \mathbf{g}_2, ..., \mathbf{g}_T)'$ and $\tilde{\mathbf{g}}_1 = \sum_{j=0}^\infty \gamma^j\mathbf{g}_{1-j}$, and $\mathbf{A} = \mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}$. The restrictions implied by $\mathbf{Q} = \sigma^{-1}\mathbf{G}\boldsymbol{\Omega}_\eta^{1/2}$ 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 $\boldsymbol{\delta} = \sigma\boldsymbol{\Omega}_\eta^{-1/2}\boldsymbol{\lambda}$, and $\mathbf{q}_t$ is the tth row of $\mathbf{Q}$. For details of the derivations see Section A.4 of the Appendix. It is also easily verified that (24) reduces to (13) when $m = 1$.

In the case where the panel data model contains exogenous regressors the form of the log-likelihood function is as in (24), with the difference that $\mathbf{v}_i$ is now defined by (21).

3 The GMM Approach

In this section we provide details of two different GMM estimators proposed in the literature for the estimation of dynamic panel data models with interactive effects. We shall then use these estimators in the Monte Carlo experiments for comparison with the transformed ML estimator proposed in this paper.

3.1 Ahn, Lee and Schmidt (2013)

Ahn et al. (2001) consider a single factor panel model (without specification of a separate fixed effect component) which they extend to the multifactor case in Ahn et al. (2013). While they consider static models with weakly exogenous variables, it is straightforward to extend their analysis to the dynamic case. As Ahn et al. (2001) is a special case of Ahn et al. (2013), we focus on the latter and consider the model

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

where $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}_{it}')'$, $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$, $\tilde{\boldsymbol{\lambda}}_i = (\alpha_i, \lambda_{1i}, ..., \lambda_{mi})'$ and $\tilde{\mathbf{f}}_t = (1, f_{1t}, ..., f_{mt})'$ are $(\tilde{m} \times 1)$ vectors with $\tilde{m} = m + 1$, and ε_{it} are cross-sectionally and temporally uncorrelated. The individual specific effects $\boldsymbol{\lambda}_i$ are allowed to be correlated with $\mathbf{x}_{it}$, while $\mathbf{x}_{it}$ is assumed to be strictly or weakly exogenous. The model in matrix notation can be written as

$$
\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \tilde{\mathbf{F}}\tilde{\boldsymbol{\lambda}}_i + \boldsymbol{\varepsilon}_i, \tag{25}
$$

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 $\tilde{\boldsymbol{\lambda}}_i$, the authors impose $\tilde{m}^2$ restrictions on the factors themselves such that $\tilde{\mathbf{F}} = (\boldsymbol{\Psi}', \mathbf{I}_{\tilde{m}})'$ where $\boldsymbol{\Psi}$ is a $(T - \tilde{m}) \times \tilde{m}$ matrix of unrestricted parameters. Let $\mathbf{H} = (\mathbf{I}_{T-\tilde{m}}, -\boldsymbol{\Psi})'$, so that $\mathbf{H}'\tilde{\mathbf{F}} = (\mathbf{I}_{T-\tilde{m}}, -\boldsymbol{\Psi})(\boldsymbol{\Psi}', \mathbf{I}_{\tilde{m}})' = \mathbf{0}_{(T-\tilde{m})\times\tilde{m}}$. Then, pre-multiplying equation (25) by $\mathbf{H}'$ removes the unobservable effects so that

$$
\mathbf{H}'\mathbf{y}_i = \mathbf{H}'\mathbf{W}_i\boldsymbol{\delta} + \mathbf{H}'\boldsymbol{\varepsilon}_i,
$$

or

$$
\begin{aligned}
\dot{\mathbf{y}}_i &= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \boldsymbol{\Psi}\ddot{\mathbf{y}}_i - \boldsymbol{\Psi}\ddot{\mathbf{W}}_i\boldsymbol{\delta} + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{\varepsilon}}_i \\
&= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \left(\mathbf{I}_{T-\tilde{m}} \otimes \ddot{\mathbf{y}}_i'\right) vec(\boldsymbol{\Psi}) - \left(vec(\ddot{\mathbf{W}}_i)' \otimes \mathbf{I}_{T-\tilde{m}}\right) vec(\boldsymbol{\delta}' \otimes \boldsymbol{\Psi}) + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{\varepsilon}}_i,
\end{aligned}
$$

where $\dot{\mathbf{y}}_i = (y_{i1}, ..., y_{i,T-\tilde{m}})'$, $\ddot{\mathbf{y}}_i = (y_{i,T-\tilde{m}+1}, ..., y_{iT})'$, $\dot{\mathbf{W}}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{i,T-\tilde{m}})'$, $\ddot{\mathbf{W}}_i = (\mathbf{w}_{i,T-\tilde{m}+1}, ..., \mathbf{w}_{iT})'$, $\boldsymbol{\Psi}' = (\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_{T-\tilde{m}})$, $\dot{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i1}, ..., \varepsilon_{i,T-\tilde{m}})'$, and $\ddot{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i,T-\tilde{m}+1}, ..., \varepsilon_{iT})'$.

The tth equation is given by

$$
y_{it} = \boldsymbol{\delta}'\mathbf{w}_{it} + \boldsymbol{\psi}_t'\ddot{\mathbf{y}}_i - \boldsymbol{\psi}_t'\ddot{\mathbf{W}}_i\boldsymbol{\delta} + v_{it}, \qquad (i = 1, ..., N; t = 1, ..., T - \tilde{m}), \tag{26}
$$

where $v_{it} = (\varepsilon_{it} - \boldsymbol{\psi}_t'\ddot{\boldsymbol{\varepsilon}}_i)$. Then, if $\mathbf{x}_{it}$ is strictly exogenous, we end up with $(T - \tilde{m})(T - \tilde{m} + 1)/2 + kT(T - \tilde{m})$ moment conditions given by $E[\mathbf{z}_{it}v_{it}] = \mathbf{0}$, for $t = 1, ..., T - \tilde{m}$, where $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-1}, \mathbf{x}_{i1}', ..., \mathbf{x}_{iT}')'$. In matrix notation the moment conditions can be written as $E[\mathbf{Z}_i'\mathbf{v}_i(\boldsymbol{\theta})] = \mathbf{0}$, where $\mathbf{Z}_i = diag(\mathbf{z}_{i1}', ..., \mathbf{z}_{i,T-\tilde{m}}')$, $\mathbf{v}_i(\boldsymbol{\theta}) = (v_{i1}, ..., v_{i,T-\tilde{m}})'$ and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\psi}')'$ with $\boldsymbol{\psi} = vec(\boldsymbol{\Psi})$.

Then the one-step and two-step GMM estimators are given respectively by

$$\hat{\boldsymbol{\theta}}_{1step} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{v}_i(\boldsymbol{\theta}) \right),$$

and

$$\hat{\boldsymbol{\theta}}_{2step} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{v}_i(\hat{\boldsymbol{\theta}}_{1step}) \mathbf{v}_i(\hat{\boldsymbol{\theta}}_{1step})' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{v}_i(\boldsymbol{\theta}) \right). \quad (27)$$

The continuous updating GMM estimator (CUE) is given by

$$\hat{\boldsymbol{\theta}}_{CUE} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{v}_i(\boldsymbol{\theta}) \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{v}_i(\boldsymbol{\theta}) \right). \quad (28)$$

The asymptotic covariance matrix of the above estimator is given, respectively, by

$$Var(\hat{\boldsymbol{\theta}}_{1step}) = N^{-1} \left(\hat{\mathbf{G}}_{1step}' \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{1step} \right)^{-1} \hat{\mathbf{G}}_{1step}' \hat{\mathbf{W}}^{-1} \hat{\boldsymbol{\Omega}}_{1step} \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{1step} \left(\hat{\mathbf{G}}_{1step}' \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{1step} \right)^{-1}$$
(29)

$$Var(\hat{\boldsymbol{\theta}}_{2step}) = N^{-1} \left(\hat{\mathbf{G}}_{2step}' \hat{\boldsymbol{\Omega}}_{2step}^{-1} \hat{\mathbf{G}}_{2step} \right)^{-1}, \quad (30)$$

and

$$Var(\hat{\boldsymbol{\theta}}_{CUE}) = N^{-1} \left(\hat{\mathbf{G}}_{CUE}' \hat{\boldsymbol{\Omega}}_{CUE}^{-1} \hat{\mathbf{G}}_{CUE} \right)^{-1}, \quad (31)$$

where $\hat{\mathbf{G}}_j = \partial \bar{\mathbf{g}}(\hat{\boldsymbol{\theta}}_j) / \partial \boldsymbol{\theta}'$ for $j = 1step, 2step, CUE$, with $\mathbf{g}_i(\hat{\boldsymbol{\theta}}_j) = \mathbf{Z}_i' \mathbf{v}_i(\hat{\boldsymbol{\theta}}_j)$ and $\bar{\mathbf{g}}(\hat{\boldsymbol{\theta}}_j) = N^{-1} \sum_{i=1}^{N} \mathbf{g}_i(\hat{\boldsymbol{\theta}}_j)$, $\hat{\mathbf{W}} = N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i$, and $\hat{\boldsymbol{\Omega}}_j = N^{-1} \sum_{i=1}^{N} \mathbf{g}_i(\hat{\boldsymbol{\theta}}_j) \mathbf{g}_i(\hat{\boldsymbol{\theta}}_j)'$. The derivatives involved in $\hat{\mathbf{G}}_j$ are computed numerically.

3.2 Nauges and Thomas (2003)

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

$$y_{it} = \mathbf{w}_{it}' \boldsymbol{\delta} + u_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T), \quad (32)$$

where $u_{it} = \alpha_i + \lambda_i f_t + \varepsilon_{it}$. It is assumed that $|\gamma| < 1$ with the initial values, y_{i0}, treated as given. It is further assumed that

$$E(\alpha_i) = 0, \ E(\lambda_i) = 0, \ E(\varepsilon_{it}) = 0,$$

and

$$E(y_{i0}\varepsilon_{it}) = 0, \ E(\alpha_i \varepsilon_{it}) = 0, \ E(\lambda_i \varepsilon_{it}) = 0, \ E(\varepsilon_{it}\varepsilon_{is}) = 0, \quad (33)$$

for $i = 1, 2, ..., N$, $t = 1, 2, ..., T$ and $t \neq s$. As a first step they first difference to eliminate α_i so that (32) becomes

$$\Delta y_{it} = \boldsymbol{\delta}' \Delta \mathbf{w}_{it} + \Delta u_{it}, \quad (34)$$

where

$$\Delta u_{it} = \lambda_i g_t + \Delta \varepsilon_{it},$$

and $g_t = \Delta f_t$. In the second step, following Holtz-Eakin et al. (1988), they perform a quasi-differencing transformation to obtain

$$\Delta u_{it} - r_t \Delta u_{i,t-1} = \Delta \varepsilon_{it} - r_t \Delta \varepsilon_{i,t-1}, \qquad (i = 1, 2, ..., N; \ t = 3, 4, ..., T),$$

10

where $r_t = g_t/g_{t-1} = (f_t - f_{t-1})/(f_{t-1} - f_{t-2})$. Using (34) it follows that

$$v_{it} = \Delta u_{it} - r_t\Delta u_{i,t-1} = (\Delta y_{it} - r_t\Delta y_{i,t-1}) - \boldsymbol{\delta}'(\Delta\mathbf{w}_{it} - r_t\Delta\mathbf{w}_{i,t-1}).$$

Under the conditions set out in (33), if x_{it} is strictly exogenous, the following $(T-2)(T-1)/2+kT$ $(T-2)$ moment conditions hold:

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

where $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-3}; \mathbf{x}'_{i1}, ..., \mathbf{x}'_{iT})'$. These moment conditions are non-linear in the parameters, as the nuisance parameters r'_ts are estimated jointly with the parameter of interest, $\boldsymbol{\delta}$. The moment conditions in matrix notation can be written as

$$E\left[\mathbf{Z}'_i\mathbf{v}_i(\boldsymbol{\theta})\right] = \mathbf{0},$$

where $\mathbf{Z}_i = diag(\mathbf{z}'_{i3}, ..., \mathbf{z}'_{iT})$ and $\mathbf{v}_i(\boldsymbol{\theta}) = (v_{i3}, ..., v_{iT})'$. Based on the above orthogonality conditions, and starting from some initial estimate of $\boldsymbol{\theta}$, in the first step a consistent GMM estimator of the parameter of interest is obtained as

$$\hat{\boldsymbol{\theta}}_{1step} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i(\boldsymbol{\theta})'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{H}\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{v}_i(\boldsymbol{\theta})\right),$$

where $\mathbf{H}$ is a matrix with 2's on the main diagonal, -1's on the first sub-diagonal and 0's elsewhere. Two-step and continuous-updating GMM estimators are obtained similarly to (27) and (28), respectively. The asymptotic covariance matrix is obtained similarly to (29), (30) and (31).

4 Monte Carlo designs

We investigate by means of Monte Carlo simulations the finite sample properties of the transformed likelihood approach and compare them to those of the GMM estimators of Ahn, Lee and Schmidt (2013, ALS) and Nauges and Thomas (2003, NT) described above. We begin by considering the simple AR(1) model followed by the ARX(1) model with an exogenous regressor.

4.1 AR(1) model with a single factor

In this case the observations on y_{it} are generated as

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

where $|\gamma| < 1$. To ensure that y_{i0} are correlated with the fixed effects, α_i, and the error terms, ξ_{it}, we assume that the AR(1) processes have started at time $t = -S$ with starting values $y_{i,-S}$. It is then easily seen that

$$y_{i0} = \frac{1-\gamma^S}{1-\gamma}\alpha_i + \gamma^S y_{i,-S} + \sum_{j=0}^{S-1}\gamma^j\xi_{i,-j},$$

and with S sufficiently large we have

$$y_{i0} \approx \frac{1}{1-\gamma}\alpha_i + \sum_{j=0}^{S-1}\gamma^j\xi_{i,-j}.$$

To deal with the initial values for each i we generate the $T + 1 + S$ observations $t = -S + 1, -S + 2, ..., 0, 1, ..., T$ using $y_{i,-S} = 0$ and discard the first $S = 50$, and use the remaining $T+1$ observations in estimation and inference.

For the unobserved common factor, f_t, we consider a determinstic and a stochastic option:

$$f_t = \begin{cases} 0 & t = -S + 1, ..., -1, 0 \\ t & t = 1, 2, ..., T \end{cases},$$

and

$$f_t = \rho_f f_{t-1} + \sqrt{1 - \rho_f^2}\varepsilon_{ft}, \ \varepsilon_{ft} \sim iidN(0, 1), \text{ for } t = -S + 1, ..., -1, 0, 1, .., T.$$

We consider a relatively persistent case where $\rho_f = 0.9$ and without loss of generality set $f_{-S} = 0$.

Under both specifications of f_t we also scale the resultant f_t values such that $T^{-1}\sum_{t=1}^T f_t^2 = 1$. The values f_t for $t = -S + 1, ..., -1, 0$ are not scaled. The scaling is done to ensure a particular average value of fit as explained below. In all experiments each f_t is generated once and the same $f_t's$ are used in all replications of a given experiment.

The factor loadings, λ_i, are generated independently of the error terms as

$$\lambda_i = \lambda + \eta_i \ \text{ with } \ \lambda = 1 \text{ and } \eta_i \sim iidN(0, 1).$$

However, the fixed effects, α_i, are allowed to be correlated with the errors by generating them as

$$\alpha_i = T^{-1}(\xi_{i1} + \xi_{i2} + ... + \xi_{iT}) + v_i = \lambda_i \bar{f} + \bar{u}_i + v_i,$$

where $\bar{f} = T^{-1}\sum_{t=1}^T f_t$, $\bar{u}_i = T^{-1}\sum_{t=1}^T u_{it}$ and $v_i \sim iidN(0, 1)$. Thus, the fixed effects are correlated with the errors in contrast to the factor loadings, λ_i, that are generated independently of all the other random variables influencing y_{it}. Note that both options of generating the unobserved factors yield a non-zero value for $\bar{f}$, and the (correlated) fixed effects specification can not be generated simply by setting $f_t = 1$. This is because our approach to dealing with the unobserved common factors rules out the factor loadings to be correlated with the errors, u_{it}, whilst we do not rule out correlation between the fixed effects and the errors.

Finally, as shown in Section A.5 of the Appendix, the average fit of the panel AR(1) model is determined by γ and does not depend on $\sigma_u^2 = Var(u_{it})$, and hence we set $\sigma_u^2 = 1$. For the key parameter of the model, γ, we consider a medium and a high value, namely $\gamma = 0.4$ and 0.8, and consider the following combinations of sample sizes, $T = \{6, 10\}$ and $N = \{150, 300, 500\}$. For the GMM estimators of Ahn et al. (2013, ALS) and Nauges and Thomas (2003, NT) we report results for the one-step, two-step and CU GMM estimators. $T = 6$ is the smallest value for which the ALS GMM estimators are computable. For inference we use the standard errors computed based on the second derivative of the log-likelihood function given in (14) for the ML estimator. For the GMM estimators, we use the conventional formulas given in (29), (30) and (31). All derivatives are evaluated numerically.

