

Short T Dynamic Panel Data Models with Individual, Time and Interactive Effects*

Kazuhiko Hayakawa
Hiroshima University

M. Hashem Pesaran
University of Southern California,
and Trinity College, Cambridge

L. Vanessa Smith
University of York

25 July 2021

Abstract

This paper proposes a transformed quasi maximum likelihood (TQML) estimator for short T dynamic fixed effects panel data models allowing for interactive effects through a multi-factor error structure. The proposed estimator is robust to the heterogeneity of the initial values and common unobserved effects, whilst at the same time allowing for standard fixed and time effects. It is applicable to both stationary and unit root cases. The order condition for identification of the number of interactive effects is established, and conditions are derived under which the parameters are almost surely locally identified. It is shown that global identification in the presence of the lagged dependent variable cannot be guaranteed. The TQML estimator is proven to be consistent and asymptotically normally distributed. A sequential multiple testing likelihood ratio procedure is also proposed for estimation of the number of factors which is shown to be consistent. Finite sample results obtained from Monte Carlo simulations show that the proposed procedure for determining the number of factors performs very well and the TQML estimator has small bias and RMSE, and correct empirical size in most settings. The practical use of the TQML approach is demonstrated by means of two empirical illustrations from the literature on cross county crime rates and cross country growth regressions.

JEL Classifications: C12, C13, C23

Keywords: short T dynamic panels, unobserved common factors, quasi maximum likelihood, interactive effects, multiple testing, sequential likelihood ratio tests, crime rate, growth regressions

*The authors would like to thank the editor, three referees, Alex Chudik and Ron Smith for helpful suggestions, Liudas Giraitis for useful discussions, and Vasilis Sarafidis for helpful comments on a preliminary version of the paper. Part of this paper was written whilst Hayakawa was visiting the University of Cambridge as a JSPS Postdoctoral Fellow for Research Abroad. He acknowledges financial support from the JSPS Fellowship and the Grant-in-Aid for Scientific Research (KAKENHI 17K03660, 20H01484, 20K20760) provided by the JSPS. Pesaran and Smith acknowledge financial support from ESRC Grant No. ES/I031626/1.

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. Both generalised 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), Ahn and Schmidt (1995), Blundell and Bond (1998), Hsiao et al. (2002), Binder et al. (2005) and Moral-Benito (2013). As a natural extension of the traditional two-way error component model, the recent literature considers the case where individual and time effects are included in a multiplicative manner.¹ Such a structure is termed *time-varying individual effects* by Ahn et al. (2001, 2013) or *interactive fixed effects* by Bai (2009), otherwise characterised as a multi-factor error structure.

Main contributions to this literature include the papers by Phillips and Sul (2007) and Sarafidis and Robertson (2009) who investigate the implications of ignoring the interactive fixed effects for the behaviour of the fixed effects and GMM estimators, respectively.² Ahn et al. (2001) consider a single factor error structure and propose a quasi-differencing approach to eliminate the factor, subsequently applying GMM to consistently estimate the parameters. The quasi-differencing transformation was originally proposed by Chamberlain (1984) and implemented by Holtz-Eakin et al. (1988) in the context of a bivariate panel autoregression. Nauges and Thomas (2003) follow the same approach, and in addition to prior first-differencing to eliminate the fixed effects, they also consider a single factor structure for the errors. Ahn et al. (2013) extend their quasi-differencing approach to a multi-factor error structure. More recently, Hayakawa (2012) proposes a GMM estimator based on the projection method to deal with short dynamic panel data models with interactive fixed effects, while Robertson and Sarafidis (2015) propose an instrumental variable estimation procedure that introduces new parameters to represent the unobserved covariances between the instruments and the unobserved factors. Comments on the latter approach are provided by Ahn (2015) and Hayakawa (2016). As an alternative to GMM, Bai (2013) proposes a quasi-maximum likelihood (QML) approach applied to the original dynamic panel data model without differencing, treating time effects as free parameters. To deal with possible correlations between the factor loadings and the regressors Bai follows Mundlak (1978) and Chamberlain (1982) and specifies linear relationships between the factor loadings and the regressors to be estimated along with the other parameters. A survey of short T panel data models with interactive effects can be found in Sarafidis and Wansbeek (2012).

Building on the work of Hsiao et al. (2002), this paper proposes a transformed QML approach (TQML), applied to the short T dynamic panel data model after first-differencing, that allows for interactive effects in addition to the standard individual and time fixed effects. In this way we directly address the empirical question of whether inclusion of individual and time effects are sufficient to deal with error cross-sectional dependence in short T panels. Our approach also accounts for heterogeneity of the initial values and the common factors in an integrated framework, and allows the initial values to be correlated with the fixed effects and other model parameters. We establish the order condition for identification of the number of interactive effects, discuss identification based on moment conditions and the likelihood framework, and finally derive conditions under which the parameters are almost surely locally identified. It is shown that global identification in the presence of the lagged dependent variable cannot be guaranteed. These results can be useful for the development of QML theory in the case of more general models. The TQML estimator is shown to be consistent and asymptotically normally distributed

¹The terms individual effects and fixed effects are used interchangeably, as are the terms interactive effects and common factors.

²For the case of panel models with interactive fixed effects when N and T are both large, see, for example, Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011), Chudik et al. (2011), and Kapetanios et al. (2011).

both for stationary and unit root cases. We also propose a sequential multiple testing likelihood ratio (MTLR) procedure to estimate the number of interactive effects and show that it delivers a consistent estimator of the true number of factors, and has the added advantage that it does not depend on an arbitrary choice of a maximum number of factors as required in the large N and T factor literature.

The theoretical results are further supported by means of extensive Monte Carlo experiments, covering both stationary and unit root cases, showing that the methods proposed for estimating the number of factors and the unknown parameters of the model perform well in most settings. It is also shown that the TQML estimator compares favourably to the QML estimator of Bai (2013) and the GMM type estimators proposed in the literature, and interestingly enough is reasonably robust to a number of important departures from its underlying assumptions. The practical use of the TQML approach is demonstrated with two empirical illustrations from the literature, focusing on the importance of allowing for interactive effects in empirical analysis. The first illustration estimates a dynamic version of the panel data model considered by Cornwell and Trumbull (1994) and Baltagi (2006) to explain the incidence of crime across counties in North Carolina; the second illustration estimates growth regressions using the recent data analysed by Acemoglu et al. (2019). In the case of both illustrations we find statistically significant evidence of interactive effects, even after allowing for fixed and time effects.

Our contribution differs from Bai (2013) in a number of important respects, despite the fact that both approaches make use of the likelihood framework. First, our procedure applies maximum likelihood estimation after first-differencing that eliminates the individual effects, whereas Bai (2013) considers the model in levels. Second, we assume the initial values, y_{i0} , $i = 1, 2, \dots, N$, follow the postulated dynamic processes from some arbitrary initial values, thus also allowing the underlying processes to have unit roots. Bai notes that "the initial observation y_{i0} may or may not follow the [considered] dynamic process" but in his analysis he follows Bhargava and Sargan (1983) and assumes (rather than derives) initial values can be modelled as linear projections on the regressors and the factor loadings. Third, we address the issue of identification of short T dynamic panel data models with a multi-factor error structure, and propose a sequential multiple testing likelihood procedure for estimating the number of factors, topics that are not addressed by Bai (2013).

The rest of this paper is organised as follows. Section 2 discusses the relation to the literature. Section 3 sets out the dynamic panel data model and its assumptions. Section 4 considers the quasi maximum likelihood estimation with details of derivations given in Appendix S.3. Identification of the number of factors and the parameters of the model are discussed in Section 5. Section 6 establishes the asymptotic properties of the TQML estimator. Section 7 presents the sequential MTLR procedure for estimating the number of factors. Section 8 describes the Monte Carlo experiments and provides finite sample results on the performance of the sequential MTLR estimator for the number of factors, and the proposed TQML estimator. Empirical illustrations are provided in Section 9. The final section presents some concluding remarks. All technical proofs are provided in the Appendix. Details of alternative GMM estimators used in the Monte Carlo experiments together with additional Monte Carlo results are provided in an online supplement.

Notations: Let $\mathbf{w} = (w_1, w_2, \dots, w_n)'$ and $\mathbf{A} = (a_{ij})$ be an $n \times 1$ vector and an $n \times n$ matrix, respectively. Denote the Euclidean norm of \mathbf{w} and the Frobenius norm of \mathbf{A} by $\|\mathbf{w}\| = (\sum_{i=1}^n w_i^2)^{1/2}$ and $\|\mathbf{A}\| = [\text{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$ respectively, and the largest and smallest eigenvalue of \mathbf{A} by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$. If $\{y_n\}_{n=1}^\infty$ is any real sequence and $\{x_n\}_{n=1}^\infty$ is a sequence of positive real numbers, then $y_n = O(x_n)$ if there exists a positive finite constant K such that $|y_n|/x_n \leq K$ for all n . $y_n = o(x_n)$ if $y_n/x_n \rightarrow 0$ as $n \rightarrow \infty$. If $\{y_n\}_{n=1}^\infty$ and $\{x_n\}_{n=1}^\infty$ are both positive sequences of real numbers, then $y_n = \Theta(x_n)$ if there exists $N_0 \geq 1$ and positive finite constants K_0 and K_1 such that $\inf_{n \geq N_0} (y_n/x_n) \geq K_0$ and $\sup_{n \geq N_0} (y_n/x_n) \leq K_1$. Positive, possibly large, fixed constants will be denoted by K , K_0 , K_1 and so on, that could take different values in different equations. c , c_{\min} and c_{\max} will also denote positive fixed constants. Small positive constants will be denoted by ϵ . $E_0(\cdot)$ denotes expectations taken under the true probability measure. \rightarrow_p and $\xrightarrow{a.s.}$ denote convergence in probability and almost sure (a.s.) convergence,

respectively. \rightarrow_d denotes convergence in distribution for fixed T and as $N \rightarrow \infty$.

2 Related literature

For the purpose of relating our approach to the literature we start with a simple dynamic panel data model with a single common factor and abstract from fixed effects. Adding more factors and fixed and time effects does not materially change the narrative. Specifically we consider the simple dynamic panel data model

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \lambda_i f_t + \varepsilon_{it}, \text{ for } t = 1, 2, 3, \dots, T, i = 1, 2, \dots, N, \quad (1)$$

where x_{it} is strictly exogenous, such that $E(x_{it}\varepsilon_{jt'}) = 0$ for all i, j, t and t' . It will be assumed that λ_i and x_i are uncorrelated and have zero means, namely $E(x_{it}) = 0$, $E(\lambda_i) = 0$, and $E(x_{it}\lambda_i) = 0$, for all i and t . These assumptions are made to simplify the derivations of rank conditions for identification and are not needed. The key assumptions are that conditional on f_t , $(y_{it}, x_{it}$ and $\varepsilon_{it})$ are cross-sectionally independent, and $f_t \neq 0$, for some t , in addition to x_{it} being strictly exogenous. $\varepsilon_{it} \sim IID(0, \sigma_i^2)$, with $\sup_i(\sigma_i^2) < c_{\max} < \infty$, and $\inf_i(\sigma_i^2) > c_{\min} > 0$. Also for the purpose of illustration we assume the initial values, y_{i0} , are obtained by projection of y_{i0} onto $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$ and f_0 , and assume the following data generating process (DGP) for y_{i0} :

$$y_{i0} = \boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + v_i, \quad i = 1, 2, \dots, N, \quad (2)$$

where $v_i \sim IID(0, \sigma_{i,v}^2)$ is distributed independently of $\{\varepsilon_{it}, t = 1, 2, \dots, T\}$ and $\sigma_{i,v}^2$ could differ from σ_i^2 . Since T is short, how initial values, y_{i0} for $i = 1, 2, \dots, N$ are generated play a crucial role in identification and estimation of the unknown parameters of interest, namely γ and β .

There are two main approaches to identification and estimation of γ and β . The first one builds on the pioneering contribution of Holtz-Eakin et al. (1988) and employs a quasi-differencing procedure to eliminate the factor loadings, λ_i , viewed as nuisance (incidental) parameters. The second approach advanced by Bai (2013) treats λ_i as free parameters and estimates them together with the factors, f_t , and the parameters of interest using the maximum likelihood approach. In what follows we consider these two approaches and highlight their main underlying assumptions, and discuss their relations to the transformed quasi-ML approach that we propose in this paper. With this in mind we also introduce a new GMM method which treats the factors, f_t , as given constants and avoids the incidental parameter problem by conditioning on

$$d_N(\boldsymbol{\lambda}) = N^{-1} \sum_{i=1}^N \lambda_i^2 \quad (3)$$

rather than the individual factor loadings, λ_i . The limiting value of $d_N(\boldsymbol{\lambda})$ as $N \rightarrow \infty$, depends on the degree of pervasiveness (strength) of the factor. In general we could have $\sum_{i=1}^N \lambda_i^2 = \Theta(N^\alpha)$, where α measures the strength of the factor. When the factor is strong $\alpha = 1$ and $\lim_{N \rightarrow \infty} [d_N(\boldsymbol{\lambda})] = \bar{d}(\boldsymbol{\lambda}) > c > 0$. But when the factor is not strong ($\alpha < 1$) $\lim_{N \rightarrow \infty} [d_N(\boldsymbol{\lambda})] = 0$. It is typically assumed that $\alpha = 1$, but it is also of interest to consider the possibility of weak factors and their implications for identification and estimation under different estimation approaches. We shall also see that once we allow the initial values to depend on the loadings, λ_i , for consistent estimation of γ all the methods we consider require the orthogonality assumption

$$E(\lambda_i \varepsilon_{it}) = 0, \text{ for all } i \text{ and } t. \quad (4)$$

In their more recent contribution, Ahn et al. (2013, ALS) consider a multi-factor panel regression where they allow a subset of the regressors to be weakly exogenous, and use lags and leads of the strictly exogenous regressors as instruments for the weakly exogenous variables (see Section 3.2 of ALS). As a result, their set up does not apply to a *pure* dynamic panel data model without any exogenous regressors.³

³ALS do not consider dynamics in their Monte Carlo experiments either.

It is, therefore, important that the properties of the GMM approach are specifically investigated for the dynamic specification in (1). In what follows we consider two alternative approaches considered in the literature to eliminate the factor loadings.

2.1 Quasi-differenced GMM estimator

The quasi-differencing idea was introduced by Holtz-Eakin et al. (1988) and has been adopted in the literature by a number of authors. Eliminating the incidental parameters λ_i from (1) by quasi first-differencing yields:

$$y_{it} - b_t y_{i,t-1} = \gamma (y_{i,t-1} - b_t y_{i,t-2}) + \beta (x_{it} - b_t x_{i,t-1}) + \nu_{it}, \text{ for } t = 2, 3, \dots, T, \quad (5)$$

where

$$\nu_{it} = \varepsilon_{it} - b_t \varepsilon_{i,t-1} \quad (6)$$

and $b_t = f_t/f_{t-1}$. The values of b_t for $t = 2, 3, \dots, T$ are treated as given unknown constants to be estimated along with γ and β . Note that b_t is invariant to the scaling of f_t , and the importance of the unobserved factor, f_t , is determined by $d_N(\boldsymbol{\lambda})$ defined by (3).

Using (5) we note that under the strict exogeneity assumption we have

$$E(x_{is} \nu_{it}) = E[x_{is} (\varepsilon_{it} - b_t \varepsilon_{i,t-1})] = 0 \text{ for } t = 1, 2, \dots, T \text{ and } s = 1, 2, \dots, T,$$

and no further assumptions concerning the factor loadings are required. But when $\beta = 0$, we need to use y_{i0} and y_{i1} as instruments and for these to be valid, we also require that

$$E(\lambda_i \varepsilon_{it}) = 0, \text{ for all } i \text{ and } t.$$

To see this note that (recall that $y_{i0} = \boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + v_i$, with $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$)

$$\begin{aligned} E(y_{i0} \nu_{it}) &= E[y_{i0} (\varepsilon_{it} - b_t \varepsilon_{i,t-1})], \text{ for } t = 2, 3, \dots \\ &= E[(\boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + v_i) (\varepsilon_{it} - b_t \varepsilon_{i,t-1})] = f_0 E[\lambda_i (\varepsilon_{it} - b_t \varepsilon_{i,t-1})]. \end{aligned}$$

Therefore, in general where $f_0 \neq 0$, it is required that $E(\lambda_i \varepsilon_{it}) = b_t E(\lambda_i \varepsilon_{i,t-1})$, which will be met for all values of b_t if $E(\lambda_i \varepsilon_{it}) = 0$, for all t . This condition is also required when we consider using y_{i1} as an instrument. In what follows we assume that $E(\lambda_i \varepsilon_{it}) = 0$, holds.

For illustrative purposes we focus on the relatively simple case where $T = 3$, and assume the available observations are $(y_{i0}, y_{it}, x_{it}, t = 1, 2, 3; i = 1, 2, \dots, N)$. Let $\mathbf{z}_i = (y_{i0}, y_{i1}, x_{i1}, x_{i2}, x_{i3})' = (\mathbf{w}'_i, \mathbf{x}'_i)'$ be the set of instruments under consideration and write the moment conditions as $E[\mathbf{m}_N(\boldsymbol{\theta}_0)] = 0$, where $\boldsymbol{\theta} = (\gamma, \beta, b_3)'$, $b_3 = f_3/f_2$, $f_2 \neq 0$, $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$, and

$$\mathbf{m}_N(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^N \mathbf{z}_i \nu_{i3}(\boldsymbol{\theta}). \quad (7)$$

Note that under quasi-differencing $f_2 \neq 0$ and $E(\lambda_i \varepsilon_{it}) = 0$ are the necessary conditions for identification. There are also other moment conditions that could be used. For example $E(y_{i0} \nu_{i2}) = 0$, and $E(x_{is} \nu_{i2}) = 0$, and $E(x_{is} \nu_{i1}) = 0$, for $s = 1, 2, 3$. But including these moment conditions involve the additional parameters, b_1 and b_2 and do not materially impact the nature of the rank conditions needed for identification of γ and β .

We first note that

$$\nu_{i3}(\boldsymbol{\theta}) = y_{i3} - (b_3 + \gamma) y_{i2} + b_3 \gamma y_{i1} - \beta x_{i3} + b_3 \beta x_{i2}, \quad (8)$$

from which it follows immediately that when $\beta = 0$ it will not be possible to distinguish between γ and b_3 , and these parameters are not identified. Notice also that in this case considering the additional moment condition $E(y_{i0} \nu_{i2}) = 0$ yields

$$E\{y_{i0} [y_{i2} - (b_2 + \gamma) y_{i1} + b_2 \gamma y_{i0} - \beta x_{i2} + b_2 \beta x_{i1}]\} = 0,$$

and when $\beta = 0$, again we have the same identification problem - we are only able to consistently estimate $b_2 + \gamma$ and γb_2 , and further a priori information is needed to distinguish between γ and b_2 . For example, if it is known that $|\gamma| < 1$ and we end up with two estimates one inside and another outside the unit circle we could then use the small root to represent γ .

Another possibility would be when it is known with *certainty* that $\beta \neq 0$. In such a case it is possible to estimate γ by GMM subject to the usual rank conditions. In the present application it is required that the 5×3 matrix \mathbf{D} and the 5×5 matrix \mathbf{S} defined by

$$\frac{\partial \mathbf{m}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \rightarrow_p \mathbf{D}, \text{ and } N \mathbf{m}_N(\boldsymbol{\theta}) \mathbf{m}_N'(\boldsymbol{\theta}) \rightarrow_p \mathbf{S},$$

are both full rank for all $\boldsymbol{\theta} \in \mathbb{R}^3$. Details of the derivations of \mathbf{D} and \mathbf{S} are provided in Section S.2 of the online supplement, where it is shown that \mathbf{S} is positive definite so long as $\boldsymbol{\Sigma}_{xx} = N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i'$ is a positive definite matrix.⁴ Also it is shown that

$$\mathbf{D} = - \begin{pmatrix} q_1 & f_0 f_2 \bar{d}(\boldsymbol{\lambda}) & \boldsymbol{\pi}_0' \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \\ q_2 & (\gamma f_0 + f_1) f_2 \bar{d}(\boldsymbol{\lambda}) & \boldsymbol{\pi}_1' \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \\ \boldsymbol{\Sigma}_{xx} [(\gamma - b) \boldsymbol{\pi}_1 + \beta \mathbf{e}_2] & \mathbf{0} & \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \end{pmatrix},$$

where \mathbf{e}_s is a 3×1 vector of zeros except for its s^{th} element which is unity,

$$\begin{aligned} q_1 &= \boldsymbol{\pi}_0' \boldsymbol{\Sigma}_{xx} [(\gamma - b) \boldsymbol{\pi}_1 + \beta \mathbf{e}_2] + f_0 [(\gamma - b) (\gamma f_0 + f_1) + f_2] \bar{d}(\boldsymbol{\lambda}) + \gamma (\gamma - b) \bar{\sigma}^2, \\ q_2 &= \boldsymbol{\pi}_1' \boldsymbol{\Sigma}_{xx} [(\gamma - b) \boldsymbol{\pi}_1 + \beta \mathbf{e}_2] + (\gamma f_0 + f_1) [(\gamma - b) (\gamma f_0 + f_1) + f_2] \bar{d}(\boldsymbol{\lambda}) + (\gamma - b) (1 + \gamma^2) \bar{\sigma}^2 \end{aligned}$$

$\boldsymbol{\pi}_1 = \gamma \boldsymbol{\pi}_0 + \beta \mathbf{e}_1$, $\bar{\sigma}^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$, and $\bar{d}(\boldsymbol{\lambda}) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \lambda_i^2$. It is now clear that \mathbf{D} does not have full rank for all values of $\boldsymbol{\theta} \in \mathbb{R}^3$. For example, as anticipated earlier, \mathbf{D} becomes rank deficient when $\bar{d}(\boldsymbol{\lambda}) = 0$, namely if the common factor, f_t , is not strong. Even when $\bar{d}(\boldsymbol{\lambda}) > 0$, the rank condition fails if $f_0 = f_1 = 0$. Recall that the validity of the moment condition (7) only requires that $f_2 \neq 0$, and is silent regarding the values of f_0 and f_1 . We have already seen that there is an identification problem when $\beta = 0$. The \mathbf{D} matrix for this case is given by

$$\mathbf{D} = - \begin{pmatrix} f_0 [(\gamma - b) (\gamma f_0 + f_1) + f_2] \bar{d}(\boldsymbol{\lambda}) + \gamma (\gamma - b) \bar{\sigma}^2 & f_0 f_2 \bar{d}(\boldsymbol{\lambda}) \\ (\gamma f_0 + f_1) [(\gamma - b) (\gamma f_0 + f_1) + f_2] \bar{d}(\boldsymbol{\lambda}) + (\gamma - b) (1 + \gamma^2) \bar{\sigma}^2 & (\gamma f_0 + f_1) f_2 \bar{d}(\boldsymbol{\lambda}) \end{pmatrix}.$$

It is easily seen that $|\mathbf{D}| = -\bar{d}(\boldsymbol{\lambda}) \bar{\sigma}^2 (\gamma - b) (\gamma f_1 - f_0) f_2$ which could take zero values for $\gamma = f_0/f_1$ and/or $\gamma = b = f_3/f_2$ even if $f_t \neq 0$ for $t = 1, 2, 3$ and the factor is strong, namely $\bar{d}(\boldsymbol{\lambda}) \neq 0$. In short, there is no guarantee that the minimand for the quasi-differenced GMM estimator will have a unique solution.

2.2 Ahn et al. (2013) GMM approach

Ahn et al. (2013) use a different normalisation to eliminate the factor loadings, λ_i . In the case of a single factor model their approach reduces to using the normalisation $f_T = 1$ to eliminate the factor loadings, λ_i . It is implicitly assumed that f_T is known a priori to be non-zero. In the case of the application considered above (with $T = 3$), setting $f_3 = 1$ yields $\lambda_i = y_{i3} - \gamma y_{i2} - \beta x_{i3} - \varepsilon_{i3}$, and we obtain (assuming $f_2 \neq 0$)

$$y_{i2} = \gamma y_{i1} + \beta x_{i2} + f_2 (y_{i3} - \gamma y_{i2} - \beta x_{i3} - \varepsilon_{i3}) + \varepsilon_{i2},$$

that can be written equivalently as (with $a = 1/f_2$)

$$v_{i3}(\boldsymbol{\psi}) = \varepsilon_{i3} - a \varepsilon_{i2} = y_{i3} - (a + \gamma) y_{i2} + \gamma a y_{i1} - (x_{i3} - a x_{i2}) \beta$$

⁴To simplify the notations in deriving \mathbf{D} we have assumed that $\sigma_{i,v}^2 = \sigma_i^2$.

which has the same form as (8), with $\psi = (\gamma, a, \beta)'$. So long as the same a priori information is imposed on whether f_t are zero or not, the manner by which λ_i is eliminated is non-consequential. As an alternative normalisation suppose that we set $f_2 = 1$, and use

$$\lambda_i = y_{i2} - \gamma y_{i1} - \beta x_{i2} - \varepsilon_{i2},$$

to eliminate λ_i from the equations for y_{i3} , for $i = 1, 2, \dots, N$. Then

$$y_{i3} = \gamma y_{i2} + \beta x_{i3} + f_3 (y_{i2} - \gamma y_{i1} - \beta x_{i2} - \varepsilon_{i2}) + \varepsilon_{i3},$$

and not-surprisingly we again arrive at (8) with $b_3 = f_3$. Therefore, the same identification issues discussed above in relation to the quasi-GMM approach would also apply to the ALS type normalisation.

For the set of nonlinear moment conditions proposed by Ahn et al. (2001, 2013), Hayakawa (2016) shows that these do not always satisfy the global identification assumption which is necessary for consistency of GMM estimation. He further shows that the same problem occurs for the moment conditions proposed by Robertson and Sarafidis (2015) and Hayakawa (2012), since their moment conditions become identical to those of Ahn et al. (2001, 2013) in some cases. The results are demonstrated for the ALS model $y_{it} = \beta' x_{it} + \lambda_i' f_t + \varepsilon_{it}$ where x_{it} is allowed to include a lagged dependent variable $y_{i,t-1}$. It readily follows from his results that for the case of a pure dynamic panel model with no additional regressors, a quadratic equation in γ arises leading generally to two solutions for γ and could lead to global identification failure.

2.3 Likelihood approach Bai (2013)

The likelihood method advanced by Bai (2013), instead of eliminating the factor loadings, treats λ_i as random variables. He considers both cases, when λ_i are distributed independently of the regressors as well as when they are modelled as linear functions of them, with the errors distributed independently over i . He proposes two estimation approaches one where he follows the approach of Bhargava and Sargan (1983) and models the initial values in terms of cross section averages of the regressors, independently of the dynamic processes generating y_{it} for $t = 1, 2, \dots, T$ and another conditional on the initial values, y_{i0} .⁵ When T is short Bai motivates and formulates the likelihood by treating the factor loadings as random and estimates their sample variance matrix to avoid the incidental parameter problem, which is what we propose to do in this paper as well. However, we differ from Bai in two respects. We explicitly model fixed effects and work with first differences of the panel regression model, thus allowing for arbitrary correlations between fixed effects and the regressors, whilst under Bai's approach the fixed effects are implicitly treated as random or are assumed to be linearly correlated with the regressors à la the Mundlak-Chambelain projection device. We also provide a more general treatment of the initial values that explicitly relates $\Delta y_{i1} = y_{i1} - y_{i0}$ to the unobserved past history of the dynamic panel under consideration that allows for initialisations from a finite past as well as unit roots. In addition we establish the condition under which the Mundlak type linear projection can be justified for the dynamic panel data model. Furthermore, we allow the regressors to share one or all of the latent factors that drives y_{it} . It is also perhaps worth noting that Bai does not provide any proofs for the short T case, and simply states that "the standard theory of the quasi-maximum likelihood applies". He also simply states that β and γ are identified subject to an order condition without considering whether the related rank condition is also met. See Section 4.1 of Bai (2013). In contrast, we provide a detailed analysis of the identification, estimation and inference problems whilst also allowing for interactive effects in the process generating the regressors. Lastly, Bai does not provide a method for selecting the number of factors when T is fixed as N tends to infinity.

⁵When T is fixed and $N \rightarrow \infty$, the conditional likelihood approach is applicable only when the initial values, $y_{i0}, i = 1, 2, \dots, N$, do not depend on γ and/or β .

With regard to accommodating fixed effects, under the method of Bai (2013) the unit-specific intercept is absorbed in the interactive factor part and treated as another factor to be estimated, so that the number of factors in this case is $\tilde{m} = m + 1$. In the dynamic AR(1) panel data model, for example, where the process has started in the distant past, the unit-specific intercept does not imply that $f_{1t} = 1$ for all t , but $f_{10} = 1/(1 - \gamma)$ whilst $f_{1t} = 1$, for $t = 1, 2, \dots$. This has bearing on what normalisation can be validly imposed on $f_{1t}, f_{2t}, \dots, f_{\tilde{m}t}$ for $t = 0, 1, \dots, T$, as discussed below. When x_{it} is included then other issues arise relating to the past values of x_{it} for $t = -1, -2, \dots$ that need to be resolved; another issue that Bai (2013) does not address explicitly, but simply assumes a process for y_{i0} . Moreover, by treating unit-specific intercepts as a factor, Bai's approach requires the use of the Mundlak-Chamberlain projection device to account for possible correlation between the corresponding loadings (the fixed effects) and x_{it} , and rules out the unit-specific intercepts to be spatially correlated and/or heteroskedastic, which could be restrictive and renders Bai's approach inconsistent. Some small sample evidence on the adverse effects of spatially correlated fixed effects on Bai's QMLE is provided in Section 8.2.3.

To illustrate the issue of normalisation, consider the panel AR(1) model

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \lambda_i f_t + \varepsilon_{it}, \text{ for } t = 1, 2, 3, \dots, T, i = 1, 2, \dots, N.$$

To simplify the analysis suppose that $|\gamma| < 1$, $\varepsilon_{it} \sim IID(0, \sigma^2)$, $\sup_s |f_{-s}| < K$, and that $\{y_{it}\}$ has started in the distant past. Then

$$y_{i0} = \frac{\alpha_i}{1 - \gamma} + \lambda_i \sum_{s=0}^{\infty} \gamma^s f_{-s} + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,-s}.$$

Suppose that $\sum_{s=0}^{\infty} \gamma^s f_{-s} = f_0^*$ exists (this follows if $|f_0^*| < K$). Then

$$y_{i0} = \frac{\alpha_i}{1 - \gamma} + \lambda_i f_0^* + v_i,$$

where $v_i = \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,-s}$, and $\sum_{s=0}^{\infty} \gamma^s f_{-s} = f_0^*$. Also, $E(v_i) = 0$, $E(v_i^2) = \frac{\sigma^2}{1 - \gamma^2} = \omega^2$. For $T = 3$

$$\begin{aligned} y_{i0} &= \frac{\alpha_i}{1 - \gamma} + \lambda_i f_0^* + v_i \\ y_{i1} &= \gamma y_{i0} + \alpha_i + \lambda_i f_1 + \varepsilon_{i1} \\ y_{i2} &= \gamma y_{i1} + \alpha_i + \lambda_i f_2 + \varepsilon_{i2} \\ y_{i3} &= \gamma y_{i2} + \alpha_i + \lambda_i f_3 + \varepsilon_{i3}. \end{aligned}$$

Bai treats the above model as a two factor model with $\mathbf{f}_t = (f_{1t}, f_{2t})'$, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \lambda_{i2})' = (\alpha_i, \lambda_i)'$ where

$$\mathbf{F} = \begin{pmatrix} f_{10} & f_{20} \\ f_{11} & f_{21} \\ f_{12} & f_{22} \\ \vdots & \vdots \\ f_{1T} & f_{2T} \end{pmatrix} = \begin{pmatrix} 1/(1 - \gamma) & f_0^* \\ 1 & f_1 \\ 1 & f_2 \\ \vdots & \vdots \\ 1 & f_T \end{pmatrix}.$$

In this application the identification restrictions used in Bai (2013), namely $\mathbf{F}^+ = (\mathbf{I}_2, \mathbf{F}_2')'$ which sets $f_{10} = \frac{1}{1 - \gamma} = 1$, $f_{11} = f_{20} = 0$, and $f_{21} = 1$, imposes an invalid restriction on the first column of \mathbf{F} . To impose valid identification restrictions, *a priori* knowledge regarding the presence of individual-specific effects and the initialisation of $\{y_{it}\}$ are needed. It is easily seen that adding time effects does not alter the above conclusions.

2.4 Bias-corrected method of moments

Bai's short T log-likelihood approach and the transformed quasi maximum likelihood (TQML) proposed in this paper estimate the moments of the factor loadings, λ_i , instead of eliminating them. To illustrate how the two approaches are related, as with the likelihood approaches we derive moment conditions without first eliminating λ_i . We refer to this as the bias-corrected method of moments as in Chudik and Pesaran (2021). For the purpose of illustration and without loss of generality we abstract from exogenous regressors and focus on the simple case where $\beta = 0$ and $\boldsymbol{\pi}_0 = 0$, and set $T = 3$. Using (1) and (2), under the orthogonality condition given by (4) we have

$$E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{it} - \gamma y_{i,t-1}) \right] = f_0 f_t E[d_N(\boldsymbol{\lambda})], \text{ for } t = 1, 2, \quad (9)$$

$$E \left[N^{-1} \sum_{i=1}^N (y_{it} - \gamma y_{i,t-1})^2 \right] = f_t^2 E[d_N(\boldsymbol{\lambda})] + E \left(N^{-1} \sum_{i=1}^N \varepsilon_{it}^2 \right), \text{ for } t = 1, 2 \quad (10)$$

where $d_N(\boldsymbol{\lambda})$ is defined by (3) and⁶

$$E \left[N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0}) (y_{i2} - \gamma y_{i1}) \right] = f_1 f_2 E[d_N(\boldsymbol{\lambda})]. \quad (11)$$

Assuming further that $E(\varepsilon_{it}^2) = \sigma_i^2$, then $E \left(N^{-1} \sum_{i=1}^N \varepsilon_{it}^2 \right) = \bar{\sigma}_N^2$, and using (10) we have

$$E \left[N^{-1} \sum_{i=1}^N (y_{i2} - \gamma y_{i1})^2 - N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0})^2 \right] = (f_2^2 - f_1^2) E[d_N(\boldsymbol{\lambda})]. \quad (12)$$

The four moment conditions (9), (11) and (12) can now be used to estimate γ . To this end it is useful to distinguish between strong and weak factor cases, namely when $d_N(\boldsymbol{\lambda}) \rightarrow \bar{d} > 0$, and $d_N(\boldsymbol{\lambda}) \rightarrow 0$, respectively. When the factor is weak we have

$$\begin{aligned} \lim_{N \rightarrow \infty} E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{it} - \gamma y_{i,t-1}) \right] &= 0, \text{ for } t = 1, 2 \\ \lim_{N \rightarrow \infty} E \left[N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0}) (y_{i2} - \gamma y_{i1}) \right] &= 0, \end{aligned}$$

and these moment conditions can be used to uniquely identify γ , even if $E(\varepsilon_{i1}^2) \neq E(\varepsilon_{i2}^2)$. This result is in contrast to the quasi-differenced GMM approach that breaks down under a weak factor scenario.

When the factor is strong, $d_N(\boldsymbol{\lambda}) > 0$, we need to use a normalisation since λ_i (or $d_N(\boldsymbol{\lambda})$ in the present context) can not be identified from f_t . Here it is convenient to set $f_2 = 1$ and eliminate $d_N(\boldsymbol{\lambda})$ from (9), (11) and (12) to yield

$$\frac{E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{i2} - \gamma y_{i1}) \right]}{E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{i1} - \gamma y_{i0}) \right]} = \frac{f_2}{f_1} = \frac{1}{f_1},$$

and

$$\frac{E \left[N^{-1} \sum_{i=1}^N (y_{i2} - \gamma y_{i1})^2 - N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0})^2 \right]}{E \left[N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0}) (y_{i2} - \gamma y_{i1}) \right]} = \frac{f_2^2 - f_1^2}{f_2 f_1} = \frac{1 - f_1^2}{f_1}.$$

⁶We also have $E \left(N^{-1} \sum_{i=1}^N y_{i0}^2 \right) = f_0^2 E[d_N(\boldsymbol{\lambda})] + E \left(N^{-1} \sum_{i=1}^N v_i^2 \right)$. But since in general $E(v_i^2) \neq E(\sigma_{it}^2)$ this moment condition does not help with identification of γ .

Further eliminating f_1 we obtain

$$\begin{aligned} & \frac{E \left[N^{-1} \sum_{i=1}^N (y_{i2} - \gamma y_{i1})^2 - N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0})^2 \right]}{E \left[N^{-1} \sum_{i=1}^N (y_{i1} - \gamma y_{i0}) (y_{i2} - \gamma y_{i1}) \right]} \\ &= \frac{E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{i2} - \gamma y_{i1}) \right]}{E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{i1} - \gamma y_{i0}) \right]} - \frac{E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{i1} - \gamma y_{i0}) \right]}{E \left[N^{-1} \sum_{i=1}^N y_{i0} (y_{i2} - \gamma y_{i1}) \right]}, \end{aligned}$$

which can be used to estimate γ . But there is no guarantee that the real solution to the above moment condition will be unique.

3 The dynamic panel data model

In this paper we consider a multi-variate and multi-factor version of (1), but explicitly allow for fixed and time effects. Although, as noted by Bai (2013) and others, heterogeneous individual effects can be implicitly allowed for in interactive factor models, standard GMM and likelihood approaches require such effects to be uncorrelated with the errors. See the orthogonality condition given by (4). But in practice most researchers start with panel data models with fixed effects, where such effects are allowed to have non-zero correlations with the errors and the regressors. Finally, by starting with a standard panel data model our estimation strategy enables the researchers to investigate the importance of allowing for (additional) interactive effects for their empirical analysis.

Accordingly we consider the following standard dynamic panel data model with time and fixed effects

$$y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta' \mathbf{x}_{it} + \boldsymbol{\eta}_i' \mathbf{f}_t + u_{it}, \text{ for } t = 1, 2, \dots, T, \text{ and } i = 1, 2, \dots, N, \quad (13)$$

where \mathbf{x}_{it} is a $k \times 1$ vector of regressors that vary both across i and t , $|\gamma| < K$, β is a $k \times 1$ vector of unknown coefficients, with $\|\beta\| < K$, and K denotes a finite positive constant. α_i and δ_t denote unit-specific fixed effects and time effects, respectively. $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{mt})'$, an $m \times 1$ vector of unobserved common factors, and $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \dots, \eta_{im})'$, an $m \times 1$ vector of associated factor loadings, with u_{it} denoting the remaining idiosyncratic error terms. This specification includes a number of models considered in the literature and reviewed in Section 1 as special cases. It also provides a direct generalisation of Hsiao and Tahmiscioglu (2008) who consider estimation of (13) with *IID* errors using the transformed MLE procedure. The explicit inclusion of time effects, δ_t , in the model also allows us, without loss of generality, to assume the factor loadings, $\boldsymbol{\eta}_i$, have zero means. Note that $\delta_t + \boldsymbol{\eta}_i' \mathbf{f}_t$ can be written equivalently as $\boldsymbol{\lambda}_i' \mathbf{f}_t$, where $\boldsymbol{\eta}_i = \boldsymbol{\lambda}_i - \boldsymbol{\lambda}$, and $\boldsymbol{\delta}_t = \boldsymbol{\lambda}' \mathbf{f}_t$, where $\boldsymbol{\lambda} = E(\boldsymbol{\lambda}_i)$.

We consider T to be fixed, and allow $N \rightarrow \infty$, under which the unit root case where $|\gamma| = 1$ is also covered. It is assumed that the observations $\{y_{i0}, y_{it}, \mathbf{x}_{it}, \text{ for } t = 1, 2, \dots, T; i = 1, 2, \dots, N\}$ are available for estimation of γ and β , which are the parameters of interest. We propose an extension of the transformed MLE by treating the unknown factors as fixed parameters to be estimated for each t , but assume the factor loadings to be random and distributed independently of the errors, u_{it} . In addition, we contribute to the analysis of identification of short T dynamic models with a multiple factor error structure, and derive order conditions for identification of m and the parameters of interest, γ and β . Initially, we develop our proposed estimation method assuming that m is known, and consider the problem of consistent estimation of m in Section 7.1.

We make the following assumptions:

Assumption 1 *The idiosyncratic errors, u_{it} , for $i = 1, 2, \dots, N$ are distributed independently across i and over t with zero means and constant variance, σ^2 , such that $0 < \sigma^2 < K$, and $\sup_{i,t} E |u_{it}|^{4+\epsilon} < K$.*

Assumption 2 The time effects, δ_t , for $t = 1, 2, \dots, T$, and the $m \times 1$ vector of factors \mathbf{f}_t , vary across t , so that $d_t = \Delta\delta_t \neq 0$ and $\mathbf{g}_t = \Delta\mathbf{f}_t \neq \mathbf{0}$ at least for some $t = 2, \dots, T$, $m < T$, and $\sup_t \|\mathbf{g}_t\| < K$ and $\sup_t |d_t| < K$. For a fixed T , δ_t and \mathbf{g}_t , for $t = 1, 2, \dots, T$ are taken as fixed constants.

Assumption 3 The unobserved $m \times 1$ factor loadings, $\boldsymbol{\eta}_i$, for $i = 1, 2, \dots, N$ are distributed independently of u_{jt} , for all i, j and t , and are independently and identically distributed across i with zero means, and a finite covariance matrix, namely, $\boldsymbol{\eta}_i \sim IID(\mathbf{0}, \boldsymbol{\Omega}_\eta)$, where $\boldsymbol{\Omega}_\eta$ is an $m \times m$ symmetric positive definite matrix with $\|\boldsymbol{\Omega}_\eta\| < K$ and $\sup_i E \|\boldsymbol{\eta}_i\|^{4+\epsilon} < K$.

Assumption 4 The unit specific fixed effects, α_i , for $i = 1, 2, \dots, N$ are allowed to be correlated with \mathbf{x}_{jt} , $\boldsymbol{\eta}_j$, and u_{jt} , for all i, j and t , and could be deterministic and uniformly bounded, $\sup_i |\alpha_i| < K$, or stochastic and uniformly bounded, $\sup_i E |\alpha_i| < K$.

Assumption 5 The first-difference of the regressors, $\Delta\mathbf{x}_{it}$, for $i = 1, 2, \dots, N$ follows the multi-factor model

$$\Delta\mathbf{x}_{it} = \boldsymbol{\delta}_{x,t} + \mathbf{E}_{i,x}\mathbf{g}_{x,t} + \mathbf{v}_{it}, \text{ for all } t = \dots - 2, -1, 0, 1, 2, \dots, \quad (14)$$

where \mathbf{v}_{it} (the idiosyncratic component) follows the general linear stationary process $\mathbf{v}_{it} = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{i,t-j}$, $\boldsymbol{\delta}_{x,t}$ is a $k \times 1$ vector of time effects, $\mathbf{g}_{x,t} = (g_{x,1t}, g_{x,2t}, \dots, g_{x,m_x t})'$ is a $m_x \times 1$ vector of common factors, $\mathbf{E}_{i,x} = (\boldsymbol{\eta}_{i1,x}, \boldsymbol{\eta}_{i2,x}, \dots, \boldsymbol{\eta}_{i,m_x,x})$ is a $k \times m_x$ matrix of loadings, with $\boldsymbol{\eta}_{ij,x}$ a $k \times 1$ vector associated with the j^{th} factor $g_{x,jt}$, $\boldsymbol{\Psi}_j$ for $j = 0, 1, \dots$ are $k \times k$ matrices of fixed constants such that $\sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_j\| < K$, $\sup_t E \|\boldsymbol{\delta}_{x,t}\| < K$, and $\sup_{j,t} E |g_{x,jt}| < K$. Furthermore, conditional on the common factors, $\mathbf{E}_{i,x}$ is distributed independently over i , and of $\boldsymbol{\eta}_i$ and $u_{it'}$ for all i, t , and t' , $E(\boldsymbol{\eta}_{ij,x} | \mathcal{I}_{\delta,g}) = \mathbf{0}$, $E(\boldsymbol{\eta}_{ij,x} \boldsymbol{\eta}_{ij',x}' | \mathcal{I}_{\delta,g}) = \mathbf{V}_j$ if $j = j'$ and $E(\boldsymbol{\eta}_{ij,x} \boldsymbol{\eta}_{ij',x}' | \mathcal{I}_{\delta,g}) = \mathbf{0}$, where $\mathcal{I}_{\delta,g} = (\boldsymbol{\delta}_{x,T}, \boldsymbol{\delta}_{x,T-1}, \boldsymbol{\delta}_{x,T-2}, \dots; \mathbf{g}_{x,T}, \mathbf{g}_{x,T-1}, \mathbf{g}_{x,T-2}, \dots)$ for all $j \neq j' = 1, 2, \dots, m_x$, $\sup_{i,j} E \|\boldsymbol{\eta}_{ij,x}\|^{4+\epsilon} < K$, $\boldsymbol{\varepsilon}_{it} \sim IID(\mathbf{0}, \mathbf{I}_k)$ with $\sup_{i,t} E \|\boldsymbol{\varepsilon}_{it}\|^{4+\epsilon} < K$ for some small $\epsilon > 0$, and $\boldsymbol{\varepsilon}_{it}$ are distributed independently of $\boldsymbol{\delta}_{x,t'}$, $\mathbf{g}_{x,t'}$, $u_{jt'}$ for all i, j, t and t' .

Remark 1 The time effects and factors in the Δy_{it} and $\Delta\mathbf{x}_{it}$ equations, namely $\Delta\delta_t \neq 0$, \mathbf{g}_t , $\boldsymbol{\delta}_{x,t}$ and $\mathbf{g}_{x,t}$, are assumed to be draws from stochastic processes, but the analysis is carried out conditional on given values of $d_t = \Delta\delta_t \neq 0$, \mathbf{g}_t , $\boldsymbol{\delta}_{x,t}$ and $\mathbf{g}_{x,t}$, over the estimation sample $t = 1, 2, \dots, T$. As it is standard in short T panels, d_t and \mathbf{g}_t , for $t = 1, 2, \dots, T$ are treated as free parameters and estimated subject to suitable normalisation restrictions. But for the derivation of the initial values, Δy_{i0} , for $i = 1, 2, \dots, N$, we require the time effects and factors for $t < 0$ to follow stable processes so that the distribution of Δy_{i0} conditional on the observed values, $\{\Delta y_{it} \text{ and } \Delta\mathbf{x}_{it}, \text{ for } t = 1, 2, \dots, T\}$, can be obtained.

Assumptions 1, 2 and 4 are standard in the literature on short T dynamic panels. Assumption 1 can be relaxed to allow for time series heteroskedasticity so that $\text{Var}(u_{it}) = \sigma_t^2$, as shown in Section S.10 of the online supplement. Bai (2013) allows for time series heteroskedasticity while the GMM framework of Ahn et al. (2013) accommodates heteroskedasticity and/or serial correlation in a static model. In our context, serial correlation in the idiosyncratic errors can be entertained by allowing for a higher order autoregressive model. Assumption 2 is innocuous and requires time effects and the factors to be time-varying, otherwise they can not be distinguished from the fixed effects. Note that the case where $\delta_t = \delta$ and/or $\mathbf{f}_t = \mathbf{f}$ for all t is already covered by the presence of the fixed-effects, α_i . Assumption 3 imposes strong restrictions on the distribution of the factor loadings, $\boldsymbol{\eta}_i$, and is required for identification of the factors and the parameters. Ahn et al. (2013) entertain the same assumption for their factor loadings, which they treat as random alongside the factors which are taken to be fixed parameters. This assumption could be somewhat relaxed as noted in what follows. In contrast, Assumption 4 does not impose any restrictions on the fixed effects, α_i , and allows them to be correlated with the regressors as well as with the composite errors, ζ_{it} . In this way, as noted above, our model specification can be viewed as a direct generalisation of the standard time and fixed effects models considered routinely in the empirical literature.

As noted above, our specification also differs from the one considered by Bai (2013) and Ahn et al. (2013) who do not model the fixed effects explicitly but assume that the fixed effects can be captured *implicitly* through the interactive effects, for example, by setting $f_{1t} = 1$. In the context of our set up, following this line of reasoning leads to a random coefficient specification, which is likely to be restrictive in practice. Bai (2013) does consider the possible dependence of η_{i1} on the regressors, using the methods of Mundlak (1978) and Chamberlain (1982), whereby it is assumed that the random components of α_i , namely η_{i1} , is given by

$$\eta_{i1} = \sum_{t=1}^T \mathbf{b}'_t [\mathbf{x}_{it} - E(\mathbf{x}_{it})] + \varepsilon_{\eta_{i1}}, \text{ for } i = 1, 2, \dots, N, \quad (15)$$

where $(\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_T)'$ is a $Tk \times 1$ vector of coefficients to be estimated and $\varepsilon_{\eta_{i1}}$ are mean zero cross-sectionally independent random variables distributed independently of $u_{jt'}$ for all i, j , and t' . This specification ensures that $E(\eta_{i1}) = 0$, as required, but depends on $E(\mathbf{x}_{it})$ which is unobserved. To make this scheme operational it is typically assumed that $E(\mathbf{x}_{it})$ is fixed so that it can be absorbed in an intercept. But in the more general context where \mathbf{x}_{it} could be non-stationary, the use of the Mundlak scheme as applied in (15) directly to \mathbf{x}_{it} could be problematic. The quasi-differenced GMM approach also allows for correlation between the regressors and the random factor loadings. In our context, possible correlation between $\boldsymbol{\eta}_i$ and the regressors $\Delta \mathbf{x}_i$ can be dealt with using the Mundlak device as set out above for the case of fixed effects, but applied to $\Delta \mathbf{x}_i$.⁷

Assumption 5 provides a general linear multi-factor time series specification for $\Delta \mathbf{x}_{it}$. This is done for convenience. We could have equally started with a model for \mathbf{x}_{it} . This assumption postulates that $\Delta \mathbf{x}_{it}$ is composed of three components, a $k \times 1$ vector of time effects, $\boldsymbol{\delta}_{x,t}$, a multifactor component with m_x common factors, $\mathbf{g}_{x,t}$, and a stationary component \mathbf{v}_{it} which is assumed to be cross-sectionally independent. The assumption that the factor loadings, $\boldsymbol{\eta}_{ij,x}$, $j = 1, 2, \dots, m_x$ have zero mean and are uncorrelated over j is made for convenience, and can be relaxed without any consequences for the subsequent analysis.

Remark 2 *Our assumptions require u_{it} and \mathbf{v}_{it} to be uncorrelated which rules out classical simultaneity and measurement errors. The assumption that u_{it} and \mathbf{v}_{it} and their factor loadings, $\boldsymbol{\eta}_i$ and $\mathbf{E}_{i,x}$, are independently distributed can, however, be relaxed by considering a vector autoregressive version of (13), where $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ is modelled jointly as in Holtz-Eakin et al. (1988) and Binder et al. (2005).*

Finally, while the composite error term, $\zeta_{it} = \boldsymbol{\eta}'_i \mathbf{f}_t + u_{it}$, in (13) is cross-sectionally heteroskedastic through the presence of the interactive effects, allowing explicitly for the same in the idiosyncratic error, u_{it} , can be pursued along the lines of Hayakawa and Pesaran (2015). These authors extend the cross-sectionally independent homoskedastic idiosyncratic errors of Hsiao et al. (2002) to the heteroskedastic case. These extensions are not considered here as they are beyond the scope of the present focus of the paper.

We follow the standard practice and eliminate the fixed effects by application of the first-difference operator to both sides of (13):

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \boldsymbol{\beta}' \Delta \mathbf{x}_{it} + d_t + \mathbf{g}'_t \boldsymbol{\eta}_i + \Delta u_{it}, \text{ for } t = 2, 3, \dots, T; \ i = 1, 2, \dots, N, \quad (16)$$

where $d_t = \Delta \delta_t \neq 0$ and $\mathbf{g}_t = \Delta \mathbf{f}_t \neq \mathbf{0}$ for some $t \geq 2$, and

$$\xi_{it} = \mathbf{g}'_t \boldsymbol{\eta}_i + \Delta u_{it}, \text{ for } t = 2, 3, \dots, T. \quad (17)$$

For $t = 1$ (16) is not defined as Δy_{i1} depends on the unobserved Δy_{i0} , which in turn depends on the past history of the regressors, $\Delta \mathbf{x}_{it}$ for $t \leq 0$ which are not observed. To derive the joint probability

⁷Though we do not pursue this idea in the present paper, we do investigate the effect of such correlations on the proposed TQML estimator in our Monte Carlo experiments, where we also consider the effect of weakly correlated factor loadings. Our findings suggest that neither of these have a significant impact on the results.

distribution of $(\Delta y_{i1}, \Delta y_{i2}, \dots, \Delta y_{iT})$ the process generating Δy_{i1} in terms of the available observations is also required. For this purpose we need to specify the data generating process of $\Delta \mathbf{x}_{it}$, which we do under Assumption 5, as well as the initialisation of $\Delta y_{i,-S+1}$, for some $S > 0$, which we formalise in the following assumption.

Assumption 6 Suppose that for each i , $\{\Delta y_{it}\}$ is started from time $t = -S + 1$, for some $S > 0$, with the initial first differences, $\Delta y_{i,-S+1}$, as random draws from a distribution such that

$$E(\Delta y_{i,-S+1} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g}) = a_S + \boldsymbol{\pi}'_S \Delta \mathbf{x}_i, \quad (18)$$

where $\Delta \mathbf{x}_i = (\Delta \mathbf{x}'_{i1}, \Delta \mathbf{x}'_{i2}, \dots, \Delta \mathbf{x}'_{iT})'$ is the $kT \times 1$ vector of observations on the regressors, $\mathcal{I}_{\delta,g} = (\boldsymbol{\delta}_{x,T}, \boldsymbol{\delta}_{x,T-1}, \boldsymbol{\delta}_{x,T-2}, \dots; \mathbf{g}_{x,T}, \mathbf{g}_{x,T-1}, \mathbf{g}_{x,T-2}, \dots)$, a_S is a fixed coefficient that allows for non-zero means, and $\boldsymbol{\pi}_S$ is the $kT \times 1$ vector of coefficients, such that $\sup_S |a_S| < K$, and $\sup_S \|\boldsymbol{\pi}_S\| < K$. Furthermore, let $\varpi_i = \Delta y_{i,-S+1} - E(\Delta y_{i,-S+1} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g})$, and suppose that $\varpi_i \sim IID(0, \sigma_\varpi^2)$, $0 < \sigma_\varpi^2 < K$, and $\sup_i E|\varpi_i|^{4+\epsilon} < K$.

Equation (18) can be viewed as a linear projection of $\Delta y_{i,-S+1}$ on the observables, $\Delta \mathbf{x}_i$, and allows the initial values, $y_{i,-S}$ and $y_{i,-S+1}$ to depend on the fixed effects, α_i , as well as other parameters. Also it is redundant if $|\gamma| < 1$ and S is sufficiently large, and does not apply if there are no regressors in (13). The main restriction here is the assumed linearity of (18). One can think of Assumption 6 as "implicitly" using Mundlak-type projections for $\Delta y_{i,-S+1}$. Using first differences allows us to make less restrictive assumptions about α_i to the extent that such assumption implicitly involves α_i .

It is possible to dispense with Assumptions 5 and 6 by postulating a model for the initial first-differences, Δy_{i1} , similar to what we assumed for y_{i0} in our discussion of the GMM approach (see equation (2)). Under the GMM approach, the moment conditions take the initial values y_{i0} (or Δy_{i1}), as given. But as we have seen a model for the initial values is required if we are to check the validity of the rank condition typically assumed when the GMM approach is used in the literature.

3.1 Modelling initial values

Given the above assumptions, we can now derive an expression for Δy_{i1} that depends on the observables and the unknown parameters only. Using (16), and starting from some arbitrary point in the past at $t = -S + 1$ with $\Delta y_{i,-S+1}$ as given we obtain the following expression

$$\Delta y_{i1} = \gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} + \tilde{d}_1 + \tilde{\mathbf{g}}_1' \boldsymbol{\eta}_i + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j}, \quad (19)$$

where $\tilde{d}_1 = \sum_{j=0}^{S-1} \gamma^j d_{1-j}$, and $\tilde{\mathbf{g}}_1 = \sum_{j=0}^{S-1} \gamma^j \mathbf{g}_{1-j}$. In the case of models without regressors Δy_{i1} is fully determined under Assumptions 1 to 3. But when the model includes regressors and $S > 2$, the distribution of Δy_{i1} also depends on the $k(S-2) \times 1$ vector of past observations $\Delta \mathbf{x}_i^0 = (\Delta \mathbf{x}'_{i0}, \Delta \mathbf{x}'_{i,-1}, \dots, \Delta \mathbf{x}'_{i,-S+3})'$, not available to the researcher. To deal with this missing observation problem, Hsiao et al. (2002) propose back-casting these missing data points from $\Delta \mathbf{x}_i$ which is observed. Following a similar procedure, we first note that under Assumption 6

$$\Delta \mathbf{x}_i^0 = \boldsymbol{\delta}_x^0 + \sum_{j=1}^{m_x} (\mathbf{g}_{x,j}^0 \otimes \boldsymbol{\eta}_{ij,x}) + \mathbf{v}_i^0, \text{ and } \Delta \mathbf{x}_i = \boldsymbol{\delta}_x + \sum_{j=1}^{m_x} (\mathbf{g}_{x,j} \otimes \boldsymbol{\eta}_{ij,x}) + \mathbf{v}_i, \quad (20)$$

where $\boldsymbol{\delta}_x^0 = (\boldsymbol{\delta}'_{x,0}, \boldsymbol{\delta}'_{x,-1}, \dots, \boldsymbol{\delta}'_{x,-S+3})'$, $\mathbf{g}_{x,j}^0 = (g_{x,j,0}, g_{x,j,-1}, \dots, g_{x,j,-S+3})'$, and $\mathbf{v}_i^0 = (\mathbf{v}'_{i0}, \mathbf{v}'_{i,-1}, \dots, \mathbf{v}'_{i,-S+3})'$, and similarly $\boldsymbol{\delta}_x = (\boldsymbol{\delta}'_{x,1}, \boldsymbol{\delta}'_{x,2}, \dots, \boldsymbol{\delta}'_{x,T})'$, $\mathbf{g}_{x,j} = (g_{x,j,1}, g_{x,j,2}, \dots, g_{x,j,T})'$, and $\mathbf{v}_i = (\mathbf{v}'_{i1}, \mathbf{v}'_{i2}, \dots, \mathbf{v}'_{iT})'$. Also

$E(\Delta \mathbf{x}_i^0) = \boldsymbol{\delta}_x^0$, $E(\Delta \mathbf{x}_i) = \boldsymbol{\delta}_x$, and using linear projections, we have⁸

$$E(\Delta \mathbf{x}_i^0 | \Delta \mathbf{x}_i) = \boldsymbol{\delta}_x^0 + \boldsymbol{\Omega}_{01} \boldsymbol{\Omega}_{11}^{-1} (\Delta \mathbf{x}_i - \boldsymbol{\delta}_x) \quad (21)$$

where

$$\boldsymbol{\Omega}_{11} = \sum_{j=1}^{m_x} (\mathbf{g}_{x,j} \mathbf{g}_{x,j}' \otimes \mathbf{V}_j) + E(\mathbf{v}_i \mathbf{v}_i'), \quad \boldsymbol{\Omega}_{01} = \sum_{j=1}^{m_x} (\mathbf{g}_{x,j}^0 \mathbf{g}_{x,j}' \otimes \mathbf{V}_j) + E(\mathbf{v}_i^0 \mathbf{v}_i').$$

Since \mathbf{v}_{it} is a stationary process with zero means and variance-covariances that do not depend on i , it then readily follows that $E(\mathbf{v}_i \mathbf{v}_i') = \boldsymbol{\Omega}_{v,11}$ and $E(\mathbf{v}_i^0 \mathbf{v}_i') = \boldsymbol{\Omega}_{v,01}$ that also do not depend on i . Now using (21) along with (18) we have

$$E\left(\gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} | \Delta \mathbf{x}_i\right) = a + \boldsymbol{\pi}' \Delta \mathbf{x}_i, \quad (22)$$

where a and $\boldsymbol{\pi}$ are fixed parameters that are complicated functions of γ and $\boldsymbol{\beta}$, the parameters of the \mathbf{x}_{it} process as well as the parameters of the initial values. Now let

$$\begin{aligned} \chi_i &= \left(\gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} \right) - E\left(\gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g} \right) \\ &= \gamma^S [\Delta y_{i,-S+1} - E(\Delta y_{i,-S+1} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g})] + \boldsymbol{\beta}' \sum_{j=0}^{S-1} \gamma^j [\Delta \mathbf{x}_{i,1-j} - E(\Delta \mathbf{x}_{i,1-j} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g})], \end{aligned} \quad (23)$$

and note that under Assumption 6 $\Delta y_{i,-S+1} - E(\Delta y_{i,-S+1} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g}) = \varpi_i \sim IID(0, \sigma_\varpi^2)$, and $\sup_i E|\varpi_i|^{4+\epsilon} < K$. Also, under Assumption 5

$$\begin{aligned} \Delta \mathbf{x}_{i,1-j} - E(\Delta \mathbf{x}_{i,1-j} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g}) &= [\mathbf{E}_{i,x} - E(\mathbf{E}_{i,x} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g})] \mathbf{g}_{x,1-j} + \mathbf{v}_{i,1-j} - E(\mathbf{v}_{i,1-j} | \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g}) \\ &= \mathbf{E}_{i,x} \mathbf{g}_{x,1-j} + \mathbf{v}_{i,1-j}, \end{aligned}$$

and overall

$$\chi_i = \gamma^S \varpi_i + \boldsymbol{\beta}' \sum_{j=0}^{S-1} \gamma^j (\mathbf{E}_{i,x} \mathbf{g}_{x,1-j} + \mathbf{v}_{i,1-j}).$$

Therefore, $\{\chi_i\}$ is a sequence of cross-sectionally independent random variables with zero means. Also in view of Assumptions 5 and 6 and by application of the Minkowski inequality to both sides of χ_i we have $\sup_i |\chi_i|^{4+\epsilon} < K$.⁹ Hence, using (22) and (23) in (19) we obtain

$$\Delta y_{i1} = d_1 + \boldsymbol{\pi}' \Delta \mathbf{x}_i + \xi_{i1}, \quad (24)$$

where $d_1 = a + \tilde{d}_1$,

$$\xi_{i1} = \tilde{\mathbf{g}}_1' \boldsymbol{\eta}_i + v_{i1}, \quad (25)$$

and

$$v_{i1} = \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j} + \chi_i. \quad (26)$$

In the analysis that follows we treat d_1 and $\boldsymbol{\pi}$ as unknown parameters to be estimated along with the parameters of interest γ and $\boldsymbol{\beta}$. We also note that $v_{i1} \sim IID(0, \omega \sigma^2)$, and v_{i1} is distributed independently

⁸This result provides an optimal linear approximation when the regressors are not normally distributed.

⁹Note that under Assumption 5 $\sup_{i,t} E \|\Delta \mathbf{x}_{it}\|^{4+\epsilon} < K$. See Lemma 1.

of $\Delta \mathbf{x}_i$ and $\boldsymbol{\eta}_i$. Further, by application of the Minkowski inequality to (26) we have $\sup_i E |v_{i1}|^{4+\epsilon} < K$, and under Assumptions 5 and 6, $\sup_i \text{Var}(\chi_i) < K$; as a result $0 < \omega_{\min} < \omega < \omega_{\max} < \infty$, where ω_{\min} and ω_{\max} are fixed constants, with ω taken as a free parameter to be estimated together with other model parameters.

Finally, using (26) we have

$$\text{Cov}(v_{i1}, \Delta u_{it}) = \begin{cases} -\sigma^2 & \text{for } t = 2 \\ 0 & \text{for } t = 3, 4, \dots, T \end{cases} \quad (27)$$

Remark 3 As noted earlier, in the case where $|\gamma| < 1$ and $S \rightarrow \infty$ we have $\Delta y_{i1} = d_1 + \boldsymbol{\pi}' \Delta \mathbf{x}_i + \xi_{i1}$, where ξ_{i1} is defined by (25), with v_{i1} given by $v_{i1} = \sum_{j=0}^{\infty} \gamma^j \Delta u_{i,1-j} + \chi_i$, and

$$\chi_i = \sum_{j=0}^{\infty} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} - E \left(\sum_{j=0}^{\infty} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} \mid \Delta \mathbf{x}_i, \mathcal{I}_{\delta,g} \right).$$

where $\mathcal{I}_{\delta,g} = (\boldsymbol{\delta}_{x,T}, \boldsymbol{\delta}_{x,T-1}, \boldsymbol{\delta}_{x,T-2}, \dots; \mathbf{g}_{x,T}, \mathbf{g}_{x,T-1}, \mathbf{g}_{x,T-2}, \dots)$. Since $\Delta \mathbf{x}_{it}$, $\boldsymbol{\eta}_i$, and $u_{it'}$ are independently distributed for all i , t and t' , it then follows that v_{i1} is distributed independently of $\boldsymbol{\eta}_i$ and $\Delta \mathbf{x}_i$, with $E(v_{i1}) = 0$, and

$$\text{Var}(v_{i1}) = \text{Var} \left(\sum_{j=0}^{\infty} \gamma^j \Delta u_{i,1-j} \right) + \text{Var}(\chi_i) = \frac{2\sigma^2}{1+\gamma} + \text{Var}(\chi_i) > 0.$$

In the case of pure AR(1) panels, we have the further parametric restriction, $\text{Var}(v_{i1}) = \frac{2\sigma^2}{1+\gamma}$, which, if imposed, can increase estimation efficiency.

3.2 The full model specification

We can now combine the processes for Δy_{i1} and Δy_{it} conditional on $\Delta y_{i,t-1}$, for $t = 2, 3, \dots, T$ to write down the quasi-likelihood function of the first-differenced model. Writing (16) and (24) in matrix notation we note that

$$\Delta \mathbf{y}_i = \Delta \mathbf{W}_i \boldsymbol{\varphi} + \boldsymbol{\xi}_i, \quad \boldsymbol{\xi}_i = \mathbf{G} \boldsymbol{\eta}_i + \mathbf{r}_i, \quad (28)$$

where $\Delta \mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, \dots, \Delta y_{iT})'$, $\Delta \mathbf{W}_i$ is the $T \times (T + Tk + k + 1)$ matrix given by

$$\Delta \mathbf{W}_i = \begin{pmatrix} 1 & 0 & \dots & 0 & \Delta \mathbf{x}'_i & 0 & 0 \\ 0 & 1 & \dots & 0 & \mathbf{0} & \Delta \mathbf{x}'_{i2} & \Delta y_{i1} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \mathbf{0} & \Delta \mathbf{x}'_{iT} & \Delta y_{i,T-1} \end{pmatrix}, \quad (29)$$

$\boldsymbol{\varphi} = (\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}', \gamma)'$ with $\mathbf{d} = (d_1, d_2, \dots, d_T)'$, $\mathbf{G}' = (\tilde{\mathbf{g}}_1, \mathbf{g}_2, \dots, \mathbf{g}_T)$, $\mathbf{r}_i = (v_{i1}, \Delta u_{i2}, \dots, \Delta u_{iT})'$, and $\boldsymbol{\xi}_i = (\tilde{\xi}_{i1}, \xi_{i2}, \dots, \xi_{iT})'$, and recall that $\tilde{\xi}_{i1} = \tilde{\mathbf{g}}_1' \boldsymbol{\eta}_i + v_{i1}$, and $\xi_{it} = \mathbf{g}_t' \boldsymbol{\eta}_i + \Delta u_{it}$, for $t = 2, 3, \dots, T$.