We report simulation results for the autoregressive parameter γ. Specifically, we report the bias and root mean square error (RMSE). In addition, we present size and power estimates. The power is computed at $\{\gamma \pm 0.10, \gamma \pm 0.05\}$ for the null values of $\gamma = \{0.4, 0.8\}$. All tests are carried out at the 5% significance level and all experiments are replicated 1,000 times.

4.1.1 Results for the AR(1) case

The simulation results for the AR(1) case are presented in Tables 1 to 4.[6],[7] In terms of bias and RMSE, the transformed ML estimator performs well for all cases. As the sample size N and/or T increases, the RMSE decreases irrespective of the value of the autoregressive parameter γ and the specification used for f_t. With regard to inference, the ML estimator performs well in that it has correct size for all combinations of N and T. Power performance is satisfactory though there is the tendency for the ML estimator to display low power for small positive departures from the null. For example, when $\gamma = 0.8$, $T = 6$ and $N = \{150, 300\}$, the power is quite low for the alternative $\gamma = 0.9$ when testing the null $\gamma = 0.8$. This tendency is also evident when f_t is generated as a time trend. Contrary to the well behaved finite sample properties of the transformed ML estimator, the performance of the GMM estimators are not generally good. In terms of bias and RMSE, the GMM estimators are substantially worse than the transformed ML estimator. With regard to size, the one-step ALS-GMM estimator displays empirical sizes close to the nominal level in many cases. However, its power is much lower as compared to that of the transformed ML estimator.

4.2 ARX(1) model with a single factor

The observations on y_{it} for the ARX(1) model are generated as

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

As in the AR(1) case, for values of $|\gamma|$ not too close to unity we set $y_{i,-S} = 0$ and note that for S sufficiently large

$$
y_{i0} \approx \frac{1}{1-\gamma}\alpha_i + \beta \sum_{j=0}^{S-1} \gamma^j x_{i,-j} + \sum_{j=0}^{S-1} \gamma^j \xi_{i,-j}.
$$

The regressors, x_{it}, are generated as

$$
x_{it} = \mu_i + \vartheta_i f_t + \breve{x}_{it}, \, , \breve{x}_{it} = \rho_x \breve{x}_{i,t-1} + \sqrt{1-\rho_x^2}\varepsilon_{it}, \tag{35}
$$

with $\breve{x}_{i,-S} = 0$, for $t = -S+1, ..., 0, 1, ..., T$, where $|\rho_x| < 1$, $\mu_i \sim iid\mathcal{N}(0, 1)$, $\varepsilon_{it} \sim iid\mathcal{N}(0, 1)$ and f_t is generated as in the AR(1) case. We set $\rho_x = 0.8$ which yields relatively persistence regressors.

We generate the factor loadings independently as

$$
\vartheta_i \sim iid\mathcal{N}(0.5, \sigma_\vartheta^2), \, \lambda_i \sim iid\mathcal{N}(0.5, \sigma_\lambda^2), \tag{36}
$$

---

[6] For the starting values in the optimization routine used to compute the ML estimators, we use $\boldsymbol{\theta}_{ini} = (\gamma_{ini}, \omega_{ini}, \mathbf{q}'_{ini})'$ with $\gamma_{ini} \sim U[-0.999, 0.999]$, $\omega_{ini} \sim U[1, 2]$ and $q_{t,ini} \sim U[-1, 1]$ where $q_{t,ini}$ is the tth element of $\mathbf{q}_{ini}$. In addition ω needs to satisfy $\omega > (T-1)/T$ since $|\boldsymbol{\Omega}| = 1 + T(\omega - 1) > 0$. Specifically, we use five such sets of random starting values and choose the largest among the maximum of the log-likelihood values as the estimate of the ML estimator. Similarly, for the one-step ALS and NT GMM estimators we use five sets of starting values $\boldsymbol{\theta}_{ini,ALS} = (\gamma_{ini}, \boldsymbol{\psi}'_{ini})'$ and $\boldsymbol{\theta}_{ini,NT} = (\gamma_{ini}, \mathbf{r}'_{ini})'$ respectively, where $\gamma_{ini} \sim U[-0.999, 0.999]$, $\psi_{t,ini} \sim U[-1, 1]$ with $\psi_{t,ini}$ the tth element of $\boldsymbol{\psi}_{ini}$, and $r_{t,ini} \sim U[-1, 1]$ with $r_{t,ini}$ the tth element of $\mathbf{r}_{ini}$. We select the smallest among the minimum values of the objective function as the estimate of the one-step ALS and NT GMM estimators. For the two-step and continuous-updating ALS and NT GMM estimators we use the one-step estimates as the starting value of the optimization routine.

[7] In certain cases, the Hessian evaluated at the global maximum for the ML estimator was not positive definite. The simulation draw for these cases was discarded and an additional draw was generated until the total number of simulations with a positive definite Hessian reached 1,000. The number of these additional draws decreased for a fixed T as N increased, and as T increased for all N.

and to ensure that the fixed effects, α_i, are correlated with the regressors, as well as with the errors, we generate them as

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

where as in the AR(1) case, $\bar{f} = T^{-1} \sum_{t=1}^{T} f_t$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$ and $v_i \sim iid\mathcal{N}(0,1)$.

We set the remaining parameters bearing in mind that in the case of ARX(1) panels the average R^2 is at least as large as γ^2. In particular, from the results for the R^2 derived in Section A.5 of the Appendix we have that

$$R_y^2 = \frac{\beta^2 Var(\check{x}_{it}) + \left[\left(N^{-1} \sum_{i=1}^{N} c_i^2 \right) \left(T^{-1} \sum_{t=1}^{T} f_t^2 \right) + \sigma^2 \right] \gamma^2}{\beta^2 Var(\check{x}_{it}) + \left(N^{-1} \sum_{i=1}^{N} c_i^2 \right) \left(T^{-1} \sum_{t=1}^{T} f_t^2 \right) + \sigma^2} \geq \gamma^2,$$

with the equality holding when $\beta = 0$ and where $c_i = \beta \vartheta_i + \lambda_i$. In view of (35) $Var(\check{x}_{it}) = 1$ and without loss of generality we set $\beta = 1$. Also, recall that $T^{-1} \sum_{t=1}^{T} f_t^2 = 1$. For comparability with the AR(1) case we set $\gamma = (0.4, 0.8)$ and determine σ^2, σ_λ^2, and σ_ϑ^2 such that $R_y^2 - \gamma^2 = 0.1$. To this end we note that

$$R_y^2 - \gamma^2 = \frac{(1 - \gamma^2)}{1 + N^{-1} \sum_{i=1}^{N} c_i^2 + \sigma^2} = 0.1.$$

Further, for sufficiently large N and noting that λ_i and ϑ_i are generated independently (see (36)) it follows that

$$N^{-1} \sum_{i=1}^{N} c_i^2 \to_p \beta^2 \sigma_\vartheta^2 + \sigma_\lambda^2 + \frac{1}{4}(1 + \beta)^2.$$

Hence with $\beta = 1$ we have

$$R_y^2 - \gamma^2 = \frac{(1 - \gamma^2)}{2 + \sigma_\vartheta^2 + \sigma_\lambda^2 + \sigma^2} = 0.1.$$

We set $\sigma_\lambda^2 = \sigma_\vartheta^2 = \sigma^2$ and using the above result we obtain

$$\sigma^2 = \frac{0.8 - \gamma^2}{0.3} > 0.$$

Finally, we consider the same combinations of T and N as in the AR(1) case, namely $T = \{6, 10\}$ and $N = \{150, 300, 500\}$, and discard the first 50 observations basing estimation on the remaining observations over the period $t = 0, 1, ...T$. Note that after first-differencing we end up with T observations for estimation of γ and β. The standard errors used for inference are based on the same formulas as those used in the AR(1) case with all derivatives computed numerically.

We report simulation results for the same set of statistics as in the AR(1) case, for both γ and β, including size and power. Power is computed for the null values of $(\gamma, \beta) = \{0.4, 1.0\}$ and $(\gamma, \beta) = \{0.8, 1.0\}$. As previously, all tests are carried out at the 5% significance level and all experiments are replicated 1,000 times.

Under strict exogeneity, for the ALS and NT GMM estimators there are so many moment conditions and using all of them causes a large finite sample bias. Hence, we use only a subset of moment conditions for the exogenous variable x_{it}. Specifically, for ALS GMM we use $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-1}, x_{it}, ..., x_{iT})'$, since $\mathbf{w}_{it}$ and $\ddot{\mathbf{W}}_{it}$ in (26) contain x_{it} and $x_{i,T-m}, ..., x_{iT}$. Similarly, for NT GMM we use $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-m-2}, x_{i1}, ..., x_{it})'$. Recall that m is the number of unobserved factors which, in the case of current experiments, is set to 1.

14

4.2.1 Results for the ARX(1) case

Simulation results for the ARX(1) model are summarized in Tables 5 to 8.[8] In terms of bias and RMSE, the results are very similar to the AR(1) case. As the sample size increases, the RMSE decreases in all cases. The sizes are close to the nominal level in all cases and, contrary to the AR(1) case, the power is reasonably high even for $\gamma = 0.8$ and $N = 150$ irrespective of the specification of f_t. The augmentation of the AR(1) model with exogenous regressors has also benefited the GMM estimators who show improved performance as compared to the results obtained for the AR(1) model. However, the transformed ML estimator continues to outperform the GMM estimators (sometimes substantially) both in terms of bias and RMSE. In terms of size, all the GMM estimators exihibit large size distortions in almost all cases. An exception is the one-step NT-GMM with $T = 6$, $\gamma = 0.8$ and $f_t \sim$AR(1). In this case, the empirical size is close to the nominal one, but power is lower than the transformed ML estimator.

4.3 AR(1) model with two factors

The observations on y_{it} for the AR(1) model are generated as

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

where $\mathbf{f}_t = (f_{1t}, f_{2t})'$ and $\boldsymbol{\lambda}_i = (\lambda_{1i}, \lambda_{2i})'$, with the initial values of y_{it} for $|\gamma| < 1$ dealt with as in the single factor case.

The unobserved common factors, $f_{\ell t}$, are generated as

$$
f_{\ell t} = \rho_{f\ell} f_{\ell, t-1} + \sqrt{1 - \rho_{f\ell}^2}\, \varepsilon_{f\ell t}, \ \varepsilon_{f\ell t} \sim iid\mathcal{N}(0, 1), \text{ for } \ell = 1, 2; t = -S+1, ..., -1, 0, 1, .., T,
$$

with $\rho_{f\ell} = 0.9$, and without loss of generality $f_{\ell, -S} = 0$. As in the single factor case, we scale the resultant $f_{\ell t}$ values such that $T^{-1} \sum_{t=1}^{T} f_{\ell t}^2 = 1$ (the past values $f_{\ell t}$ for $t = -S+1, ..., -1, 0$ are not scaled) to ensure a particular average value of fit.

The factor loadings, $\boldsymbol{\lambda}_i = (\lambda_{1i}, \lambda_{2i})'$ are generated independently of the error terms and all other variables influencing y_{it} as

$$
\lambda_{\ell i} = \lambda + \eta_{\ell i}, \quad \text{with} \ \ \lambda = 1 \text{ and } \eta_{\ell i} \sim iid\mathcal{N}(0, 1).
$$

The fixed effects, α_i, are allowed to be correlated with the errors by generating them as

$$
\alpha_i = T^{-1}(\xi_{i1} + \xi_{i2} + ... + \xi_{iT}) + v_i = \lambda_{1i} \bar{f}_1 + \lambda_{2i} \bar{f}_2 + \bar{u}_i + v_i,
$$

where $\bar{f}_\ell = T^{-1} \sum_{t=1}^{T} f_{\ell t}$, $\ell = 1, 2$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and $v_i \sim iid\mathcal{N}(0, 1)$.

As mentioned earlier, since the average fit of the panel AR(1) model is solely determined by γ (a result which holds irrespective of the number of factors) we set $\sigma_u^2 = 1$.

---

[8] As starting values, in the case of the ML estimation we use $\boldsymbol{\theta}_{ini} = (\boldsymbol{\varphi}_{ini}', \omega_{ini}, \mathbf{q}_{ini}')'$ with $\boldsymbol{\varphi}_{ini} = (b_{ini}, \boldsymbol{\pi}_{ini}', \gamma_{ini}, \beta_{ini})'$, where b_{ini} and $\boldsymbol{\pi}_{ini}$ are obtained as the OLS estimates of (19), $\beta_{ini} \sim U[0, 1]$, and the remaining parameters are generated as in the AR(1) case using five sets of starting values. For the one-step ALS and NT GMM estimators we use $\boldsymbol{\theta}_{ini,ALS} = (\gamma_{ini}, \beta_{ini}, \boldsymbol{\psi}_{ini}')'$ and $\boldsymbol{\theta}_{ini,NT} = (\gamma_{ini}, \beta_{ini}, \mathbf{r}_{ini}')'$ respectively, where $\beta_{ini} \sim U[0, 1]$, and the remaining parameters are generated as in the AR(1) case using five sets of starting values. For the CUE, for both ALS and NT we use the parameter estimates obtained from the one-step GMM.

4.3.1 Results for the AR(1) case

Simulation results for the AR(1) model are provided in Tables 9 and 10. Since the single factor results showed that the GMM estimators do not work well, we consider here the transformed ML estimator only. From the tables we find that the behaviour of the proposed estimator for the two factor case is similar to that of the single factor case. In particular, the bias of the transformed ML estimator is very small and RMSE decreases as N increases. In terms of inference, sizes are close to the nominal level and power is relatively high except for some cases with $\gamma = 0.8$.

4.4 ARX(1) model with two factors

The dependent variable, y_{it}, for the ARX(1) model is generated as

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

The regressors, x_{it}, are generated as

$$
x_{it} = \mu_i + \boldsymbol{\vartheta}_i' \mathbf{f}_t + \breve{x}_{it}, \ , \breve{x}_{it} = \rho_x \breve{x}_{i,t-1} + \sqrt{1 - \rho_x^2} \varepsilon_{it}, \tag{37}
$$

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

$$
\vartheta_{\ell i} \sim iidN(0.5, \sigma_{\ell\vartheta}^2), \ \lambda_{\ell i} \sim iidN(0.5, \sigma_{\ell\lambda}^2), \ \ell = 1, 2, \tag{38}
$$

and to ensure that the fixed effects, α_i, are correlated with the regressors, as well as with the errors, as in the single factor case we generate them as

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

where $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$, and the remaining parameters are set as in the two factor AR(1) model.

In setting the remaining parameters, using results in Section A.5 of the Appendix, for the two factor case we have

$$
R_y^2 = \frac{\beta^2 Var(\breve{x}_{it}) + \left[\left(N^{-1} \sum_{i=1}^N c_{1i}^2\right) \left(T^{-1} \sum_{t=1}^T f_{1t}^2\right) + \left(N^{-1} \sum_{i=1}^N c_{2i}^2\right) \left(T^{-1} \sum_{t=1}^T f_{2t}^2\right) + \sigma^2 \right] \gamma^2}{\beta^2 Var(\breve{x}_{it}) + \left(N^{-1} \sum_{i=1}^N c_{1i}^2\right) \left(T^{-1} \sum_{t=1}^T f_{1t}^2\right) + \left(N^{-1} \sum_{i=1}^N c_{2i}^2\right) \left(T^{-1} \sum_{t=1}^T f_{2t}^2\right) + \sigma^2} \geq \gamma^2,
$$

where $c_{\ell i} = \beta \vartheta_{\ell i} + \lambda_{\ell i}$, $\ell = 1, 2$. From (37) we have that $Var(\breve{x}_{it}) = 1$ and we set $\beta = 1$. For comparability with the AR(1) case $\gamma = (0.4, 0.8)$ and σ^2, $\sigma_{\ell\lambda}^2$, and $\sigma_{\ell\vartheta}^2$, $\ell = 1, 2$, are determined such that $R_y^2 - \gamma^2 = 0.1$, as in the single factor case. Thus, recalling that $T^{-1} \sum_{t=1}^T f_{\ell t}^2 = 1$

$$
R_y^2 - \gamma^2 = \frac{1 - \gamma^2}{1 + N^{-1} \sum_{i=1}^N c_{1i}^2 + N^{-1} \sum_{i=1}^N c_{2i}^2 + \sigma^2} = 0.1,
$$

and for sufficiently large N since $\boldsymbol{\lambda}_i$ and $\boldsymbol{\vartheta}_i$ are generated independently (see (38)) we have

$$
N^{-1} \sum_{i=1}^N c_{\ell i}^2 \to_p \beta^2 \sigma_{\ell\vartheta}^2 + \sigma_{\ell\lambda}^2 + \frac{1}{4}(1 + \beta)^2, \text{ for } \ell = 1, 2.
$$

For $\beta = 1$ we then obtain

$$
R_y^2 - \gamma^2 = \frac{1 - \gamma^2}{3 + \sigma_{1\vartheta}^2 + \sigma_{1\lambda}^2 + \sigma_{2\vartheta}^2 + \sigma_{2\lambda}^2 + \sigma^2} = 0.1.
$$

Setting $\sigma_{1\vartheta}^2 = \sigma_{1\lambda}^2 = \sigma_{2\vartheta}^2 = \sigma_{2\lambda}^2 = \sigma^2$ and using the above result yields

$$
\sigma^2 = \frac{0.7 - \gamma^2}{0.5} > 0.
$$

4.4.1 Results for the ARX(1) case

Simulation results for the ARX(1) model are provided in Tables 11 and 12. As in the AR(1) case only the transformed ML estimator is considered. The results show that bias is very small and that RMSE decreases as N and T increase. In addition, size is close to its nominal value and power is high in all cases.