In using the first-differenced specification (28), it is first worth noting that despite the presence of common factors in Δy_{it} and $\Delta \mathbf{x}_{it}$, the composite errors, $\boldsymbol{\xi}_i$, and the regressors $\Delta \mathbf{x}_i = (\Delta \mathbf{x}'_{i1}, \Delta \mathbf{x}'_{i2}, \dots, \Delta \mathbf{x}'_{iT})$ are independently distributed over i , conditional on $\boldsymbol{\delta}_{x,t}$ and $\mathbf{g}_{x,t}$. This follows since under the above assumptions the cross sectional-variation of $\Delta \mathbf{x}_i$, given by (20), is governed by \mathbf{v}_i and $\{\boldsymbol{\eta}_{ij,x}, \text{ for } j = 1, 2, \dots, m_x\}$ that are assumed to be distributed independently of $\boldsymbol{\eta}_i$ and Δu_{it} for all i and t (see Assumption 5)). For future reference it is also convenient to partition $\Delta \mathbf{W}_i$, as $\Delta \mathbf{W}_i = (\Delta \mathbf{Z}_i, \Delta \mathbf{y}_{i,-1})$ and write (28) as

$$\Delta \mathbf{y}_i = \Delta \mathbf{Z}_i \boldsymbol{\delta} + \Delta \mathbf{y}_{i,-1} \gamma + \boldsymbol{\xi}_i, \quad (30)$$

where $\boldsymbol{\delta} = (\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}')'$.

4 Transformed quasi maximum likelihood estimation

Consider the panel data model given by (28) and note that under Assumption 1, and using (25) and (27), we have (recall also that $v_{i1} \sim IID(0, \omega\sigma^2)$)

$$E(\mathbf{r}_i \mathbf{r}_i') = \sigma^2 \mathbf{\Omega}, \quad (31)$$

where

$$E(\mathbf{r}_i \mathbf{r}_i') = \sigma^2 \begin{pmatrix} \omega & -1 & & 0 \\ -1 & 2 & \ddots & 0 \\ & & \ddots & \\ 0 & & & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} = \sigma^2 \mathbf{\Omega}, \quad (32)$$

and $\mathbf{\Omega} = \mathbf{\Omega}(\omega)$. Since $|\mathbf{\Omega}| = 1 + T(\omega - 1)$, ω needs to satisfy $\omega > 1 - \frac{1}{T}$ to ensure that $\mathbf{\Omega}$ is positive definite. Also, since $\boldsymbol{\eta}_i$ and \mathbf{r}_i are independently distributed, conditional on $\boldsymbol{\delta}_{x,t}$ and $\mathbf{g}_{x,t}$ we have

$$Var(\boldsymbol{\xi}_i) = \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) = \sigma^2 \mathbf{\Omega} + \mathbf{G} \mathbf{\Omega}_\eta \mathbf{G}' = \sigma^2 (\mathbf{\Omega} + \mathbf{Q} \mathbf{Q}') \quad (33)$$

where $\mathbf{Q} = (1/\sigma) \mathbf{G} \mathbf{\Omega}_\eta^{1/2}$, $rank(\mathbf{Q}) = m$, and $\boldsymbol{\psi} = (\omega, \sigma^2, vec(\mathbf{Q}))'$.

Our parameters of primary interest are given by $\boldsymbol{\varphi} = (\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}', \gamma)' = (\boldsymbol{\delta}', \gamma)'$, with the interactive effects treated as nuisance parameters. In consequence, we shall also focus on conditions under which $\boldsymbol{\varphi}_0$ the true value of $\boldsymbol{\varphi}$, can be identified, globally or locally. We are only interested in controlling for the latent interactive effects, and not in their interpretation. This is reflected in the above specification of \mathbf{Q} , the parameter associated with such effects. Given that $\mathbf{Q} \mathbf{Q}'$ is of reduced rank $m < T$, it is not possible to identify \mathbf{Q} without additional restrictions. This is because for any orthonormal $m \times m$ matrix \mathbf{C} , $\mathbf{Q} \mathbf{Q}' = \mathbf{Q}^* \mathbf{Q}^{*'} where $\mathbf{Q}^* = \mathbf{Q} \mathbf{C}$. To avoid such non-trivial identification $m(m-1)/2$ restrictions need to be imposed on \mathbf{Q} .¹⁰ The number of non-redundant parameters in \mathbf{Q} is then $mT - m(m-1)/2$ (see also Hayashi et al. (2007, p.507)).$

The quasi-log-likelihood of the transformed model (28) is given by

$$\ell_N(\boldsymbol{\theta}) = \ell_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi}) = -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2} \sum_{i=1}^N \boldsymbol{\xi}_i'(\boldsymbol{\delta}, \gamma) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\delta}, \gamma) \quad (34)$$

$$\boldsymbol{\xi}_i(\boldsymbol{\delta}, \gamma) = \Delta \mathbf{y}_i - \Delta \mathbf{Z}_i \boldsymbol{\delta} - \Delta \mathbf{y}_{i,-1} \gamma \quad (35)$$

and it is assumed that $\boldsymbol{\varphi} = (\boldsymbol{\delta}', \gamma)'$ does not depend on $\boldsymbol{\psi}$. For fixed m and T , the above log-likelihood function depends on a fixed number of unknown parameters, which are collected in the $[T(m+k+1) - m(m-1)/2 + k + 3] \times 1$ vector $\boldsymbol{\theta} = (\boldsymbol{\delta}', \gamma, \boldsymbol{\psi}')'$.¹¹

5 Identification

We begin our identification analysis by focusing on the identification of \mathbf{d} and γ in the panel AR(1) model before turning to the general likelihood framework allowing also for exogenous regressors. Prior to this, for identification of the number of interactive effects we derive the order condition on m and T , and show that $m_{\max} = T - 2$ is an important input in the determination of m_0 , the true value of m . We also show that the same order condition applies irrespective of whether the model contains exogenous regressors.

¹⁰Note that $m(m+1)/2$ restrictions are imposed by expressing $\mathbf{G} \mathbf{\Omega}_\eta \mathbf{G}'$ as $\mathbf{Q} \mathbf{Q}'$. For the m^2 restrictions typically imposed on $Var(\mathbf{G} \boldsymbol{\eta}_i)$ in traditional factor analysis an additional $m(m-1)/2$ restrictions need to be placed on \mathbf{Q} .

¹¹In the Monte Carlo and empirical applications that follow the TQML estimates are obtained by maximising a concentrated version of the likelihood function in (34). This is derived using an eigenvalue approach which greatly simplifies the computations. For details see Section S.3 of the online supplement.

5.1 Order condition

We first consider the order condition on m and T associated with the AR(1) model. Using (16) and (24), we have

$$\begin{aligned}\Delta y_{i1} &= d_1 + \tilde{\mathbf{g}}_1' \boldsymbol{\eta}_i + v_{i1}, \\ \Delta y_{it} - \gamma \Delta y_{i,t-1} &= d_t + \mathbf{g}_t' \boldsymbol{\eta}_i + \Delta u_{it}, \text{ for } t = 2, 3, \dots, T,\end{aligned}\quad (36)$$

which can be written as $\mathbf{B}(\gamma) \Delta \mathbf{y}_i = \mathbf{d} + \mathbf{G} \boldsymbol{\eta}_i + \mathbf{r}_i = \mathbf{d} + \boldsymbol{\xi}_i$, for $i = 1, 2, \dots, N$, where $\mathbf{d} = (d_1, \dots, d_T)'$, $\Delta \mathbf{y}_i$ and $\boldsymbol{\xi}_i$ are as defined above, and

$$\mathbf{B}(\gamma) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -\gamma & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -\gamma & 1 \end{pmatrix}. \quad (37)$$

Note also that, $|\mathbf{B}(\gamma)|=1$, and

$$\mathbf{B}^{-1}(\gamma) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \gamma & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \gamma^{T-1} & \dots & \gamma & 1 \end{pmatrix}, \quad (38)$$

and hence $\Delta \mathbf{y}_i = \mathbf{a} + \mathbf{B}^{-1}(\gamma) \boldsymbol{\xi}_i$, where

$$\mathbf{a} = \mathbf{B}^{-1}(\gamma) \mathbf{d} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \gamma & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \gamma^{T-1} & \dots & \gamma & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_T \end{pmatrix} = \begin{pmatrix} d_1 \\ \gamma d_1 + d_2 \\ \vdots \\ \gamma^{T-1} d_1 + \gamma^{T-2} d_2 + \dots + \gamma d_{T-1} + d_T \end{pmatrix}. \quad (39)$$

The parameters associated with this model are $\boldsymbol{\theta} = (\mathbf{d}', \gamma, \boldsymbol{\psi}')' = (\mathbf{d}', \boldsymbol{\varrho})'$ with $\boldsymbol{\varrho} = (\gamma, \boldsymbol{\psi}')'$, and recall $\boldsymbol{\psi} = (\omega, \sigma^2, \text{vec}(\mathbf{Q})')'$. In deriving the order condition on m and T for the AR(1) model, and the ARX(1) that follows, it suffices to focus on the identification of $\boldsymbol{\varrho}$ as none of the remaining parameters of either model depend on m .

For the AR(1) model since \mathbf{d} is a $T \times 1$ unrestricted parameter vector, then \mathbf{a} is also unrestricted, namely knowing \mathbf{a} will not help with identification γ , or any of the remaining parameters in $\boldsymbol{\psi}$. Hence, the identification of $\boldsymbol{\varrho} = (\gamma, \boldsymbol{\psi}')'$ can only come from the $T(T+1)/2$ distinct elements of $\text{Var}(\Delta \mathbf{y}_i) = \boldsymbol{\Sigma}_{\Delta y}$ which are given by

$$\begin{aligned}\boldsymbol{\Sigma}_{\Delta y} &= \mathbf{B}(\gamma)^{-1} \text{Var}(\boldsymbol{\xi}_i) \mathbf{B}'(\gamma)^{-1} \\ &= \sigma^2 \mathbf{B}(\gamma)^{-1} (\boldsymbol{\Omega} + \mathbf{Q} \mathbf{Q}') \mathbf{B}'(\gamma)^{-1} = \boldsymbol{\Sigma}(\boldsymbol{\varrho}),\end{aligned}\quad (40)$$

where $\boldsymbol{\Sigma}_{\Delta y}$ can be consistently estimated. Since \mathbf{Q} enters $\boldsymbol{\Sigma}(\boldsymbol{\varrho})$ as $\mathbf{A} = \mathbf{Q} \mathbf{Q}'$, we need to consider the unknown elements of the symmetric matrix \mathbf{A} under different rank conditions. First it is clear that if \mathbf{A} has full rank, namely if $\text{rank}(\mathbf{A}) = T$, then $\boldsymbol{\varrho}$ cannot be identified. Therefore, to identify $\boldsymbol{\varrho}$, we must have $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{Q}) = m < T$. Recall also from Section 4 that the number of non-redundant elements of \mathbf{Q} is given by $mT - m(m-1)/2$. The order condition necessary for identification of $\boldsymbol{\varrho}$ is then given by

$$T(T+1)/2 \geq 3 + Tm - m(m-1)/2. \quad (41)$$

This order condition is satisfied if $T \geq 3$, for $m = 0, 1, 2, \dots, m_{\max}$ where m_{\max} is the largest value of m that satisfies (41), that is $m_{\max} = T - 2$.

Consider now the more general case where the panel AR(1) model also contains exogenous regressors. For this case note that the system of equations (28) can be written equivalently as

$$\Delta \mathbf{y}_i = \mathbf{a} + \widetilde{\Delta \mathbf{X}}_i(\gamma) \phi + \mathbf{B}^{-1}(\gamma) \xi_i, \quad (42)$$

where \mathbf{a} , $\mathbf{B}^{-1}(\gamma)$ and ξ_i are as defined above, $\phi = (\pi', \beta')'$, $\widetilde{\Delta \mathbf{X}}_i(\gamma) = \mathbf{B}^{-1}(\gamma) \Delta \mathbf{X}_i$, and $\Delta \mathbf{X}_i$ is the $T \times (Tk + k)$ matrix of observations on the exogenous regressors defined by

$$\Delta \mathbf{X}_i = \begin{pmatrix} \Delta \mathbf{x}'_i & \mathbf{0} \\ \mathbf{0} & \Delta \mathbf{x}'_{i2} \\ \vdots & \vdots \\ \mathbf{0} & \Delta \mathbf{x}'_{iT} \end{pmatrix}. \quad (43)$$

The parameters associated with the ARX(1) model in (42) are $\theta = (\mathbf{d}', \phi', \gamma, \psi')' = (\mathbf{d}', \phi', \varrho')'$, with ψ as defined earlier. Here, as above, \mathbf{d} and ϕ are unrestricted parameters in the sense that knowing them will not help identification of ϱ since $\Sigma(\varrho)$ does not depend on \mathbf{d} and ϕ . But it is already established that identification of γ is based on the covariance of $\mathbf{B}^{-1}(\gamma) \xi_i$, which is given by $\Sigma(\varrho) = \sigma^2 \mathbf{B}(\gamma)^{-1} (\Omega + \mathbf{Q}\mathbf{Q}') \mathbf{B}'(\gamma)^{-1}$ if the order condition (41) is met. Hence, it follows that the same order condition given by (41) continues to hold in the case of the ARX(1) model.

5.2 Rank condition

Subject to the order condition, (41), being satisfied we now consider if the mapping

$$\Sigma_{\Delta y} = \sigma^2 \mathbf{B}(\gamma)^{-1} (\Omega + \mathbf{Q}\mathbf{Q}') \mathbf{B}'(\gamma)^{-1},$$

provides a unique solution for γ , in terms of $\Sigma_{\Delta y}$. The moment conditions implicit in this mapping can also be obtained explicitly using (36). To simplify the exposition we use \mathbf{g}_1 for $\tilde{\mathbf{g}}_1$, abstract from exogenous regressors and set $T = 3$ which implies $m_{max} = T - 2 = 1$, and assume that the observations y_{i0} , y_{i1} , y_{i2} , and y_{i3} are available for the units $i = 1, 2, \dots, N$. We have the following relations

$$\begin{aligned} \Delta y_{i1} &= d_1 + g_1 \eta_i + v_{i1}, \\ \Delta y_{i2} - \gamma \Delta y_{i1} &= d_2 + g_2 \eta_i + \Delta u_{i2}, \\ \Delta y_{i3} - \gamma \Delta y_{i2} &= d_3 + g_3 \eta_i + \Delta u_{i3}. \end{aligned}$$

It is clear that d_1 is identified since $d_1 = E(\Delta y_{i1})$, and can be consistently estimated by $\hat{d}_{1N} = N^{-1} \sum_{i=1}^N \Delta y_{i1}$. To identify d_2 and d_3 we need to know γ . But since $d_t = E(\Delta y_{it} - \gamma \Delta y_{i,t-1})$, we can eliminate d_t from the above equations to obtain

$$\Delta y_{i1} - E(\Delta y_{i1}) = g_1 \eta_i + v_{i1}, \quad (44)$$

$$[\Delta y_{i2} - E(\Delta y_{i2})] - \gamma [\Delta y_{i1} - E(\Delta y_{i1})] = g_2 \eta_i + \Delta u_{i2}, \quad (45)$$

$$[\Delta y_{i3} - E(\Delta y_{i3})] - \gamma [\Delta y_{i2} - E(\Delta y_{i2})] = g_3 \eta_i + \Delta u_{i3}. \quad (46)$$

Recall that $v_{i1} \sim IID(0, \omega\sigma^2)$, $\Delta u_{it} \sim IID(0, 2\sigma^2)$ for $t = 2, 3$, $E(\Delta u_{i2} v_{i1}) = E(\Delta u_{i2} \Delta u_{i3}) = -\sigma^2$, and $E(\Delta u_{i3} v_{i1}) = 0$. Furthermore, by assumption η_i is distributed independently of $(v_{i1}, \Delta u_{i2}, \Delta u_{i3})$. Here we assume the factor, g_t , is strong and set $\sigma_\eta^2 = 1$. Using (44)-(46) we obtain the moment conditions

$$m_{11} = \sigma_{11} - (g_1^2 + \omega\sigma^2) = 0, \quad (47)$$

$$m_{22} = \sigma_{22} - 2\gamma\sigma_{12} + \gamma^2\sigma_{11} - (g_2^2 + 2\sigma^2) = 0, \quad (48)$$

$$m_{33} = \sigma_{33} - 2\gamma\sigma_{23} + \gamma^2\sigma_{22} - (g_3^2 + 2\sigma^2) = 0, \quad (49)$$

$$m_{12} = \sigma_{12} - \gamma\sigma_{11} - (g_1 g_2 - \sigma^2) = 0, \quad (50)$$

$$m_{13} = \sigma_{13} - \gamma\sigma_{12} - g_1 g_3 = 0, \quad (51)$$

$$m_{23} = \sigma_{23} - (\sigma_{13} + \sigma_{22})\gamma + \gamma^2\sigma_{12} - (g_2 g_3 - \sigma^2) = 0, \quad (52)$$

where

$$\sigma_{tt'} = \text{Cov}(\Delta y_{it}, \Delta y_{it'}) = E\{[\Delta y_{it} - E(\Delta y_{it})][\Delta y_{it'} - E(\Delta y_{it'})]\}, \quad \forall t, t' = 1, 2, 3.$$

As γ only enters equations (48)-(52), the moment condition in (47), $m_{11}(\boldsymbol{\theta}) = 0$, is not informative about γ but can be used to identify ω . The five equations (48)-(52) can then be solved for the unknowns $\boldsymbol{\theta} = (\gamma, \sigma^2, g_1, g_2, g_3)'$, with global identification requiring that the solution to $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{m}(\boldsymbol{\theta}) = (m_{22}, m_{33}, m_{12}, m_{13}, m_{23})'$, is unique in terms of $\sigma_{tt'}$, which can be estimated consistently (as $N \rightarrow \infty$) by $\hat{\sigma}_{tt'} = \frac{1}{N} \sum_{i=1}^N (\Delta y_{it} - \Delta \bar{y}_t)(\Delta y_{it'} - \Delta \bar{y}_{t'})$, where $\Delta \bar{y}_t = N^{-1} \sum_{i=1}^N \Delta y_{it}$.

A unique solution for γ can be obtained if $g_1 = 0$, but not more generally when $g_1 \neq 0$. To see this note that when $g_1 = 0$, using (51) we have $\sigma_{13} - \gamma\sigma_{12} = 0$, and γ , d_1 , d_2 and d_3 are uniquely identified, by

$$\begin{aligned} \gamma &= \frac{E\{[\Delta y_{i1} - E(\Delta y_{i1})][\Delta y_{i3} - E(\Delta y_{i3})]\}}{E\{[\Delta y_{i1} - E(\Delta y_{i1})][\Delta y_{i2} - E(\Delta y_{i2})]\}}, \\ E(\Delta y_{i1}) &= d_1, E(\Delta y_{i2}) = d_2 + \gamma d_1, E(\Delta y_{i3}) = d_3 + \gamma d_2 + \gamma^2 d_3. \end{aligned}$$

The remaining moment conditions can also be used to identify σ^2 and ω , as well as g_2 and g_3 if the sign of g_2 is set *a priori*.¹² But as soon as it is assumed that Δy_{i1} also depends on η_i (i.e. $g_1 \neq 0$), then the resultant moment conditions need not have a unique solution. In general the rank condition required for a unique solution is given by $\text{rank}(\partial \mathbf{m}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}') = 5$, where

$$\frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = - \begin{pmatrix} 2(\sigma_{12} - \gamma\sigma_{11}) & 2 & 0 & 2g_2 & 0 \\ 2(\sigma_{23} - \gamma\sigma_{22}) & 2 & 0 & 0 & 2g_3 \\ \sigma_{11} & -1 & g_2 & g_1 & 0 \\ \sigma_{12} & 0 & g_3 & 0 & g_1 \\ \sigma_{13} + \sigma_{22} - 2\gamma\sigma_{12} & 1 & 0 & g_3 & g_2 \end{pmatrix}.$$

It is clear that the rank condition is not met if $g_2 = g_3 = 0$, since in this case g_1 cannot be identified. Using (44) and noting that $\text{Var}(v_{i1}) = \omega\sigma^2$, then $\text{Var}(\Delta y_{i1}) = g_1^2\sigma_\eta^2 + \omega\sigma^2$, and even if one sets $\sigma_\eta^2 = 1$ this moment condition can not be used to identify both ω and g_1 . To identify g_1 , moment conditions for observations 2 and 3 must be used.

5.3 Identification in the likelihood setting

We now turn to the general likelihood framework allowing also for exogenous regressors. Recall $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\psi}')' = (\boldsymbol{\delta}', \gamma, \boldsymbol{\psi}')'$, with $\boldsymbol{\delta} = (\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}')'$, $\boldsymbol{\psi} = (\omega, \sigma^2, \mathbf{q}')'$ and $\mathbf{q} = \text{vec}(\mathbf{Q})$, where $\boldsymbol{\delta}$ collects the parameters associated with the initial values, the regressors, $\Delta \mathbf{x}_i$, and the time-effects, and as defined earlier, $\boldsymbol{\varrho} = (\gamma, \boldsymbol{\psi}')'$ collects the non-linear parameters. Consider the average log-likelihood function defined by (34) expressed as

$$\bar{\ell}_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi}) = N^{-1} \ell_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2N} \sum_{i=1}^N \boldsymbol{\xi}_i'(\boldsymbol{\delta}, \gamma) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\delta}, \gamma), \quad (53)$$

where $\boldsymbol{\xi}_i(\boldsymbol{\delta}, \gamma)$ is given by (35).

We require the following additional assumption.

Assumption 7 (i) $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_\delta \times \boldsymbol{\Theta}_\gamma \times \boldsymbol{\Theta}_\psi$, where $\boldsymbol{\Theta}_\delta = \boldsymbol{\Theta}_d \times \boldsymbol{\Theta}_\pi \times \boldsymbol{\Theta}_\beta$ and $\boldsymbol{\Theta}_\psi = \boldsymbol{\Theta}_\omega \times \boldsymbol{\Theta}_\sigma \times \boldsymbol{\Theta}_q$, with $\boldsymbol{\Theta}_d$, $\boldsymbol{\Theta}_\pi$, $\boldsymbol{\Theta}_\beta$ and $\boldsymbol{\Theta}_q$ compact subsets of \mathbb{R}^{n_d} , \mathbb{R}^{n_π} , \mathbb{R}^{n_β} , and \mathbb{R}^{n_q} , respectively; $\boldsymbol{\Theta}_\gamma$, $\boldsymbol{\Theta}_\omega$ and $\boldsymbol{\Theta}_\sigma$ are compact subsets of \mathbb{R} , where $n_d = T$, $n_\pi = kT$, $n_\beta = k$, and $n_q = Tm - m(m-1)/2$;

¹²Equations m_{22} and m_{23} , for example, can be used to globally identify g_2 and g_3 respectively, once the sign of g_2 is fixed. See also the related discussion, for example, in Bai and Ng (2013) in the case of the pure factor model.

$\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}'_0, \boldsymbol{\psi}'_0)' = (\boldsymbol{\delta}'_0, \gamma_0, \boldsymbol{\psi}'_0)'$ lies in the interior of $\boldsymbol{\Theta}$ (ii) the likelihood $\bar{\ell}_N(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ and for some $c_{\max} > c_{\min} > 0$, $c_{\min} \leq \inf_{\boldsymbol{\psi} \in \boldsymbol{\Theta}_{\boldsymbol{\psi}}} \lambda_{\min}[\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})] < \sup_{\boldsymbol{\psi} \in \boldsymbol{\Theta}_{\boldsymbol{\psi}}} \lambda_{\max}[\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})] \leq c_{\max}$, (iii) $\mathbf{A}(\boldsymbol{\psi}) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E \left(\Delta \mathbf{W}'_i \boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})^{-1} \Delta \mathbf{W}_i \right)$ is positive definite almost surely uniformly on $\boldsymbol{\psi} \in \boldsymbol{\Theta}_{\boldsymbol{\psi}}$.¹³

Assumption 7(i) is standard and rules out parameter values on the boundary of the parameter space. The eigenvalue conditions on $\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})$ in Assumption 7 (ii) ensure that $\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})$ is uniformly bounded.¹⁴ Assumption 7(iii) is required for identification of $\boldsymbol{\delta}_0$ and γ_0 , and also implies that $\mathbf{A}_z(\boldsymbol{\psi})$ and $\alpha_y(\boldsymbol{\psi})$, defined by

$$\mathbf{A}_z(\boldsymbol{\psi}) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E \left(\Delta \mathbf{Z}'_i \boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})^{-1} \Delta \mathbf{Z}_i \right) \text{ and } \alpha_y(\boldsymbol{\psi}) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E \left(\Delta \mathbf{y}'_{i-1} \boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})^{-1} \Delta \mathbf{y}_{i-1} \right),$$

are strictly positive definite uniformly on $\boldsymbol{\psi} \in \boldsymbol{\Theta}_{\boldsymbol{\psi}}$, where $\Delta \mathbf{Z}_i$ is the matrix of time dummies and observations on $\Delta \mathbf{x}_i$, and $\Delta \mathbf{y}_{i-1} = (0, \Delta y_{i1}, \dots, \Delta y_{iT-1})'$, as defined by the partition of $\Delta \mathbf{W}_i$ in (29). For γ we need to distinguish between the case where S is fixed (namely initialisation is from a finite past) and when $S \rightarrow \infty$. Under the former, it is only required that $|\gamma| < K$, which includes the unit root case ($|\gamma| = 1$). Under the latter (when $S \rightarrow \infty$), we must have $|\gamma| < 1$.

Given Assumptions 1-7, the global identification condition requires $f(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi}) = \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi})]$ to attain a unique maximum at $\boldsymbol{\theta}_0 = (\boldsymbol{\delta}_0, \gamma_0, \boldsymbol{\psi}_0) \in \boldsymbol{\Theta}$.

Using results (A.25) and (A.26) in Lemma 4, we have

$$\bar{\ell}_N(\boldsymbol{\delta}_0, \gamma_0, \boldsymbol{\psi}_0) - \bar{\ell}_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi}) \xrightarrow{a.s.} \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\boldsymbol{\delta}_0, \gamma_0, \boldsymbol{\psi}_0) - \bar{\ell}_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi})], \quad (54)$$

where

$$2 \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\boldsymbol{\delta}_0, \gamma_0, \boldsymbol{\psi}_0) - \bar{\ell}_N(\boldsymbol{\delta}, \gamma, \boldsymbol{\psi})] = (\boldsymbol{\delta} - \boldsymbol{\delta}_0)' \mathbf{A}_z(\boldsymbol{\psi}) (\boldsymbol{\delta} - \boldsymbol{\delta}_0) + (\gamma - \gamma_0)^2 \alpha_y(\boldsymbol{\psi}) + w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0), \quad (55)$$

$$w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) = \chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + 2(\gamma - \gamma_0) \kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0), \quad (56)$$

and

$$\chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = \text{tr} \left[\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})^{-1} \boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi}_0) \right] - \ln(|\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi}_0)| / |\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})|) - T. \quad (57)$$

Also

$$\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = \text{tr} \left\{ [\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi}) - \boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi}_0)] \boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})^{-1} \boldsymbol{\Gamma}(\gamma_0) \right\}, \quad (58)$$

where $\boldsymbol{\Gamma}(\gamma_0)$ is the lower triangular matrix with zero diagonal elements

$$\boldsymbol{\Gamma}(\gamma_0) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_0^{T-3} & \gamma_0^{T-4} & \cdots & 0 & 0 \\ \gamma_0^{T-2} & \gamma_0^{T-3} & \cdots & 1 & 0 \end{pmatrix}. \quad (59)$$

To investigate identification of the parameters of interest, namely $\boldsymbol{\delta}_0$ and γ_0 , we first write (55) more compactly as

$$f(\boldsymbol{\varepsilon}_{\delta}, \boldsymbol{\varepsilon}_{\gamma}, \boldsymbol{\varrho}, \boldsymbol{\varrho}_0) = \boldsymbol{\varepsilon}'_{\delta} \mathbf{A}_z(\boldsymbol{\psi}) \boldsymbol{\varepsilon}_{\delta} + \alpha_y(\boldsymbol{\psi}) \boldsymbol{\varepsilon}_{\gamma}^2 + w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0).$$

¹³All expectations are taken with respect to the true parameter vector $\boldsymbol{\theta}_0$, even when not explicitly denoted by $E_0(\cdot)$.

¹⁴Note also that $\boldsymbol{\Sigma}_{\xi}(\boldsymbol{\psi})$ is positive definite for every $\boldsymbol{\psi} \in \boldsymbol{\Theta}_{\boldsymbol{\psi}}$, when the order condition is met and $\omega > 1 - \frac{1}{T}$. Recall that under the latter $\boldsymbol{\Omega}$ is a positive definite matrix and \mathbf{Q} is rank deficient, and under Assumption 1, $0 < \sigma^2 < K$.

We also note that by the information inequality $f(\varepsilon_\delta, \varepsilon_\gamma, \boldsymbol{\varrho}, \boldsymbol{\varrho}_0) \geq 0$ for all values of $\varepsilon_\delta, \varepsilon_\gamma, \boldsymbol{\varrho}$, and $\boldsymbol{\varrho}_0$. Global identification of $\boldsymbol{\delta}_0$ and γ_0 , requires that $f(\varepsilon_\delta, \varepsilon_\gamma, \boldsymbol{\varrho}, \boldsymbol{\varrho}_0) = 0$ solves uniquely for $\varepsilon_\delta = \mathbf{0}$, and $\varepsilon_\gamma = 0$, for all values of $\boldsymbol{\varrho}$ and $\boldsymbol{\varrho}_0$. Furthermore, we have that

$$f(\varepsilon_\delta, \varepsilon_\gamma, \boldsymbol{\varrho}, \boldsymbol{\varrho}_0) = \varepsilon'_\delta \mathbf{A}_z(\boldsymbol{\psi}) \varepsilon_\delta + \alpha_y(\boldsymbol{\psi}) \varepsilon_\gamma^2 + w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) \geq \lambda_{\min}[\mathbf{A}_z(\boldsymbol{\psi})] \varepsilon'_\delta \varepsilon_\delta + \alpha_y(\boldsymbol{\psi}) \varepsilon_\gamma^2 + w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) \geq 0. \quad (60)$$

It is now easily established that $\boldsymbol{\delta}_0$ and γ_0 are globally identified if $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) \geq 0$ for all values of γ and $\boldsymbol{\psi}$. Note that since the right hand side of (60) is non-negative, then if $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) \geq 0$ we must also have $\lambda_{\min}[\mathbf{A}_z(\boldsymbol{\psi})] \varepsilon'_\delta \varepsilon_\delta \geq 0$ and $\alpha_y(\boldsymbol{\psi}) \varepsilon_\gamma^2 \geq 0$. Then condition $f(\varepsilon_\delta, \varepsilon_\gamma, \boldsymbol{\varrho}, \boldsymbol{\varrho}_0) = 0$ can occur if and only if

$$\lambda_{\min}[\mathbf{A}_z(\boldsymbol{\psi})] \varepsilon'_\delta \varepsilon_\delta = 0, \text{ and } \alpha_y(\boldsymbol{\psi}) \varepsilon_\gamma^2 = 0, \quad (61)$$

noting further that, if $\lambda_{\min}[\mathbf{A}_z(\boldsymbol{\psi})] \varepsilon'_\delta \varepsilon_\delta > 0$ and/or $\alpha_y(\boldsymbol{\psi}) \varepsilon_\gamma^2 > 0$, then $f(\varepsilon_\delta, \varepsilon_\gamma, \boldsymbol{\varrho}, \boldsymbol{\varrho}_0) > 0$ for sure, so long as $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) \geq 0$. It now follows that since by Assumption 7(iii) $\lambda_{\min}[\mathbf{A}_z(\boldsymbol{\psi})] > 0$, and $\alpha_y(\boldsymbol{\psi}) > 0$, then conditions in (61) hold if and only if $\varepsilon_\delta = 0$ and $\varepsilon_\gamma = 0$, and the desired result is established.

But in general it is not possible to be sure that $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0)$ is non-negative. Consider now $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0)$, and note that its second component, $\chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0)$, can be written as

$$\chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = \text{tr}[\mathbf{B}] - \ln(\mathbf{B}) - T,$$

where $\mathbf{B} = \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)$ which is a positive definite matrix, and using result 10 on p.44 of Lütkepohl (1996) we have that $\chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) \geq 0$.

Also,

$$\begin{aligned} \kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) &= \text{tr} \left\{ [\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) - \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)] \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\Gamma}(\gamma_0) \right\} \\ &= \text{tr} \left\{ [\mathbf{I}_T - \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}] \boldsymbol{\Gamma}(\gamma_0) \right\} \\ &\geq \text{tr}[\mathbf{I}_T - \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}] \lambda_{\min}[\boldsymbol{\Gamma}(\gamma_0)], \end{aligned} \quad (62)$$

and since $\lambda_{\min}[\boldsymbol{\Gamma}(\gamma_0)] = 0$, then $\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) \geq 0$, as well. Overall, for values of $\boldsymbol{\varrho} \neq \boldsymbol{\varrho}_0$, $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) = \chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + 2(\gamma - \gamma_0) \kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0)$ is ensured to be non-negative only if $(\gamma - \gamma_0) > 0$, otherwise the second term of $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0)$ could become sufficiently large and negative such that $w(\boldsymbol{\varrho}, \boldsymbol{\varrho}_0) < 0$. Therefore, to ensure global identification of $\boldsymbol{\delta}_0$ and γ_0 for all values of γ and $\boldsymbol{\psi}$ it is required that $\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = 0$. But as shown in the example below, this can occur if the distribution of the initial first differences, Δy_{i1} does not depend on the latent factor, which renders Δy_{i1} uncorrelated with Δy_{it} , for $t \geq 2$.

Finally, it is worth noting that even if $\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = 0$, global identification of $\boldsymbol{\psi}_0$ will involve additional restrictions on $\boldsymbol{\psi}$, since $\chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = 0$ only ensures equality of eigenvalues of $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})$ and $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)$, which does not necessarily imply that $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Identification of $\boldsymbol{\psi}_0$ is ensured if $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})$ and $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)$ commute, as the simple example below illustrates.

Remark 4 *The above results clearly highlight the fact that in general it is not possible to guarantee global identification in the presence of the lagged dependent variable. Allowing for regressors $\Delta \mathbf{x}_i$ and the associated initial values component, as well as the time effects, do not alter this conclusion. These results are also in line with the moment condition based identification results discussed earlier. Further insights into conditions related to global identification using the likelihood framework are illustrated by the example that follows.*

Example 1 *To keep the illustration as simple as possible we consider the panel data model without fixed effects given by*

$$\begin{aligned} y_{i1} &= \lambda_i f_1 + v_i, \\ y_{i2} &= \gamma y_{i1} + \lambda_i f_2 + u_{i2}, \end{aligned}$$

for $i = 1, 2, \dots, N$, and assume that λ_i , v_i , and u_{i2} are cross-sectionally, and mutually independent, have zero means, with variances, σ_λ^2 , σ_v^2 and σ_2^2 , respectively. As shown earlier global identification is possible when the initial values, y_{i1} , do not depend on the common factor. It is clear that in this model γ_0 , the true value of γ , is not identified, unless $f_1 = 0$. Under this restriction γ_0 is identified using the moment condition $E_0[y_{i1}(y_{i2} - \gamma y_{i1})] = 0$. Consider now the application of the likelihood approach to this simple model under $f_1 = 0$. In this case

$$\Sigma_\xi(\psi) = \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma^2 \end{pmatrix},$$

with $\psi = (\sigma_v^2, \sigma^2)$, $\sigma^2 = f_2^2 \sigma_\lambda^2 + \sigma_2^2 > 0$. Using (55) we have

$$2 \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\gamma_0, \psi_0) - \bar{\ell}_N(\gamma, \psi)] = (\gamma - \gamma_0)^2 \alpha_y(\psi) + \chi(\psi, \psi_0) + 2(\gamma - \gamma_0) \kappa(\psi, \psi_0), \quad (63)$$

where $\alpha_y(\psi) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E(\mathbf{y}'_{i,-1} \Sigma_\xi(\psi)^{-1} \mathbf{y}_{i,-1})$, $\mathbf{y}_{i,-1} = (0, y_{i1})$, which simplifies to $\alpha_y(\psi) = \sigma^{-2} \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N y_{i1}^2 = \sigma_v^2 / \sigma^2$,

$$\begin{aligned} \chi(\psi, \psi_0) &= \text{tr} [\Sigma_\xi(\psi)^{-1} \Sigma_\xi(\psi_0)] - \ln(|\Sigma_\xi(\psi_0)| / |\Sigma_\xi(\psi)|) - 2 \\ &= \left[\frac{\sigma_0^2}{\sigma^2} - \ln \left(\frac{\sigma_0^2}{\sigma^2} \right) - 1 \right] + \left[\frac{\sigma_{0,v}^2}{\sigma_v^2} - \ln \left(\frac{\sigma_{0,v}^2}{\sigma_v^2} \right) - 1 \right] \geq 0, \end{aligned}$$

and

$$\begin{aligned} \kappa(\psi, \psi_0) &= \text{tr} \left\{ [\Sigma_\xi(\psi) - \Sigma_\xi(\psi_0)] \Sigma_\xi(\psi)^{-1} \Gamma(\gamma_0) \right\} \\ &= -\text{tr} \left\{ \Sigma_\xi(\psi_0) \Sigma_\xi(\psi)^{-1} \Gamma(\gamma_0) \right\} \\ &= -\text{tr} \left\{ \begin{pmatrix} \frac{\sigma_{0,v}^2}{\sigma_v^2} & 0 \\ 0 & \frac{\sigma_0^2}{\sigma^2} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right\} = 0. \end{aligned}$$

Hence,

$$2 \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\gamma_0, \psi_0) - \bar{\ell}_N(\gamma, \psi)] = (\gamma - \gamma_0)^2 \alpha_y(\psi) + \chi(\psi, \psi_0).$$

We further have that $\frac{\sigma_0^2}{\sigma^2} - \ln \left(\frac{\sigma_0^2}{\sigma^2} \right) - 1 \geq 0$ and $\frac{\sigma_{0,v}^2}{\sigma_v^2} - \ln \left(\frac{\sigma_{0,v}^2}{\sigma_v^2} \right) - 1 \geq 0$, with equalities holding if and only if $\sigma^2 = \sigma_0^2$ and $\sigma_v^2 = \sigma_{0,v}^2$, respectively. Note also that in this simple example the matrices $\Sigma_\xi(\psi)$ and $\Sigma_\xi(\psi_0)$ commute. It then follows that we must also have $\gamma = \gamma_0$ if and only if $\alpha_y(\psi) > 0$. In fact, the diagonality of $\Sigma_\xi(\psi)$ and $\alpha_y(\psi) = \sigma_v^2 / \sigma^2 > 0$ are both necessary and sufficient for global identification of γ_0 . A similar outcome also follows if we allow for fixed effects and work with the first-differenced version of the panel. But for the first-differenced version we need $T = 3$ with $g_1 = 0$. The likelihood approach can now be applied to

$$\begin{aligned} \Delta y_{i1} &= \Delta u_{i1}, \\ \Delta y_{i2} - \gamma \Delta y_{i1} &= g_2 \eta_i + \Delta u_{i2}, \\ \Delta y_{i3} - \gamma \Delta y_{i2} &= g_3 \eta_i + \Delta u_{i3}. \end{aligned}$$

Since due to first-differencing $\text{Cov}(\Delta u_{i1}, \Delta u_{i2}) = -\sigma^2$, to ensure the diagonality of $\Sigma_\xi(\psi)$ for this application, noting that $\text{Cov}(\Delta u_{i1}, g_3 \eta_i + \Delta u_{i3}) = 0$, it is sufficient to apply the likelihood approach to

$$\begin{aligned} \Delta y_{i1} &= \Delta u_{i1}, \\ \Delta y_{i3} - \gamma \Delta y_{i2} &= g_3 \eta_i + \Delta u_{i3}, \end{aligned}$$

with $\xi_i(\gamma) = (\Delta y_{i1}, \Delta y_{i3} - \gamma \Delta y_{i2})'$. In this set up one can first obtain a uniquely consistent estimator using Δy_{i1} and $\Delta y_{i3} - \gamma \Delta y_{i2}$, and then use this consistent estimator as initial value for a more efficient ML estimation that also makes use of the relations $\Delta y_{i2} - \gamma \Delta y_{i1} = g_2 \eta_i + \Delta u_{i2}$, for $i = 1, 2, \dots, N$.

5.3.1 Local identification

As global identification of δ_0 and γ_0 on the parameter space Θ cannot be guaranteed, we proceed by considering a restriction of Θ on which identification and consistency will be shown.¹⁵ To this end we introduce the following definition:

Definition 1 Let $\mathcal{N}_\epsilon(\varrho_0)$ be a set in the closed neighbourhood of ϱ_0 defined by

$$\mathcal{N}_\epsilon(\varrho_0) = \{\varrho \in \Theta_\gamma \times \Theta_\psi : \|\varrho - \varrho_0\| \leq \epsilon\}$$

for some $\epsilon > 0$, such that

$$w(\varrho, \varrho_0) = \chi(\psi, \psi_0) + 2(\gamma - \gamma_0)\kappa(\psi, \psi_0) \geq 0, \quad (64)$$

for all values of $\gamma \in \Theta_\gamma$ and $\psi \in \Theta_\psi$ where Θ_γ is a compact subset of \mathbb{R} and $\Theta_\psi = \Theta_\omega \times \Theta_\sigma \times \Theta_q$, with Θ_ω and Θ_σ compact subsets of \mathbb{R} , and Θ_q a compact subset of \mathbb{R}^{n_q} , with $n_q = Tm - m(m-1)/2$.

In view of the local nature of our analysis, from hereon we consider the more restricted parameter space as set out in the following assumption.

Assumption 8 $\theta \in \Theta_\epsilon = \Theta_\delta \times \mathcal{N}_\epsilon(\varrho_0)$, where $\Theta_\delta = \Theta_d \times \Theta_\pi \times \Theta_\beta$ and $\mathcal{N}_\epsilon(\varrho_0)$ is specified in Definition 1, $\Theta_\psi = \Theta_\omega \times \Theta_\sigma \times \Theta_q$, with $\Theta_d, \Theta_\pi, \Theta_\beta$ and Θ_q compact subsets of $\mathbb{R}^{n_d}, \mathbb{R}^{n_\pi}, \mathbb{R}^{n_\beta}$, and \mathbb{R}^{n_q} , respectively; Θ_ω and Θ_σ are compact subsets of \mathbb{R} , where $n_d = T, n_\pi = kT, n_\beta = k$, and $n_q = Tm - m(m-1)/2$; $\mathcal{N}_\epsilon(\varrho_0)$ is given in Definition 1, Θ_ϵ is a compact subset of \mathbb{R}^{n_θ} with $n_\theta = 3 + T(k+1) + k + Tm - m(m-1)/2$; and $\theta_0 = (\varphi'_0, \psi'_0)' = (\delta'_0, \gamma_0, \psi'_0)'$ lies in the interior of Θ_ϵ .

We now have:

Proposition 1 Consider the model given by (13), with the associated log-likelihood function for first-differences given by (34). Suppose that Assumptions 1-7(ii),(iii) and 8, as well as the order condition (41) hold. Then δ_0 and γ_0 are almost surely (locally) identified on Θ_ϵ .

The proof follows noting that for all values of $\theta \in \Theta_\epsilon$, condition (64) is met and hence local identification of δ_0 and γ_0 is established using (60). In what follows we also assume that ψ_0 is locally identified under suitable additional restrictions on $\Sigma_\xi(\psi)$ such that $\chi(\psi, \psi_0) = 0$ ensures that $\psi_0 = \psi$. Note that under local identification of γ_0 and δ_0 we also have $\chi(\psi, \psi_0) = 0$, but as noted earlier this by itself does not necessarily ensure that $\psi = \psi_0$. Under local identification of γ_0 and δ_0 , in order for ψ_0 to also be locally identified it is further required that $\Sigma_\xi(\psi_0)$ and $\Sigma_\xi(\psi)$ have the same eigenvectors and eigenvalues, and this is ensured if on Θ_ϵ the two matrices $\Sigma_\xi(\psi_0)$ and $\Sigma_\xi(\psi)$ commute, as previously mentioned.

6 Asymptotic properties of the transformed QML estimator

The analysis of consistency and asymptotic normality of the TQML estimator, $\hat{\theta} = \arg \max_{\theta \in \Theta_\epsilon} \bar{\ell}_N(\theta)$, now follows by application of standard results from the literature. Almost sure local consistency of $\hat{\theta}$ follows, for example, from a straightforward adaptation of Theorem 9.3.1 of Davidson (2000). Specifically under Assumptions 1-7(ii),(iii), and 8 we have: (i) Θ_ϵ as a subset of Θ is compact, (ii) setting $\bar{C}_N(\theta) = -2\bar{\ell}_N(\theta)$, and $\bar{C}(\theta) = E_0[\bar{C}_N(\theta)]$, $\bar{C}_N(\theta) \xrightarrow{a.s.} \bar{C}(\theta)$ uniformly on Θ_ϵ as shown in the proof of 2 in the

¹⁵This approach is typical in the time series literature under QMLE theory, see for example Lumsdaine (1996) for the GARCH model, Allen et al. (2008) for the case of the Logarithmic Autoregressive Conditional Duration model, Kristensen and Rahbek (2010) for nonlinear error-correction models, and Han and Kristensen (2014) for GARCH-X models with stationary and nonstationary covariates, among others.

Appendix, (iii) θ_0 , an interior point of Θ_ϵ , is the unique minimum of $\bar{C}(\theta)$ on Θ_ϵ by Proposition 1 and given that $\Sigma_\xi(\psi_0)$ and $\Sigma_\xi(\psi)$ commute. Therefore, all three conditions of Theorem 9.3.1 of Davidson are satisfied and $\hat{\theta} \xrightarrow{a.s.} \theta_0$ on the set Θ_ϵ .

The asymptotic distribution of $\hat{\theta}$ is derived by taking a Taylor expansion of $\frac{\partial \bar{\ell}_N(\hat{\theta})}{\partial \theta} = \mathbf{0}$ at θ_0 and checking the asymptotic behaviour of the score function, $\bar{s}_N(\theta) = \frac{\partial \bar{\ell}_N(\theta)}{\partial \theta}$, and Hessian matrix, $\mathbf{H}_N(\theta) = -\frac{\partial^2 \bar{\ell}_N(\theta)}{\partial \theta \partial \theta'}$. If $E_0 \left[\frac{\partial \bar{\ell}_N(\theta_0)}{\partial \theta} \right] = \mathbf{0}$ and $\mathbf{H}_N(\check{\theta}) \xrightarrow{a.s.} \mathbf{H}(\theta_0)$, the asymptotic normality of the TQML estimator will follow from the mean value theorem:

$$\mathbf{0} = \sqrt{N} \bar{s}_N(\hat{\theta}) = \sqrt{N} \bar{s}_N(\theta_0) - \mathbf{H}_N(\check{\theta}) \sqrt{N}(\hat{\theta} - \theta_0) \quad (65)$$

where $\check{\theta}$ lies between $\hat{\theta}$ and θ_0 . The resultant asymptotic distribution is summarised in the following theorem:

Theorem 2 *Consider the dynamic panel data model with interactive effects given by (13). Suppose that Assumptions 1-7(ii),(iii) and 8, as well as the order condition (41) and Proposition 1 hold, and that $\Sigma_\xi(\psi_0)$ and $\Sigma_\xi(\psi)$ commute. Denote the TQML estimator of θ_0 by $\hat{\theta} = \arg \max_{\theta \in \Theta_\epsilon} \bar{\ell}_N(\theta)$, where $\bar{\ell}_N(\theta)$ is given by (53). Then, $\hat{\theta}$ is almost surely locally consistent for θ_0 on Θ_ϵ and*

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow_d N \left[\mathbf{0}, \mathbf{H}^{-1}(\theta_0) \mathbf{J}(\theta_0) \mathbf{H}^{-1}(\theta_0) \right], \quad (66)$$

where $\mathbf{H}(\theta_0) = \lim_{N \rightarrow \infty} E_0 \left[-\frac{\partial^2 \bar{\ell}_N(\theta_0)}{\partial \theta \partial \theta'} \right]$ and $\mathbf{J}(\theta_0) = \lim_{N \rightarrow \infty} E_0 \left[N \frac{\partial \bar{\ell}_N(\theta_0)}{\partial \theta} \frac{\partial \bar{\ell}_N(\theta_0)}{\partial \theta'} \right]$ are assumed to exist and be positive definite.

When $\xi_i(\varphi_0)$ is Gaussian $\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow_d N \left[\mathbf{0}, \mathbf{H}^{-1}(\theta_0) \right]$. A consistent estimator for the variance in (66) can be obtained by substituting $\hat{\theta}$ for θ_0 in the expressions for $\mathbf{J}(\theta_0)$ and $\mathbf{H}(\theta_0)$.

Remark 5 *Since in general we do not have global identification, in practice when computing the proposed TQML estimator it is advisable that a number of different initial parameter vectors are considered in the optimisation procedure to ensure, as far as possible, that the resultant estimates correspond to the locally consistent maximum.*

7 Estimating the number of factors

There are a number of studies that provide information criteria for selecting the number of factors including Bai and Ng (2002), Onatski (2010), Kapetanios (2010), Ahn and Horenstein (2013), among others. However, these are not applicable to short T panel data sets, and require both N and T to be large. In the case of short T panels Ahn et al. (2013) estimate the true number of factors, m_0 , within a GMM framework using the Sargan-Hansen misspecification statistic in a sequential manner, as well as information criteria. To ensure consistency of the selected number of factors under the former case, following Bauer and Hackl (1988) and Cragg and Donald (1997), Ahn et al. (2013) choose the significance level b_N such that $b_N \rightarrow 0$ and $-\ln(b_N)/N \rightarrow 0$ as $N \rightarrow \infty$. Using simulations they find that the sequential method could produce better estimates if the significance level depends also on T (in addition to N), when the regressors and $\eta'_i \mathbf{f}_t$ are not highly correlated, but do not provide theoretical details on how best to allow for T as well as N in their selection procedure. In what follows we consider a sequential likelihood ratio (LR) testing procedure, but adjust the significance level of the tests to take account of the multiple testing nature of the procedure in terms of T , as well as adjusting the size of the tests in terms of N to ensure consistency of the selected number of factors. We provide a formal theory that should be of general interest for the analysis of short T factor models.

7.1 A sequential multiple testing likelihood ratio procedure for estimating the number of factors

We first consider the problem of testing $H_0: m = m_0$ against $H_1: m = m_{\max}$, where m_{\max} is the largest value of m that satisfies the order condition (41), namely $m_{\max} = T - 2$. This is in contrast to the problem of selecting m in the case of large N and T factor models where it is often based on an arbitrary choice of m_{\max} . Under H_0 , the maximised log-likelihood function, $\ell_N(\hat{\theta}_{m_0})$, is computed by maximising (53) subject to r_0 over-identifying restrictions given by

$$r_0 = T(T+1)/2 - 3 - [Tm_0 - m_0(m_0 - 1)/2]. \quad (67)$$

Denote the exactly identified estimator of θ (under H_1) by $\hat{\theta}_{m_{\max}}$ with its dimension $n_{\theta}^* = 3 + T(k+1) + k + (T-2)(T+3)/2$, and the constrained estimator of θ under $H_0: m = m_0 < T - 2$ by $\hat{\theta}_{m_0}$. The latter estimator is obtained under $\mathbf{r}(\theta_0) = \mathbf{0}$, where $\mathbf{r}(\theta_0)$ is the $r_0 \times 1$ vector of restrictions on $\ell_N(\theta)$, the log-likelihood function defined by (34), implied by setting $m = m_0$. The LR statistic for testing $H_0: m = m_0$ against $H_1: m = m_{\max} = T - 2$, is then given by

$$\mathcal{LR}_N(m_0, m_{\max}) = 2 \left[\ell_N(\hat{\theta}_{m_{\max}}) - \ell_N(\hat{\theta}_{m_0}) \right], \text{ for } m_0 = 0, 1, 2, \dots, T - 3. \quad (68)$$

The following theorem provides the asymptotic distribution of \mathcal{LR}_N under the null and η -local alternatives, the latter to be defined below.

Theorem 3 *Consider the dynamic panel data model given by (13), and suppose that Theorem 2 holds. Denote the constrained TQML estimator of θ obtained under $H_0: m = m_0$ by $\hat{\theta}_{m_0}$ and its unconstrained estimator by $\hat{\theta}_{m_{\max}}$, where $m_{\max} = T - 2$. Also let the restrictions imposed under H_0 be given by $\mathbf{r}(\theta_0) = \mathbf{0}$, where $\mathbf{r}(\theta_0)$ is the $r_0 \times 1$ vector function of θ implied by setting $m = m_0$ where $r_0 = T(T+1)/2 - 3 - [Tm_0 - m_0(m_0 - 1)/2]$. Then: (a) under the null $H_0: m = m_0$ (or equivalently under $\mathbf{r}(\theta_0) = \mathbf{0}$), the log-likelihood ratio statistic \mathcal{LR}_N , defined by (68), has the following asymptotic distribution (for a fixed T , and as $N \rightarrow \infty$)*

$$\mathcal{LR}_N \rightarrow_d \sum_{j=1}^{r_0} w_j z_j^2, \quad (69)$$

where $z_j \sim \text{IIDN}(0, 1)$, w_1, w_2, \dots, w_{r_0} are the strictly positive eigenvalues of the symmetric matrix

$$\mathbf{A}_0 = \mathbf{J}_0^{1/2} \mathbf{H}_0^{-1} \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0^{1/2}, \quad (70)$$

with $\mathbf{J}_0 = \mathbf{J}(\theta_0)$, $\mathbf{H}_0 = \mathbf{H}(\theta_0)$, and $\mathbf{R}_0 = \mathbf{R}(\theta_0)$, where $\mathbf{R}(\theta_0) = \partial \mathbf{r}(\theta_0) / \partial \theta'$ is of dimension $(r_0 \times n_{\theta}^*)$ with $n_{\theta}^* = 3 + T(k+1) + k + (T-2)(T+3)/2$, such that $\text{rank}[\mathbf{R}(\theta_0)] = r_0$, (b) furthermore, under η -local alternatives $H_{1N}: \theta_{1N} = \theta_0 + N^{-\eta/2} \kappa$, where κ is a $n_{\theta}^* \times 1$ vector of constants such that $\kappa' \kappa > 0$ and $0 < \eta < 1$, we have

$$\frac{N^{-(1-\eta)/2} \mathcal{LR}_N - N^{(1-\eta)/2} \kappa' \mathbf{S}_c \kappa}{2 \sqrt{\kappa' \mathbf{S}_b' \mathbf{S}_b \kappa}} \underset{a}{\sim} N(0, 1), \quad (71)$$

where \mathbf{S}_c and $\mathbf{S}_b' \mathbf{S}_b$ are symmetric positive definite matrices defined by

$$\mathbf{S}_c = \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0, \quad (72)$$

and

$$\mathbf{S}_b' \mathbf{S}_b = \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0 \mathbf{H}_0^{-1} \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0, \quad (73)$$

respectively.

Remark 6 It is worth noting that the concept of η -local alternatives extends the standard Pitman sequence of local alternatives where η is set to $\eta = 1$. By considering alternatives that tend towards the null at a slower rate, with $\eta < 1$, we are able to allow both Types I and II errors to tend to zero.

Remark 7 Note that the non-zero eigenvalues of \mathbf{A}_0 (given by (70)) are also the eigenvalues of $(\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')$. Hence, if $\mathbf{J}_0 = \mathbf{H}_0$, this matrix becomes equal to \mathbf{I}_{r_0} and we have $w_i = 1, (i = 1, 2, \dots, r_0)$, which yields the familiar result

$$\mathcal{LR}_N \rightarrow_d \chi_{r_0}^2, \text{ under } \mathbf{r}(\boldsymbol{\theta}_0) = \mathbf{0},$$

where $\chi_{r_0}^2$ is a central chi-squared variate with r_0 degrees of freedom.

Theorem 3 shows that the use of the LR test in the non-Gaussian setting is non-standard and requires an explicit derivation when $H_0 : \mathbf{r}(\boldsymbol{\theta}_0) = \mathbf{0}$. Furthermore, even in the standard case the use of the sequential LR procedure for the estimation of m is subject to the multiple testing problem and does not guarantee that m_0 , the true value of m , will be estimated consistently. This is a well known problem in the sequential testing literature. In this paper, we deal with both of these problems by letting the overall size of the sequential LR tests decline with N at a suitable rate, which we show yields the desired result even if the underlying individual LR tests are non-standard.

Proposition 2 Suppose that the assumptions of Theorem 3 hold, and that under the null hypothesis H_0 the LR test statistic \mathcal{LR}_N given by (68) is distributed as $\sum_{i=1}^{r_0} w_i \chi_i^2(1)$, where the weights $w_1 \geq w_2 \geq \dots \geq w_{r_0} > 0$ are finite constants, and $\chi_i^2(1)$ for $i = 1, 2, \dots, h$ are independently distributed central chi-squared variates with 1 degree of freedom. Denote the type I error probability of the test by α_N , and the critical value of the test by $c_N^2(r_0)$. If $c_N^2(r_0) \rightarrow \infty$ as $N \rightarrow \infty$, then $\lim_{N \rightarrow \infty} \alpha_N = 0$.

Corollary 1 Under the assumptions of Theorem 3, define the critical value of the test by $c_N^2(r_0)$ with $c_N^2(r_0) \rightarrow \infty$ as $N \rightarrow \infty$, and the type II error probability by β_N . For all η -local alternatives $H_{1N} : \boldsymbol{\theta}_{1N} = \boldsymbol{\theta}_0 + N^{-\eta/2} \boldsymbol{\kappa}$, with $\boldsymbol{\kappa}' \boldsymbol{\kappa} > 0$,

$$\lim_{N \rightarrow \infty} \beta_N = \lim_{N \rightarrow \infty} \Pr[\mathcal{LR}_N \leq c_N^2(r_0) | H_{1N}] = 0, \quad (74)$$

so long as $\eta < 1$, and $N^{-(1-\eta)} c_N^2(r_0) \rightarrow 0$, as $N \rightarrow \infty$.

Remark 8 From Proposition 2 and Corollary 1 it follows that if $c_N^2(r_0) \rightarrow \infty$ as $N \rightarrow \infty$ such that $N^{-(1-\eta)} c_N^2(r_0) \rightarrow 0$, then $\lim_{N \rightarrow \infty} \alpha_N = \lim_{N \rightarrow \infty} \beta_N = 0$, assuming that the relevant Hessian matrices are non-singular and the restrictions are full rank. To see that both these conditions are met if $\alpha_N = p/N^\delta$ with δ a finite positive constant, using (A.56) in the Appendix we have that

$$\frac{c_N^2(r_0)}{N^{(1-\eta)}} \leq \frac{2w_1 r_0 \ln\left(\frac{r_0}{\alpha_N}\right)}{N^{(1-\eta)}} = \frac{2w_1 r_0 \ln\left(\frac{r_0 N^\delta}{p}\right)}{N^{(1-\eta)}} = O\left(\frac{\delta \ln(N)}{N^{(1-\eta)}}\right), \quad (75)$$

Since $\ln(N) \rightarrow \infty$ as $N \rightarrow \infty$, then for any $\delta > 0$ it follows that $c_N^2(r_0) \rightarrow \infty$ as $N \rightarrow \infty$. Also, as $N \rightarrow \infty$, then $\ln(N)/N^{(1-\eta)} \rightarrow 0$, so long as η is not too close to unity, and it will be surely met if η is close to $1/2$. Hence $c_N^2(r_0)/N^{(1-\eta)} \rightarrow 0$ as $c_N^2(r_0) \rightarrow \infty$ with $N \rightarrow \infty$.

Remark 9 When α_N is set as $\alpha_N = p/N^\delta$, the parameter p ($0 < p < 1$) can be viewed as the nominal size of the test. The Neyman-Pearson case is obtained if we set $\delta = 0$. The case of $\delta > 0$ relates to the Chernoff test procedure that aims at minimizing $\Pr(H_0)\alpha_N + \Pr(H_1)\beta_N$, where $0 < \Pr(H_0) < 1$ and $0 < \Pr(H_1) < 1$ are prior probabilities of H_0 and H_1 , respectively. When N is finite the solution to this problem depends on the prior probabilities. But in the case of chi-squared tests, we have $\Pr(H_0)\alpha_N + \Pr(H_1)\beta_N \rightarrow 0$ as $N \rightarrow \infty$, irrespective of the prior probabilities $\Pr(H_0)$ and $\Pr(H_1)$, so long as $\alpha_N = p/N^\delta$ for $\delta > 0$ and $p > 0$.

Remark 10 *In finite samples the choice of p and δ can matter, though for moderate values of N the choice of p is likely to be of second order importance. In the simulation results that follow we set $\delta = 1$ and $p = 5\%$, and investigate the robustness of the results to other choices of p .*

Theorem 3 together with Corollary 1 and Proposition 2, can now be used to develop a sequential procedure for estimating (selecting) m that accounts for the multiple testing nature of the approach, and is consistent for the true number of factors m_0 . Consistency is ensured as long as Proposition 2 and Corollary 1 both hold, which in conjunction with Remark 8 effectively requires the size of the sequential LR tests to decline with N .

As m_0 is unknown and could be $T - 2$, we assume the sequential procedure involves $T - 2$ separate tests, although in some applications we might end up stopping the sequential procedure having carried out a fewer number of tests than $T - 2$. Let the hypotheses of interest be $H_{0,T-2}, H_{1,T-2}, \dots, H_{T-3,T-2}$ (the total available as determined by the order condition) with the first and second subscripts denoting the number of factors specified under the null and alternative hypothesis respectively, and write the $T - 2$ LR tests as

$$\Pr [\mathcal{LR}_N (m_0 = t - 1, m_{\max} = T - 2) > CV_{N,t-1,T-2} | H_{t-1,T-2}] \leq p_{N,t-1,T-2}, \text{ for } t = 1, 2, \dots, T - 2,$$

where $\mathcal{LR}_N (m_0, m_{\max})$ is given by (68), $CV_{N,t-1,T-2}$ is the critical value for the test of $H_{t-1,T-2}$, and $p_{N,t-1,T-2}$ is the realised p -value for $H_{t-1,T-2}$.

The sequential testing procedure begins by using the likelihood ratio statistic \mathcal{LR}_N to test $H_{0,T-2}$, that is the null hypothesis $m = 0$ against the alternative $m = T - 2$. If the null hypothesis is rejected, one proceeds to test $H_{1,T-2}$, that is the null hypothesis $m = 1$ against the alternative $m = T - 2$, and so forth. This sequential process is continued until the \mathcal{LR}_N test fails to reject the null hypothesis associated with $H_{m_0,m_{\max}}$. The estimated number of factors, \hat{m} , is then equal to the number of factors specified under the null hypothesis associated with this event of non-rejection. If \mathcal{LR}_N rejects the null hypothesis associated with all $H_{0,T-2}, H_{1,T-2}, \dots, H_{T-3,T-2}$ then \hat{m} is set equal to $T - 2$.

The overall size of the test is given by the family-wise error rate (FWER) defined by

$$FWER_N = \Pr \left[\bigcup_{t=1}^{T-2} (\mathcal{LR}_N (m_0 = t - 1, m_{\max} = T - 2) > CV_{N,t-1,T-2} | H_{t-1,T-2}) \right].$$

Suppose that we wish to control $FWER_N$ to lie below a pre-determined value, p . An exact solution to this problem depends on the nature of the dependence across the underlying tests, which is generally difficult to obtain. But one could derive bounds on $FWER_N$ using, for example, the Bonferroni (1936) or Holm (1979) procedures. Both of these procedures are valid for all possible degrees of dependence across the individual tests, and as a result tend to be conservative in the sense that the actual size will be lower than the overall target size of p . Using the union bound we have

$$\begin{aligned} & \Pr \left\{ \bigcup_{t=1}^{T-2} [\mathcal{LR}_N (m_0 = t - 1, m_{\max} = T - 2) > CV_{N,t-1,T-2} | H_{t-1,T-2}] \right\} \\ & \leq \sum_{t=1}^{T-2} \Pr (\mathcal{LR}_N (m_0 = t - 1, m_{\max} = T - 2) > CV_{N,t-1,T-2} | H_{t-1,T-2}) \leq \sum_{t=1}^{T-2} p_{N,t-1,T-2}. \end{aligned}$$

Hence, to obtain $FWER_N \leq p$, it is sufficient to set $p_{N,t-1,T-2} \leq p/(T - 2)$. To ensure consistency of the sequential LR procedure, in line with the earlier discussion and the theorem that follows, $p/(T - 2)$ is further adjusted so that $\alpha_N = p/N(T - 2)$.¹⁶ The individual critical values, $CV_{N,t-1,T-2}$ for performing the sequential MTLR procedure are based on the critical values of the χ^2 distribution, namely

¹⁶Substituting for $\alpha_N = p/N(T - 2)$ in (75), it is easy to see that the required conditions that ensure consistency of the test continue to be satisfied.

$\chi_{r_0}^2 [p/N(T-2)]$, where $p/N(T-2)$ is the right-tail probability of the individual tests and $r_0 = T(T+1)/2 - 3 - [Tm_0 - m_0(m_0-1)/2]$.¹⁷

Local consistency of \hat{m} for m_0 on Θ_ϵ is established in the following theorem.

Theorem 4 *Let \hat{m} be the number of factors obtained using the sequential likelihood ratio procedure based on the statistic \mathcal{LR}_N given by (68), for which Theorem 3, Corollary 1 and Proposition 2 hold. Then \hat{m} is almost surely locally consistent for m_0 on Θ_ϵ .*

8 Small sample properties of the transformed QML estimator

In this section, we investigate the finite sample properties of the proposed estimator using Monte Carlo (MC) simulations. We start by presenting the MC design.