5 Conclusion

In this paper we proposed the transformed maximum likelihood estimator for short dynamic panel data models with interactive fixed effects. This is a natural extension of Hsiao, Pesaran, and Tahmiscioglu (2002) to incorporate a factor structure in the error, while retaining the advantages of the transformed likelihood approach. Monte Carlo simulations were carried out to investigate the finite sample behaviour of the proposed estimator and to compare its performance with several GMM estimators available in the literature. The simulation results showed that the ML estimator performs well in finite samples and outperforms the GMM estimators in almost all cases considered. In our analysis we assumed that the number of factors is known. Estimating the number of factors in the current setting where T is short and N tends to infinity is a topic for future research.

Table 1: Bias($\times 100$) and RMSE($\times 100$) for the AR(1) model with a single factor ($T = 6$)

	$T = 6, \gamma = 0.4, f_t \sim$AR(1)					
	$N = 150$		$N = 300$		$N = 500$	
Estimator	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)
ML	0.34	6.26	0.01	4.27	-0.16	3.31
ALS(1step)	-17.20	33.97	-16.65	30.94	-17.82	29.22
ALS(2step)	-15.94	32.21	-16.14	29.60	-16.62	27.89
ALS(CUE)	-16.51	33.99	-14.99	29.36	-17.08	28.87
NT(1step)	-58.44	60.78	-60.38	61.31	-61.05	61.62
NT(2step)	-57.94	60.32	-60.58	61.38	-61.31	61.76
NT(CUE)	-64.14	66.30	-65.22	65.72	-65.35	65.64
	$T = 6, \gamma = 0.8, f_t \sim$AR(1)					
	$N = 150$		$N = 300$		$N = 500$	
Estimator	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)
ML	-0.14	7.35	0.04	5.63	0.15	4.71
ALS(1step)	-34.26	48.24	-28.17	39.85	-27.57	37.15
ALS(2step)	-35.26	49.17	-29.50	40.56	-28.70	37.97
ALS(CUE)	-33.98	50.34	-27.33	40.68	-26.74	37.55
NT(1step)	-59.14	85.47	-65.45	89.77	-71.95	93.70
NT(2step)	-56.65	82.48	-60.66	84.75	-66.43	87.70
NT(CUE)	-57.22	83.70	-58.49	81.90	-61.448	83.58
	$T = 6, \gamma = 0.4, f_t \sim trend$					
	$N = 150$		$N = 300$		$N = 500$	
Estimator	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)
ML	0.23	8.12	-0.02	5.45	0.05	4.16
ALS(1step)	-19.00	33.04	-19.87	27.96	-19.84	24.86
ALS(2step)	-18.62	32.37	-19.63	27.68	-19.11	23.95
ALS(CUE)	-18.83	34.47	-19.58	28.67	-18.95	24.38
NT(1step)	-12.74	54.24	-20.47	58.26	-28.74	60.59
NT(2step)	-13.36	55.40	-21.22	59.06	-29.30	61.14
NT(CUE)	-18.64	62.90	-24.75	62.83	-31.45	63.35
	$T = 6, \gamma = 0.8, f_t \sim trend$					
	$N = 150$		$N = 300$		$N = 500$	
Estimator	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)
ML	-3.21	13.36	-1.68	10.36	-0.43	8.34
ALS(1step)	-34.76	51.98	-34.05	52.22	-35.43	53.05
ALS(2step)	-37.44	54.36	-36.70	55.01	-36.99	54.37
ALS(CUE)	-35.42	56.23	-34.47	55.11	-35.13	54.39
NT(1step)	-49.95	63.27	-61.37	73.58	-71.26	81.07
NT(2step)	-50.83	64.44	-61.79	74.83	-72.50	82.81
NT(CUE)	-60.29	80.38	-72.93	89.58	-83.87	96.77

Notes: y_{it} is generated as $y_{it} = \alpha_i + \gamma y_{i,t-1} + \xi_{it}$, $\xi_{it} = \lambda_i f_t + u_{it}$, $u_{it} \sim iid\mathcal{N}(0, \sigma^2)$, $i = 1, 2, ..., N$; $t = -49, 48, ...0, 1, ..., T$, with $y_{i,-50} = 0$ and $\sigma^2 = 1$. The factor is generated as: $f_t = \rho_f f_{t-1} + \sqrt{1 - \rho_f^2} \varepsilon_{ft}$, $\varepsilon_{ft} \sim iid\mathcal{N}(0, 1)$, for $t = -49, 48, ...0, 1, ..., T$, with $f_{-50} = 0$, and $\rho_f = 0.9$, in the case where $f_t \sim$AR(1); $f_t = 0$ for all $t = -49, 48, ...0$, and $f_t = t$ for $1, 2, ..., T$, in the case where $f_t \sim trend$. Under both specifications of f_t, the resultant f_t values are scaled such that $T^{-1} \sum_{t=1}^{T} f_t^2 = 1$. The values of f_t for $t = -49, 48, ...0$ are not scaled. The factor loadings, λ_i, are generated as $\lambda_i = \lambda + \eta_i$ with $\lambda = 1$ and $\eta_i \sim iidN(0, 1)$. The fixed effects, α_i, are generated as $\alpha_i = T^{-1}(\xi_{i1} + \xi_{i2} + ... + \xi_{iT}) + v_i = \lambda_i \bar{f} + \bar{u}_i + v_i$, where $\bar{f} = T^{-1} \sum_{t=1}^{T} f_t$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and $v_i \sim iid\mathcal{N}(0, 1)$. Each f_t is generated once and the same $f_t's$ are used throughout the replications. The first 50 observations are discarded. ML is the proposed transformed maximum likelihood estimator. ALS(j) and NT(j) with $j = 1step, 2step, CUE$ are the one step, two step and continuous updating GMM estimators of Ahn et al. (2013), and Nauges and Thomas (2003), respectively. All experiments are based on 1,000 replications.

Table 2: Bias(×100) and RMSE(×100) ($T = 10$) for the AR(1) model with a single factor

	$N = 150$		$N = 300$		$N = 500$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
Estimator	(×100)	(×100)	(×100)	(×100)	(×100)	(×100)
$T = 10, \gamma = 0.4, f_t \sim$AR(1)						
ML	0.30	4.47	0.01	3.11	-0.09	2.28
ALS(1step)	15.18	23.26	10.06	18.99	6.69	15.39
ALS(2step)	12.45	20.40	7.03	15.88	3.58	11.42
ALS(CUE)	11.66	19.92	3.68	13.94	0.91	9.66
NT(1step)	-35.40	41.02	-43.60	44.20	-47.11	47.48
NT(2step)	-41.82	47.14	-51.89	52.29	-55.49	55.69
NT(CUE)	-56.39	61.07	-61.56	61.62	-61.87	61.90
$T = 10, \gamma = 0.8, f_t \sim$AR(1)						
ML	0.30	6.01	0.12	4.59	0.06	3.63
ALS(1step)	-5.35	9.68	-5.01	9.08	-4.51	9.84
ALS(2step)	-8.12	11.72	-7.58	11.15	-6.71	11.49
ALS(CUE)	-3.92	10.87	-3.49	9.64	-3.04	10.27
NT(1step)	2.32	31.97	0.98	36.27	-0.05	40.53
NT(2step)	-5.33	34.26	-5.70	37.08	-5.72	40.21
NT(CUE)	-7.64	37.24	-9.28	39.41	-10.04	39.70
$T = 10, \gamma = 0.4, f_t \sim trend$						
ML	0.21	4.18	0.21	4.18	-0.09	2.28
ALS(1step)	-9.06	12.94	-9.06	12.94	6.69	15.39
ALS(2step)	-9.59	13.61	-9.59	13.61	3.58	11.42
ALS(CUE)	-11.77	15.28	-11.77	15.28	0.91	9.66
NT(1step)	-24.10	52.37	-24.10	52.37	-47.11	47.48
NT(2step)	-27.60	55.35	-27.60	55.35	-55.49	55.69
NT(CUE)	-24.64	61.13	-24.64	61.13	-61.87	61.90
$T = 6, \gamma = 0.8, f_t \sim trend$						
ML	-0.10	6.92	0.15	5.39	-0.06	4.24
ALS(1step)	-10.74	18.61	-11.93	21.10	-13.84	23.54
ALS(2step)	-11.89	16.89	-12.85	19.63	-17.46	24.11
ALS(CUE)	-10.15	20.96	-12.66	24.69	-16.94	28.01
NT(1step)	-46.57	60.17	-60.55	72.32	-75.19	83.16
NT(2step)	-49.44	63.01	-63.07	75.26	-78.42	87.33
NT(CUE)	-56.11	77.72	-77.00	93.18	-95.74	105.41

See notes to Table 1.

Table 3: Size(%) and power(%) for the AR(1) model with a single factor ($T = 6$)

$T = 6,\ \gamma = 0.4,\ f_t \sim AR(1)$

	$N = 150$					$N = 300$					$N = 500$				
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ γ	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
ML	38.1	14.5	5.4	12.1	40.0	64.2	23.2	4.7	20.8	65.5	83.8	35.1	4.8	28.6	85.6
ALS(1step)	6.1	4.8	3.7	3.8	4.3	8.5	5.9	4.8	4.2	4.5	13.0	7.0	4.1	3.1	3.2
ALS(2step)	15.9	13.7	11.6	11.1	10.9	26.0	18.8	14.0	11.9	11.6	33.6	23.2	15.9	12.6	11.9
ALS(CUE)	13.4	11.2	10.1	9.0	8.6	17.0	11.0	8.3	6.6	6.1	25.0	17.1	11.4	8.1	7.7
NT(1step)	92.3	89.4	86.0	82.1	77.0	99.9	99.8	99.1	97.8	96.0	100.0	100.0	100.0	99.9	99.6
NT(2step)	95.6	94.0	91.7	89.1	83.6	100.0	99.7	99.3	98.9	98.5	100.0	100.0	100.0	100.0	99.8
NT(CUE)	99.0	99.0	98.8	97.8	95.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

$T = 6,\ \gamma = 0.8,\ f_t \sim AR(1)$

	$N = 150$					$N = 300$					$N = 500$				
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ γ	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
ML	30.9	15.6	5.8	2.3	5.6	42.2	20.7	6.4	1.7	12.0	55.4	25.8	4.8	2.6	47.3
ALS(1step)	8.1	6.4	5.3	4.4	3.7	8.8	7.1	5.6	4.6	3.5	8.8	7.7	6.2	4.6	3.6
ALS(2step)	19.1	16.1	13.9	11.8	10.0	18.6	15.7	12.7	10.9	8.3	18.4	15.8	12.8	10.5	8.7
ALS(CUE)	15.1	12.8	10.9	9.4	7.8	15.9	13.6	11.8	9.7	7.8	15.6	13.8	11.5	10.0	8.0
NT(1step)	56.1	55.6	55.2	54.5	54.0	63.4	63.3	62.9	62.7	62.7	69.2	69.0	68.9	68.9	69.7
NT(2step)	59.1	58.6	58.5	58.1	57.6	64.3	63.9	63.9	63.8	64.2	69.4	69.3	69.3	69.3	70.5
NT(CUE)	54.1	53.8	53.5	53.3	53.4	57.8	57.7	57.7	57.5	57.7	57.6	57.6	57.6	57.7	59.1

$T = 6,\ \gamma = 0.4,\ f_t \sim trend$

	$N = 150$					$N = 300$					$N = 500$				
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ γ	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
ML	28.7	14.4	6.0	9.6	23.9	45.4	18.9	5.0	13.7	43.6	66.3	24.0	5.8	21.0	67.6
ALS(1step)	6.4	5.4	4.1	3.3	2.6	9.4	6.5	3.8	2.0	1.2	18.5	13.2	8.0	4.5	2.2
ALS(2step)	14.1	10.8	9.9	8.2	6.8	19.7	15.2	10.5	7.6	5.3	27.1	19.3	13.0	8.3	5.0
ALS(CUE)	9.8	8.2	7.3	5.6	4.1	15.0	10.6	7.2	4.7	3.1	22.0	15.2	10.0	5.9	3.1
NT(1step)	47.0	45.5	43.1	40.6	38.2	67.2	67.5	68.7	71.8	74.0	78.3	84.1	90.2	93.7	95.7
NT(2step)	48.8	47.7	47.5	47.4	46.9	67.7	69.3	71.5	75.5	78.8	79.5	85.8	91.0	94.5	96.5
NT(CUE)	56.6	55.7	56.3	56.7	58.6	69.6	71.4	74.5	78.7	81.8	79.3	85.5	91.5	94.8	96.6

$T = 6,\ \gamma = 0.8,\ f_t \sim trend$

	$N = 150$					$N = 300$					$N = 500$				
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ γ	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
ML	22.5	16.0	10.3	5.3	2.3	21.9	15.1	7.9	3.1	1.3	25.2	13.4	5.1	1.7	1.4
ALS(1step)	4.1	3.2	2.8	1.8	1.1	6.0	4.5	3.0	2.3	1.8	7.3	5.0	3.7	2.8	2.0
ALS(2step)	13.5	11.5	8.1	6.8	5.0	15.5	13.3	11.5	9.7	7.2	20.4	16.5	14.2	10.7	8.3
ALS(CUE)	10.0	7.7	5.9	4.6	3.9	13.3	11.3	9.3	7.6	6.1	15.1	12.2	9.5	7.4	6.6
NT(1step)	32.2	31.7	30.8	30.1	28.2	46.6	45.8	44.6	43.7	42.9	60.2	59.6	59.1	57.8	56.9
NT(2step)	31.8	31.0	30.8	30.4	29.4	46.4	45.6	44.9	44.0	42.9	59.5	59.0	58.5	57.8	56.6
NT(CUE)	41.4	41.0	41.0	40.4	40.0	52.8	52.7	52.4	52.0	51.6	64.6	64.4	64.3	64.0	63.6

See notes to Table 1.