8.1 Monte Carlo design

The observations on y_{it} are generated assuming $k = 1$ (one exogenous regressor) and m_0 unobserved factors as

$$y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}, \quad (76a)$$

$$\zeta_{it} = \sum_{\ell=1}^{m_0} \eta_{\ell i} f_{\ell t} + u_{it} = \boldsymbol{\eta}'_i \mathbf{f}_t + u_{it}, \quad (76b)$$

for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$. Together with the initial observation for $t = 0$ which will be set below, this yields T observations for estimation after first-differencing. The fixed effects α_i are generated as $\alpha_i \sim IIDN(0, 1)$. The factor loadings, $\boldsymbol{\eta}_i = (\eta_{1i}, \eta_{2i}, \dots, \eta_{m_0, i})'$ are generated as

$$\eta_{\ell i} \sim IIDN\left(0, \frac{\kappa^2}{m_0}\right), \quad \ell = 1, 2, \dots, m_0. \quad (77)$$

We have scaled the variance of $\eta_{\ell i}$, $\sigma_{\eta_\ell}^2$, by $1/m_0$ to ensure that the relative importance of the factor component of ζ_{it} is not affected by the choice of m_0 . We also consider the case where $m_0 = 0$ for which we set $Var(\eta_{\ell i}) = 0$ for all ℓ . The strength of the factors is controlled by the parameter κ^2 .

The idiosyncratic errors, u_{it} , for $t = 0, 1, \dots, T$ and $i = 1, 2, \dots, N$ are generated as $u_{it} \sim IID_{\sqrt{12}}^{-\sigma}(\chi_6^2 - 6)$ where χ_6^2 is a chi-square variate with six degrees of freedom. The regressors, x_{it} , for $i = 1, 2, \dots, N$ are generated as

$$x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} + v_{it}, \quad v_{it} = \rho_x v_{i,t-1} + (1 - \rho_x^2)^{1/2} \varepsilon_{it}, \quad \text{for } t = 1, 2, \dots, T, \quad (78)$$

with $\rho_x = 0.95$, and $\varepsilon_{it} \sim IIDN(0, \sigma_{\varepsilon}^2)$. We set m_x at $m_x = 2$, but consider different values of m_0 . In this way we allow for interactive effects in the $\{x_{it}\}$ processes for all values of m_0 , including when $m_0 = 0$. We draw v_{i0} from the steady state distribution of v_{it} , namely $v_{i0} \sim IIDN(0, \sigma_{v_i}^2)$, for $i = 1, 2, \dots, N$. This in turn ensures that $Var(v_{it}) = \sigma_{v_i}^2$. These error variances are drawn as $\sigma_{v_i}^2 \sim IID \frac{1}{4}(\chi_2^2 + 2)\sigma_v^2$, thus ensuring that $E(\sigma_{v_i}^2) = \sigma_v^2$. The factor loadings in the x_{it} equations, $\vartheta_{\ell i}$, are generated as $\vartheta_{\ell i} \sim IIDN(0, \sigma_{\vartheta_\ell}^2)$, for $\ell = 1, 2, \dots, m_x$. To establish that the fit of the model is not affected by the number of factors (m_0 and m_x) in what follows we set $\sigma_{\vartheta_\ell}^2 = \sigma_v^2/m_x$, for all ℓ . Finally, we set $\alpha_{xi} = \alpha_i + v_i$,

¹⁷Note that $\Pr(\sum_{i=1}^{r_0} w_i z_i^2 < c_N^2) \leq \Pr(w_i \sum_{i=1}^{r_0} z_i^2 < c_N^2) = \Pr(\sum_{i=1}^{r_0} z_i^2 < (c_N^2/w_i))$ and for $c_N^2 \rightarrow \infty$ as $N \rightarrow \infty$, $\lim_{N \rightarrow \infty} \Pr(\sum_{i=1}^{r_0} w_i z_i^2 < c_N^2) = 0$. Hence, using the critical values of $\sum_{i=1}^{r_0} z_i^2$ instead of $\sum_{i=1}^{r_0} w_i z_i^2 < c_N^2$ will still deliver a consistent estimator of m .

where $v_i \sim IIDN(0, 1)$, for all i . This specification ensures that the fixed effects, α_i , are correlated with the regressors, x_{it} .

We generate the time effects, δ_t , and unobserved common factors, $f_{\ell t}$, as $\delta_t = \frac{1}{2}(t^2 - t)$, for $t = 1, 2, \dots, T$, and

$$f_{\ell t} = \rho_{\ell f} f_{\ell, t-1} + (1 - \rho_{\ell f}^2)^{1/2} \varepsilon_{f\ell t}, \quad \varepsilon_{f\ell t} \sim IIDN(0, 1), \text{ for } \ell = 1, 2, \dots, m_0, \text{ and } t = 1, 2, \dots, T, \quad (79)$$

with $\rho_{f\ell} = \rho_f = 0.5$, and $f_{\ell, 0} = 0$ for $\ell = 1, 2, \dots, m_0$. Setting the initial values of $f_{\ell t}$ to zero is not restrictive since any non-zero sample means for the $f'_{\ell t}$ s would be absorbed by the values of the fixed effects, α_i , and the estimation results would be invariant to the choice of $f_{\ell, 0}$.

To investigate the performance of our proposed estimator and its robustness to the relative importance of the common factors in the generation of y_{it} , we calibrate the variance of x_{it} relative to the regression noise, ζ_{it} , as well as the variance of the factors $\boldsymbol{\eta}'_i \mathbf{f}_t$ to the idiosyncratic components, u_{it} . More specifically we consider the following ratios

$$\lambda_{f, NT} = \frac{N^{-1} \sum_{i=1}^N \boldsymbol{\eta}'_i \left(T^{-1} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}'_t \right) \boldsymbol{\eta}_i}{N^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^N u_{it}^2}, \quad (80)$$

$$\lambda_{x, NT} = \frac{N^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^N (x_{it} - \alpha_{xi})^2}{N^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^N \zeta_{it}^2}, \quad (81)$$

and to simplify the derivations we re-scale the values of the factors such that they are orthonormalised, namely

$$T^{-1} \sum_{t=1}^T f_{\ell t} = 0, \quad T^{-1} \sum_{t=1}^T f_{\ell t}^2 = 1, \quad T^{-1} \sum_{t=1}^T f_{\ell t} f_{\ell' t} = 0, \text{ for all } \ell \text{ and } \ell \neq \ell'. \quad (82)$$

Under the above scaling and using (77) we have (for any finite T) and as $N \rightarrow \infty$

$$\lambda_f = \lim_{N \rightarrow \infty} \kappa_{f, NT} = \frac{E(\boldsymbol{\eta}'_i \boldsymbol{\eta}_i)}{\sigma^2} = \frac{\kappa^2}{\sigma^2}. \quad (83)$$

Similarly, using (78) and (76b) we have

$$\begin{aligned} \lambda_x &= \frac{\lim_{N \rightarrow \infty} \left[N^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^N (\sum_{\ell=1}^{m_x} v_{i\ell} f_{\ell t} + v_{it})^2 \right]}{\lim_{N \rightarrow \infty} \left[N^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^N (\boldsymbol{\eta}'_i \mathbf{f}_t + u_{it})^2 \right]} \\ &= \frac{2\sigma_v^2}{\kappa^2 + \sigma^2} = \frac{2\sigma_v^2/\sigma^2}{1 + \kappa^2/\sigma^2}. \end{aligned} \quad (84)$$

To control the ratios λ_f and λ_x , without loss of generality, we set $\sigma^2 = 1$, and consider the values of $\kappa^2 = \{1/4, 1/2, 1, 2\}$ and $\sigma_v^2 = \{1/2, 1, 3/2\}$. These combinations allow us to examine the extent to which the small sample results are dependent on κ^2 and σ_v^2 that measure the relative importance of the unobserved common factors, \mathbf{f}_t , and the idiosyncratic components of x_{it} .

To set the initial values, $\{y_{i0}; i = 1, 2, \dots, N\}$, we distinguish between the case where $|\gamma| < 1$, and the unit-root case where $\gamma = 1$. Under the former, for each i , we generate y_{i0} from the steady state distribution of $\{y_{it}\}$, and set¹⁸

$$y_{i0} = \mu_{i0} + \sigma_{i0} (u_{i0}/\sigma), \text{ for } i = 1, 2, \dots, N \quad (85)$$

where

$$\mu_{i0} = \frac{\alpha_i + \beta \alpha_{xi}}{1 - \gamma}, \quad \sigma_{i0}^2 = \frac{\sigma^2 + a_x \beta^2 \sigma_{iv}^2 + a_f a_i}{1 - \gamma^2}, \quad (86)$$

¹⁸For the derivation of μ_{i0} and σ_{i0} see Section S.4 of the online supplement.

$$a_x = \frac{1 + \gamma\rho_x}{1 - \gamma\rho_x}, a_f = \frac{1 + \gamma\rho_f}{1 - \gamma\rho_f}, \quad (87)$$

$$a_i = \sum_{\ell=1}^{m_0} \eta_{\ell i}^2 + \beta^2 \sum_{\ell=1}^{m_x} \vartheta_{\ell i}^2 + 2\beta \sum_{\ell=1}^{\min(m_0, m_x)} \eta_{\ell i} \vartheta_{\ell i}, \quad (88)$$

and u_{i0} is generated as above. To check the robustness of our MC analysis to the choice of the initial values, we also consider generating y_{it} with μ_{i0} and σ_{i0} in (85) replaced by $\kappa_1\mu_{i0}$ and $\kappa_2\sigma_{i0}$ and experiment with the values of $\kappa_1, \kappa_2 = 1.2, 0.8$. For the remaining parameters we consider $\beta = 0$ (the pure autoregressive case) and $\beta = 1$, and experiment with medium and high values of γ , namely $\gamma = 0.4$ and 0.8 .

In the unit root case ($\gamma = 1$) we avoid incidental parameters in first differences by first generating the first-differences and then cumulating them to obtain y_{it} from some arbitrary values for y_{i0} . The first-differences are generated as

$$\Delta y_{i1} = \Delta \delta_1 + \beta \Delta x_{i1} + \Delta \zeta_{i1}, \quad (89)$$

$$\Delta y_{it} = \Delta \delta_t + \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta \zeta_{it}, \quad t = 2, 3, \dots, T, \quad (90)$$

with $\Delta y_{i0} = 0$, for $i = 1, 2, \dots, N$. The regressors and error processes are generated as above.

8.2 Monte Carlo results

We begin by reporting on the performance of the sequential MTLR procedure for estimating m_0 , the true number of latent factors. We then report on the bias and root mean square error (RMSE) of the TQML estimator of the parameters (γ and β), as well as size and power using the number of factors estimated by the MTLR procedure. Throughout we consider the parameter choices $\gamma_0 = \{0.4, 0.8\}$ and $\beta_0 = 1$, the sample size configurations $T = \{5, 10\}$ and $N = \{100, 300, 500, 1000\}$, and values of $m_0 = \{0, 1, 2\}$. Thereafter, we provide results comparing the TQML estimator with the QML estimator of Bai (2013), which we denote by Bai-QML, and separately with the GMM quasi-difference (QD) and first-difference (FD) estimator of ALS (where the latter takes the first-difference prior to applying the quasi-difference approach by Ahn et al. (2013)), assuming m_0 is known.¹⁹ Finally, we turn to the unit root case ($\gamma_0 = 1$), and end with a summary discussion of the main results from our robustness analysis. In the paper we focus on the baseline case where $\kappa^2 = \sigma_v^2 = 1$; results for other values of $\kappa^2 = \{1/4, 1/2, 2\}$ and $\sigma_v^2 = \{1/2, 3/2\}$ are provided in the online supplement and are discussed only briefly to save space. Further, we only report results for non-Gaussian errors. The results for the case of Gaussian errors are available upon request.

All panel regressions related to the TQML approach are estimated including both individual and time effects as well as an intercept, and regressors (in the case of the ARX(1) model), associated with Δy_{i1} , as in (24). Time effects are explicitly included in the regressions for the Bai-QML estimator while for the GMM regressions deviation from cross section averages is taken prior to estimation to remove the time effects; for both these set of regressions the individual effects are subsumed within the interactive effects. For further details related to the computation of the quasi-log-likelihood for Bai-QML and the GMM estimators see Sections S.5 and S.6 respectively of the online supplement. Not surprisingly, the conditional QML estimator of Bai (2013) did not perform well given that under our MC design y_{i0} depends on the model's unknown parameters, and therefore is not included.

Unless otherwise stated, the sequential MTLR procedure is implemented using the $\mathcal{LR}_N(m_{\max}, m_0)$ statistic for testing $m = m_0 = \{0, 1, 2, \dots, T - 3\}$ against $m = m_{\max} = T - 2$, with significance level $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$, using the critical values of the chi-square distribution with degrees of freedom as given by (67). The standard errors used for inference are based on equation (66) with all derivatives computed numerically. All tests are carried out at the 5% significance level and all experiments are replicated 2000 times.

¹⁹To simplify the comparisons we thought it more instructive to base our comparisons assuming that m_0 is known. Also, as will be seen, under our approach m is generally well estimated.

8.2.1 Selecting the number of factors

Table 1 reports the number of times (in %) that the estimated number of factors, \hat{m} , is equal to the true number of factors, m_0 , following the sequential MTLR procedure outlined in Section 7.1. The results refer to the baseline case where $\kappa^2 = \sigma_v^2 = 1$ and show that \hat{m} performs well for most parameter values and sample sizes. Even when $N = 100$, the true number of factors is estimated quite precisely except for the ARX(1) panel data model when $T = 5$ and $m_0 = 2$. However, by the time N reaches 300 the probability of selecting the true number of factors approaches 100%, across all parameter values. The results for other values of κ^2 and σ_v^2 are given in Tables A1(i) and A1(ii) in the online supplement. As to be expected, the empirical frequency of correctly selecting m_0 declines as the value of κ^2 (which measures the strength of the factors relative to the idiosyncratic error) is reduced for small N . However, as N increases the probability of selecting the true number of factors improves and approaches 100%, as to be expected given the consistency of the proposed procedure. Table A1(ii) further shows that the performance of \hat{m} is not that much affected as other values of σ_v^2 are considered.

8.2.2 Performance of the TQML estimator

We next consider the small sample performance of the TQML estimator of γ and β , after estimating m by the sequential MTLR procedure.

AR(1) For this panel data model, bias, RMSE, and empirical size for the TQML estimator of γ are reported in Table 2. The overall performance of the bias and RMSE is favourable with a few exceptions when $T = 5$, $N \leq 100$ and $m_0 = 2$. Specifically when $\gamma_0 = 0.4$, we need N larger than 100, particularly if $m_0 = 2$. The bias and size distortions are more serious when $\gamma_0 = 0.8$, and much larger sample sizes are required. However, as predicted by the asymptotic theory, the results improve as N increases. The performance of the TQML estimator improves considerably as T is increased to $T = 10$, and evidence of size distortions is limited to a few cases where $m_0 = 0$ and $\gamma_0 = 0.8$, and $N \leq 300$. The results for all combinations of $\kappa^2 = \{1/4, 1/2, 1, 2\}$ and $\sigma_v^2 = \{1/2, 1, 3/2\}$ are reported in Tables A2(i) and A2(ii) in the online supplement. As with the estimation of m discussed above, the performance of the TQML estimator deteriorates as κ^2 is reduced towards zero, and large sample sizes (N and/or T) are required for satisfactory outcomes in the case of the AR(1) specification. The power functions in Figure 1 show that overall the power is satisfactory. While power is low when $\gamma_0 = 0.8$ for small N , it improves as N increases. Power functions across alternative values of κ^2 are shown in Figures A3(i), A3(iv) and A3(vii) in the online supplement. The shape of these functions becomes quite distorted if the factors are very weak relative to the signal (namely for small values of κ^2), particularly when $T = 5$ and $\gamma_0 = 0.8$, or $\gamma_0 = 0.4$ and $m_0 = 2$.

ARX(1) Simulation results for the ARX(1) panel data model are provided in Table 3, and show the much better small sample performance as compared to the AR(1) model. This seems to be primarily due to the additional source of variations from the regressor. The bias and RMSE for the estimators of γ and β are both very small in all cases, and empirical sizes are also close to their nominal levels. In addition, as shown in Figure 2, power is reasonably high. From Table A2(iii) in the online supplement we also note that biases are very small across all values of κ^2 . As κ^2 reduces, the RMSE of γ increases while that of β decreases. Differences in RMSE across κ^2 for each of these parameters tends to decrease as N increases. Furthermore, Table A2(iv) shows that empirical sizes behave well across all values of κ^2 with only a couple of exceptions for $N = 100$ and smaller values of κ^2 . Power functions across the different values of κ^2 , as shown in Figures A3(ii)-A3(iii), A3(v)-A3(vi) and A3(viii)-A3(ix) of the online supplement, are similar to those of Figure 2 given below for $\kappa^2 = 1$. Results for the other values of σ_v^2 (namely 1/2 and 3/2) are very similar to those of $\sigma_v^2 = 1$, and are available upon request.

8.2.3 Comparison of TQML with alternative estimators

We begin by presenting results for the TQML and Bai-QML estimators followed by the GMM estimator proposed by ALS, for the AR(1) panel data model initially and for the ARX(1) subsequently. The GMM estimators we consider include the quasi-difference and first-difference ALS one step and two step estimators, denoted by QD1, QD2, FD1 and FD2, respectively.

For the comparison of TQML with the Bai-QML estimator, we provide results both for the *IID* specification of fixed effects (used in the Monte Carlo design of Section 8.1), namely $\alpha_i \sim IIDN(0, 1)$, as well as for spatially correlated fixed effects. Under the latter the $N \times 1$ vector of fixed effects, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)'$, is generated as the first-order spatial autoregressive process

$$\alpha = \rho_\alpha \mathbf{W}\alpha + \varepsilon_\alpha, \text{ or } \alpha = (\mathbf{I}_N - \rho_\alpha \mathbf{W})^{-1} \varepsilon_\alpha,$$

with heteroskedastic errors $\varepsilon_\alpha = (\varepsilon_{\alpha,1}, \varepsilon_{\alpha,2}, \dots, \varepsilon_{\alpha,N})'$, where $\rho_\alpha = 0.9$,

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 1/2 & 0 & 1/2 & 0 & & 0 \\ 0 & 1/2 & 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \ddots & 1/2 & 0 \\ \vdots & & & & 1/2 & 0 & 1/2 \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}, \quad (91)$$

and for each i , $\varepsilon_{\alpha,i}$ are drawn as $IIDN(0, \sigma_{\varepsilon_{\alpha,i}}^2)$ with

$$\sigma_{\varepsilon_{\alpha,i}}^2 = \begin{cases} 1, & \text{for } i = 1, 2, \dots, N/2, \\ 2, & \text{for } i = N/2 + 1, \dots, N. \end{cases}$$

The TQML estimator is fully robust to the way fixed effects for y_{it} and x_{it} are generated (random or correlated). Hence, the results for TQML under both the above fixed effect specifications are identical, and it does not matter which fixed effects specification is used.²⁰ Also, since the GMM estimators first eliminate the fixed effects (GMM based on QD does it implicitly) to save space for the GMM estimators we show results only under the IID specification of the fixed effects.

In line with the discussion in Section 2.3, for the Bai-QML estimator the Mundlak-Chamberlain projection of the regressor equation fixed effects, α_{xi} , (recall $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IIDN(0, 1)$, for all i) on \mathbf{x}_i is used to deal with the dependence of α_{xi} on α_i . In addition, the number of factors included in the regressions for the computation of the Bai-QML estimator is $\tilde{m}_0 = m_0 + 1$, and following Bai (2013) we use the factor normalisation $\mathbf{F}^+ = (\mathbf{I}_{\tilde{m}}, \mathbf{F}'_2)'$. The same number of factors is used for the QD GMM estimators given that the individual effects are subsumed within the interactive effects, while m_0 is used for the FD GMM estimators that employ prior first-differencing; the same normalisation is also used on the factor matrix for these estimators.

AR(1): TQML and Bai-QML Table 4 reports the bias and RMSE of γ for the TQML and Bai-QML estimators in the case of the AR(1) panel data model. Results show that the Bai-QML estimator performs very poorly in terms of bias and RMSE for both values of γ_0 . The same is true with regard to size as seen from the results for the two estimators summarised in Table 5, which show the Bai-QML estimator to have severe size distortions. The poor performance of the Bai-QML estimator in terms of bias, RMSE and size is, on the whole, more pronounced when the fixed effects are spatially correlated

²⁰The use of the Mundlak-Chamberlain projection for the Bai-QML estimator helps in the present MC design because α_{ix} for $i = 1, 2, \dots, N$ are generated as a linear function of α_i . It would not have helped if the α_{ix} were generated as a general (for example quadratic) function of α_i . The TQML estimator is not affected even if α_i and α_{ix} are non-linearly related.

compared to the *IID* case. In contrast, the TQML estimator performs well, requiring larger values of N only when $\{T = 5, \gamma_0 = 0.8\}$ to reduce the mild biases observed for this parameter combination. The small size distortion of TQML in the case of $\{N = 100, T = 5, \gamma_0 = 0.8\}$, vanish when larger values of $N = \{300, 500\}$ are considered. Power functions for $\{T = 5, N = 500\}$, are shown in Figures 3a and 3b for the *IID* and spatially correlated fixed effects, respectively. The TQML estimator shows satisfactory power as the distance from the null hypothesis increases, with the power curves exhibiting slight asymmetry for the case of $\gamma_0 = 0.8$. While power appears higher for the Bai-QML estimator, this is accompanied by evidence of large size distortions which are higher for the case of $\gamma_0 = 0.8$, and even more so when the fixed effects are spatially correlated as compared to the *IID* results.

AR(1): TQML and GMM In comparing the TQML and GMM estimators, results for the AR(1) panel data model are only reported for $T = 10$ as the GMM estimators are not computable for the case of $T = 5$ due to failure of the order condition. Results in Table 6 show that the TQML estimator performs substantially better than the GMM estimators in terms of bias and RMSE. When $\gamma = 0.8$, the GMM estimators, especially FD1 and FD2, perform very poorly possibly due to weak instruments, whereas the TQML estimator has small bias and RMSE. With regard to size shown in Table 7, the GMM estimators display substantial size distortions while the TQML estimator has empirical size close to the nominal value, except for the case where $\gamma = 0.8$ and $N = 100$.

ARX(1): TQML and Bai-QML Bias and RMSE of γ and β for the ARX(1) panel data model are given in Table 8. While the bias is generally small for the TQML estimator for γ and β across all parameter combinations, for the Bai-QML estimator these are larger, and much more so in the case of spatially correlated fixed effects. The same holds for the RMSE. Empirical size is reported in Table 9, which shows the TQML estimator to have little size distortions for all parameter configurations, even for $T = 5$ and $N = 100$. In contrast, the performance of the Bai-QML estimator varies considerably depending on the parameter values, the number of latent factors, and the way fixed effects are generated. For example, Bai-QML shows little size distortion when $T = 5$, $N > 100$ and $\gamma_0 = 0.4$. But significant size distortions occur when $\gamma_0 = 0.8$, and the extent of these become more pronounced under spatially correlated fixed effects, and as m_0 is increased to 2. Power functions for $T = 5$ and $N = 500$ are shown in Figures 4 and 5 for the *IID* and spatially correlated fixed effects, respectively. These figures show that under *IID* fixed effects the Bai-QML estimator for γ exhibits similar power performance compared to the TQML estimator when $\gamma_0 = 0.4$, and moderately lower power for $\gamma_0 = 0.8$. For β , when $m_0 = 1$ across both values of γ_0 , the Bai-QML estimator continues to show lower power as compared to TQML, which becomes extremely lower in the case of $m_0 = 2$. The picture is qualitatively similar for spatially correlated fixed effects, however with more pronounced power discrepancies. For $\gamma_0 = 0.8$ the large size distortions for γ and β do not allow for a meaningful power comparison of the two estimators. The power performance of the TQML estimator is satisfactory across all parameter combinations.

ARX(1): TQML and GMM Table 10 reports the bias and RMSE of γ and β for the TQML and GMM estimators and shows that the TQML estimator has better small sample properties both in terms of bias and RMSE. The same also follows if we consider the size of the tests based on these estimators summarised in Table 11. For the GMM estimators, the performance crucially depends on the specific values of γ_0 , m_0 , N and T , and there is no GMM estimator that performs well for all combinations, which is in contrast to the TQML estimator that performs well for all cases considered. For instance, when $T = 5$, FD1 and FD2 tend to have correct empirical sizes when N is large. However, they tend to have large size distortions when T is increased to $T = 10$ for $m_0 = 1$. QD2 and FD2 tend to have larger size distortions than QD1 and FD1. This is partly due to the downward bias of the standard errors used in

the two-step estimators.²¹

8.2.4 The unit root case

The results for the unit root case are very similar to those already reported for the stationary case. Table 12 reports the number of factors correctly selected (in %) by the sequential MTLR procedure when $\gamma_0 = 1$. As can be seen, the results are uniformly good for all values of m_0 , N and T . Also the effects of deviating from the baseline values of κ^2 and σ_v^2 on the empirical frequency of correctly selecting the true number of factors are similar to the stationary case. See Tables B1(i) and B1(ii) in the online supplement. The results for bias, RMSE and size of the TQML estimator when $\gamma_0 = 1$ are summarised in Tables 13 and 14 for the AR(1) and ARX(1) panel data models, respectively. These show that the bias and RMSE are reasonably small, and the empirical size for γ is slightly below the nominal value. The effects of deviating from the baseline value of κ^2 are reported in Tables B2(i) and B2(ii) of the online supplement, and show that the bias and RMSE become smaller as the value of κ^2 is reduced, which is different from the stationary case. Power is also reasonably high as shown in Figures 6 and 7 for the AR(1) and ARX(1) panel data models, respectively, when $\kappa^2 = 1$. The power plots for other values of κ^2 , namely $\{1/4, 1/2, 2\}$, are given in Figures B3(i), B3(iv), and B3(vii) of the online supplement for the AR(1) model, and Figures B3(ii)-B3(iii), B3(v)-B3(vi) and B3(viii)-B3(ix) for the ARX(1).

8.2.5 Robustness of baseline MC results

Lastly we investigate the performance of our selection and estimation strategy under a number of deviations from the baseline model. Specifically, we consider the following scenarios: (i) initial values that deviate from the steady state distribution, whereby y_{i0} is generated as in (85) but with means and variances given by $\kappa_1\mu_{i0}$ and $\kappa_2\sigma_{i0}$, with $\kappa_1, \kappa_2 = 1.2, 0.8$; (ii) implementing the sequential MTLR procedure with different p -values, namely $p = \{0.01, 0.10\}$, instead of our baseline value of $p = 0.05$; (iii) factor loadings that are correlated with the regressors; and (iv) factor loadings that are mutually weakly correlated. Further details on the data generating process for the last two cases and related results can be found in Section S.9 of the online supplement.

As shown in Tables C1(i)-C1(iii) of the online supplement, deviating the initial values from those of the steady state distribution has only a limited effect on the results with the performance of our estimator remaining reasonably good overall. The only effect observed is for the AR(1) panel data model for which size distortions are slightly more pronounced for $T = 5$, $\gamma_0 = 0.8$ and $N \leq 500$ as compared to the case where y_{i0} are drawn from the steady state distribution. For the rest of the results, including those of the ARX(1) model bias and RMSE values are still reasonably small with empirical sizes close to their nominal value across all parameter configurations.

Regarding the use of alternative values of p in implementing the MTLR test, as can be seen from Tables C2(i)-C2(iii) for $p = 0.01$ and Tables C2(iv)-C2(vi) for $p = 0.10$, the results are very similar and in some cases even better than those obtained in Tables 1-3 for $p = 0.05$.

When the factor loadings are correlated with the regressor, from Tables C3(i)-C3(iii) of the online supplement, we find that the sequential MTLR procedure estimates the number of factors very precisely across all parameters, the bias is sufficiently small, and empirical size is close to the nominal level, with one exception, namely, when $N = 100$, $T = 5$ and $\gamma_0 = 0.8$ for the AR(1) model. When the factor loadings are weakly correlated, as shown in Tables C4(i)-C4(iii) in the online supplement, the results are very similar to those in Tables 1-3 where such correlation is absent. The same also applies if we consider the estimates for the ARX(1) model.

²¹Since both QD2 and FD2 are nonlinear GMM estimators, it is not straightforward to apply the Windmeijer (2005) correction.

9 Empirical illustrations

We investigate the importance of allowing for interactive effects in empirical analysis by applying our selection and estimation strategy to two empirical problems addressed in the literature. In the first illustration we estimate a dynamic version of the model considered by Cornwell and Trumbull (1994) and subsequently by Baltagi (2006), to explain the incidence of crime across $N = 90$ counties in North Carolina over the period 1981-1987 ($T = 6$). In the second illustration, we use the data set recently analysed by Acemoglu et al. (2019) to estimate output regressions on a balanced panel of $N = 82$ countries using $T = 5$ five-year time intervals over the period 1981-2005. All regressions include both individual and time effects, plus the regressors associated with the initial observation of the dependent variable in first differences. The presence of interactive effects is investigated by first estimating m , the number of unobserved factors, subject to $m_{\max} = T - 2$. Results are presented for the parameters of interest, namely the coefficient of the lagged dependent variable and the regressors; estimates for the remaining parameters (such as time effects) are available upon request.

9.1 Cross county crime rate regressions

The crime rate in county i , year t (y_{it}) is explained by the deterrent variables, namely the probability of arrest ($P_{it,A}$), the probability of conviction given arrest ($P_{it,C}$), the probability of a prison sentence given a conviction ($P_{it,P}$), average prison sentence in days (S_{it}), and a number of other variables such as population density ($Density_{it}$), percent young male (YM_{it}), the wage rates in manufacturing (WMF_{it}), and the wage rate in transportation, utilities and communication industries ($WTUC_{it}$).²² The panel regressions estimated by Cornwell and Trumbull (1994) and Baltagi (2006) are static and could be misspecified since jurisdictions with high crime rates in one year are likely to continue to have high crime rates into the near future. By including lagged crime rates ($y_{i,t-1}$) in the model we account for the possible persistence of crime rates over time, and by allowing for unobserved common effects we take account of possible persistence and spill-over effects of crimes across counties.

To investigate the importance of the interactive effects we first estimated m (the number of latent factors) using the proposed sequential MTLR procedure, with the nominal value of the test, p , set to 5%, and the maximum number of factors, $m_{\max} = T - 2 = 4$ (see Section 7.1). We obtain $\hat{m} = 3$ and reject the null hypothesis that the panel regressions are not subject to interactive effects, despite the fact that they include country and year fixed effects. The estimate of m is reasonably robust to the choice of p values and we obtain the same estimate ($\hat{m} = 3$) if we set $p = 10\%$, although setting $p = 1\%$ yields $\hat{m} = 2$. In Table 15 we report the results for $\hat{m} = 3$, along with the estimates without interactive effects (with $m = 0$). We first note that irrespective of whether we allow for interactive effects or not, there is clear evidence of dynamics and the coefficient of the lagged crime rate is highly significant, even though when we allow for interactive effects this coefficient falls from 0.501 to 0.402, but remains highly significant. Amongst the $\mathbf{x}_{it} = (P_{it,A}, P_{it,C}, P_{it,P}, S_{it}, Density_{it}, YM_{it}, WMF_{it}, WTUC_{it})'$ variables, only the deterrent variables and the wage rate in manufacturing are statistically significant once we allow for interactive effects. The results are similar when we do not allow for interactive effects, with the exception of the $WTUC_{it}$ variable which is marginally significant when $m = 0$. It is also worth noting that all the estimated coefficients that are statistically significant have the correct signs when $\hat{m} = 3$.

²²Cornwell and Trumbull (1994) and Baltagi (2006) consider a number of other variables such as wage rates in other industries and the number of police, which we exclude to simplify the exposition.

Table 15: Dynamic panel estimates of crime rates (y_{it}) across 90 counties in North Carolina over the period 1981-1987
($T = 6, N = 90$)

Explanatory Variables ($y_{i,t-1}, \mathbf{x}_{it}$)	$\hat{m} = 3$	$m = 0$
Lagged crime rate ($y_{i,t-1}$)	0.402*** (0.108)	0.501*** (0.086)
Probability of arrest ($P_{it,A}$)	-0.301*** (0.072)	-0.221*** (0.070)
Probability of conviction given arrest ($P_{it,C}$)	-0.193*** (0.032)	-0.147*** (0.055)
Probability of prison given conviction ($P_{it,P}$)	-0.154*** (0.042)	-0.137*** (0.051)
Severity of punishment (S_{it})	-0.093*** (0.035)	-0.130*** (0.048)
Population density ($Density_{it}$)	0.172 (0.459)	0.148 (0.430)
Wage: transportation, utilities & communication ($WTUC_{it}$)	0.016 (0.019)	0.033* (0.019)
Wage: manufacturing ($WMFG_{it}$)	-0.563*** (0.158)	-0.431*** (0.105)
Percent young male (YM_{it})	0.839 (0.694)	0.601 (0.664)

Note: The estimates allow for county and year fixed effects. T is the number of time periods used in TQML estimation after first differencing. \hat{m} is the latent factors estimated using the sequential MTLR procedure described in Section 7.1 with $m_{\max} = T - 2 = 4$ and $\alpha_N = 0.05/(N(T - 2))$. Figures in parentheses are standard errors that are computed according to equation (66). ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

9.2 Cross country growth regressions

There is a large empirical literature on cross country growth regressions, using cross section as well as panel data sets. Examples include Barro (1991), Mankiw et al. (1992), Sala-i-Martin (1996), Islam (1995), Caselli et al. (1996) and Lee et al. (1997, 1998). Our application is closest to the panel regressions by Islam (1995) and Caselli et al. (1996) who estimate dynamic panel regressions with time and fixed effects using log GDP per capita at five-year time intervals. A similar approach is also used by Acemoglu et al. (2019) who focus on the effect of democracy on GDP per capita. However, none of these studies allow for interactive effects. In our empirical application we regress log GDP per capita (y_{it}) measured over five-year intervals on $y_{i,t-1}$, log investment-output ratio, log total factor productivity (TFP), log trade share in GDP, log infant mortality, and a dichotomous democracy variable. As noted above, the data set used covers $N = 82$ countries with $T = 5$ five-yearly periods spanning 1981-2005.²³

For this illustration the number of latent factors (m) was estimated to be $\hat{m} = 2$, using the sequential MTLR procedure with $p = 5\%$ and $m_{\max} = T - 2 = 3$. The same result was obtained setting $p = 1\%$ and 10%. The parameter estimates together with their standard errors for $\hat{m} = 2$ and $m = 0$ are summarised in Table 16. As can be seen, allowing for interactive effects substantially lowers the degree of output persistence from 0.583 to 0.246, raises the coefficient of log TFP from 0.547 to 0.870, and increases the size and significance of the coefficient of infant mortality on output from -0.042 (and not significant) to -0.075 (and highly significant). The negative and significant effect of infant mortality on GDP is also found in similar growth regressions by Somé et al. (2019). They explore the impact of healthcare on economic growth in Africa, but do not allow for error cross-sectional dependence in their analysis. The trade share and democracy variables both have a positive sign though are found to be insignificant. The

²³For further information on the data and related sources see Acemoglu et al. (2019).

latter finding is in line with recent results by Jacob and Osang (2018) who perform a dynamic panel analysis using GMM for a sample of more than 160 countries based on $T = 10$ five year averages. In contrast Acemoglu et al. (2019) find that democracy does cause GDP using an annual panel data of $T = 50$ observations without allowing for interactive effects. The only parameter estimate which has not been affected by the inclusion of interactive effects is the coefficient of the investment-output ratio, which is estimated at 0.078 when $m = 0$ as compared to 0.071 when $\hat{m} = 2$.

The empirical illustrations provided suggest that allowing for error cross-sectional dependence in dynamic panels could be important and ought to be considered in applied research.

Table 16: Dynamic panel regressions for cross country log per capita output equations (y_{it}) (1981-2005, five yearly $T = 5, N = 82$)

Explanatory Variables	$\hat{m} = 2$	$m = 0$
Lagged log GDP per capita ($y_{i,t-1}$)	0.246*** (0.063)	0.583*** (0.042)
Log investment output ratio (INV_{it})	0.071*** (0.014)	0.078*** (0.018)
Log total factor productivity (TFP_{it})	0.870*** (0.051)	0.547*** (0.059)
Log trade share in GDP ($Trade_{it}$)	0.010 (0.019)	0.047** (0.021)
Log infant mortality	-0.075*** (0.029)	-0.042 (0.027)
Democracy indicator	0.012 (0.014)	0.008 (0.017)

Note: \hat{m} is the estimated number of factors using the sequential MTLR procedure described in Section 7.1 with $m_{\max} = T - 2 = 3$ and $\alpha_N = 0.05/(N(T - 2))$. See also the note to Table 15.

10 Conclusion

This paper proposes a quasi maximum likelihood estimator for short dynamic panel data models with unobserved multiple common factors, where individual and time fixed effects are also explicitly included. This provides a natural extension of Hsiao et al. (2002) to panel data models with a multi-factor error structure. Our contribution can also be viewed as extending the standard dynamic panel data models with fixed and time effects, routinely used in the empirical literature, to allow for error cross sectional dependence through interactive effects.

We have also contributed to the literature on short T factor models with regard to identification and estimation of the number of unobserved factors, as well as parameter identification. Our proposed sequential multiple testing likelihood ratio (MTLR) procedure can be particularly relevant to the analysis of short T factor models. Monte Carlo results provide small sample evidence in support of the proposed TQML estimator and show that the sequential MTLR procedure performs very well in selecting the number of unobserved factors in most settings. The same is also true for the performance of the TQML estimator in terms of bias, RMSE and empirical size, and power. Empirical illustrations involving cross country crime and growth regressions suggest that allowing for interactive effects in dynamic panels could be important and ought to be considered in applied work.

Although we allow the error variances to vary across units through the differences in factor loadings, it is assumed that the unit specific errors are cross sectionally homoskedastic, which is rather restrictive. However, our theoretical derivations can be readily adapted to cover the heteroskedastic error case, as was done in the recent paper by Hayakawa and Pesaran (2015) for models without unobserved common factors. It would also be interesting to extend the analysis to panel VAR models with interactive effects.

References

- Acemoglu, D. and Naidu, S. and Restrepo, P. and Robinson, J.A. (2019). Democracy does cause growth. *Journal of Political Economy* 127, 47–100.
- Ahn, S.C. (2015). Comment on "IV estimation of panels with factor residuals" by D. Robertson and V. Sarafidis. *Journal of Econometrics* 185, 542–544.
- Ahn, S.C. and Horenstein, A.R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203–1227.
- Ahn, S.C. and Lee, H.Y. and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101, 219–255.
- (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174, 1–14.
- Ahn, S.C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5–27.
- Allen, D., and Chan, F., and McAleer, M. and Peiris, M.S. (2008). Finite sample properties of the QMLE for the log-ACD model: Application to australian stocks. *Journal of Econometrics* 147, 163–185.
- Anderson, T.W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598–606.
- Arellano, M. and Bond, S. (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 Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29–51.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.
- (2013). Likelihood approach to dynamic panel models with interactive effects. mimeo.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- (2013). Principal components estimation and identification of static factors. *Journal of Econometrics* 176, 18–29.
- Bailey, N. and Pesaran, M.H. and Smith, L.V. (2019). A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics* 208, 507–534.
- Baltagi, B. (2006). Estimating an economic model of crime using panel data from north carolina. *Journal of Applied Econometrics* 21, 543–547.
- Barro, R.J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics* 106, 407–443.
- Bauer, P., Pötscher, B.M. and Hackl, P. (1988). Model selection by multiple test procedures. *Statistics* 19, 39–44.
- Bhargava, A. and Sargan, J. (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635–1659.

- Binder, M. and Hsiao, C. and Pesaran, M.H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795–837.
- Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.
- Bonferroni, C.E. (1936). Statistical theory of classes and calculation of probabilities. *Pub R Ist Superiore Sci Econ Commere Firenze* 8, 36–62.
- Caselli, F. and Esquivel, G. and Lefort, F. (1996). Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth* 1, 363–389.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.
- (1984). Panel data. In *Handbook of Econometrics*, Volume 2, Chapter 22, pp. 1248–1318. North-Holland.
- Chudik, A. and Kapetanios, G. and Pesaran, M.H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica* 86, 1479–1512.
- Chudik, A. and Pesaran, M.H. (2021). An augmented Anderson-Hsiao estimator for dynamic short-T panels. *Mimeo*.
- Chudik, A. and Pesaran, M.H. and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal* 14, C45–C90.
- Cornwell, C. and Trumbull, W.N. (1994). Estimating the economic model of crime with panel data. *The Review of Economics and Statistics* 76, 360–366.
- Cragg, J.G. and Donald, S.G. (1997). Inferring the rank of a matrix. *Journal of Econometrics* 76, 223–250.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.
- (2000). *Econometric Theory*. Wiley-Blackwell.
- Han, H. and Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Hayakawa, K. (2012). GMM estimation of a short dynamic panel data model with interactive fixed effects. *Journal of the Japan Statistical Society* 42, 109–123.
- (2016). Identification problem of GMM estimators for short panel data models with interactive fixed effects. *Economics Letters* 139, 22–26.
- Hayakawa, K. and Pesaran, M.H. (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics* 188, 111–134.
- Hayashi, K. and Bentler, P.M. and Yuan, K-H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal* 14, 505–526.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.

- Holtz-Eakin, D. and Newey, W.K. and Rosen, H.S. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56, 1371–1395.
- Hsiao, C. and Pesaran, M.H. and Tahmiscioglu, K.A. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107–150.
- Hsiao, C. and Tahmiscioglu, K.A. (2008). Estimation of dynamic panel data models with both individual and time specific effects. *Journal of Statistical Planning and Inference* 138, 2698–2721.
- Islam, N. (1995). Growth empirics: A panel data approach. *The Quarterly Journal of Economics* 110, 1127–1170.
- Jacob, J. and Osang, T. (2018). Democracy and growth: A dynamic panel data study. *The Singapore Economic Review*, 1–40.
- Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics* 28, 397–409.
- Kapetanios, G. and Pesaran, M.H. and Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics* 160, 326–348.
- Kristensen, D. and Rahbek, A.C. (2010). Likelihood-based inference for cointegration with nonlinear error-correction. *Journal of Econometrics* 158, 78–94.
- Lee, K. and Pesaran, M.H. and Smith, R.P. (1997). Growth and convergence in a multi-country empirical stochastic solow model. *Journal of Applied Econometrics* 12, 358–392.
- Lee, K. and Pesaran, M.H. and Smith, R. (1998). Growth empirics: A panel data approach—a comment. *The Quarterly Journal of Economics* 113, 319–323.
- Lee, L.-F. and Yu, J. (2016). Identification of spatial durbin panel models. *Journal of Applied Econometrics* 31, 133–162.
- Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.
- Lütkepohl, H. (1996). *Handbook of matrices*. New York: John Wiley.
- Magnus, J.R. and Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Oxford: John Wiley & Sons.
- Mankiw, N.G. and Romer, D. and Weil, D.N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics* 107, 407–437.
- Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics* 31, 451–472.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* 46, 69–85.
- Nauges, C. and Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique* 70, 53–75.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92, 1004–1016.
- Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with cross section dependence. *Econometrica* 74, 967–1012.

- Pesaran, M.H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182–202.
- Phillips, P.C.B. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 127, 162–188.
- Pötscher, B.M. and Prucha, I.R. (2001). Basic elements of asymptotic theory. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, pp. 201–229. Oxford: Blackwell.
- Robertson, D. and Sarafidis, V. (2015). IV estimation of panels with factor residuals. *Journal of Econometrics* 185, 526–541.
- Sala-i-Martin, X.X. (1996). The classical approach to convergence analysis. *The Economic Journal* 106, 1019–1036.
- Sarafidis, V. and Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal* 12, 62–81.
- Sarafidis, V. and Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31, 483–531.
- Somé, J. and Pasali, S. and Kaboine, M. (2019). Exploring the impact of healthcare on economic growth in Africa. *Applied Economics and Finance* 6, 45–57.
- White, H. (2001). *Asymptotic Theory for Econometricians*. San Diego: Academic Press.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25–51.

Tables and Figures for the Monte Carlo Results

Table 1: Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure ($\kappa^2 = \sigma_v^2 = 1$)

	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
m_0	0	1	2	0	1	2	0	1	2	0	1	2
N	AR(1)											
100	99.4	99.7	88.9	99.2	99.8	96.3	99.5	99.6	99.7	99.7	99.5	99.7
300	99.8	100.0	100.0	99.8	100.0	100.0	99.8	100.0	100.0	99.8	100.0	100.0
500	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.9	99.9	100.0
1000	99.9	100.0	100.0	99.9	100.0	100.0	99.7	100.0	100.0	99.6	100.0	100.0
	ARX(1)											
100	99.7	98.7	31.0	99.6	99.2	33.0	99.3	99.6	99.7	99.4	99.6	99.7
300	100.0	100.0	99.5	99.9	100.0	99.5	100.0	100.0	99.9	100.0	99.9	99.9
500	99.9	99.9	100.0	99.9	99.9	100.0	99.9	100.0	100.0	99.9	100.0	100.0
1000	99.9	99.9	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: y_{it} is generated as $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}$, $\zeta_{it} = \sum_{\ell=1}^{m_0} \eta_{\ell i} f_{\ell t} + u_{it} = \boldsymbol{\eta}'_i \mathbf{f}_t + u_{it}$, for $i = 1, 2, \dots, N$; $t = 1, \dots, T$, with $y_{i0} = \mu_{i0} + \sigma_{i0}(u_{i0}/\sigma)$ where $\mu_{i0} = (\alpha_i + \beta \alpha_{xi})/(1 - \gamma)$ and $\sigma_{i0}^2 = (\sigma^2 + a_x \beta^2 \sigma_v^2 + a_f a_i)/(1 - \gamma^2)$. In addition, $a_x = (1 + \gamma \rho_x)/(1 - \gamma \rho_x)$, $a_f = (1 + \gamma \rho_f)/(1 - \gamma \rho_f)$ and $a_i = \sum_{\ell=1}^{m_0} \eta_{\ell i}^2 + \beta^2 \sum_{\ell=1}^{m_x} \vartheta_{\ell i}^2 + 2\beta \sum_{\ell=1}^{\min(m_0, m_x)} \eta_{\ell i} \vartheta_{\ell i}$, where $\eta_{\ell i} \sim IIDN(0, \frac{\kappa^2}{m_0})$, $\ell = 1, 2, \dots, m_0$, $\vartheta_{\ell i} \sim IIDN(0, \sigma_{\vartheta \ell}^2)$, for $\ell = 1, 2, \dots, m_x$, with $\sigma_{\vartheta \ell}^2 = \sigma_v^2/m_x$, for all ℓ , $\rho_x = 0.95$, $m_x = 2$, and $\beta = 1$. The idiosyncratic errors are generated as $u_{it} \sim IID \frac{\sigma}{\sqrt{12}}(\chi_6^2 - 6)$ for $i = 1, 2, \dots, N$; $t = 0, 1, \dots, T$ where χ_6^2 is a chi-square variate with 6 degrees of freedom and $\sigma^2 = 1$. The fixed effects are generated as $\alpha_i \sim IIDN(0, 1)$. The regressors, x_{it} , for $i = 1, 2, \dots, N$ are generated as $x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} + v_{it}$, with $v_{it} = \rho_x v_{i,t-1} + (1 - \rho_x^2)^{1/2} \varepsilon_{it}$, for $t = 1, 2, \dots, T$, $\varepsilon_{it} \sim IIDN(0, \sigma_v^2)$, $v_{i0} \sim IIDN(0, \sigma_v^2)$, for $i = 1, 2, \dots, N$, with $\sigma_v^2 \sim IID \frac{1}{4}(\chi_2^2 + 2)\sigma_v^2$ and $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IIDN(0, 1)$, for all i . Each f_t is generated once and the same f_t 's are used throughout the replications. In the AR(1) case $\beta = 0$ and under $m_0 = 0$, ζ_{it} collapses to u_{it} .

Table 2: Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m} ($\kappa^2 = 1$)

	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
N	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)
$m_0 = 0$												
100	0.42	8.69	6.2	0.65	12.29	21.3	-0.03	3.76	6.5	1.94	7.90	16.4
300	-0.03	4.26	5.4	1.42	9.26	19.2	-0.04	2.18	5.1	0.68	4.62	8.7
500	0.03	3.22	4.8	1.46	7.80	14.6	-0.01	1.70	5.9	0.26	3.09	6.7
1000	0.00	2.29	4.5	1.02	6.07	12.1	-0.01	1.22	5.4	0.18	2.24	5.7
$m_0 = 1$												
100	0.41	9.39	5.1	1.42	12.99	19.6	-0.05	4.20	6.1	0.23	4.64	4.9
300	-0.09	4.99	5.1	1.00	9.04	11.9	0.02	2.38	4.5	0.08	2.41	4.7
500	0.05	3.68	3.9	0.96	7.12	7.1	-0.06	1.90	6.0	0.01	1.88	5.4
1000	0.04	2.67	4.7	0.61	5.08	4.7	-0.01	1.32	4.9	0.00	1.30	4.2
$m_0 = 2$												
100	4.09	16.38	11.5	1.82	16.38	19.8	-0.08	5.12	5.8	0.19	5.32	5.3
300	0.20	4.99	3.9	1.38	4.99	10.3	0.04	2.81	4.6	0.08	2.66	4.0
500	0.05	3.81	3.1	0.98	3.81	6.3	-0.10	2.16	4.9	-0.09	2.06	4.7
1000	0.02	2.62	3.3	0.45	2.62	4.4	0.00	1.59	4.7	0.01	1.44	4.0

See the note to Table 1.

Table 3: Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\kappa^2 = \sigma_v^2 = 1$)

		$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
		Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
γ													
N	$m_0 = 0$												
100		-0.15	3.45	5.9	-0.07	3.02	6.6	-0.06	1.95	5.4	-0.03	1.37	5.8
300		-0.04	1.97	5.6	-0.05	1.71	6.1	0.08	1.14	5.3	0.04	0.77	5.1
500		0.02	1.47	5.1	0.00	1.27	4.4	-0.01	0.86	4.5	0.00	0.58	4.3
1000		-0.05	1.08	5.1	-0.03	0.93	5.8	0.00	0.62	4.9	0.00	0.42	5.8
$m_0 = 1$													
100		0.09	4.30	5.1	0.23	4.74	5.2	-0.10	2.15	6.0	-0.07	1.54	6.5
300		-0.05	2.39	4.4	-0.02	2.56	5.1	0.03	1.20	5.2	0.02	0.83	4.0
500		0.01	1.83	3.8	0.02	1.92	3.9	-0.02	0.92	5.5	-0.01	0.65	5.1
1000		-0.04	1.35	4.5	-0.02	1.41	4.5	0.01	0.67	5.4	0.00	0.46	5.4
$m_0 = 2$													
100		0.37	4.70	5.8	0.47	4.99	4.7	-0.09	2.33	5.8	-0.05	1.59	5.9
300		0.03	2.46	4.1	0.07	2.63	4.8	-0.06	1.33	5.4	-0.02	0.91	4.8
500		0.07	1.94	3.6	0.10	2.10	4.6	-0.03	0.98	4.3	-0.01	0.69	4.7
1000		0.05	1.39	3.6	0.05	1.47	4.2	0.02	0.70	4.3	0.01	0.48	4.1
β													
$m_0 = 0$													
100		-0.06	4.44	5.6	-0.06	4.55	5.4	-0.01	3.04	6.5	-0.02	3.02	6.6
300		0.02	2.53	5.7	0.01	2.58	5.8	-0.05	1.73	6.0	-0.03	1.71	6.0
500		0.04	1.92	5.2	0.04	1.97	5.2	0.00	1.34	5.7	0.00	1.33	5.6
1000		0.00	1.38	5.0	0.00	1.40	4.9	0.01	0.96	5.6	0.01	0.95	5.8
$m_0 = 1$													
100		-0.01	5.99	5.6	0.06	6.16	5.5	0.09	3.98	6.3	0.07	3.98	6.2
300		-0.15	3.39	4.9	-0.14	3.46	4.9	0.01	2.29	6.0	0.02	2.28	5.6
500		0.09	2.65	5.5	0.09	2.70	5.3	0.00	1.74	5.2	0.00	1.72	5.2
1000		0.05	1.88	5.5	0.06	1.91	5.7	0.03	1.21	4.4	0.04	1.20	4.7
$m_0 = 2$													
100		0.27	8.33	6.5	0.41	8.56	5.8	0.15	6.27	4.9	0.13	6.24	5.0
300		0.18	4.62	5.2	0.20	4.67	5.3	0.09	3.63	5.3	0.08	3.61	5.4
500		0.11	3.55	5.0	0.14	3.63	5.0	0.02	2.85	5.7	0.01	2.84	5.9
1000		-0.06	2.51	4.9	-0.05	2.55	5.2	0.04	1.96	5.3	0.05	1.95	5.3

See the note to Table 1.

Table 4: Bias($\times 100$) and RMSE($\times 100$) of γ for the TQML and Bai-QML estimators in the case of the AR(1) panel data model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

$T = 5, \gamma_0 = 0.4$						$T = 5, \gamma_0 = 0.8$						
Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)			
TQML		Bai-QML	TQML		Bai-QML	TQML		Bai-QML	TQML		Bai-QML	
		<i>IID</i> <i>spatial</i>			<i>IID</i> <i>spatial</i>			<i>IID</i> <i>spatial</i>			<i>IID</i> <i>spatial</i>	
γ												
N	$m_0 = 1$					$m_0 = 1$						
100	0.60	11.16	44.34	9.44	22.94	50.47	1.52	12.52	17.20	13.00	15.34	18.62
300	0.04	3.79	37.16	5.04	12.89	46.07	1.51	10.92	17.33	9.32	14.16	18.51
500	-0.01	3.08	33.53	3.87	11.25	43.76	0.96	9.54	17.01	7.36	13.33	18.38
1000	0.05	2.69	30.01	2.70	9.96	41.23	0.53	8.20	16.86	5.01	12.13	18.67
N	$m_0 = 2$					$m_0 = 2$						
100	0.52	9.02	41.71	9.69	21.17	49.13	1.34	10.04	17.70	12.86	14.47	18.54
300	0.01	4.16	33.97	4.97	13.71	44.17	1.20	8.70	18.05	9.03	13.06	18.61
500	0.16	3.76	32.06	3.78	12.27	42.88	1.12	7.69	18.35	7.17	12.03	18.76
1000	-0.11	3.40	28.35	2.69	11.70	40.29	0.27	6.09	18.50	4.91	10.63	18.87
$T = 10, \gamma_0 = 0.4$												
Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)			
TQML		Bai-QML	TQML		Bai-QML	TQML		Bai-QML	TQML		Bai-QML	
		<i>IID</i> <i>spatial</i>			<i>IID</i> <i>spatial</i>			<i>IID</i> <i>spatial</i>			<i>IID</i> <i>spatial</i>	
γ												
N	$m_0 = 1$					$m_0 = 1$						
100	-0.06	17.36	41.58	4.37	31.54	51.83	0.26	1.45	15.86	4.80	35.39	20.89
300	-0.05	15.79	43.83	2.46	30.46	52.12	0.03	-0.73	16.74	2.48	40.26	19.70
500	0.00	14.77	44.31	1.86	32.80	52.37	0.06	-0.06	17.05	1.83	39.96	19.59
1000	-0.03	17.36	45.23	1.32	29.83	51.74	-0.02	-0.31	17.07	1.33	38.67	19.83
N	$m_0 = 2$					$m_0 = 2$						
100	-0.06	8.87	39.48	5.12	53.23	51.44	0.18	-10.56	14.18	5.08	46.78	17.13
300	-0.11	6.93	41.71	2.82	55.25	51.57	-0.01	-9.14	14.79	2.75	44.75	16.83
500	-0.09	10.08	43.39	2.16	52.25	51.20	-0.04	-11.78	15.25	2.11	47.13	17.19
1000	0.04	7.58	43.13	1.57	53.65	51.11	0.05	-10.87	15.03	1.48	46.24	16.74

Note: α_i , $i = 1, \dots, N$, are the fixed effects in the y_{it} equation given by (13) in the absence of regressors. Under *IID* these are generated as $\alpha_i \sim IIDN(0, 1)$ and under *spatial* as spatially correlated according to $\alpha = (\mathbf{I}_N - \rho_\alpha \mathbf{W})^{-1} \epsilon_\alpha$ with heteroskedastic errors $\epsilon_\alpha = (\epsilon_{\alpha,1}, \epsilon_{\alpha,2}, \dots, \epsilon_{\alpha,N})'$, where $\rho_\alpha = 0.9$, the matrix \mathbf{W} is specified as in (91) and for each i , $\epsilon_{\alpha,i}$ are drawn as $IDN(0, \sigma_{\epsilon_{\alpha,i}}^2)$ with $\sigma_{\epsilon_{\alpha,i}}^2 = 1$, for $i = 1, 2, \dots, N/2$, and $\sigma_{\epsilon_{\alpha,i}}^2 = 2$, for $N/2 + 1, \dots, N$. TQML is invariant to how the fixed effects are generated. The factor normalisation for the Bai-QML estimator is based on $\mathbf{F} = (\mathbf{I}_{\bar{m}}, \mathbf{F}'_2)'$. See also the note to Table 1.

Table 5: Size($\times 100$) of γ for the TQML and Bai-QML estimators in the case of the AR(1) panel data model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
TQML	Bai-QML			TQML	Bai-QML		TQML	Bai-QML		TQML	Bai-QML	
	<i>IID</i>	<i>spatial</i>			<i>IID</i>	<i>spatial</i>		<i>IID</i>	<i>spatial</i>		<i>IID</i>	<i>spatial</i>
γ												
N	$m_0 = 1$			$m_0 = 1$			$m_0 = 1$			$m_0 = 1$		
100	4.6	35.1	76.5	21.3	79.6	72.7	6.5	15.4	40.2	7.2	37.1	44.3
300	5.0	14.8	65.6	12.3	69.5	69.6	5.8	18.3	44.6	5.0	44.6	46.2
500	5.4	13.9	60.5	8.8	61.9	67.1	5.3	21.2	47.2	4.8	47.8	43.5
1000	4.8	12.0	55.5	4.9	51.9	64.4	5.3	21.7	48.1	4.6	47.4	45.8
N	$m_0 = 2$			$m_0 = 2$			$m_0 = 2$			$m_0 = 2$		
100	4.8	29.65	70.4	18.1	65.9	78.2	4.8	47.35	63.8	4.4	54.7	72.7
300	4.0	15.90	57.9	10.1	56.9	75.8	4.9	54.80	68.5	4.7	59.7	78.1
500	2.7	13.65	55.5	6.3	49.0	76.5	3.7	55.85	69.5	5.1	65.1	80.8
1000	3.6	14.60	49.4	4.3	39.7	75.7	5.3	57.70	69.6	4.7	66.5	80.7

See the note to Table 4.

Table 6: Bias($\times 100$) and RMSE($\times 100$) of γ for the TQML and GMM estimators in the case of the AR(1) panel data model, using the true number of factors, m_0 ($T = 10, \kappa^2 = 1$)

Bias ($\times 100$)						RMSE ($\times 100$)				
	TQML	GMM				TQML	GMM			
		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2
m_0	1									
N	$\gamma_0 = 0.4$									
100	-0.06	47.59	46.28	-77.87	-71.71	4.37	48.52	47.71	79.19	73.47
300	-0.05	48.22	45.18	-67.05	-55.28	2.46	49.30	47.25	68.19	56.85
500	0.00	47.26	42.83	-62.18	-48.23	1.86	48.63	45.64	62.83	49.40
1000	-0.03	44.17	37.98	-55.13	-39.34	1.32	46.17	42.08	55.69	40.28
	$\gamma_0 = 0.8$									
100	0.26	17.82	17.85	-103.25	-100.24	4.80	17.86	17.89	104.33	102.19
300	0.03	17.83	17.74	-89.22	-77.41	2.48	18.18	18.07	90.14	79.44
500	0.06	17.57	17.44	-81.44	-65.55	1.83	18.90	18.81	82.30	67.37
1000	-0.02	17.50	17.35	-72.58	-52.73	1.33	18.87	18.82	73.30	54.20
m_0	2									
N	$\gamma_0 = 0.4$									
100	-0.06	36.71	36.04	-31.72	-28.39	5.12	42.41	42.49	56.67	55.29
300	-0.11	31.22	29.25	-11.99	-7.84	2.82	40.23	38.88	37.23	32.67
500	-0.09	25.70	23.64	-1.81	0.31	2.16	36.29	34.28	23.75	19.81
1000	0.04	16.64	14.62	2.66	2.90	1.57	28.58	26.14	10.95	8.99
	$\gamma_0 = 0.8$									
100	0.18	14.76	14.79	-97.44	-97.95	5.08	22.92	23.33	110.76	112.19
300	-0.01	15.15	15.00	-68.59	-67.07	2.75	23.47	23.67	89.36	88.73
500	-0.04	16.02	15.94	-46.19	-43.19	2.11	21.06	21.08	71.95	69.03
1000	0.05	14.93	14.81	-27.04	-23.18	1.48	22.68	22.72	53.52	48.06

Note: GMM QD1, QD2, FD1 and FD2 are the quasi-difference and first-difference ALS one step and two step estimators respectively computed as described in Section II. See also the note to Table 1.

Table 7: Size($\times 100$) of γ for the TQML and GMM estimators in the case of the AR(1) panel data model, using the true number of factors, m_0 ($T = 10, \kappa^2 = 1$)

TQML GMM						TQML GMM							
		QD1		QD2	FD1	FD2			QD1		QD2	FD1	FD2
m_0	1						2						
N	$\gamma_0 = 0.4$												
100	6.5	95.5	98.4	97.7	100.0	4.8	73.9	81.8	51.5	71.0			
300	5.8	95.3	98.7	97.9	100.0	4.9	64.2	70.2	34.1	50.2			
500	5.3	95.1	99.6	97.8	100.0	3.7	54.5	61.8	22.5	38.0			
1000	5.3	92.2	99.5	97.8	100.0	5.3	41.1	48.4	15.0	27.3			
$\gamma_0 = 0.8$													
100	7.2	99.8	100.0	98.8	100.0	4.4	95.8	97.3	80.2	86.4			
300	5.0	100.0	100.0	98.3	100.0	4.7	96.7	97.2	62.1	72.0			
500	4.8	99.9	100.0	98.2	100.0	5.1	96.8	97.3	46.6	58.3			
1000	4.6	99.8	100.0	98.7	100.0	4.7	95.4	96.3	32.4	43.8			

See the note to Table 6.

Table 8: Bias($\times 100$) and RMSE($\times 100$) of γ and β for the TQML and Bai-QML estimators in the case of the ARX(1) panel data model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

$T = 5, \gamma_0 = 0.4$							$T = 5, \gamma_0 = 0.8$							
Bias($\times 100$)				RMSE($\times 100$)			Bias($\times 100$)				RMSE($\times 100$)			
TQML		Bai-QML		TQML	Bai-QML		TQML		Bai-QML		TQML	Bai-QML		
		<i>IID</i>	<i>spatial</i>						<i>IID</i>	<i>spatial</i>			<i>IID</i>	<i>spatial</i>
γ														
N	$m_0 = 1$													
100	0.09	1.59	1.98		4.28	6.95	8.49	0.23	2.31	4.90		4.74	5.75	7.93
300	-0.05	0.07	0.35		2.39	2.82	4.21	-0.02	0.87	3.26		2.56	3.58	6.35
500	0.02	0.12	0.25		1.82	2.27	3.16	0.02	0.60	2.81		1.91	2.82	5.86
1000	-0.04	0.05	0.07		1.35	1.43	1.76	-0.02	0.41	2.25		1.41	2.13	5.21
$m_0 = 2$														
100	0.22	2.96	3.41		4.48	10.35	12.18	0.41	2.41	5.38		4.89	6.47	8.97
300	0.03	0.46	1.23		2.46	3.98	7.24	0.07	1.05	4.21		2.63	3.94	7.57
500	0.07	0.28	1.07		1.94	3.08	6.83	0.09	0.57	3.82		2.10	2.97	7.02
1000	0.05	0.23	0.88		1.39	2.70	6.12	0.05	0.31	3.59		1.47	2.00	6.79
β														
$m_0 = 1$														
100	-0.01	0.57	0.81		5.98	7.45	7.72	0.06	0.98	1.18		6.16	7.25	8.50
300	-0.15	0.06	0.05		3.39	3.83	4.17	-0.14	0.47	0.51		3.46	4.10	5.97
500	0.09	0.08	0.08		2.65	3.03	3.21	0.10	0.36	0.30		2.70	3.21	5.15
1000	0.05	0.01	0.00		1.87	2.03	2.02	0.06	0.17	0.23		1.91	2.19	3.97
$m_0 = 2$														
100	0.25	0.88	1.43		8.30	31.26	34.36	0.38	-0.14	-3.77		8.51	30.55	34.03
300	0.17	0.97	-0.30		4.61	17.16	16.92	0.20	0.53	-4.01		4.66	17.68	27.14
500	0.11	0.68	-0.54		3.55	13.22	13.82	0.14	0.60	-4.37		3.63	13.61	19.30
1000	-0.06	0.01	-0.75		2.51	9.16	10.46	-0.05	0.02	-6.41		2.55	9.41	23.16
$T = 10, \gamma_0 = 0.4$							$T = 10, \gamma_0 = 0.8$							
Bias($\times 100$)				RMSE($\times 100$)			Bias($\times 100$)				RMSE($\times 100$)			
TQML		Bai-QML		TQML	Bai-QML		TQML		Bai-QML		TQML	Bai-QML		
		<i>IID</i>	<i>spatial</i>						<i>IID</i>	<i>spatial</i>			<i>IID</i>	<i>spatial</i>
γ														
N	$m_0 = 1$													
100	-0.10	3.46	1.18		2.15	11.42	7.30	-0.07	1.16	2.81		1.53	18.51	6.01
300	0.03	3.27	1.12		1.20	10.17	6.93	0.02	2.50	2.59		0.82	14.31	6.72
500	-0.02	3.62	1.18		0.92	10.05	6.98	-0.01	1.45	2.21		0.65	17.73	5.35
1000	0.01	3.46	0.80		0.67	9.05	5.80	0.00	2.25	2.22		0.46	14.20	5.22
$m_0 = 2$														
100	-0.10	5.66	12.20		2.33	21.98	24.14	-0.06	-0.04	6.56		1.58	20.16	10.11
300	-0.06	5.67	10.44		1.33	22.84	21.98	-0.02	0.75	6.70		0.91	21.30	9.85
500	-0.03	5.09	9.37		0.98	25.66	20.77	-0.01	-0.61	6.46		0.69	21.96	10.77
1000	0.02	5.80	8.99		0.70	21.67	20.72	0.01	0.54	6.63		0.48	18.43	9.96
β														
$m_0 = 1$														
100	0.10	-1.25	-0.79		3.98	12.39	7.21	0.07	3.32	-1.73		3.98	40.08	12.53
300	0.01	-1.10	-1.01		2.29	9.96	6.54	0.02	2.57	-2.84		2.28	38.77	11.99
500	0.00	-1.69	-1.00		1.74	10.30	6.24	0.00	2.47	-2.42		1.72	30.21	9.75
1000	0.03	-1.48	-0.52		1.21	8.03	5.38	0.04	1.14	-2.27		1.20	26.98	9.32
$m_0 = 2$														
100	0.15	0.88	-11.76		6.27	35.52	32.59	0.15	4.18	-12.85		6.25	77.77	38.09
300	0.09	1.48	-10.90		3.63	32.50	25.17	0.08	5.91	-14.21		3.61	68.38	40.00
500	0.02	0.58	-9.52		2.85	32.43	24.73	0.01	9.55	-14.12		2.84	85.11	38.50
1000	0.04	1.33	-8.51		1.96	36.25	29.94	0.05	9.65	-13.29		1.95	87.15	44.42

Note: α_i , $i = 1, \dots, N$, are the fixed effects in the y_{it} equation given by (13). The regressor equation fixed effects, α_{xi} , are generated as $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IIDN(0, 1)$ for all i . For the Bai-QML estimator the Mundlak-Chamberlain projection of α_{xi} on the regressors is used to deal with the dependence of α_{xi} on α_i . See also the note to Tables 1 and 4.