Table 4: Size(%) and power(%) for the AR(1) model with a single factor ($T = 10$)

$T = 10$, $\gamma = 0.4$, $f_t \sim$ AR(1)

Estimators \ γ	Power(H_1) 0.30	0.35	Size 0.40	Power(H_1) 0.45	0.50	Power(H_1) 0.30	0.35	Size 0.40	Power(H_1) 0.45	0.50	Power(H_1) 0.30	0.35	Size 0.40	Power(H_1) 0.45	0.50
		$N=150$					$N=300$					$N=500$			
ML	62.5	20.2	6.3	21.6	66.9	89.2	37.9	5.7	38.0	91.1	99.1	58.4	3.2	54.8	99.7
ALS(1step)	13.6	16.3	20.0	22.4	25.3	12.1	15.3	20.9	25.8	29.8	11.6	14.2	19.0	23.7	29.3
ALS(2step)	55.8	60.4	66.9	70.6	75.5	51.3	47.2	50.7	58.4	68.2	58.5	39.9	34.9	45.2	63.9
ALS(CUE)	24.9	25.9	29.7	31.7	33.0	36.2	29.2	24.0	25.0	30.0	48.2	30.1	18.7	20.7	36.5
NT(1step)	94.3	91.0	86.8	78.4	68.3	99.9	99.7	98.8	97.7	93.4	100.0	100.0	100.0	99.9	99.8
NT(2step)	99.1	98.9	98.3	96.7	94.6	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0
NT(CUE)	98.4	98.4	98.3	98.4	98.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

$T = 10$, $\gamma = 0.8$, $f_t \sim$ AR(1)

Estimators \ γ	Power(H_1) 0.70	0.75	Size 0.80	Power(H_1) 0.85	0.90	Power(H_1) 0.70	0.75	Size 0.80	Power(H_1) 0.85	0.90	Power(H_1) 0.70	0.75	Size 0.80	Power(H_1) 0.85	0.90
		$N=150$					$N=300$					$N=500$			
ML	37.0	18.1	4.7	3.9	15.8	53.8	24.9	4.8	5.0	54.5	72.1	33.8	5.0	13.4	84.5
ALS(1step)	25.6	16.5	4.6	1.1	4.1	26.2	16.6	7.3	2.5	4.9	29.1	17.0	7.5	4.6	9.0
ALS(2step)	79.5	74.2	61.3	45.0	35.5	77.2	70.9	59.1	43.8	35.8	72.7	66.7	58.3	45.1	37.0
ALS(CUE)	29.0	23.3	18.0	15.0	17.6	37.7	28.8	21.9	15.7	20.7	38.1	30.3	23.9	16.8	21.3
NT(1step)	10.3	9.9	8.6	7.8	7.4	11.8	11.4	11.3	11.0	10.8	13.2	13.1	13.1	13.1	13.9
NT(2step)	21.1	18.9	17.2	18.1	21.1	18.0	16.6	15.6	15.2	19.5	18.2	17.4	16.1	16.4	22.9
NT(CUE)	25.8	23.2	21.3	21.9	27.4	23.7	21.7	20.9	21.3	26.9	19.8	18.1	17.6	19.4	25.0

$T = 10$, $\gamma = 0.4$, $f_t \sim trend$

Estimators \ γ	Power(H_1) 0.30	0.35	Size 0.40	Power(H_1) 0.45	0.50	Power(H_1) 0.30	0.35	Size 0.40	Power(H_1) 0.45	0.50	Power(H_1) 0.30	0.35	Size 0.40	Power(H_1) 0.45	0.50
		$N=150$					$N=300$					$N=500$			
ML	66.7	23.1	5.5	23.6	71.2	92.6	40.6	4.8	42.6	93.9	99.1	63.3	4.7	58.7	99.9
ALS(1step)	23.3	13.7	6.7	2.6	2.5	41.3	32.0	16.6	5.7	0.9	48.2	44.4	30.9	11.8	1.7
ALS(2step)	49.2	35.4	20.2	11.6	10.2	69.4	59.0	36.5	15.7	6.6	79.6	75.3	53.1	21.0	4.9
ALS(CUE)	30.2	22.8	14.9	9.0	6.5	45.3	38.9	26.6	11.9	4.7	54.1	51.5	38.5	16.4	4.7
NT(1step)	89.3	90.4	90.4	89.2	86.1	99.9	99.8	99.8	99.6	99.5	100.0	100.0	100.0	100.0	100.0
NT(2step)	97.6	97.8	97.6	97.1	96.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
NT(CUE)	80.2	80.6	80.6	80.7	80.7	97.3	97.3	97.3	97.3	97.3	99.8	99.8	99.8	99.8	99.8

$T = 10$, $\gamma = 0.8$, $f_t \sim trend$

Estimators \ γ	Power(H_1) 0.70	0.75	Size 0.80	Power(H_1) 0.85	0.90	Power(H_1) 0.70	0.75	Size 0.80	Power(H_1) 0.85	0.90	Power(H_1) 0.70	0.75	Size 0.80	Power(H_1) 0.85	0.90
		$N=150$					$N=300$					$N=500$			
ML	31.0	15.5	4.5	2.1	11.3	44.0	19.9	5.4	6.6	39.4	64.1	26.5	4.8	14.5	61.4
ALS(1step)	6.5	2.8	1.0	0.3	0.8	10.5	4.5	1.8	1.2	2.0	9.4	5.2	2.3	2.0	3.3
ALS(2step)	41.9	35.4	25.6	18.1	13.8	39.0	31.2	23.1	16.2	15.0	42.6	38.8	31.9	23.0	18.0
ALS(CUE)	19.4	16.1	14.2	13.2	11.9	25.6	22.2	19.3	16.6	16.8	32.8	29.7	26.7	24.6	23.2
NT(1step)	47.0	46.0	44.8	43.3	41.8	67.8	67.3	66.5	65.7	64.2	82.4	81.9	81.2	80.6	80.2
NT(2step)	46.7	45.3	44.6	43.5	43.0	63.6	62.8	62.3	61.5	60.4	79.3	79.1	78.6	77.9	77.4
NT(CUE)	47.9	47.4	47.3	47.0	46.9	66.8	66.7	66.5	66.2	66.2	82.2	82.1	82.1	82.1	82.1

See notes to Table 1.

Table 5: Bias($\times100$) and RMSE($\times100$) for the ARX(1) model with a single factor ($T = 6$)

	$T=6, \gamma=0.4, \beta=1.0, f_t \sim$AR(1)											
	$N=150$				$N=300$				$N=500$			
	γ		β		γ		β		γ		β	
Estimators	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)
ML	-0.19	4.29	-0.05	7.41	0.03	3.01	0.05	5.26	-0.05	2.30	0.01	4.08
ALS(1step)	0.81	16.60	-3.11	17.23	-0.87	11.56	-1.59	11.85	-1.71	8.16	-0.86	8.51
ALS(2step)	2.10	17.47	-5.20	19.40	1.69	11.82	-4.80	13.42	1.76	8.32	-4.47	9.97
ALS(CUE)	2.46	22.58	-7.08	23.84	-0.70	15.87	-4.44	15.66	-2.51	11.59	-2.93	11.07
NT(1step)	-3.70	32.16	4.70	14.58	8.52	24.51	7.37	12.02	15.80	20.97	8.52	10.84
NT(2step)	-6.23	35.96	4.09	16.28	6.58	27.22	6.96	12.90	14.76	21.98	8.07	11.01
NT(CUE)	17.04	41.42	0.60	22.77	26.21	35.06	4.72	13.98	29.58	31.63	5.70	10.90
	$T=6, \gamma=0.8, \beta=1.0, f_t \sim$AR(1)											
	$N=150$				$N=300$				$N=500$			
	γ		β		γ		β		γ		β	
Estimators	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)
ML	-0.06	2.38	-0.07	4.33	-0.10	1.74	0.09	3.14	-0.01	1.32	0.01	2.42
ALS(1step)	-1.35	5.39	3.03	8.29	-2.11	4.17	4.02	6.85	-2.32	3.16	4.29	5.56
ALS(2step)	-0.33	5.20	1.52	7.97	-0.65	3.55	2.02	5.78	-0.67	2.28	2.23	4.26
ALS(CUE)	-0.44	5.90	0.81	8.44	-1.15	3.93	1.87	5.76	-1.21	2.66	1.99	4.32
NT(1step)	-1.37	14.76	0.51	6.36	5.89	12.26	0.83	4.57	9.39	11.85	0.79	3.65
NT(2step)	-2.45	17.17	0.38	7.20	5.32	13.55	0.69	5.06	9.06	12.37	0.60	3.97
NT(CUE)	8.50	18.47	-0.09	8.50	14.21	16.97	0.14	5.51	16.50	17.16	0.01	4.22
	$T=6, \gamma=0.4, \beta=1.0, f_t \sim trend$											
	$N=150$				$N=300$				$N=500$			
	γ		β		γ		β		γ		β	
Estimators	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)
ML	-0.07	5.82	-0.28	8.96	0.07	3.98	-0.08	6.27	-0.07	2.95	0.01	4.77
ALS(1step)	9.07	36.37	-13.00	40.27	5.61	36.84	-8.65	40.08	3.71	37.09	-6.38	39.69
ALS(2step)	10.86	36.14	-16.42	41.33	10.57	34.73	-15.78	39.41	10.68	34.05	-15.65	38.38
ALS(CUE)	1.58	43.38	-7.24	48.41	-0.94	39.84	-2.94	42.45	-2.13	39.03	-1.24	40.83
NT(1step)	54.52	55.27	-4.96	16.36	59.27	59.32	-5.65	12.47	59.81	59.82	-6.11	10.67
NT(2step)	55.16	56.00	-6.97	20.08	59.47	59.52	-8.04	15.34	59.87	59.87	-8.68	13.38
NT(CUE)	55.52	57.40	-7.78	27.23	59.31	59.43	-7.13	18.07	59.82	59.83	-6.68	14.17
	$T=6, \gamma=0.8, \beta=1.0, f_t \sim trend$											
	$N=150$				$N=300$				$N=500$			
	γ		β		γ		β		γ		β	
Estimators	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)	Bias ($\times100$)	RMSE ($\times100$)
ML	-0.09	3.10	-0.14	4.88	-0.11	2.18	0.04	3.44	-0.01	1.65	0.01	2.64
ALS(1step)	4.65	10.21	-4.82	14.47	4.23	9.69	-3.85	13.19	3.66	9.12	-2.99	12.24
ALS(2step)	6.54	10.13	-6.98	13.73	6.95	9.82	-7.09	12.28	7.41	9.79	-7.20	11.61
ALS(CUE)	4.71	10.38	-5.36	14.56	4.38	9.43	-4.24	11.94	4.06	8.87	-3.51	10.83
NT(1step)	18.93	19.50	1.40	6.85	19.86	19.87	1.49	4.93	19.90	19.90	1.44	3.99
NT(2step)	17.83	18.92	1.24	7.84	19.77	19.80	1.58	5.42	19.89	19.89	1.52	4.26
NT(CUE)	17.64	19.22	1.67	9.08	19.66	19.71	2.08	6.04	19.84	19.85	2.12	4.70

Notes: y_{it} is generated as $y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}$, $\xi_{it} = \lambda_i f_t + u_{it}$, $u_{it} \sim iid\mathcal{N}(0, \sigma^2)$, $i = 1, 2, ..., N; t = -49, 48, ...0, 1, ..., T$, with $y_{i,-50} = 0$ and $x_{it} = \mu_i + \vartheta_i f_t + \breve{x}_{it}$, $\breve{x}_{it} = \rho_x \breve{x}_{i,t-1} + \sqrt{1 - \rho_x^2} \varepsilon_{it}$, with $\breve{x}_{i,-50} = 0$, for $t = -49, 48, ...0, 1, ..., T$, where $\rho_x = 0.8$, $\mu_i \sim iid\mathcal{N}(0,1)$, and $\varepsilon_{it} \sim iid\mathcal{N}(0,1)$. The factor f_t is generated as in the AR(1) case (see notes to Table 1). The factor loadings, ϑ_i and λ_i, are generated as $\vartheta_i \sim iid\mathcal{N}(0.5, \sigma_\vartheta^2)$ and $\lambda_i \sim iid\mathcal{N}(0.5, \sigma_\lambda^2)$, respectively. The fixed effects, α_i, are generated as $\alpha_i = T^{-1} \sum_{t=1}^{T} x_{it} + \lambda_i \bar{f} + \bar{u}_i + v_i$, where $\bar{f} = T^{-1} \sum_{t=1}^{T} f_t$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and $v_i \sim iid\mathcal{N}(0,1)$. The remaining parameters are set at $\beta = 1$, $\sigma_\lambda^2 = \sigma_\vartheta^2 = \sigma^2$, with $\sigma^2 = (0.8 - \gamma^2)/0.3$. Each f_t is generated once and the same $f_t's$ are used throughout the replications. The first 50 observations are discarded. ML is the proposed maximum likelihood estimator. ALS(j) and NT(j) with $j = 1step, 2step, CUE$ are the one step, two step and continuous updating GMM estimators of Ahn et al. (2013), and Nauges and Thomas (2003), respectively. All experiments are based on 1,000 replications.

Table 6: Bias(×100) and RMSE(×100) for the ARX(1) model with a single factor ($T = 10$)

$T = 10, \gamma = 0.4, \beta = 1.0, f_t \sim \text{AR}(1)$

| | N = 150 | | | | N = 300 | | | | N = 500 | | | |
| | γ | | β | | γ | | β | | γ | | β | |
Estimators	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)
ML	-0.03	2.58	0.12	5.53	-0.10	1.89	0.07	3.99	-0.04	1.42	0.07	3.06
ALS(1step)	0.62	6.01	-6.28	11.10	0.67	4.60	-6.03	9.11	0.44	3.36	-5.74	7.82
ALS(2step)	0.90	7.87	-10.61	16.20	1.55	5.34	-10.43	13.14	2.04	4.16	-10.81	12.34
ALS(CUE)	2.11	12.64	-9.92	21.91	0.00	8.24	-6.73	13.39	-1.29	5.51	-5.27	9.33
NT(1step)	21.94	26.73	7.16	11.51	30.58	31.45	6.63	9.17	33.75	34.16	6.39	8.10
NT(2step)	20.08	30.04	6.69	14.92	31.21	32.96	6.50	10.48	34.83	35.68	6.21	8.62
NT(CUE)	20.45	39.77	3.98	22.90	27.29	34.94	6.80	13.07	28.77	32.16	7.23	10.48

$T = 10, \gamma = 0.8, \beta = 1.0, f_t \sim \text{AR}(1)$

| | N = 150 | | | | N = 300 | | | | N = 500 | | | |
| | γ | | β | | γ | | β | | γ | | β | |
Estimators	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)
ML	-0.04	1.24	0.08	3.03	-0.02	0.89	0.02	2.17	-0.02	0.67	0.03	1.67
ALS(1step)	-0.10	1.75	1.32	4.59	-0.06	1.24	1.32	3.37	-0.09	1.00	1.32	2.82
ALS(2step)	0.21	2.44	-0.54	5.70	0.63	1.59	-0.68	3.54	0.76	1.31	-0.76	2.71
ALS(CUE)	0.58	3.78	-1.37	7.84	0.69	1.94	-0.66	4.08	0.59	1.36	-0.49	2.83
NT(1step)	11.26	15.39	3.56	5.66	17.35	17.89	3.37	4.63	19.17	19.26	3.36	4.12
NT(2step)	4.78	17.16	2.96	7.32	14.38	16.35	3.01	5.04	17.34	17.84	2.94	4.02
NT(CUE)	4.07	20.70	1.50	10.35	12.56	17.06	2.78	5.55	15.27	16.99	2.86	4.30

$T = 10, \gamma = 0.4, \beta = 1.0, f_t \sim trend$

| | N = 150 | | | | N = 300 | | | | N = 500 | | | |
| | γ | | β | | γ | | β | | γ | | β | |
Estimators	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)
ML	0.09	3.21	-0.10	6.01	-0.12	2.21	0.00	4.62	0.01	1.76	0.06	3.35
ALS(1step)	15.28	29.71	-22.34	36.05	15.80	31.79	-22.74	37.69	16.93	33.49	-23.36	39.17
ALS(2step)	14.48	30.29	-23.94	39.68	16.55	32.08	-26.34	41.32	19.12	33.90	-28.93	43.57
ALS(CUE)	6.18	31.25	-13.34	41.53	2.84	32.14	-10.09	40.13	4.31	33.29	-11.11	40.90
NT(1step)	53.85	54.45	-0.06	11.72	59.56	59.58	-0.89	8.76	59.90	59.90	-1.21	6.59
NT(2step)	53.07	54.25	1.05	18.25	59.49	59.51	-0.16	10.97	59.90	59.90	-0.40	7.74
NT(CUE)	43.81	52.07	2.39	26.83	53.89	55.78	7.22	15.00	56.43	57.15	8.31	11.77

$T = 10, \gamma = 0.8, \beta = 1.0, f_t \sim trend$

| | N = 150 | | | | N = 300 | | | | N = 500 | | | |
| | γ | | β | | γ | | β | | γ | | β | |
Estimators	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)	Bias (×100)	RMSE (×100)
ML	-0.03	1.69	0.02	3.25	-0.03	1.18	-0.02	2.45	0.01	0.89	0.04	1.77
ALS(1step)	3.77	7.11	-3.49	11.63	2.26	5.44	-1.04	8.92	1.34	3.97	0.69	6.82
ALS(2step)	4.25	7.03	-5.68	11.12	4.00	5.72	-5.10	8.33	3.72	4.81	-4.35	6.56
ALS(CUE)	3.13	7.89	-4.75	12.76	2.25	5.35	-3.04	7.81	1.73	3.78	-1.81	5.36
NT(1step)	19.51	19.64	3.53	5.91	19.90	19.90	3.89	5.18	19.90	19.90	4.06	4.80
NT(2step)	16.30	18.01	2.61	7.80	19.81	19.82	2.67	5.06	19.90	19.90	2.78	4.08
NT(CUE)	11.21	19.33	2.28	10.71	17.92	18.66	3.48	6.24	18.96	19.17	3.71	5.03

See notes to Table 5.

Table 7a: Size(%) and power(%) for the ARX(1) model with a single factor ($T = 6$, $f_t \sim$AR(1))

$T = 6, \gamma = 0.4, \beta = 1.0, f_t \sim$AR(1)

	N = 150					N = 300					N = 500				
			γ					γ					γ		
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ γ	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
ML	68.2	25.0	5.8	21.6	65.6	91.9	38.9	7.1	42.5	92.2	99.0	60.6	5.9	61.4	99.0
ALS1(1step)	37.2	22.1	15.2	16.7	26.3	48.7	27.1	15.2	15.8	29.8	62.4	32.6	15.6	15.0	37.2
ALS1(2step)	45.3	32.8	25.6	28.8	38.9	44.1	28.2	19.7	27.9	48.6	47.9	25.7	18.0	35.1	63.2
ALS1(CUE)	56.3	48.2	41.7	40.0	46.0	62.0	47.1	37.7	36.1	39.5	70.1	50.7	35.1	30.5	39.8
NT1(1step)	29.5	31.6	35.2	38.1	41.4	19.8	26.8	35.6	45.5	57.5	16.7	30.4	46.1	61.3	73.0
NT1(2step)	41.6	43.4	45.8	49.4	54.9	33.8	40.0	48.7	53.6	61.8	25.4	38.2	51.3	62.3	72.5
NT1(CUE)	54.7	59.8	65.0	71.6	75.7	57.1	66.6	74.8	82.5	88.7	61.9	77.6	88.0	93.2	96.9
			β					β					β		
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	26.3	10.6	3.9	9.7	23.3	47.6	14.7	4.2	15.9	46.3	68.0	21.2	4.7	22.0	68.4
ALS1(1step)	21.7	15.4	11.5	12.7	17.5	27.3	14.4	8.5	12.1	22.7	34.8	14.3	7.5	12.5	30.2
ALS1(2step)	34.0	24.5	20.4	21.8	27.3	44.1	26.9	15.8	14.1	21.9	60.5	32.8	16.7	11.8	23.0
ALS1(CUE)	41.1	32.4	29.5	30.2	34.4	43.3	29.4	20.7	21.5	29.9	50.7	28.8	17.1	16.5	31.0
NT1(1step)	13.3	12.8	14.4	21.0	34.2	10.1	11.0	22.1	39.7	61.2	6.6	10.4	32.1	60.1	84.6
NT1(2step)	23.7	21.5	24.2	32.1	44.5	17.5	17.1	26.8	44.5	62.0	11.4	14.2	35.0	61.3	81.9
NT1(CUE)	32.9	30.4	34.1	39.4	46.6	24.3	21.5	27.1	38.5	53.4	20.8	17.4	30.6	47.4	70.1

$T = 6, \gamma = 0.8, \beta = 1.0, f_t \sim$AR(1)

	N = 150					N = 300					N = 500				
			γ					γ					γ		
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ γ	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
ML	99.8	56.1	5.1	56.8	97.2	100.0	86.2	6.1	81.6	100.0	100.0	97.3	6.4	95.7	100.0
ALS1(1step)	92.9	55.9	14.6	18.7	62.6	99.3	81.5	16.5	23.4	80.7	99.9	95.7	21.2	27.8	94.4
ALS1(2step)	91.2	55.8	19.8	36.2	77.4	98.8	71.7	13.9	43.6	94.0	100.0	83.8	11.4	62.8	99.8
ALS1(CUE)	90.9	59.8	24.7	37.0	74.9	98.6	76.4	19.1	39.4	90.4	99.9	88.4	16.6	52.3	98.5
NT1(1step)	9.2	6.6	5.0	4.0	6.3	2.7	1.6	1.6	7.1	16.7	0.3	0.3	6.2	17.3	30.9
NT1(2step)	14.3	11.8	10.4	10.9	15.1	5.4	4.3	4.5	15.6	27.2	0.9	0.7	11.5	24.1	36.2
NT1(CUE)	8.7	8.2	6.8	11.6	28.9	2.2	2.0	5.8	35.6	58.6	0.0	0.2	32.6	64.4	78.6
			β					β					β		
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	61.9	21.1	5.1	19.1	62.0	87.8	34.6	5.4	37.4	89.8	98.8	54.0	4.6	55.0	98.5
ALS1(1step)	24.0	11.2	12.5	31.8	58.1	28.1	9.4	19.2	54.8	84.9	37.0	5.3	21.4	75.1	97.9
ALS1(2step)	40.9	20.8	19.1	34.0	60.1	50.3	17.5	17.0	47.1	82.6	68.5	16.8	15.4	62.5	95.6
ALS1(CUE)	43.0	23.8	21.6	33.8	57.1	52.5	19.4	17.5	46.7	81.6	70.5	20.3	14.9	59.6	94.4
NT1(1step)	32.7	11.7	5.1	14.7	40.2	54.0	12.9	5.5	25.4	65.2	73.1	22.7	4.8	37.5	83.4
NT1(2step)	42.4	21.7	13.2	23.9	48.1	58.2	20.8	10.6	30.5	66.3	75.2	27.2	8.4	39.5	82.3
NT1(CUE)	44.7	26.7	18.0	25.0	45.0	59.6	24.9	12.4	27.6	58.9	76.7	31.6	9.2	32.2	76.3

See notes to Table 5.

Table 7b: Size(%) and power(%) for the ARX(1) model with a single factor ($T = 6$, $f_t \sim trend$)

	$T = 6, \gamma = 0.4, \beta = 1.0, f_t \sim trend$													
	N = 150					N = 300					N = 500				
	γ														
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ γ	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
ML	47.1	18.9	6.3	13.5	43.2	71.6	28.1	5.4	26.5	73.2	91.7	41.5	4.4	37.0	92.1
ALS1(1step)	59.9	51.0	44.1	38.9	38.6	79.5	67.4	56.8	47.0	40.4	90.6	81.7	71.1	57.9	46.5
ALS1(2step)	68.4	60.2	53.6	48.9	47.9	72.9	62.5	52.2	44.7	43.7	79.9	67.9	54.2	45.4	42.8
ALS1(CUE)	87.3	82.3	76.5	70.7	64.0	95.3	90.3	83.3	73.6	62.4	98.5	97.2	92.1	83.6	69.6
NT1(1step)	78.8	85.2	88.6	91.6	94.0	99.2	99.8	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
NT1(2step)	87.6	90.6	92.4	94.3	95.9	99.5	99.6	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
NT1(CUE)	90.8	92.5	93.3	94.5	95.0	98.8	99.1	99.3	99.5	99.6	100.0	100.0	100.0	100.0	100.0
	β														
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	21.7	9.3	6.9	8.9	20.3	38.5	13.0	5.9	13.6	35.9	53.5	18.5	4.6	19.5	56.7
ALS1(1step)	37.0	36.3	36.5	38.6	41.9	36.6	37.7	43.0	50.1	57.5	39.0	43.3	52.7	64.0	74.1
ALS1(2step)	44.0	43.0	45.0	46.8	51.4	39.5	38.8	41.3	47.0	53.6	40.3	38.1	41.8	49.2	57.2
ALS1(CUE)	51.8	55.0	59.0	64.8	70.2	46.7	54.4	63.5	72.4	81.2	51.7	62.1	73.3	85.2	92.4
NT1(1step)	22.1	14.9	10.2	9.3	10.4	37.0	21.8	11.6	7.9	10.3	52.3	31.6	15.3	9.7	11.3
NT1(2step)	37.1	30.1	23.4	20.6	22.6	50.4	35.5	24.4	16.5	15.5	64.2	44.6	28.5	17.6	15.6
NT1(CUE)	42.3	39.6	38.3	37.5	37.1	48.9	39.0	29.9	26.2	27.7	55.4	41.9	30.4	24.4	25.9
	$T = 6, \gamma = 0.8, \beta = 1.0, f_t \sim trend$														
	N = 150					N = 300					N = 500				
	γ														
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ γ	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
ML	92.1	40.4	6.1	39.0	89.0	100.0	68.0	5.8	64.8	99.3	100.0	87.2	5.7	86.6	99.9
ALS1(1step)	82.9	66.3	49.1	52.1	69.8	94.9	79.8	48.9	52.1	78.5	98.2	88.8	48.0	54.1	87.1
ALS1(2step)	75.0	65.1	57.1	69.3	87.1	87.1	66.0	56.2	79.4	97.3	91.5	67.1	62.6	89.8	99.5
ALS1(CUE)	80.6	71.6	58.5	62.1	76.8	92.2	79.8	52.8	60.7	87.4	97.9	86.1	49.6	64.5	94.4
NT1(1step)	0.5	0.4	0.3	0.3	0.7	0.0	0.0	0.0	0.0	2.9	0.0	0.0	0.0	1.4	40.8
NT1(2step)	0.4	0.3	0.3	0.4	2.0	0.0	0.0	0.0	0.1	8.6	0.0	0.0	0.1	4.2	55.8
NT1(CUE)	1.1	1.0	0.6	0.7	2.5	0.0	0.0	0.0	0.1	8.5	0.0	0.0	0.0	4.1	53.7
	β														
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	57.9	18.8	5.7	17.3	54.8	82.1	33.7	5.2	31.1	84.0	97.1	47.1	5.2	48.9	96.5
ALS1(1step)	50.8	43.4	38.8	34.4	36.0	50.9	43.0	47.9	59.2	61.4	51.4	40.0	52.0	77.8	78.4
ALS1(2step)	66.1	53.6	43.0	37.0	34.5	75.6	58.2	48.7	46.1	46.8	85.4	63.9	53.4	57.7	57.8
ALS1(CUE)	60.3	48.9	43.6	42.4	46.1	63.1	50.1	45.0	51.4	58.7	68.7	46.6	47.7	69.1	73.1
NT1(1step)	26.3	9.6	5.0	17.8	41.8	44.9	11.0	5.5	27.6	66.8	63.5	16.2	7.5	43.4	86.5
NT1(2step)	37.0	18.4	15.1	26.7	49.5	47.9	15.5	10.1	34.7	70.0	65.0	19.8	11.0	47.4	86.5
NT1(CUE)	38.7	21.1	21.3	32.4	51.4	46.4	16.0	14.5	38.8	71.8	59.1	18.0	14.1	51.6	89.0

See notes to Table 5.

Table 8a: Size(%) and power(%) for the ARX(1) model with a single factor ($T = 10$, $f_t \sim$AR(1))

	$T = 10, \gamma = 0.4, \beta = 1.0, f_t \sim$AR(1)														
	$N = 150$					$N = 300$					$N = 500$				
			γ												
	Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)	
Estimators \ γ	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
ML	97.1	47.3	5.2	48.5	96.4	100.0	79.0	6.0	74.0	99.9	100.0	93.5	4.5	93.8	100.0
ALS1(1step)	45.0	15.2	6.8	21.7	54.5	68.0	22.3	7.7	33.9	76.7	87.4	33.3	7.1	45.0	90.0
ALS1(2step)	70.0	50.6	45.2	59.9	77.6	75.4	38.5	30.5	61.7	89.9	83.4	37.1	28.2	76.9	97.9
ALS1(CUE)	72.1	64.7	60.6	66.4	73.8	79.9	63.5	48.8	51.2	72.7	92.0	72.6	39.7	41.1	78.8
NT1(1step)	36.5	53.8	64.6	75.7	82.1	79.5	91.9	96.6	98.8	99.5	98.2	99.6	99.9	100.0	100.0
NT1(2step)	67.0	72.6	77.0	81.4	85.8	81.8	91.4	96.0	97.2	98.6	96.3	98.3	99.5	99.9	100.0
NT1(CUE)	83.0	84.5	84.0	84.3	83.0	76.5	81.4	85.0	87.2	88.6	80.3	86.2	91.7	95.1	96.4
			β												
	Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)	
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	42.4	13.9	4.2	14.6	43.7	72.2	23.9	4.9	24.3	71.1	90.1	35.8	5.3	37.0	91.2
ALS1(1step)	48.2	29.9	13.7	5.2	7.2	73.1	44.3	17.7	7.5	10.2	86.4	59.1	23.4	7.5	16.2
ALS1(2step)	82.2	69.6	55.2	44.5	40.3	92.3	79.6	58.4	35.6	25.0	98.1	90.9	70.8	42.5	20.0
ALS1(CUE)	71.8	65.5	60.6	59.2	59.1	76.2	60.3	47.1	41.8	43.6	84.5	63.7	41.1	28.9	42.1
NT1(1step)	9.6	10.5	19.1	37.5	58.3	11.7	10.1	24.2	52.7	80.5	15.9	8.9	32.2	70.7	93.3
NT1(2step)	49.2	45.2	49.6	60.3	71.3	31.0	28.3	40.8	62.0	81.5	28.5	19.9	44.0	71.4	91.6
NT1(CUE)	63.9	63.7	63.6	67.8	72.6	43.9	43.1	52.3	64.1	76.6	34.5	32.8	53.3	72.7	89.2
	$T = 10, \gamma = 0.8, \beta = 1.0, f_t \sim$AR(1)														
	$N = 150$					$N = 300$					$N = 500$				
			γ												
	Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)	
Estimators \ γ	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
ML	100.0	98.9	4.5	96.7	100.0	100.0	100.0	4.6	100.0	100.0	100.0	100.0	5.0	100.0	100.0
ALS1(1step)	99.1	79.9	4.0	75.0	99.1	98.8	96.6	4.2	95.4	98.8	99.0	98.7	5.0	98.3	99.0
ALS1(2step)	100.0	87.4	40.0	89.3	99.8	100.0	96.4	23.4	99.5	100.0	100.0	99.6	24.8	100.0	100.0
ALS1(CUE)	94.6	77.8	51.0	83.6	95.6	100.0	91.3	33.5	97.9	100.0	100.0	99.0	24.8	99.8	100.0
NT1(1step)	2.8	1.9	7.5	36.8	53.0	0.0	1.6	63.0	81.3	90.1	0.0	67.8	93.5	98.8	99.7
NT1(2step)	23.2	45.7	53.7	59.2	64.5	2.4	40.8	61.7	75.1	83.9	0.2	67.0	82.1	92.3	96.9
NT1(CUE)	27.1	60.1	72.9	73.6	74.3	6.8	50.7	66.7	73.5	77.9	2.3	62.6	73.2	80.5	87.8
			β												
	Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)		Power(H₁)		Size	Power(H₁)	
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	91.0	37.3	5.0	39.4	91.6	99.6	65.1	5.9	65.5	99.3	100.0	84.9	5.1	85.5	100.0
ALS1(1step)	49.9	15.2	6.1	32.0	72.7	79.0	21.8	7.2	51.6	94.2	93.5	34.5	8.4	72.2	98.8
ALS1(2step)	83.5	57.1	39.4	52.2	77.7	96.3	64.1	21.7	48.8	92.0	99.8	78.1	17.7	56.9	98.2
ALS1(CUE)	79.7	61.7	46.7	55.0	72.7	94.3	63.3	27.0	51.8	89.4	99.1	74.2	19.4	62.2	98.4
NT1(1step)	30.3	7.0	13.2	51.5	86.5	54.7	7.5	19.4	75.8	98.2	75.2	9.3	28.5	92.5	99.9
NT1(2step)	61.9	42.8	49.3	70.3	86.7	70.0	25.8	33.5	76.8	96.5	85.2	23.5	32.1	88.5	99.9
NT1(CUE)	70.7	59.5	57.7	66.6	78.2	69.5	34.4	38.5	72.6	94.7	82.3	27.8	36.0	86.2	99.7

See notes to Table 5.