Table 9: Size($\times 100$) of γ and β for the TQML and Bai-QML estimators in the case of the ARX(1) panel data model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

		$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
		TQML		Bai-QML	TQML		Bai-QML	TQML		Bai-QML	TQML		Bai-QML
		<i>IID</i>		<i>spatial</i>	<i>IID</i>		<i>spatial</i>	<i>IID</i>		<i>spatial</i>	<i>IID</i>		<i>spatial</i>
γ													
N	$m_0 = 1$												
100		5.1	12.8	11.9	5.2	37.9	55.4	6.0	18.8	9.2	6.5	38.2	29.1
300		4.4	4.5	6.3	5.1	17.5	34.8	5.2	19.2	7.4	4.0	45.6	24.8
500		3.7	5.0	5.6	3.9	12.4	31.0	5.5	20.6	7.4	5.1	45.2	22.1
1000		4.5	4.6	5.1	4.5	8.4	25.0	5.4	20.7	7.0	5.4	47.7	20.9
$m_0 = 2$													
100		4.9	14.9	14.2	4.4	31.0	46.3	5.8	25.4	31.9	5.9	41.4	37.5
300		4.1	5.7	8.2	4.8	15.7	38.8	5.4	35.2	27.1	4.8	50.6	38.1
500		3.6	5.7	8.3	4.6	10.9	34.4	4.3	36.3	24.8	4.7	51.4	39.1
1000		3.6	5.2	7.5	4.2	6.5	31.4	4.3	40.4	24.6	4.1	52.6	39.6
β													
$m_0 = 1$													
100		5.6	7.7	8.1	5.5	8.9	13.4	6.3	13.2	8.0	6.2	10.0	11.8
300		4.9	4.2	5.4	4.9	6.6	13.8	6.0	16.6	7.4	5.6	13.3	14.8
500		5.5	5.6	5.9	5.3	6.5	16.6	5.2	18.3	6.2	5.2	16.4	13.2
1000		5.5	3.9	4.0	5.7	5.3	16.6	4.4	18.4	6.8	4.7	20.3	13.9
$m_0 = 2$													
100		6.1	19.4	18.4	5.7	16.9	21.5	4.9	13.1	26.4	5.0	14.5	22.8
300		5.1	7.2	6.0	5.2	9.5	22.8	5.3	18.9	23.9	5.4	19.8	26.2
500		5.0	6.1	7.0	5.0	7.7	23.4	5.7	21.9	21.5	5.9	21.0	28.2
1000		4.9	4.1	5.9	5.2	5.6	23.9	5.3	26.0	22.5	5.3	26.3	31.2

See the note to Table 8.

Table 10: Bias($\times 100$) and RMSE($\times 100$) of γ and β for the TQML and GMM estimators in the case of the ARX(1) model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

$T = 5, \gamma_0 = 0.4$											$T = 5, \gamma_0 = 0.8$												
Bias($\times 100$)						RMSE($\times 100$)					Bias($\times 100$)						RMSE($\times 100$)						
TQML		GMM				TQML		GMM			TQML		GMM				TQML		GMM				
		QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2
γ																							
N	$m_0 = 1$																						
100		0.09	17.79	17.50	-12.27	-7.22	4.28	20.97	20.70	14.84	10.08	0.23	6.53	6.56	-13.27	-7.43	4.74	7.64	7.41	15.41	9.75		
300		-0.05	13.06	12.97	-5.74	-2.36	2.39	17.65	17.30	7.67	4.16	-0.02	6.09	6.07	-5.83	-2.10	2.56	7.11	6.79	7.83	3.78		
500		0.02	9.46	9.37	-3.18	-1.24	1.82	14.83	14.46	5.24	2.76	0.02	5.72	5.61	-3.13	-1.04	1.91	6.76	6.40	5.28	2.51		
1000		-0.04	4.88	4.82	-1.65	-0.65	1.35	10.67	10.28	3.58	1.90	-0.02	5.04	4.89	-1.59	-0.54	1.41	6.15	5.81	3.61	1.75		
$m_0 = 2$																							
100		0.22	4.44	4.43	-2.59	-1.69	4.48	11.80	12.14	8.92	8.00	0.41	2.29	2.20	-2.99	-1.97	4.89	6.69	6.93	8.91	8.10		
300		0.03	2.90	2.72	-1.05	-0.65	2.46	7.98	8.09	5.21	4.48	0.07	2.21	2.16	-1.10	-0.62	2.63	5.19	5.26	5.12	4.45		
500		0.07	2.38	2.31	-0.63	-0.29	1.94	6.84	6.87	4.14	3.48	0.09	2.19	2.19	-0.56	-0.35	2.10	4.65	4.64	4.02	3.41		
1000		0.05	1.26	1.20	-0.20	-0.16	1.39	4.90	4.96	2.76	2.47	0.05	2.03	1.99	-0.22	-0.20	1.47	4.03	4.05	2.66	2.43		
β																							
$m_0 = 1$																							
100		-0.01	-7.21	-6.90	-5.79	-3.95	5.98	15.03	15.98	10.19	8.69	0.06	-5.52	-4.75	-6.64	-4.17	6.16	10.86	10.61	10.69	8.59		
300		-0.15	-4.56	-4.08	-2.88	-1.49	3.39	11.48	12.45	5.94	4.54	-0.14	-4.34	-3.33	-3.12	-1.37	3.46	8.07	7.44	6.11	4.41		
500		0.09	-2.20	-1.85	-1.44	-0.68	2.65	9.32	10.07	4.44	3.37	0.10	-2.81	-2.06	-1.52	-0.58	2.70	6.76	6.04	4.49	3.28		
1000		0.05	-0.48	-0.24	-0.61	-0.28	1.87	6.61	7.19	3.10	2.30	0.06	-1.32	-0.57	-0.64	-0.22	1.91	5.59	5.01	3.13	2.25		
$m_0 = 2$																							
100		0.25	5.01	3.88	-1.87	-1.28	8.30	24.95	25.10	29.55	26.36	0.38	2.18	1.88	-0.39	-0.22	8.51	23.47	23.79	29.23	26.58		
300		0.17	2.96	2.45	-1.23	-0.26	4.61	14.87	14.81	15.97	14.12	0.20	1.19	1.07	-1.09	-0.18	4.66	13.43	13.41	15.97	13.98		
500		0.11	2.73	2.43	-0.86	-0.07	3.55	11.39	11.53	12.32	10.49	0.14	1.50	1.31	-0.56	0.15	3.63	10.21	10.18	12.52	10.56		
1000		-0.06	1.21	1.02	-0.77	-0.42	2.51	8.37	8.25	8.84	7.48	-0.05	0.78	0.74	-0.45	-0.25	2.55	7.57	7.61	8.79	7.62		
$T = 10, \gamma_0 = 0.4$											$T = 10, \gamma_0 = 0.8$												
Bias($\times 100$)						RMSE($\times 100$)					Bias($\times 100$)						RMSE($\times 100$)						
TQML		GMM				TQML		GMM			TQML		GMM				TQML		GMM				
		QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2
γ																							
N	$m_0 = 1$																						
100		-0.10	-	-	-	-	2.15	-	-	-	-	-0.07	-	-	-	-	1.53	-	-	-	-		
300		0.03	23.02	20.54	-36.09	-29.68	1.20	23.20	20.83	39.30	32.20	0.02	9.00	8.38	-30.82	-25.17	0.82	9.02	8.42	38.40	30.66		
500		-0.02	23.53	20.57	-31.31	-23.92	0.92	23.68	20.84	36.43	27.44	-0.01	9.08	8.34	-22.91	-17.22	0.65	9.10	8.38	33.64	23.99		
1000		0.01	23.54	20.25	-22.35	-15.88	0.67	23.68	20.55	31.55	21.37	0.00	9.17	8.41	-13.70	-9.64	0.46	9.19	8.46	27.53	17.10		
$m_0 = 2$																							
100		-0.10	20.28	19.91	1.39	1.39	2.33	21.79	21.58	5.99	5.75	-0.06	8.85	8.79	-2.28	-2.15	1.58	9.07	9.03	5.30	4.99		
300		-0.06	14.83	13.15	1.59	1.18	1.33	18.67	18.42	3.44	2.51	-0.02	8.36	7.78	-0.08	-0.05	0.91	8.66	8.17	2.65	1.64		
500		-0.03	8.70	6.57	1.16	0.72	0.98	13.32	13.16	2.63	1.68	-0.01	7.69	6.73	0.11	0.05	0.69	8.17	7.37	2.07	1.13		
1000		0.02	2.06	0.63	0.61	0.33	0.70	4.38	3.40	1.81	1.02	0.01	5.25	3.91	0.10	0.06	0.48	6.43	5.19	1.49	0.69		
β																							
$m_0 = 1$																							
100		0.10	-	-	-	-	3.98	-	-	-	-	0.07	-	-	-	-	3.98	-	-	-	-		
300		0.01	-14.30	-10.49	-25.76	-20.57	2.29	15.53	12.16	27.96	22.35	0.02	-17.89	-13.42	-29.17	-23.31	2.28	18.46	14.37	34.13	27.09		
500		0.00	-15.12	-10.59	-24.08	-17.52	1.74	16.08	12.02	26.81	19.45	0.00	-18.52	-13.21	-25.03	-17.96	1.72	19.01	14.25	31.59	22.36		
1000		0.03	-14.75	-9.88	-19.83	-13.18	1.21	15.62	11.10	24.54	15.68	0.04	-19.25	-13.74	-18.97	-12.15	1.20	19.73	14.94	28.02	17.16		
$m_0 = 2$																							
100		0.15	-14.47	-14.06	-0.75	-0.97	6.27	20.18	20.19	12.90	12.64	0.15	-18.71	-18.27	-0.17	-0.21	6.25	21.62	21.42	12.59	12.31		
300		0.09	-11.44	-10.77	-0.70	-0.71	3.63	19.80	20.12	8.52	7.14	0.08	-16.43	-13.14	0.17	0.07	3.61	18.99	16.88	8.34	6.98		
500		0.02	-6.51	-5.81	-0.41	-0.38	2.85	16.45	15.48	7.30	5.41	0.01	-13.57	-8.91	0.19	0.03	2.84	16.99	13.91	7.22	5.34		
1000		0.04	-0.65	-0.48	-0.28	-0.15	1.96	9.53	5.32	5.22	3.51	0.05	-6.33	-1.72	-0.01	-0.01	1.95	11.87	7.71	5.20	3.49		

Note: GMM QD1, QD2, FD1 and FD2 are the quasi-difference and first- difference ALS one step and two step estimators respectively computed as described in Section II of the online supplement. "-" signifies that results are not available which is due to the number of moment conditions exceeding the sample size. See also the note to Table 1.

Table 11: Size($\times 100$) of γ and β for the TQML and GMM estimators in the case of the ARX(1) panel data model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

		$T = 5, \gamma_0 = 0.4$					$T = 5, \gamma_0 = 0.8$					$T = 10, \gamma_0 = 0.4$					$T = 10, \gamma_0 = 0.8$							
		TQML	GMM				TQML	GMM				TQML	GMM				TQML	GMM						
			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2			QD1	QD2	FD1	FD2
N	γ																							
100	$m_0 = 1$	5.1	87.1	89.3	41.2	42.2	5.2	93.0	95.8	48.8	48.0	6.0	-	-	-	-	6.5	-	-	-	-			
300		4.4	69.3	70.9	23.5	17.3	5.1	89.5	90.8	24.6	17.3	5.2	99.9	100.0	96.5	99.8	4.0	100.0	100.0	96.6	100.0			
500		3.7	54.2	55.8	13.7	9.9	3.9	85.9	86.8	13.0	9.5	5.5	99.9	99.9	97.1	100.0	5.1	100.0	100.0	96.8	100.0			
1000		4.5	34.4	35.7	10.0	8.7	4.5	77.4	77.6	9.9	8.9	5.4	100.0	100.0	96.0	100.0	5.4	100.0	100.0	96.4	100.0			
$m_0 = 2$																								
100		4.9	21.3	26.8	5.7	10.6	4.4	38.0	42.6	5.8	10.3	5.8	93.7	98.0	9.2	74.0	5.9	97.9	99.3	11.7	78.2			
300		4.1	17.2	20.6	3.2	6.7	4.8	40.4	42.6	4.5	6.3	5.4	72.1	75.3	9.8	34.5	4.8	95.4	96.1	5.4	29.3			
500		3.6	17.4	19.8	3.0	5.4	4.6	39.8	41.7	3.0	5.6	4.3	51.1	48.7	9.6	23.6	4.7	90.5	89.9	5.1	19.1			
1000		3.6	9.5	11.7	2.3	4.4	4.2	35.6	37.2	2.1	4.2	4.3	18.7	16.1	7.1	13.7	4.1	69.4	65.1	4.6	11.1			
β																								
$m_0 = 1$																								
100		5.6	36.8	48.9	15.2	20.7	5.5	22.1	31.1	18.0	21.1	6.3	-	-	-	-	6.2	-	-	-	-			
300		4.9	45.1	53.3	10.3	11.3	4.9	33.5	35.1	10.5	11.0	6.0	89.0	89.3	92.4	96.1	5.6	98.9	98.0	83.5	91.2			
500		5.5	41.0	48.6	8.3	8.8	5.3	36.2	36.2	7.5	8.5	5.2	93.2	92.0	88.0	93.5	5.2	98.8	96.5	74.5	84.7			
1000		5.5	29.5	34.2	5.5	7.9	5.7	39.3	42.4	5.5	7.4	4.4	94.5	93.3	78.3	84.0	4.7	98.4	95.9	64.3	76.1			
$m_0 = 2$																								
100		6.1	15.5	20.0	10.2	17.6	5.7	11.7	18.2	10.0	18.0	4.9	52.8	83.6	8.3	65.1	5.0	69.0	90.5	8.7	64.4			
300		5.1	12.6	16.0	6.3	12.3	5.2	10.2	13.4	6.4	11.5	5.3	52.6	63.5	7.2	25.8	5.4	75.4	75.1	6.6	25.1			
500		5.0	11.8	13.6	6.0	8.7	5.0	8.8	10.3	5.9	9.0	5.7	34.9	40.9	7.1	19.6	5.9	69.0	62.6	7.1	19.1			
1000		4.9	10.1	10.9	6.3	8.4	5.2	8.3	10.3	6.7	9.3	5.3	11.8	16.3	5.6	11.6	5.3	41.8	34.3	5.3	11.7			

See the note to Table 10.

Table 12: Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure when $\gamma_0 = 1$ ($\kappa^2 = \sigma_v^2 = 1$)

		$T = 5$			$T = 10$		
m_0		0	1	2	0	1	2
N	AR(1)						
100	99.5	99.6	96.5	99.5	99.6	99.6	
300	99.8	99.9	100.0	100.0	99.9	100.0	
500	99.8	100.0	100.0	100.0	99.9	100.0	
1000	99.9	100.0	100.0	100.0	99.9	100.0	
	ARX(1)						
100	99.6	99.9	97.2	99.3	99.7	99.8	
300	100.0	100.0	100.0	100.0	100.0	99.9	
500	99.9	100.0	100.0	100.0	100.0	100.0	
1000	100.0	100.0	100.0	100.0	99.9	100.0	

Note: First-differences are generated as $\Delta y_{it} = \Delta \delta_t + \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta \zeta_{it}$, $t = 2, 3, \dots, T$, with $\Delta \zeta_{it} = \sum_{\ell=1}^{m_0} \eta_{\ell i} \Delta f_{\ell t} + \Delta u_{it} = \eta'_i \Delta \mathbf{f}_t + \Delta u_{it}$, $\Delta y_{i1} = \Delta \delta_1 + \beta \Delta x_{i1} + \Delta \zeta_{i1}$ and $\Delta y_{i0} = 0$, for $i = 1, 2, \dots, N$, and $\gamma = \beta = 1$. The first-differences are then cumulated and y_{it} is obtained using arbitrary values for y_{i0} . The idiosyncratic errors are generated as $u_{it} \sim IID_{\sqrt{12}}(\chi_6^2 - 6)$ for $i = 1, 2, \dots, N$; $t = 0, 1, \dots, T$ where χ_6^2 is a chi-square variate with 6 degrees of freedom and $\sigma^2 = 1$. The fixed effects are generated as $\alpha_i \sim IIDN(0, 1)$ and the factor loadings are specified as $\eta_{\ell i} \sim IIDN(0, \frac{\kappa^2}{m_0})$, $\ell = 1, 2, \dots, m_0$. The regressors, x_{it} , for $i = 1, 2, \dots, N$ are generated as $x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} + v_{it}$, $v_{it} = \rho_x v_{i,t-1} + (1 - \rho_x^2)^{1/2} \varepsilon_{it}$, for $t = 1, 2, \dots, T$, with $\rho_x = 0.95$, $m_x = 2$, $\vartheta_{\ell i} \sim IIDN(0, \sigma_{\vartheta \ell}^2)$, for $\ell = 1, 2, \dots, m_x$, and $\sigma_{\vartheta \ell}^2 = \sigma_v^2 / m_x$ for all ℓ , $\varepsilon_{it} \sim IIDN(0, \sigma_{v_i}^2)$, $v_{i0} \sim IIDN(0, \sigma_{v_i}^2)$, for $i = 1, 2, \dots, N$, with $\sigma_{v_i}^2 \sim IID(\frac{1}{4}(\chi_2^2 + 2)\sigma_v^2)$ and $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IIDN(0, 1)$, for all i . The remaining parameters are generated as described in Section 8.1. Each f_t is generated once and the same f_t 's are used throughout the replications. In the AR(1) case $\beta = 0$ and under $m_0 = 0$, ζ_{it} collapses to u_{it} .

Table 13: Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m} , when $\gamma_0 = 1$ ($\kappa^2 = 1$)

	$T = 5$			$T = 10$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
$m_0 = 0$						
100	-1.49	2.74	3.8	-0.53	1.24	3.3
300	-0.89	1.69	3.1	-0.33	0.50	4.2
500	-0.67	1.08	2.6	-0.26	0.37	2.5
1000	-0.53	1.25	2.4	-0.20	0.33	3.0
$m_0 = 1$						
100	-2.99	5.70	5.4	-0.61	1.01	3.0
300	-1.83	3.43	4.9	-0.39	0.95	2.8
500	-1.34	2.25	3.7	-0.31	0.46	2.9
1000	-0.97	1.64	3.4	-0.24	0.33	2.4
$m_0 = 2$						
100	-3.00	5.09	5.1	-0.61	1.01	3.8
300	-1.70	2.93	3.9	-0.39	0.95	2.3
500	-1.37	2.30	3.2	-0.31	0.46	2.4
1000	-0.99	1.65	3.3	-0.24	0.33	2.1

See the note to Table 12.

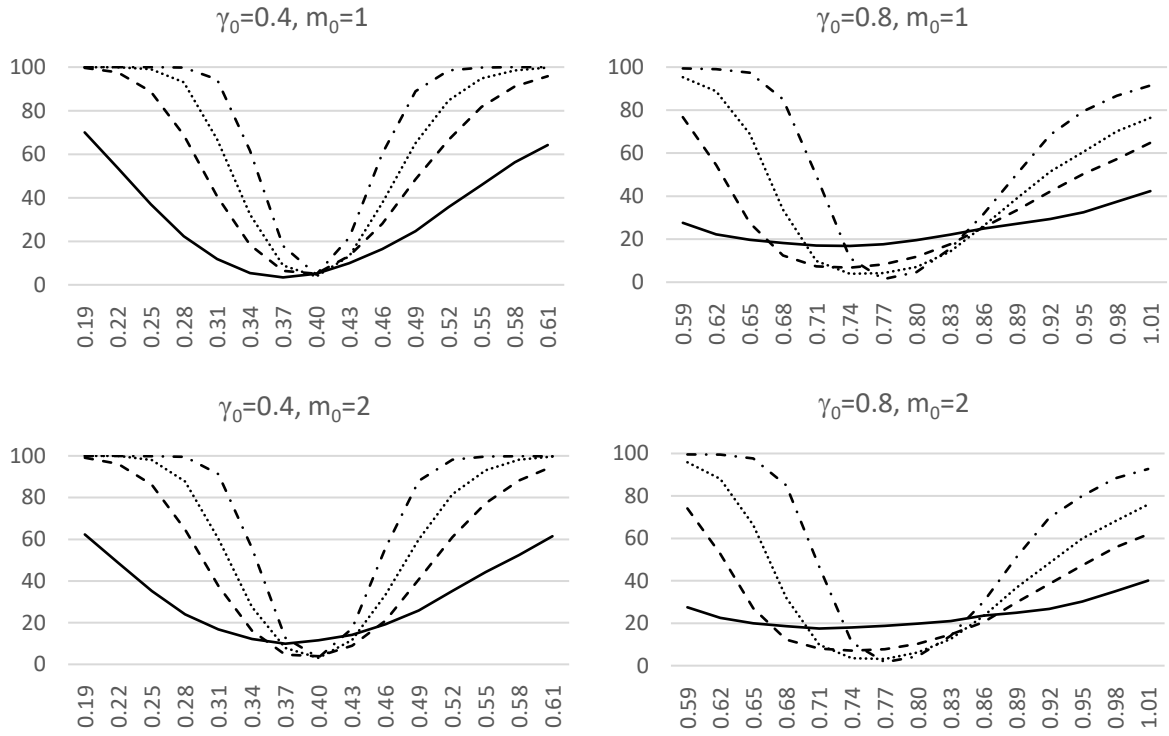
Table 14: Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} , when $\gamma_0 = 1$ ($\kappa^2 = \sigma_v^2 = 1$)

	$T = 5$			$T = 10$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
γ						
$m_0 = 0$						
100	-1.28	2.17	3.7	-0.43	0.67	3.3
300	-0.77	1.27	3.4	-0.26	0.37	2.1
500	-0.58	0.94	3.2	-0.22	0.30	2.5
1000	-0.46	0.70	3.3	-0.18	0.23	2.9
$m_0 = 1$						
100	-2.00	3.46	3.9	-0.53	0.84	3.6
300	-1.24	2.05	2.3	-0.31	0.46	2.3
500	-0.97	1.61	2.3	-0.26	0.37	2.8
1000	-0.75	1.23	3.5	-0.20	0.26	2.2
$m_0 = 2$						
100	-2.02	3.52	3.5	-0.50	0.80	2.4
300	-1.19	2.06	3.0	-0.32	0.47	2.1
500	-0.97	1.61	2.5	-0.27	0.39	2.5
1000	-0.71	1.16	2.8	-0.20	0.26	2.0
β						
$m_0 = 0$						
100	-0.58	4.47	5.5	-0.13	3.01	6.2
300	-0.30	2.55	5.0	-0.09	1.72	5.6
500	-0.21	1.94	4.0	-0.05	1.33	5.3
1000	-0.18	1.39	4.4	-0.03	0.95	4.8
$m_0 = 1$						
100	-0.97	5.95	4.5	-0.02	3.95	6.0
300	-0.69	3.38	4.2	-0.04	2.27	5.3
500	-0.36	2.62	4.5	-0.05	1.72	4.5
1000	-0.27	1.87	4.4	0.00	1.20	3.8
$m_0 = 2$						
100	-0.59	8.26	5.1	0.28	6.25	5.2
300	-0.29	4.61	4.5	0.17	3.60	5.0
500	-0.27	3.56	3.9	0.09	2.83	5.8
1000	-0.34	2.54	4.6	0.11	1.95	4.7

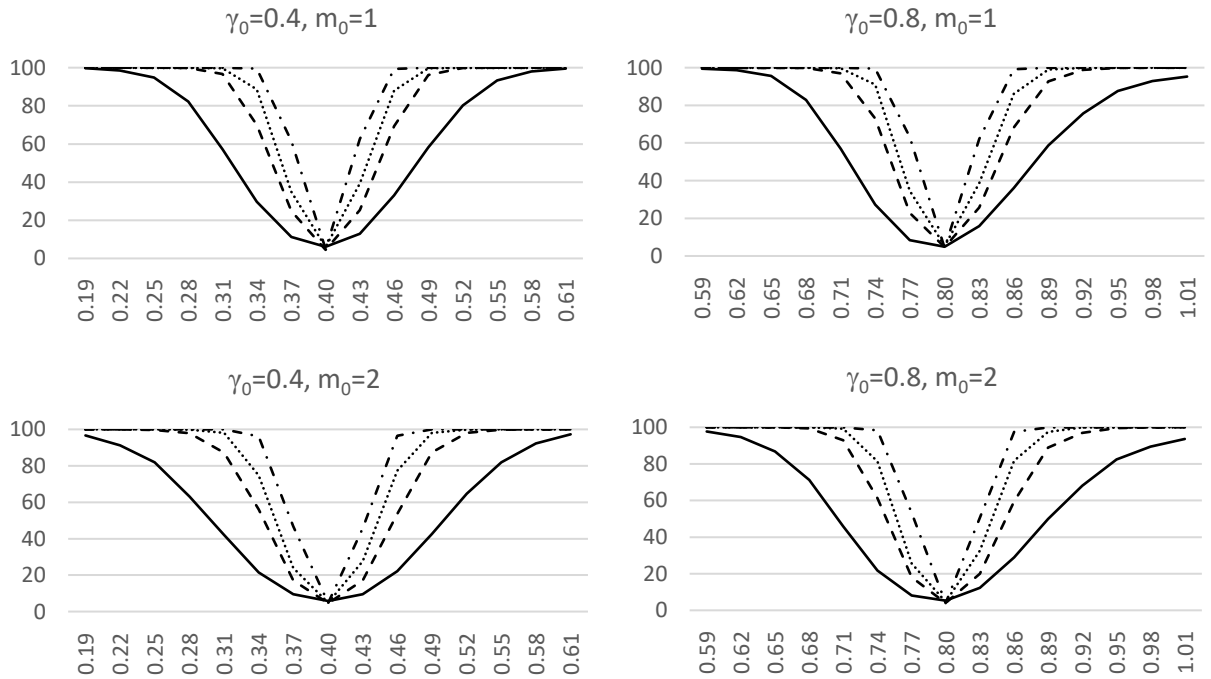
See the note to Table 12.

Figure 1: Power functions for γ in the case of the AR(1) panel data model with different values of m and N

Panel A: $T=5$



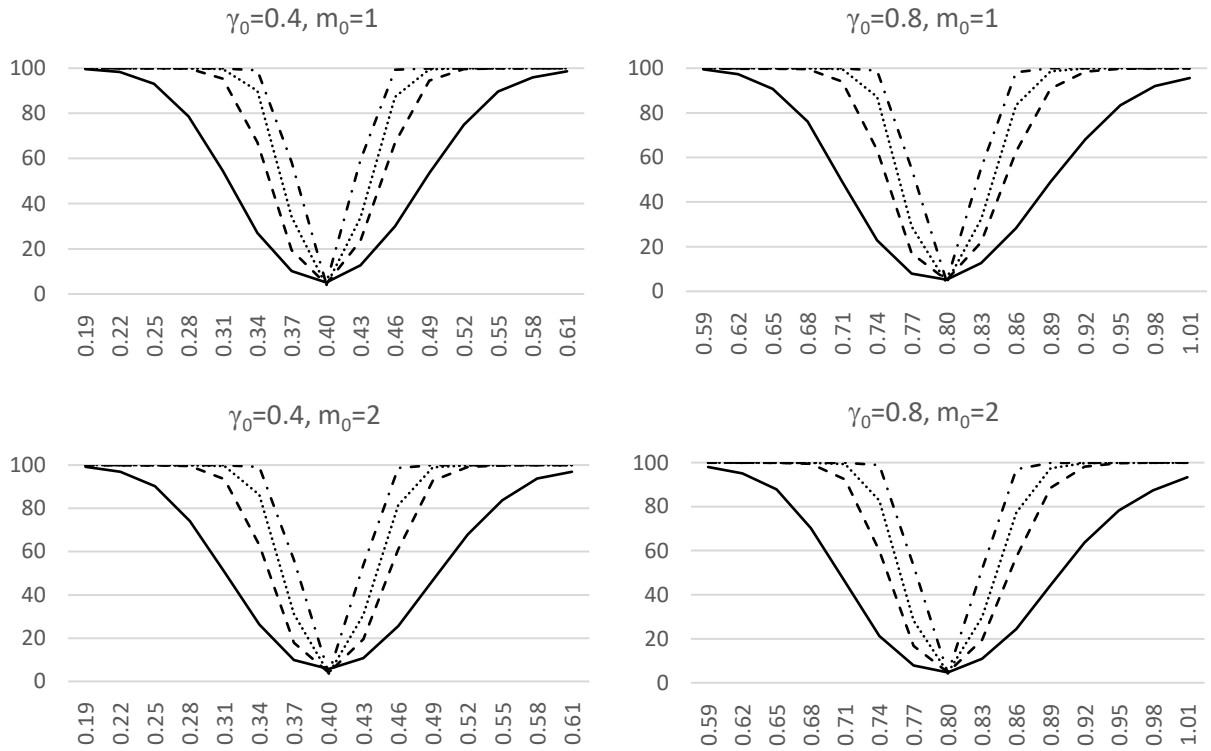
Panel B: $T=10$



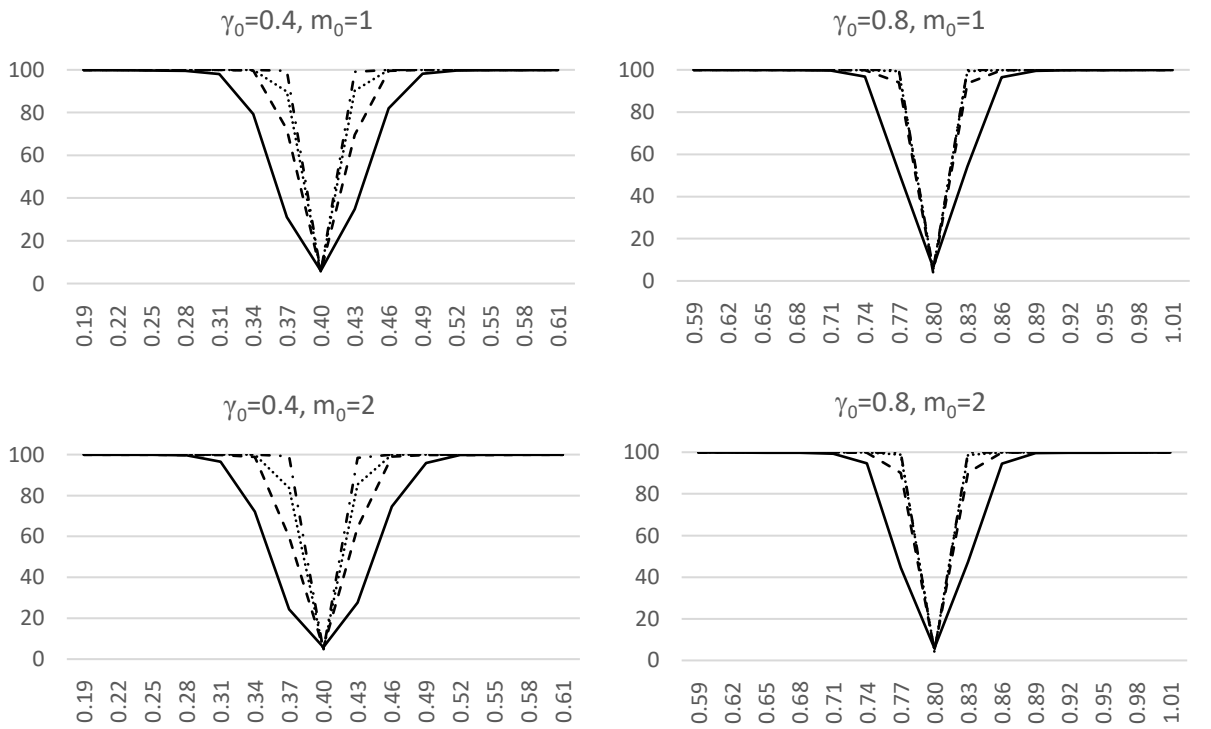
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table 1.