Table 8b: Size(%) and power(%) for the ARX(1) model with a single factor ($T = 10, f_t \sim trend$)

$T = 10, \gamma = 0.4, \beta = 1.0, f_t \sim trend$

	N = 150					N = 300					N = 500				
	γ														
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ γ	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
ML	85.3	33.3	4.4	34.2	88.4	99.5	62.1	4.3	58.0	99.8	100.0	80.7	4.9	83.2	100.0
ALS1(1step)	47.1	37.2	36.1	43.7	53.6	62.1	48.5	42.1	44.1	56.2	79.1	62.3	47.5	46.2	55.5
ALS1(2step)	74.6	68.3	64.1	67.1	72.9	75.2	62.1	56.4	61.0	71.9	83.8	65.1	53.1	59.1	77.5
ALS1(CUE)	84.2	77.5	73.7	70.6	68.7	93.7	90.9	82.5	71.2	54.6	93.5	92.9	89.8	77.6	53.0
NT1(1step)	96.4	98.6	99.4	99.6	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
NT1(2step)	96.3	97.6	98.4	99.0	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
NT1(CUE)	90.4	91.7	92.0	92.8	92.8	96.2	97.2	98.1	98.1	98.1	98.7	98.8	99.2	99.4	99.4
	β														
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	38.2	13.0	5.0	11.0	37.3	63.8	23.9	6.6	22.3	63.0	83.9	33.4	5.3	34.0	86.7
ALS1(1step)	53.2	45.1	37.9	33.5	32.4	57.7	45.9	39.9	36.4	39.8	58.2	48.3	42.1	41.1	47.3
ALS1(2step)	74.1	68.2	63.5	61.7	60.2	75.6	66.5	55.7	49.9	50.5	83.3	69.4	57.1	49.2	51.4
ALS1(CUE)	65.5	65.1	64.9	68.5	69.6	44.7	49.2	58.9	69.8	80.4	40.4	48.4	64.0	78.6	88.0
NT1(1step)	21.3	12.2	9.5	12.5	21.7	33.7	16.8	9.5	13.6	28.4	49.4	21.5	8.0	14.3	35.2
NT1(2step)	55.7	52.7	51.7	53.1	58.0	46.1	32.4	27.1	33.7	47.2	51.5	30.3	19.0	28.4	49.7
NT1(CUE)	63.1	62.0	65.0	66.4	70.0	36.9	38.7	48.6	61.5	70.6	23.2	27.4	46.2	67.8	85.1

$T = 10, \gamma = 0.8, \beta = 1.0, f_t \sim trend$

	N = 150					N = 300					N = 500				
	γ														
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ γ	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
ML	100.0	86.7	5.6	83.8	100.0	100.0	99.6	5.6	98.1	100.0	100.0	100.0	5.5	100.0	100.0
ALS1(1step)	90.3	47.4	28.1	70.4	92.3	96.6	69.1	21.4	80.7	97.1	97.7	84.2	16.0	91.0	97.7
ALS1(2step)	93.5	68.7	63.5	89.9	99.3	97.1	56.4	63.7	99.3	100.0	99.3	52.2	75.0	99.9	100.0
ALS1(CUE)	90.0	72.5	70.7	82.6	92.8	97.9	75.9	51.8	88.4	99.8	99.5	84.7	43.4	96.9	100.0
NT1(1step)	0.1	0.1	0.1	3.8	49.6	0.0	0.0	0.2	54.7	99.3	0.0	0.0	23.8	99.7	100.0
NT1(2step)	1.5	7.3	59.3	78.6	84.6	0.0	0.0	61.5	99.1	99.9	0.0	0.4	95.0	100.0	100.0
NT1(CUE)	12.5	19.6	63.7	76.4	78.9	0.5	0.5	49.6	83.5	89.4	0.0	0.6	81.0	92.0	95.5
	β														
	Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)			Power(H₁)	Size	Power(H₁)		
Estimators \ β	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
ML	86.6	33.8	5.8	34.2	87.5	98.5	58.3	6.6	56.6	98.9	100.0	79.7	5.0	82.1	100.0
ALS1(1step)	44.8	28.4	22.1	32.6	51.6	55.6	26.9	17.9	42.9	78.2	65.9	20.4	14.9	59.7	94.7
ALS1(2step)	85.8	70.9	53.9	53.3	64.4	96.9	79.8	44.0	34.2	66.0	99.9	91.3	42.5	25.8	77.4
ALS1(CUE)	78.0	68.0	60.6	61.3	67.3	89.8	64.7	38.7	46.3	77.4	97.2	71.7	27.0	50.3	90.0
NT1(1step)	28.1	6.6	12.7	44.0	82.4	44.5	7.8	22.8	75.5	98.3	62.0	6.1	34.8	93.4	99.9
NT1(2step)	61.8	46.7	46.8	63.8	82.8	67.4	27.7	30.7	71.0	94.9	81.0	22.8	28.3	84.5	99.6
NT1(CUE)	66.0	59.4	59.2	66.3	78.1	59.3	29.7	41.1	74.7	93.1	69.6	21.0	41.6	87.9	99.4

See notes to Table 5.

Table 9: Bias($\times$100) and RMSE($\times$100) of the transformed ML estimator for the AR(1) model with two factors

| | $T=6, \gamma=0.4$ | | $T=10, \gamma=0.4$ | |
| | Bias | RMSE | Bias | RMSE |
N	($\times$100)	($\times$100)	($\times$100)	($\times$100)
150	0.04	6.91	-0.27	5.95
300	0.05	4.81	-0.12	4.12
500	-0.10	3.53	-0.02	3.19
	$T=6, \gamma=0.8$		$T=10, \gamma=0.8$	
	Bias	RMSE	Bias	RMSE
N	($\times$100)	($\times$100)	($\times$100)	($\times$100)
150	-1.47	7.56	-2.27	8.60
300	-0.59	5.79	-0.90	6.33
500	-0.01	5.03	-0.20	4.94

Notes: y_{it} is generated as $y_{it} = \alpha_i + \gamma y_{i,t-1} + \xi_{it}$, $\xi_{it} = \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + u_{it}$, $u_{it} \sim iid\mathcal{N}(0,\sigma^2)$, $i = 1, 2, ..., N; t = -49, 48, ...0, 1, ..., T$, with $y_{i,-50} = 0$ and $\sigma^2 = 1$. The factors are generated as: $f_{\ell t} = \rho_{f\ell} f_{\ell,t-1} + \sqrt{1 - \rho_{f\ell}^2} \varepsilon_{f\ell t}$, $\varepsilon_{f\ell t} \sim iid\mathcal{N}(0,1)$, $t = -49, 48, ...0, 1, ..., T$, with $f_{\ell,-50} = 0$, and $\rho_{f\ell} = 0.9$ for $\ell = 1, 2$. The resultant $f_{\ell t}$ values are scaled such that $T^{-1} \sum_{t=1}^{T} f_{\ell t}^2 = 1$. The values of f_t for $t = -49, 48, ...0$ are not scaled. The factor loadings are generated as $\lambda_{\ell i} = \lambda + \eta_{\ell i}$, with $\lambda = 1$ and $\eta_{\ell i} \sim iidN(0,1)$ for $\ell = 1, 2$. The fixed effects, α_i, are generated as $\alpha_i = T^{-1}(\xi_{i1} + \xi_{i2} + ... + \xi_{iT}) + v_i = \lambda_{1i}\bar{f}_1 + \lambda_{2i}\bar{f}_2 + \bar{u}_i + v_i$, where $\bar{f}_\ell = T^{-1} \sum_{t=1}^{T} f_{\ell t}$, $\ell = 1, 2$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and $v_i \sim iid\mathcal{N}(0,1)$. Each $f_{\ell t}$ is generated once and the same $f'_{\ell t}s$ are used throughout the replications for $\ell = 1, 2$. The first 50 observations are discarded. ML is the proposed maximum likelihood estimator. All experiments are based on 1,000 replications.

Table 10: Size(%) and power(%) of the transformed ML estimator for the AR(1) model with two factors

| | $T=6, \gamma=0.4$ | | | | | $T=10, \gamma=0.4$ | | | | |
| | Power(H$_1$) | | Size | Power(H$_1$) | | Power(H$_1$) | | Size | Power(H$_1$) | |
$N \setminus \gamma$	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
150	37.8	16.7	5.0	9.4	30.3	41.5	18.6	5.6	9.2	34.6
300	59.9	22.5	5.2	16.5	57.7	65.5	25.2	5.0	17.5	66.8
500	79.0	30.5	4.5	26.3	81.3	84.1	34.7	4.3	31.8	88.7
	$T=6, \gamma=0.8$					$T=10, \gamma=0.8$				
	Power(H$_1$)		Size	Power(H$_1$)		Power(H$_1$)		Size	Power(H$_1$)	
$N \setminus \gamma$	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
150	34.8	17.0	6.7	1.7	1.4	32.6	17.8	7.7	5.0	9.1
300	41.7	20.9	5.6	0.7	5.3	39.5	18.9	5.1	4.2	13.2
500	50.8	23.7	5.5	1.3	25.3	47.6	20.0	3.8	4.7	33.4

See notes to Table 9.

Table 11: Bias($\times$100) and RMSE($\times$100) of the transformed ML estimator for the ARX(1) model with two factors

	$T = 6$				$T = 10$			
	$\gamma = 0.4$		$\beta = 1.0$		$\gamma = 0.4$		$\beta = 1.0$	
N	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)
150	-0.16	3.82	0.36	5.86	0.00	2.05	0.01	4.14
300	-0.09	2.64	-0.17	4.27	0.03	1.42	-0.02	3.03
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
	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)	($\times$100)
150	0.01	1.46	0.07	2.32	0.01	0.69	0.02	1.57
300	-0.02	1.04	0.01	1.67	0.01	0.52	-0.03	1.18
500	-0.03	0.77	-0.04	1.31	-0.02	0.39	0.01	0.86

Notes: y_{it} is generated as $y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}$, $\xi_{it} = \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + u_{it}$, $u_{it} \sim iid\mathcal{N}(0, \sigma^2)$, $i = 1, 2, ..., N; t = -49, 48, ...0, 1, ..., T$, with $y_{i,-50} = 0$ and $x_{it} = \mu_i + \vartheta_{1i} f_{1t} + \vartheta_{2i} f_{2t} + \breve{x}_{it}$, $\breve{x}_{it} = \rho_x \breve{x}_{i,t-1} + \sqrt{1 - \rho_x^2} \varepsilon_{it}$, with $\breve{x}_{i,-50} = 0$, for $t = -49, 48, ...0, 1, ..., T$, where $\rho_x = 0.8$, $\mu_i \sim iid\mathcal{N}(0, 1)$, and $\varepsilon_{it} \sim iid\mathcal{N}(0, 1)$. The factors $f_{\ell t}$, $\ell = 1, 2$, are generated as in the AR(1) case (see notes to Table 1). The factor loadings are generated as $\vartheta_{\ell i} \sim iid\mathcal{N}(0.5, \sigma_{\ell\vartheta}^2)$ and $\lambda_{\ell i} \sim iid\mathcal{N}(0.5, \sigma_{\ell\lambda}^2)$ for $\ell = 1, 2$, respectively. The fixed effects, α_i, are generated as $\alpha_i = \bar{x}_i + \lambda_{1i} \bar{f}_1 + \lambda_{2i} \bar{f}_2 + \bar{u}_i + v_i$, where $\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$, $T^{-1} \sum_{t=1}^{T} f_{\ell t}^2 = 1$, $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and $v_i \sim iid\mathcal{N}(0, 1)$. The remaining parameters are set at $\beta = 1$, $\sigma_{1\vartheta}^2 = \sigma_{1\lambda}^2 = \sigma_{2\vartheta}^2 = \sigma_{2\lambda}^2 = \sigma^2$, with $\sigma^2 = (0.7 - \gamma^2)/0.5$. Each $f_{\ell t}$ is generated once and the same $f_{\ell t}'s$ are used throughout the replications for $\ell = 1, 2$. The first 50 observations are discarded. ML is the proposed maximum likelihood estimator. All experiments are based on 1,000 replications.

Table 12: Size(%) and power(%) of the transformed ML estimator for the ARX(1) model with two factors

	$T = 6$, $\gamma = 0.4$, $\beta = 1.0$					$T = 10$, $\gamma = 0.4$, $\beta = 1.0$				
			γ							
	Power(H$_1$)	Size		Power(H$_1$)		Power(H$_1$)	Size		Power(H$_1$)	
$N \setminus \gamma$	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40	0.45	0.50
150	78.5	27.9	4.9	30.5	73.7	99.8	71.0	6.0	70.8	99.7
300	97.4	49.6	5.1	47.4	96.1	100.0	93.9	4.9	93.8	100.0
500	100.0	71.9	4.4	68.3	99.8	100.0	99.6	5.4	99.5	100.0
			β							
	Power(H$_1$)	Size		Power(H$_1$)		Power(H$_1$)	Size		Power(H$_1$)	
$N \setminus \gamma$	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
150	36.6	10.9	4.6	14.0	40.3	68.7	23.6	5.3	23.7	68.4
300	69.1	24.8	5.6	20.6	64.2	92.2	39.8	7.1	41.2	91.6
500	86.5	34.2	5.5	33.5	85.6	99.1	60.3	5.3	62.0	99.2
	$T = 6$, $\gamma = 0.8$, $\beta = 1.0$					$T = 10$, $\gamma = 0.8$, $\beta = 1.0$				
			γ							
	Power(H$_1$)	Size		Power(H$_1$)		Power(H$_1$)	Size		Power(H$_1$)	
$N \setminus \gamma$	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
150	100.0	93.5	4.8	91.6	100.0	100.0	100.0	4.0	100.0	100.0
300	100.0	99.9	5.7	99.3	100.0	100.0	100.0	5.3	100.0	100.0
500	100.0	100.0	4.4	100.0	100.0	100.0	100.0	4.9	100.0	100.0
			β							
	Power(H$_1$)	Size		Power(H$_1$)		Power(H$_1$)	Size		Power(H$_1$)	
$N \setminus \gamma$	0.90	0.95	1.00	1.05	1.10	0.90	0.95	1.00	1.05	1.10
150	98.9	54.5	4.6	56.2	98.8	100.0	87.1	5.1	89.4	100.0
300	100.0	83.5	4.5	85.0	100.0	100.0	99.1	6.3	99.4	100.0
500	100.0	96.7	4.7	97.0	100.0	100.0	100.0	4.6	100.0	100.0

See notes to Table 11.

Appendix

A.1   Proof of Theorem 1

The mean of Δy_{i1} conditional on $\Delta y_{i,-S+1}$ and $\Delta x_{i,1-j}, (j = 0, 1, 2...)$ is given by

$$
\begin{aligned}
\zeta_{i1} &= E\left(\Delta y_{i1}|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...\right) \\
&= \gamma^S \Delta y_{i,-S+1} + \beta \sum_{j=0}^{S-1} \gamma^j \Delta x_{i,1-j} + \tilde{g}_{1S} E(\lambda_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...) \\
&= \gamma^S \Delta y_{i,-S+1} + \beta \Delta x_{i1} + \beta \sum_{j=1}^{S-1} \gamma^j \Delta x_{i,1-j} + \lambda \tilde{g}_{1S} + \tilde{g}_{1S} E(\eta_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...).
\end{aligned}
$$

Conditional on $\Delta \mathbf{x}_i = (\Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT})'$ we have

$$
\begin{aligned}
E(\zeta_{i1}|\Delta \mathbf{x}_i) &= E(\gamma^S \Delta y_{i,-S+1} + \beta \Delta x_{i1}|\Delta \mathbf{x}_i) + \beta \sum_{j=1}^{S-1} \gamma^j E(\Delta x_{i,1-j}|\Delta \mathbf{x}_i) + \lambda \tilde{g}_{1S} \\
&\quad + \tilde{g}_{1S} E\left[E(\eta_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...)|\Delta \mathbf{x}_i\right] \\
&= \gamma^S \tilde{b} + \beta \Delta x_{i1} + \beta \sum_{j=1}^{S-1} \gamma^j \left(b_j + \boldsymbol{\pi}_j' \Delta \mathbf{x}_i\right) + \lambda \tilde{g}_{1S} + \tilde{g}_{1S}(h + \boldsymbol{\varphi}' \Delta \mathbf{x}_i) \\
&= \left(\gamma^S \tilde{b} + \beta \sum_{j=1}^{S-1} \gamma^j b_j + \lambda \tilde{g}_{1S} + \tilde{g}_{1S} h\right) + \left(\beta \mathbf{e}_1 + \beta \sum_{j=1}^{S-1} \gamma^j \boldsymbol{\pi}_j + \tilde{g}_{1S} \boldsymbol{\varphi}\right)' \Delta \mathbf{x}_i \\
&= b + \boldsymbol{\pi}' \Delta \mathbf{x}_i,
\end{aligned}
$$

where $\mathbf{e}_1 = (1, 0, .., 0)'$ and the following results are used

$$
\begin{aligned}
E(\Delta x_{i,1-j}|\Delta \mathbf{x}_i) &= b_j + \boldsymbol{\pi}_j' \Delta \mathbf{x}_i, \qquad (j = 0, 1, 2...) \\
E\left[E(\eta_i|\Delta y_{i,-m+1}, \Delta x_{i1}, \Delta x_{i0}, ...)|\Delta \mathbf{x}_i\right] &= h + \boldsymbol{\varphi}' \Delta \mathbf{x}_i.
\end{aligned}
$$

Then

$$
\begin{aligned}
\Delta y_{i1} &= \zeta_{i1} + \tilde{g}_{1S} \left[\lambda_i - E(\lambda_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...)\right] + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j} \\
&= E(\zeta_{i1}|\Delta \mathbf{x}_i) + \left[\zeta_{i1} - E(\zeta_{i1}|\Delta \mathbf{x}_i)\right] + \tilde{g}_{1S} \left[\lambda_i - E(\lambda_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...)\right] + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j} \\
&= b + \boldsymbol{\pi}' \Delta \mathbf{x}_i + v_{i1},
\end{aligned}
$$

where

$$
v_{i1} = \left[\zeta_{i1} - E(\zeta_{i1}|\Delta \mathbf{x}_i)\right] + \tilde{g}_{1S} \left[\lambda_i - E(\lambda_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...)\right] + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j},
$$

and

$$
\begin{aligned}
\zeta_{i1} - E(\zeta_{i1}|\Delta \mathbf{x}_i) &= \gamma^S \left[\Delta y_{i,-S+1} - E(\Delta y_{i,-S+1}|\Delta \mathbf{x}_i)\right] + \beta \sum_{j=0}^{S-1} \gamma^j \left[\Delta x_{i,1-j} - E(\Delta x_{i,1-j}|\Delta \mathbf{x}_i)\right] \\
&\quad + \tilde{g}_{1S} \left\{E(\lambda_i|\Delta y_{i,-S+1}, \Delta x_{i1}, \Delta x_{i0}, ...) - E\left[E(\lambda_i|\Delta y_{i,-m+1}, \Delta x_{i1}, \Delta x_{i0}, ...)|\Delta \mathbf{x}_i\right]\right\}.
\end{aligned}
$$

A.2 Derivation of the log-likelihood

Here we show how (8) is derived from (7). Using

$$\left(\boldsymbol{\Omega}+\phi\mathbf{g}\mathbf{g}'\right)^{-1} = \boldsymbol{\Omega}^{-1} - \frac{\phi\boldsymbol{\Omega}^{-1}\mathbf{g}\mathbf{g}'\boldsymbol{\Omega}^{-1}}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)},$$

and

$$\left|\boldsymbol{\Omega}+\phi\mathbf{g}\mathbf{g}'\right| = |\boldsymbol{\Omega}|\left(1+\phi\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right),$$

the log-likelihood function (7) can be written as

$$
\begin{aligned}
\ell\left(\boldsymbol{\psi}\right) = & -\frac{NT}{2}\ln\left(2\pi\right) - \frac{TN}{2}\ln(\sigma^2) - \frac{N}{2}\ln|\boldsymbol{\Omega}| - \frac{N}{2}\ln(1+\phi\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}) \\
& -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma - \lambda\mathbf{g}\right)'\left[\boldsymbol{\Omega}^{-1} - \frac{\phi\boldsymbol{\Omega}^{-1}\mathbf{g}\mathbf{g}'\boldsymbol{\Omega}^{-1}}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)}\right]\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma - \lambda\mathbf{g}\right) \quad \text{(A.1)}
\end{aligned}
$$

where $|\boldsymbol{\Omega}| = 1 + T\left(\omega - 1\right)$. Let $\mathbf{v}_i = \mathbf{v}_i(\gamma) = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma$, and note that

$$
\begin{aligned}
& \sum_{i=1}^{N}\left(\mathbf{v}_i - \lambda\mathbf{g}\right)'\left[\boldsymbol{\Omega}^{-1} - \frac{\phi\boldsymbol{\Omega}^{-1}\mathbf{g}\mathbf{g}'\boldsymbol{\Omega}^{-1}}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)}\right]\left(\mathbf{v}_i - \lambda\mathbf{g}\right) \\
= & \sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{\phi\sum_{i=1}^{N}\left(\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)^2}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)} + N\lambda^2\left[\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g} - \frac{\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)^2}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)}\right] \\
& -2N\lambda\left[\mathbf{g}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}} - \frac{\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)}\right] \\
= & \sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{\phi\sum_{i=1}^{N}\left(\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)^2 - N\lambda^2\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right) + 2N\lambda\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)} \\
= & \sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{N\phi\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{g} - N\lambda^2\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right) + 2N\lambda\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)}{1+\phi\left(\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right)},
\end{aligned}
$$

where $\bar{\mathbf{v}} = N^{-1}\sum_{i=1}^{N}\mathbf{v}_i$, and $\mathbf{B}_N = \mathbf{B}_N(\gamma) = N^{-1}\sum_{i=1}^{N}\mathbf{v}_i(\gamma)\mathbf{v}_i'(\gamma)$. Therefore, the log-likelihood function, (A.1), can be written as

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

A.3 Derivatives of the log-likelihood function

We give the analytical formulas of the first and second derivatives of the log-likelihood function (10). Note that

$$|\boldsymbol{\Omega}\left(\omega\right)| = g\left(\omega\right) = 1 + T\left(\omega - 1\right),$$

and

$$
\boldsymbol{\Omega}^{-1} = \frac{1}{g\left(\omega\right)}
\begin{pmatrix}
\begin{array}{c|cccc}
T & T-1 & \dots & 2 & 1 \\
\hline
T-1 & (T-1)\omega & \dots & 2\omega & \omega \\
T-2 & & & & \\
& & & & \\
2 & 2\omega & 2\left[(T-2)\omega - (T-3)\right] & (T-2)\omega - (T-3) \\
1 & \omega & \dots & (T-2)\omega - (T-3) & (T-1)\omega - (T-2)
\end{array}
\end{pmatrix}
$$

$$\frac{\partial \mathbf{\Omega}^{-1}}{\partial \omega} = \frac{-1}{g\left(\omega\right)^2} \begin{pmatrix} T^2 & T(T-1) & T(T-2) & \cdots & T \\ T(T-1) & (T-1)^2 & (T-1)(T-2) & \cdots & (T-1) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ T & (T-1) & (T-2) & \cdots & 1 \end{pmatrix} = \frac{-1}{g\left(\omega\right)^2} \mathbf{\Phi}.$$

Also

$$\frac{\partial \ln |\mathbf{\Omega}|}{\partial \omega} = \frac{T}{1 + T(\omega - 1)} = \frac{T}{g\left(\omega\right)}.$$

Using the above expressions the first derivatives are given by

$$N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\varphi}} = \frac{1}{\sigma^2}\left[N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\mathbf{\Omega}^{-1}\mathbf{v}_i - \frac{\left(N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\mathbf{\Omega}^{-1}\mathbf{q}\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{v}_i\right) + \kappa\overline{\Delta\mathbf{W}}'\mathbf{\Omega}^{-1}\mathbf{q}}{1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}\right],$$

$$\begin{aligned} N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \omega} &= -\frac{1}{2}\frac{T}{g\left(\omega\right)} + \frac{\mathbf{q}'\mathbf{\Phi}\mathbf{q}}{2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)g\left(\omega\right)^2} + \frac{1}{2\sigma^2 g\left(\omega\right)^2}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\mathbf{\Phi}\mathbf{v}_i\right] \\ &\quad + \frac{\left(\mathbf{q}'\mathbf{\Phi}\mathbf{q}\right)\left(\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}\right)}{2\sigma^2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)^2 g\left(\omega\right)^2} \\ &\quad + \frac{-\mathbf{q}'\mathbf{\Phi}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} - \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Phi}\mathbf{q} + \kappa^2\mathbf{q}'\mathbf{\Phi}\mathbf{q} - 2\kappa\mathbf{q}'\mathbf{\Phi}\bar{\mathbf{v}}}{2\sigma^2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)g\left(\omega\right)^2}, \end{aligned}$$

$$N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\mathbf{\Omega}^{-1}\mathbf{v}_i - \frac{\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}}{1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}\right],$$

$$N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \kappa} = \frac{1}{\sigma^2}\left[\frac{-\kappa\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q} + \mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}}{1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}\right],$$

$$\begin{aligned} N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \mathbf{q}} &= -\frac{\mathbf{\Omega}^{-1}\mathbf{q}}{1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}} - \left[\frac{\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\mathbf{\Omega}^{-1}\bar{\mathbf{v}}}{\sigma^2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)^2}\right]\mathbf{\Omega}^{-1}\mathbf{q} \\ &\quad + \frac{\mathbf{\Omega}^{-1}\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{\Omega}^{-1}\mathbf{q} + \kappa\mathbf{\Omega}^{-1}\bar{\mathbf{v}}}{\sigma^2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)}. \end{aligned}$$

The second derivatives are as follows:

$$N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'} = \frac{1}{\sigma^2}\left[-N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\mathbf{\Omega}^{-1}\Delta\mathbf{W}_i + \frac{\left(N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i\mathbf{\Omega}^{-1}\mathbf{q}\mathbf{q}'\mathbf{\Omega}^{-1}\Delta\mathbf{W}_i'\right)}{1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}}\right],$$

$$\begin{aligned} N^{-1}\frac{\partial \ell\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\varphi}\partial \omega} &= \frac{-1}{\sigma^2 g\left(\omega\right)^2}\left[N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\mathbf{\Phi}\mathbf{v}_i\right] \\ &\quad - \left[\frac{\mathbf{q}'\mathbf{\Phi}\mathbf{q}}{\sigma^2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)^2 g\left(\omega\right)^2}\right]\left[\left(N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\mathbf{\Omega}^{-1}\mathbf{q}\mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{v}_i\right) + \kappa\overline{\Delta\mathbf{W}}'\mathbf{\Phi}\mathbf{q}\right] \\ &\quad - \frac{1}{\sigma^2\left(1 + \mathbf{q}'\mathbf{\Omega}^{-1}\mathbf{q}\right)g\left(\omega\right)^2}\left[\begin{array}{c} -\mathbf{q}'\mathbf{\Phi}\left(N^{-1}\sum_{i=1}^{N}\mathbf{v}_i\Delta\mathbf{W}_i'\right)\mathbf{\Omega}^{-1}\mathbf{q} \\ -\mathbf{q}'\mathbf{\Omega}^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{v}_i\Delta\mathbf{W}_i'\right)\mathbf{\Phi}\mathbf{q} - \kappa\overline{\Delta\mathbf{W}}'\mathbf{\Phi}\mathbf{q} \end{array}\right], \end{aligned}$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\sigma^2\partial\boldsymbol{\varphi}} = \frac{-1}{\sigma^4}\left[N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{\left(N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{q}\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i\right) + \kappa\overline{\Delta\mathbf{W}}'\boldsymbol{\Omega}^{-1}\mathbf{q}}{1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}}\right],$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\kappa\partial\boldsymbol{\varphi}'} = \frac{-\mathbf{q}'\boldsymbol{\Omega}^{-1}\overline{\Delta\mathbf{W}}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\varphi}\partial\mathbf{q}'} = \frac{-\left(N^{-1}\sum_{i=1}^{N}\left(\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i + \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{q}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\right) + \kappa\overline{\Delta\mathbf{W}}'\boldsymbol{\Omega}^{-1}\right)}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)}$$
$$+ \frac{2\left(N^{-1}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{q}\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i + \kappa\overline{\Delta\mathbf{W}}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\right)}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\omega\partial\omega} = \frac{T^2}{2g\left(\omega\right)^2} - \left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\right)\left(\frac{-\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} + \left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)\left(2Tg\left(\omega\right)\right)}{2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2 g\left(\omega\right)^4}\right) - \frac{T}{\sigma^2 g\left(\omega\right)^3}\left(N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Phi}\mathbf{v}_i\right)$$
$$+ \frac{\left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\right)\left[-\mathbf{q}'\boldsymbol{\Phi}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Phi}\mathbf{q} + \kappa^2\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} - 2\kappa\mathbf{q}'\boldsymbol{\Phi}\bar{\mathbf{v}}\right]}{2\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2 g\left(\omega\right)^4}$$
$$- \frac{\left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\right)\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\left\{-\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} + \left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)\left(Tg\left(\omega\right)\right)\right\}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^3 g\left(\omega\right)^4}$$
$$+ \frac{\mathbf{q}'\boldsymbol{\Phi}\mathbf{B}_N\boldsymbol{\Phi}\mathbf{q}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)g\left(\omega\right)^4}$$
$$+ \frac{\left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} + \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Phi}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} + 2\kappa\mathbf{q}'\boldsymbol{\Phi}\bar{\mathbf{v}}\right)\left\{-\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} + \left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)\left(2Tg\left(\omega\right)\right)\right\}}{2\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2 g\left(\omega\right)^4},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\sigma^2\partial\omega} = \frac{-1}{2\sigma^4 g\left(\omega\right)^2}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Phi}\mathbf{v}_i\right]$$
$$- \left[\frac{\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}}{2\sigma^4\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2 g\left(\omega\right)^2}\right]\left[\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right]$$
$$- \frac{1}{2\sigma^4\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)g\left(\omega\right)^2}\left[-\mathbf{q}'\boldsymbol{\Phi}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Phi}\mathbf{q} + \kappa^2\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} - 2\kappa\mathbf{q}'\boldsymbol{\Phi}\bar{\mathbf{v}}\right],$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\kappa\partial\omega} = \frac{\left(\kappa\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} - \mathbf{q}'\boldsymbol{\Phi}\bar{\mathbf{v}}\right)\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right) + \left(-\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + \mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\right)}{\sigma^2 g\left(\omega\right)^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\mathbf{q}\partial\omega} = \frac{\boldsymbol{\Phi}\mathbf{q}\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right) - \left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\right)\left(\boldsymbol{\Omega}^{-1}\mathbf{q}\right)}{g\left(\omega\right)^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2}$$
$$+ \frac{\boldsymbol{\Phi}\mathbf{q}\left[\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right] + \mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\left[\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\boldsymbol{\Omega}^{-1}\mathbf{q} + \kappa\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right]}{\sigma^2 g\left(\omega\right)^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2}$$
$$- \frac{2\left(\mathbf{q}'\boldsymbol{\Phi}\mathbf{q}\right)\left[\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + 2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right]\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)\boldsymbol{\Omega}^{-1}\mathbf{q}}{\sigma^2 g\left(\omega\right)^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^4}$$
$$+ \frac{-\boldsymbol{\Phi}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Phi}\mathbf{q} + \kappa^2\boldsymbol{\Phi}\mathbf{q} - \kappa\boldsymbol{\Phi}\bar{\mathbf{v}}}{\sigma^2 g\left(\omega\right)^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)}$$
$$- \frac{\left(-\mathbf{q}'\boldsymbol{\Phi}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Phi}\mathbf{q} + \kappa^2\mathbf{q}'\boldsymbol{\Phi}\mathbf{q} - 2\kappa\mathbf{q}'\boldsymbol{\Phi}\bar{\mathbf{v}}\right)\left(\boldsymbol{\Omega}^{-1}\mathbf{q}\right)}{\sigma^2 g\left(\omega\right)^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\sigma^2\partial\sigma^2} = \frac{T}{2\sigma^4} - \frac{1}{\sigma^6}\left[N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \frac{\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}+2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}}\right],$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\kappa\partial\sigma^2} = \frac{\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} - \mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{\sigma^4\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\mathbf{q}\partial\sigma^2} = \left[\frac{\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}+2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{\sigma^4\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2}\right]\boldsymbol{\Omega}^{-1}\mathbf{q} - \frac{\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\boldsymbol{\Omega}^{-1}\mathbf{q}+\kappa\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}}{\sigma^4\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\kappa\partial\kappa} = \frac{-\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)},$$

$$N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\mathbf{q}\partial\kappa} = \frac{\left(-2\kappa\boldsymbol{\Omega}^{-1}\mathbf{q}+\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right) + 2\left(\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q} - \mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\boldsymbol{\Omega}^{-1}\mathbf{q}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2},$$

$$\begin{aligned}
N^{-1}\frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\mathbf{q}\partial\mathbf{q}'} = {}& -\frac{\boldsymbol{\Omega}^{-1}\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right) - 2\boldsymbol{\Omega}^{-1}\mathbf{q}\mathbf{q}'\boldsymbol{\Omega}^{-1}}{\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2} \\
& -\frac{2\boldsymbol{\Omega}^{-1}\mathbf{q}\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}+\kappa\bar{\mathbf{v}}'\boldsymbol{\Omega}^{-1}\right)}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2} \\
& +\frac{\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}+2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\boldsymbol{\Omega}^{-1}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2} \\
& +\frac{4\left(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}+2\kappa\mathbf{q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\boldsymbol{\Omega}^{-1}\mathbf{q}\mathbf{q}'\boldsymbol{\Omega}^{-1}}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^3} \\
& +\frac{\left(\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1} - \kappa^2\boldsymbol{\Omega}^{-1}\right)}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)} - \frac{\left(\boldsymbol{\Omega}^{-1}\mathbf{B}_N\boldsymbol{\Omega}^{-1}\mathbf{q} - \kappa^2\boldsymbol{\Omega}^{-1}\mathbf{q}+\kappa\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}\right)\left(2\mathbf{q}'\boldsymbol{\Omega}^{-1}\right)}{\sigma^2\left(1+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}\right)^2}.
\end{aligned}$$

A.4 Derivation of the log-likelihood function in the multifactor case

Under Assumption 8 we can rewrite model (22) as

$$\begin{aligned}
y_{it} &= \alpha_i + \gamma y_{i,t-1} + \mathbf{f}_t'\boldsymbol{\lambda}_i + u_{it} \\
&= \alpha_i + \gamma y_{i,t-1} + \mathbf{f}_t'\boldsymbol{\lambda} + \mathbf{f}_t'\boldsymbol{\eta}_i + u_{it}.
\end{aligned}$$

Eliminating the individual effects by first-differencing yields

$$\begin{aligned}
\Delta y_{it} &= \gamma\Delta y_{i,t-1} + \mathbf{g}_t'\boldsymbol{\lambda}_i + \Delta u_{it} \\
&= \gamma\Delta y_{i,t-1} + \mathbf{g}_t'\boldsymbol{\lambda} + \mathbf{g}_t'\boldsymbol{\eta}_i + \Delta u_{it} \quad \text{for } t = 2, 3, ..., T. \quad (\text{A.2})
\end{aligned}$$

Under Assumption 1, by recursive substitution, we have the following expression for $t = 1$

$$\Delta y_{i1} = \boldsymbol{\lambda}_i'\tilde{\mathbf{g}}_1 + v_{i1}, \quad (\text{A.3})$$

where $\tilde{\mathbf{g}}_1 = \sum_{j=0}^{\infty}\gamma^j\mathbf{g}_{1-j}$, $v_{i1} = \sum_{j=0}^{\infty}\gamma^j\Delta u_{i,1-j}$ with $E(v_{i1}) = 0$ and $var(v_{i1}) = \sigma^2\omega$. Using (23) in (A.2), this equation together with (A.3) can be written as

$$\begin{aligned}
\Delta y_{i1} &= \boldsymbol{\lambda}'\tilde{\mathbf{g}}_1 + \boldsymbol{\eta}_i'\tilde{\mathbf{g}}_1 + v_{i1} \\
\Delta y_{it} &= \gamma\Delta y_{i,t-1} + \boldsymbol{\lambda}'\mathbf{g}_t + \boldsymbol{\eta}_i'\mathbf{g}_t + \Delta u_{it}, \quad (t = 2, 3, ..., T).
\end{aligned}$$

In matrix notation the above system of equations can be expressed as

$$\Delta\mathbf{y}_i = \Delta\mathbf{W}_i\gamma + \mathbf{G}\boldsymbol{\lambda} + \boldsymbol{\xi}_i, \tag{A.4}$$

where $\Delta\mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$, $\Delta\mathbf{W}_i = (0, \Delta y_{i1}, ..., \Delta y_{i,T-1})'$, $\mathbf{G} = (\tilde{\mathbf{g}}_1, \mathbf{g}_2, ..., \mathbf{g}_T)'$ and $\boldsymbol{\xi}_i = \mathbf{G}\boldsymbol{\eta}_i + \mathbf{r}_i$ with $\mathbf{r}_i = (v_{i1}, \Delta u_{i2}, ..., \Delta u_{iT})'$.

In equation (A.4) $\boldsymbol{\lambda}$ is not separately identified from the elements of $\mathbf{G}$. Thus, defining the identity matrix $\mathbf{I}_m = \frac{1}{\sigma}\boldsymbol{\Omega}_\eta^{1/2}\sigma\boldsymbol{\Omega}_\eta^{-1/2}$, where recall from Assumption 8 that $\boldsymbol{\Omega}_\eta$ is a positive definite matrix, we can write

$$\mathbf{G}\boldsymbol{\lambda} = \mathbf{G}\frac{1}{\sigma}\boldsymbol{\Omega}_\eta^{1/2}\sigma\boldsymbol{\Omega}_\eta^{-1/2}\boldsymbol{\lambda} = \mathbf{Q}\boldsymbol{\kappa},$$

where $\mathbf{Q} = (1/\sigma)\mathbf{G}\boldsymbol{\Omega}_\eta^{1/2}$ and $\kappa = \sigma\boldsymbol{\Omega}_\eta^{-1/2}\boldsymbol{\lambda}$.

Recall further that $E(\mathbf{r}_i\mathbf{r}_i') = \sigma^2\boldsymbol{\Omega}$ and since $\boldsymbol{\eta}_i$ and u_{it} are independently distributed we have

$$Var(\boldsymbol{\xi}_i) = \sigma^2\boldsymbol{\Omega} + \mathbf{G}\boldsymbol{\Omega}_\eta\mathbf{G}' = \sigma^2\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right).$$

Hence, the log-likelihood function of the transformed model (A.4) is given by

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &= -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma^2) - \frac{N}{2}\ln\left|\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right| \\
&\quad -\frac{1}{2\sigma^2}\sum_{i=1}^N (\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma - \mathbf{Q}\boldsymbol{\kappa})'\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1}(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma - \mathbf{Q}\boldsymbol{\kappa}).
\end{aligned} \tag{A.5}$$

For a fixed T, the above log-likelihood function depends only on a fixed number of unknown parameters, $\boldsymbol{\theta} = \left(\gamma, \omega, \sigma^2, \boldsymbol{\kappa}', vec(\mathbf{Q})'\right)'$.

To obtain the ML estimators, since $\boldsymbol{\Omega}$ is a positive definite matrix and $\mathbf{Q}\mathbf{Q}'$ is rank deficinet (recall that by assumption $m < T$), we first note that

$$\left|\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right| = \left|\boldsymbol{\Omega}\right|\left|\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\right|,$$

and using the Woodbury matrix identity

$$\begin{aligned}
\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1} &= \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{Q}(\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1} \\
&= \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1},
\end{aligned} \tag{A.6}$$

where

$$\mathbf{A} = \mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}.$$

Using these results the log-likelihood function in (A.5) can be written as

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &\propto -\frac{NT}{2}\ln(\sigma^2) - \frac{N}{2}\ln|\boldsymbol{\Omega}| - \frac{N}{2}\ln\left|\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\right| - \frac{1}{2\sigma^2}\sum_{i=1}^N (\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma - \mathbf{Q}\boldsymbol{\kappa})' \\
&\quad \times \left[\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\right](\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma - \mathbf{Q}\boldsymbol{\kappa}),
\end{aligned} \tag{A.7}$$

with $|\boldsymbol{\Omega}| = 1 + T(\omega - 1)$. Further, since

$$\sum_{i=1}^{N} (\mathbf{v}_i - \mathbf{Q}\boldsymbol{\kappa})' \left[\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\right] (\mathbf{v}_i - \mathbf{Q}\boldsymbol{\kappa})$$

$$= \sum_{i=1}^{N} \mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \sum_{i=1}^{N} \mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}} + N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}$$

$$- N\bar{\mathbf{v}}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\boldsymbol{\kappa} + N\bar{\mathbf{v}}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\boldsymbol{\kappa} + N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\boldsymbol{\kappa} - N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\boldsymbol{\kappa}$$

$$= \sum_{i=1}^{N} \mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - \sum_{i=1}^{N} \mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - 2N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}} + 2N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}$$

$$+ N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\boldsymbol{\kappa} - N\boldsymbol{\kappa}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\boldsymbol{\kappa},$$

where $\mathbf{v}_i = \mathbf{v}_i(\gamma) = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\gamma$, and $\bar{\mathbf{v}} = N^{-1}\sum_{i=1}^{N}\mathbf{v}_i$, (A.7) can be written as

$$N^{-1}\ell(\boldsymbol{\theta}) \quad \propto \quad -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln|\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}|$$

$$-\frac{1}{2\sigma^2} \left\{ \begin{array}{c} N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i \\ +\boldsymbol{\kappa}'[\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}(\mathbf{I}_m - \mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})]\boldsymbol{\kappa} - 2\boldsymbol{\kappa}'[(\mathbf{I}_m - \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1})\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}] \end{array} \right\}$$

Note that the first two terms in the barckets using (A.6) can be written as

$$N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i - N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{v}_i = N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i.$$

Hence

$$N^{-1}\ell(\boldsymbol{\theta}) \quad \propto \quad -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln|\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}| \tag{A.8}$$

$$-\frac{1}{2\sigma^2} \left\{ \begin{array}{c} N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i \\ +\boldsymbol{\kappa}'[\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}(\mathbf{I}_m - \mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})]\boldsymbol{\kappa} - 2\boldsymbol{\kappa}'[(\mathbf{I}_m - \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1})\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}] \end{array} \right\}.$$

Also

$$\mathbf{I}_m - \mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q} = \mathbf{I}_m - \mathbf{A}^{-1}\left(\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q} - \mathbf{I}_m\right) = \mathbf{I}_m - \mathbf{A}^{-1}\left(\mathbf{A} - \mathbf{I}_m\right) = \mathbf{A}^{-1},$$

$$\mathbf{I}_m - \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1} = \mathbf{I}_m - \left(\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q} - \mathbf{I}_m\right)\mathbf{A}^{-1} = \mathbf{I}_m - \left(\mathbf{A} - \mathbf{I}_m\right)\mathbf{A}^{-1} = \mathbf{A}^{-1},$$

and

$$\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q} = \mathbf{I}_m - \mathbf{A}^{-1} = \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}.$$

The log-likelihood in (A.8) then becomes

$$N^{-1}\ell(\boldsymbol{\theta}) \quad \propto \quad -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\ln|\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}| \tag{A.9}$$

$$-\frac{1}{2\sigma^2} \left\{ \begin{array}{c} N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i \\ +\boldsymbol{\kappa}'\left(\mathbf{I}_m - \mathbf{A}^{-1}\right)\boldsymbol{\kappa} - 2\boldsymbol{\kappa}'\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}] \end{array} \right\}.$$

Setting the partial derivative of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\kappa}$ to zero, it now readily follows that

$$\left(\mathbf{I}_m - \mathbf{A}^{-1}\right)\hat{\boldsymbol{\kappa}} = \mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}},$$

which yields (recall that $\mathbf{Q}$ has the full column rank of m)

$$\hat{\boldsymbol{\kappa}} = (\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}. \tag{A.10}$$

Next, taking partial derivatives with respect to σ^2 and solving out for this we have

$$
\begin{aligned}
T\hat{\sigma}^2 &= N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega}+\mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i \\
&\quad + \hat{\boldsymbol{\kappa}}'\left(\mathbf{I}_m - \mathbf{A}^{-1}\right)\hat{\boldsymbol{\kappa}} - 2\hat{\boldsymbol{\kappa}}'\mathbf{A}^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}}.
\end{aligned} \tag{A.11}
$$

Substituting for $\hat{\boldsymbol{\kappa}}$ from (A.10) in (A.11) now yields

$$\hat{\sigma}^2 = T^{-1}\left\{ \begin{array}{c} N^{-1}\sum_{i=1}^{N}\mathbf{v}_i'\left(\boldsymbol{\Omega}+\mathbf{Q}\mathbf{Q}'\right)^{-1}\mathbf{v}_i \\ -\bar{\mathbf{v}}'\boldsymbol{\Omega}^{-1}\mathbf{Q}\mathbf{A}^{-1}(\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\boldsymbol{\Omega}^{-1}\bar{\mathbf{v}} \end{array} \right\}. \tag{A.12}$$

Finally, substituting (A.10) and (A.12) into (A.9) we obtain

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

where $\boldsymbol{\theta} = \left(\gamma, \omega, vec(\mathbf{Q})'\right)'$. Recall that, if required, $(\boldsymbol{\Omega}+\mathbf{Q}\mathbf{Q}')^{-1}$ can be expanded in terms of $\boldsymbol{\Omega}$ using the Woodbury matrix identity in (A.6).

A.5 Derivation of R_y^2

Consider the panel data model

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \xi_{it}, \quad \xi_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it},$$

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

where $\mathbf{f}_t = (f_{1t}, ..., f_{mt})'$, $\boldsymbol{\lambda}_i' = (\lambda_{1i}, ..., \lambda_{mi})'$, $\boldsymbol{\vartheta}_i = (\vartheta_{1i}, ..., \vartheta_{mi})'$, $|\gamma| < 1$ and $|\rho_x| < 1$. Due to the dependence of x_{it} and ξ_{it} on the same unobserved factors, $\mathbf{f}_t = (f_{1t}, ..., f_{mt})'$, the regressors and the errors of the above regression are correlated. Following Pesaran and Smith (1994) we base the measurement of R^2 on the following reduced form regressions

$$y_{it} = d_i + \gamma y_{i,t-1} + \beta \breve{x}_{it} + \breve{\xi}_{it}, \ \breve{\xi}_{it} = \mathbf{c}_i'\mathbf{f}_t + u_{it}, \tag{A.13}$$

where

$$d_i = \alpha_i + \beta\mu_i \text{ and } \mathbf{c}_i = \beta\boldsymbol{\vartheta}_i + \boldsymbol{\lambda}_i.$$

It is clear that in (A.13) the regressors, $\breve{x}_{it}$, and the errors, $\breve{\xi}_{it}$, are uncorrelated and standard formula for R^2 can be used. But to deal with the heterogeneity across the different equations in the panel we use the following average measure of fit

$$R_y^2 = 1 - \frac{N^{-1}\sum_{i=1}^{N}Var(\breve{\xi}_{it})}{N^{-1}\sum_{i=1}^{N}Var(y_{it})}.$$

Using the above results, and noting that u_{it} and ε_{it} are uncorrelated with f_t, it readily follows that

$$
\begin{aligned}
Var(\breve{\xi}_{it}) &= \mathbf{c}_i' Var(\mathbf{f}_t)\mathbf{c}_i + \sigma^2, \\
Var(y_{it}) &= \frac{\beta^2 Var(\breve{x}_{it}) + Var(\breve{\xi}_{it})}{1 - \gamma^2}.
\end{aligned}
$$

If we assume that the elements of $\mathbf{f}_t$ are mutually orthogonal and have zero means we have

$$
R_y^2 = \frac{\beta^2 Var(\breve{x}_{it}) + \left[\sum_{\ell=1}^m \left\{ \left(N^{-1}\sum_{i=1}^N c_{\ell i}^2\right)\left(T^{-1}\sum_{t=1}^T f_{\ell t}^2\right)\right\} + \sigma^2\right]\gamma^2}{\beta^2 Var(\breve{x}_{it}) + \sum_{\ell=1}^m \left\{ \left(N^{-1}\sum_{i=1}^N c_{\ell i}^2\right)\left(T^{-1}\sum_{t=1}^T f_{\ell t}^2\right)\right\} + \sigma^2}. \tag{A.14}
$$

It is easily seen that $R_y^2 \geq \gamma^2$ with the equality holding only if $\beta = 0$, namely when an AR(1) specification is considered.

References

Ahn, S. C., Y.H. Lee, and P. Schmidt (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics 101*, 219-255.

Ahn, S.C., Y.H. Lee, and P. Schmidt (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics 174*, 1-14.

Anderson, T.W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*, 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*, 277-297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29-52.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*, 1229-1279.

Bai, J. (2013). Likelihood approach to dynamic panel models with interactive effects. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2332992.

Binder, M., C. Hsiao, and M.H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory 21*, 795-837.

Blundell, R. and S.R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115-143.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics 18*, 5-46.

Chamberlain, G. (1984). 'Panel Data', in *Handbook of Econometrics*, Volume 2, eds. Z. Griliches and M. Intriligator, Amsterdam: North-Holland, 1247-1318.

Chudik, A., M.H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *The Econometrics Journal 14*, C45-C90.

Elhorst, J.P. (2005). Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis 37*, 85-106.

Elhorst, J.P. (2010). Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics 40*, 272-282.

Hayakawa, K. and M. H. Pesaran (2012). Robust standard errors in transformed likelihood estimation of dynamic panel data models. CESifo Working Paper Series 3850, CESifo Group Munich.

Holtz-Eakin, D., W. Newey, and H.S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*,1371-95.

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*, 107-150.

Jacobs, J.P.A.M, J.E. Ligthart, and H. Vrijburg (2009). Dynamic panel data models featuring endogenous interaction and spatially correlated errors. CentER Discussion Paper Series No. 2009-92.

Kapetanios, G., M.H. Pesaran, and T. Yagamata (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics 160*, 326-348.

Lee, L. and J. Yu (2010). Some recent developments in spatial panel data models. *Regional Science and Urban Economics 40*, 255-271.

MaCurdy, T.E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*, 83-114.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*, 69-85.

Nauges, C. and A. Thomas (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique 70*, 53-74.

Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*, 967-1012.

Pesaran, M.H. and R. J. Smith (1994). A generalized R^2 criterion for regression models estimated by the instrumental variables method. *Econometrica 62*, 705-710.

Pesaran, M.H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*, 182-202.

Robertson, D. and V. Sarafidis (2013). IV estimation of panels with factor residuals. Cambridge Working Paper in Economics No. 1321.

Phillips P. C. B. and D. Sul (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics 137*, 162-188.

Sarafidis, V. and D. Robertson (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *The Econometrics Journal 12*, 62-81.

Sarafidis V., and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews 31*, 483-531.

Su, L. and Z. Yang (2007). QML estimation of dynamic panel data models with spatial errors. Singapore Management University, Manuscript.