Figure 2a: Power functions for γ in the case of the ARX(1) panel data model with different values of m and N

Panel A: $T=5$



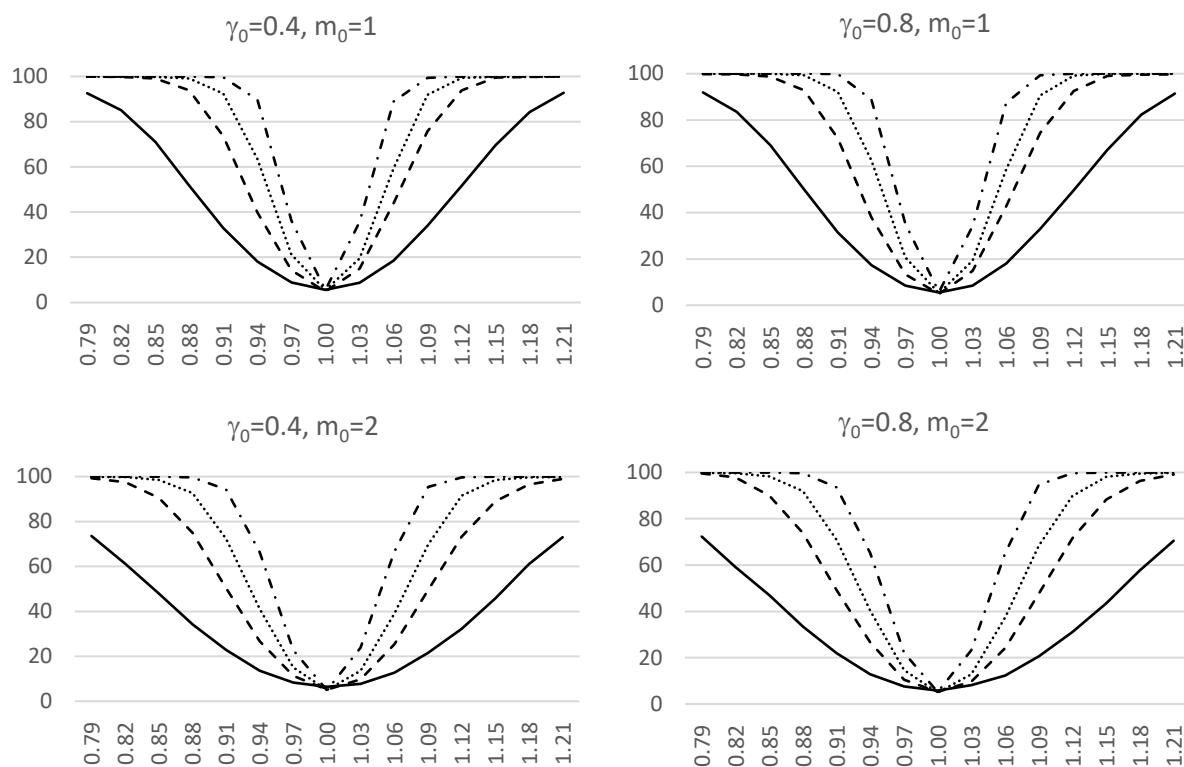
Panel B: $T=10$



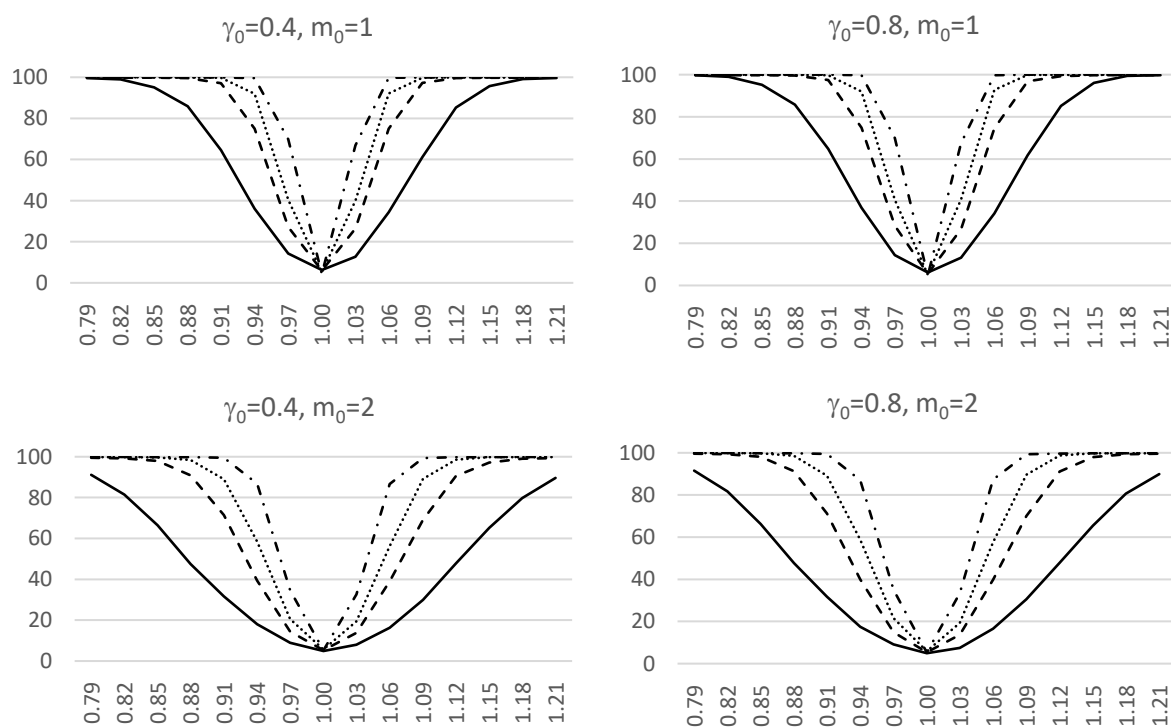
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13). See also the note to Table 1.

Figure 2b: Power functions for β in the case of the ARX(1) panel data model with different values of m and N

Panel A: $T=5$



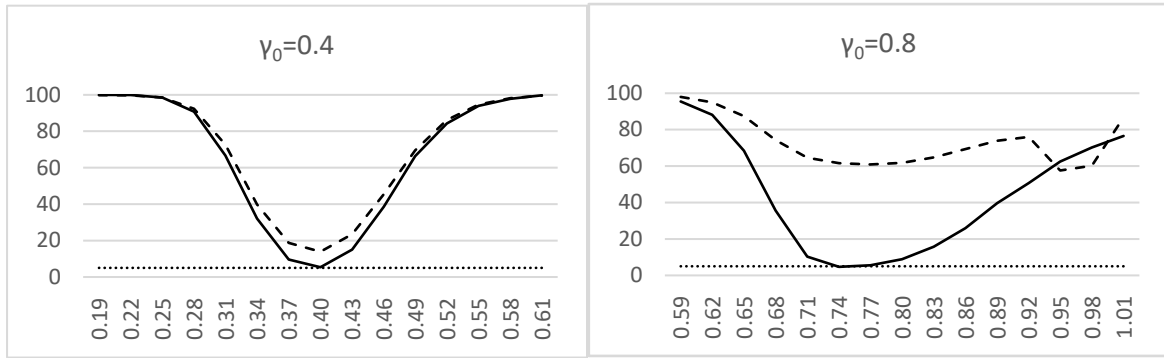
Panel B: $T=10$



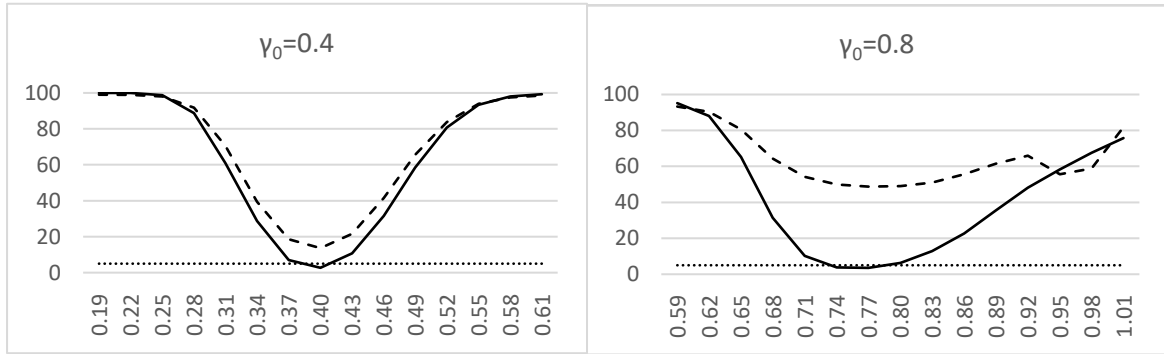
Note: — $N=100$ - - - $N=300$ $N=500$ - . - $N=1000$. β is the coefficient of the \mathbf{x}_{it} regressors in (13). See also the note to Figure 2a.

Figure 3a: Power functions for γ in the case of the AR(1) panel data model with $T=5$, $N=500$, $m=\tilde{m}_0=m_0+1$, and $\alpha_i \sim \text{IIDN}(0,1)$

Panel A: $m_0=1$



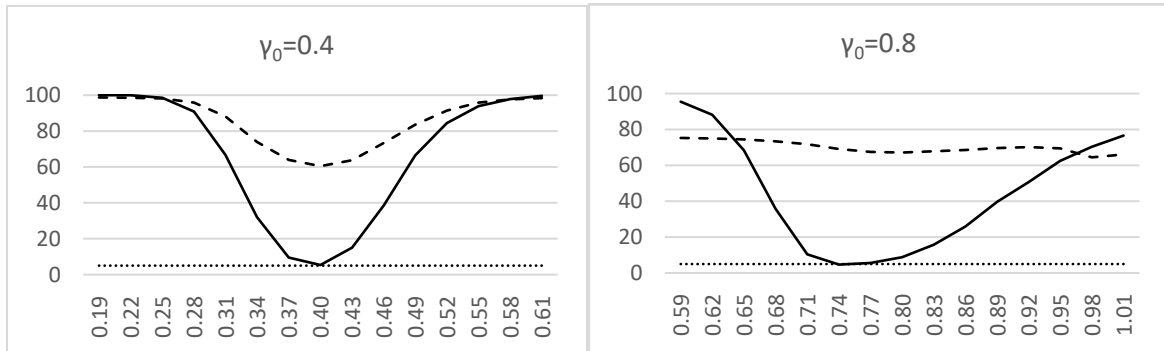
Panel B: $m_0=2$



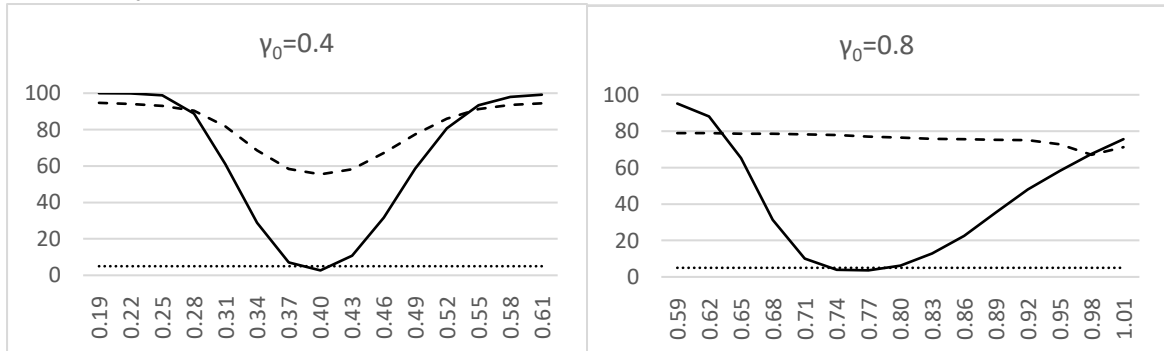
Note: — TQML - - - Bai_QML 5% nominal value. α_i are the fixed effects and γ is the coefficient of the lagged dependent variable in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table 1.

Figure 3b: Power functions for γ in the case of the AR(1) panel data model with $T=5$, $N=500$, $m=\tilde{m}_0=m_0+1$, and α_i spatially correlated

Panel A: $m_0=1$



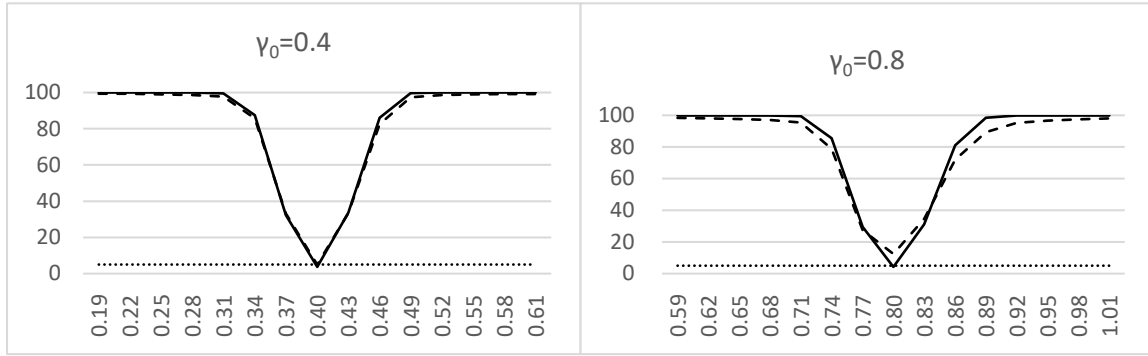
Panel B: $m_0=2$



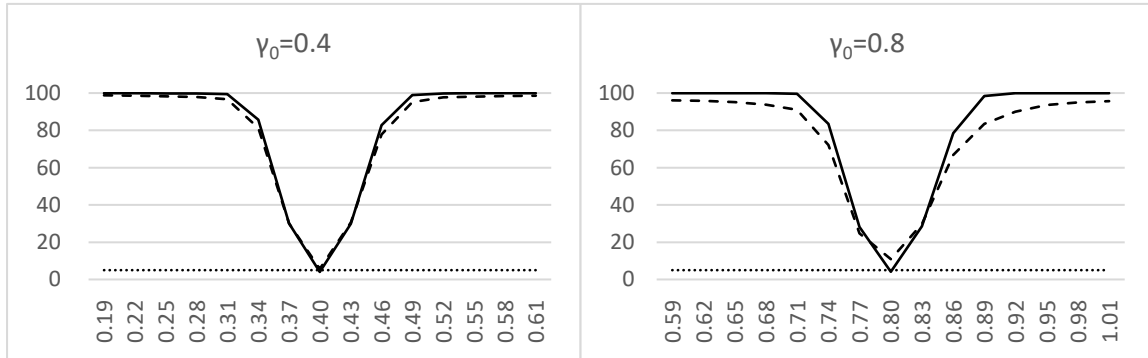
Note: — TQML - - - Bai_QML 5% nominal value. See also the note to Figure 3a.

Figure 4a: Power functions for γ in the case of the ARX(1) panel data model with $T=5$, $N=500$, $\beta_0=1$, $m=\tilde{m}_0=m_0+1$, and $\alpha_i \sim \text{IIDN}(0,1)$

Panel A: $m_0=1$



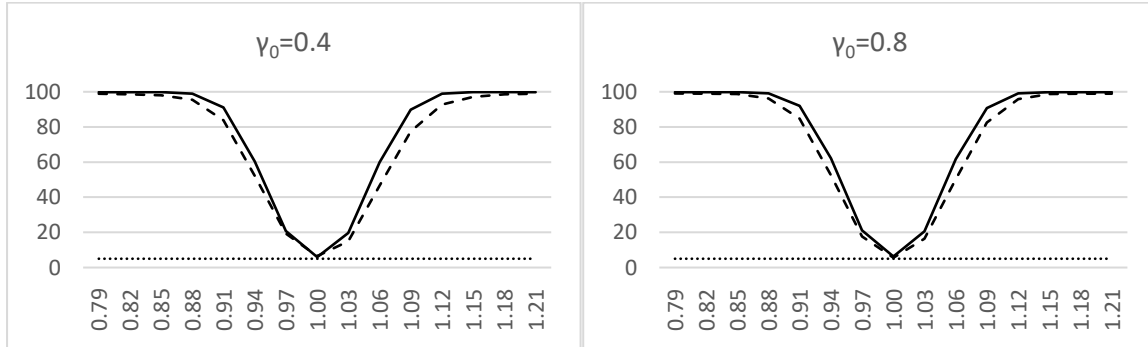
Panel B: $m_0=2$



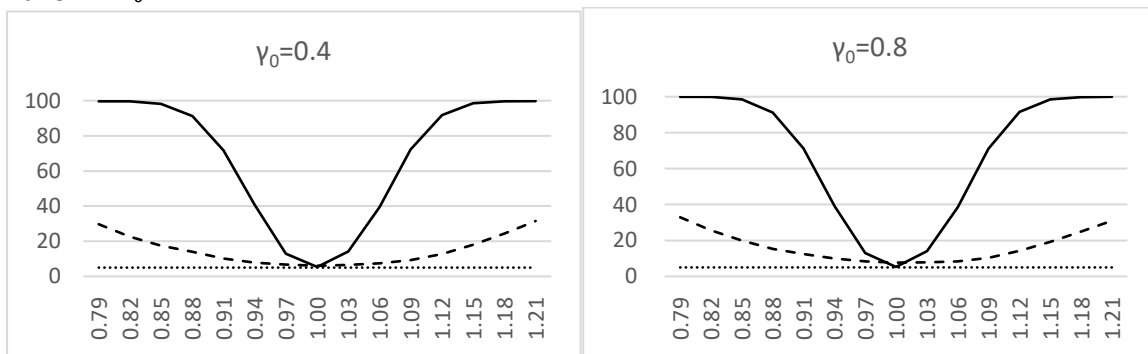
Note: — TQML - - - Bai_QML 5% nominal value. α_i are the fixed effects and γ is the coefficient of the lagged dependent variable in (13). See also the note to Table 1.

Figure 4b: Power functions for β in the case of the ARX(1) panel data model with $T=5$, $N=500$, $\beta_0=1$, $m=m_0$ with $\alpha_i \sim \text{IIDN}(0,1)$

Panel A: $m_0=1$



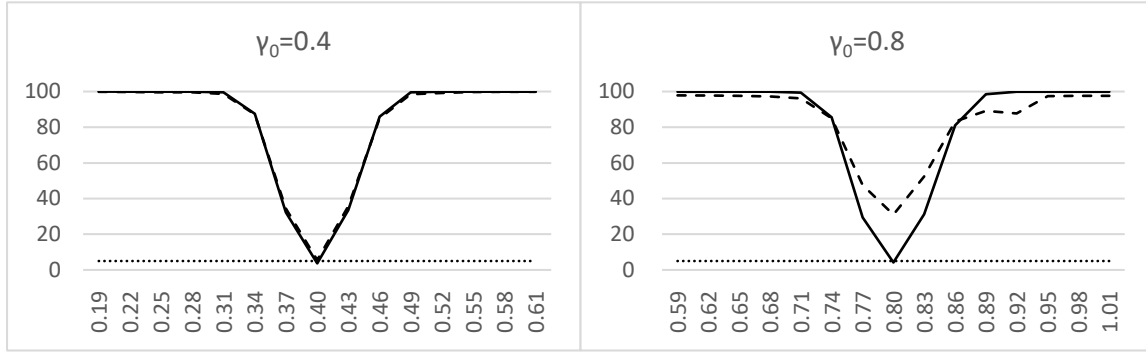
Panel B: $m_0=2$



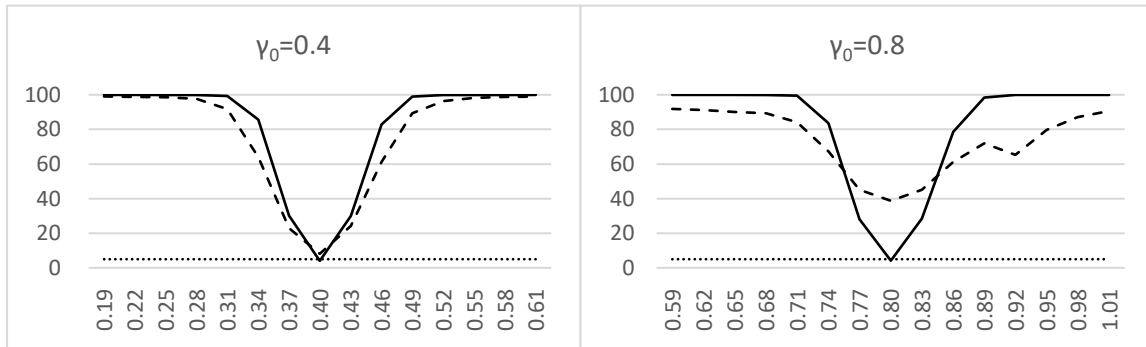
Note: — TQML - - - Bai_QML 5% nominal value. α_i are the fixed effects and β is the coefficient of the \mathbf{x}_{it} regressors in (13). See also the note to Table 1.

Figure 5a: Power functions for γ in the case of the ARX(1) panel data model with $T=5$, $N=500$, $\beta_0=1$, $m=\tilde{m}_0=m_0+1$, and α_i spatially correlated

Panel A: $m_0=1$



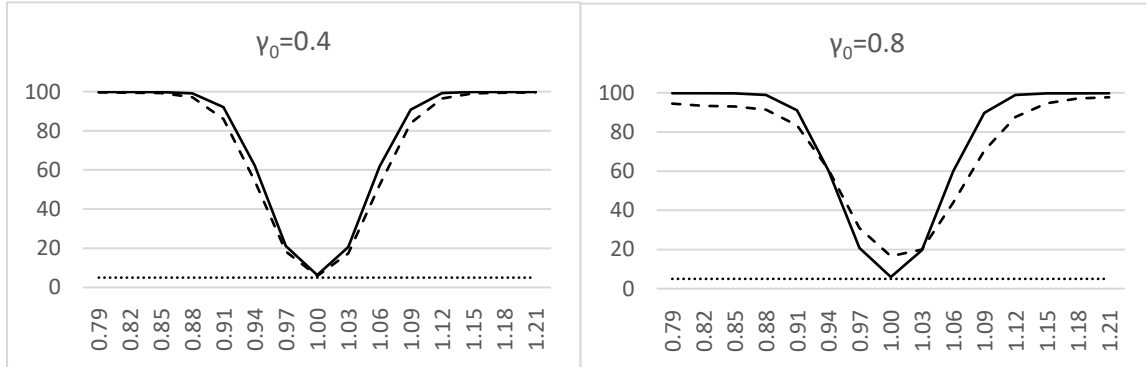
Panel B: $m_0=2$



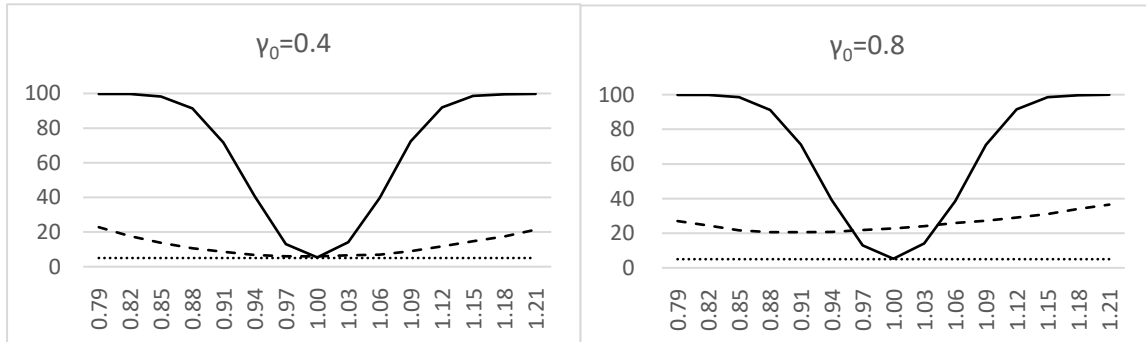
Note: — TQML - - - Bai_QML 5% nominal value. α_i are the fixed effects and γ is the coefficient of the lagged dependent variable in (13). See also the note to Table 1.

Figure 5b: Power functions for β in the case of the ARX(1) panel data model with $T=5$, $N=500$, $\beta_0=1$, $m=\tilde{m}_0=m_0+1$, and α_i spatially correlated

Panel A: $m_0=1$



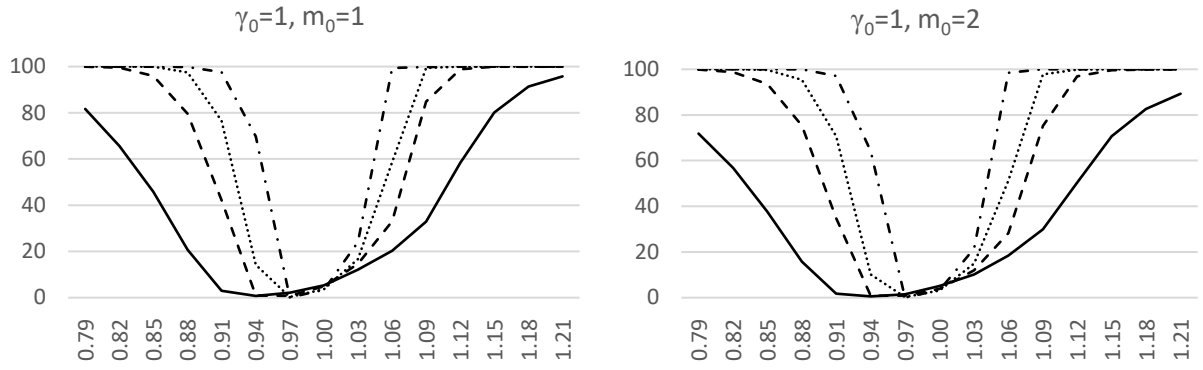
Panel B: $m_0=2$



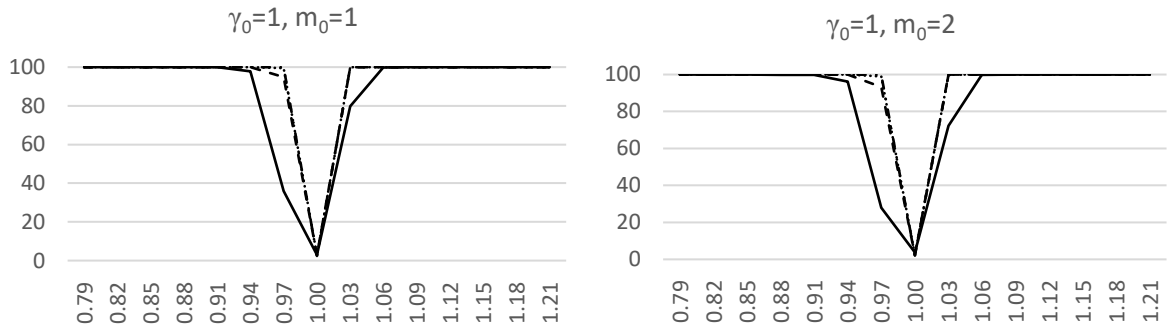
Note: — TQML - - - Bai_QML 5% nominal value. α_i are the fixed effects and β is the coefficient of the \mathbf{x}_{it} regressors in (13). See also the note to Table 1.

Figure 6: Power functions for γ in the case of the AR(1) panel data model with different values of m and N

Panel A: $T=5$



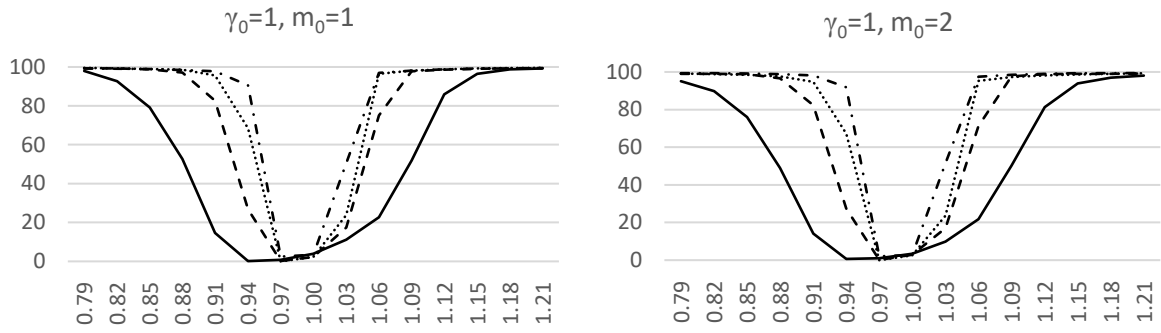
Panel B: $T=10$



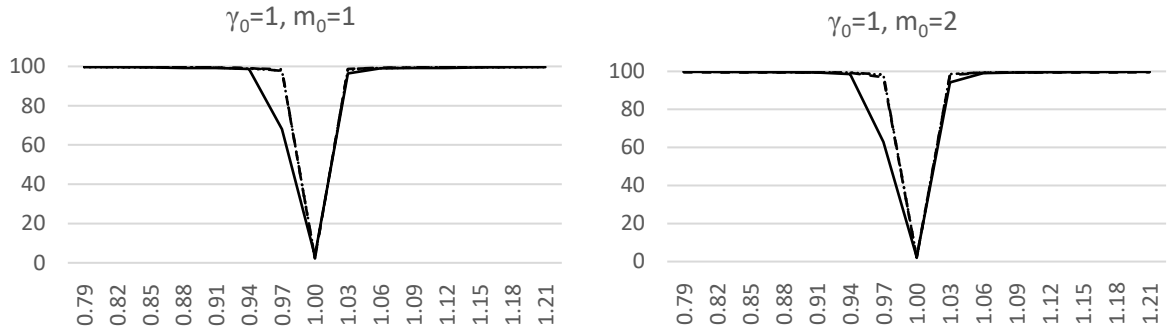
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table 4.

Figure 7a: Power functions for γ in the case of the ARX(1) panel data model with different values of m and N

Panel A: $T=5$



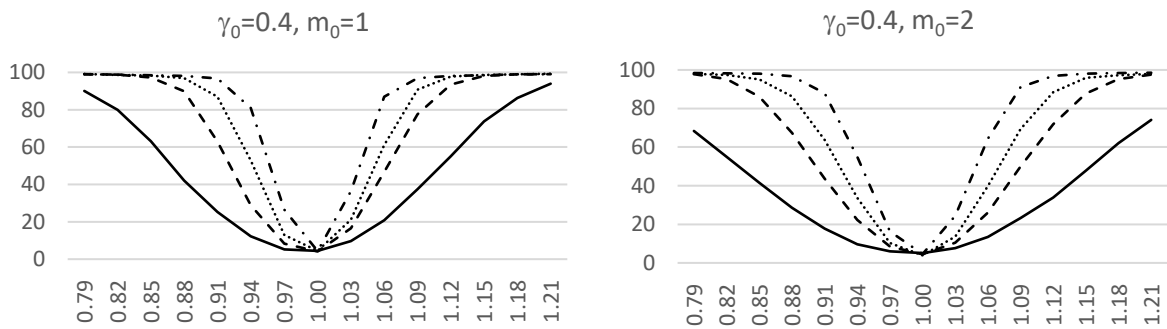
Panel B: $T=10$



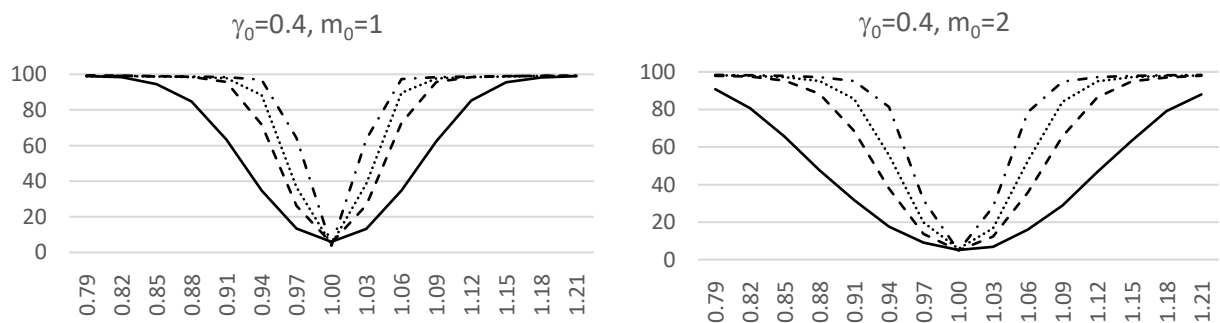
Note: — $N=100$ --- $N=300$ $N=500$ -.- $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13). See also the note to Table 4.

Figure 7b: Power functions for estimation of β in the ARX(1) model with different values of m and N

Panel A: $T=5$



Panel B: $T=10$



Note: — $N=100$ --- $N=300$ $N=500$ -.- $N=1000$. β is the coefficient of the \mathbf{x}_{it} regressors in (13). See also the note to Figure 7a.

Appendix

A.1 Lemmas and their proofs

Lemma 1 Consider the composite random variable, ξ_{it} , $i = 1, 2, \dots, N$, for $t = 1$ defined by (25), and for $t = 2, 3, \dots, T$ defined by (17). Then under Assumptions 1, 2, 3, 5, and 6, the following moment conditions hold.²⁴

$$\sup_i E \left(|\xi_{it}|^{4+\epsilon} \right) < K, \text{ for } t = 1, 2, \dots, T, \quad (\text{A.1})$$

and

$$\sup_{i,t} E \left(\|\Delta \mathbf{x}_{it}\|^{4+\epsilon} \right) < K. \quad (\text{A.2})$$

Proof. Result (A.1) follows by applying Minkowski's inequality to the elements of $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{iT})'$. Specifically, for $t = 2, 3, \dots, T$, $\xi_{it} = \mathbf{g}'_t \boldsymbol{\eta}_i + \Delta u_{it}$ and we have

$$\begin{aligned} \left(E |\xi_{it}|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}} &= \left(E |\mathbf{g}'_t \boldsymbol{\eta}_i + \Delta u_{it}|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}} \\ &\leq \left(E |\mathbf{g}'_t \boldsymbol{\eta}_i|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}} + \left(E |\Delta u_{it}|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}} \\ &\leq \|\mathbf{g}_t\| \left(E \|\boldsymbol{\eta}_i\|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}} + \left(E |\Delta u_{it}|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}}. \end{aligned}$$

Under Assumptions 1, 2 and 3 $\sup_t \|\mathbf{g}_t\| < K$, $\sup_i E \|\boldsymbol{\eta}_i\|^{4+\epsilon} < K$ and $\sup_{i,t} E |\Delta u_{it}|^{4+\epsilon} < K$. Similarly for $t = 1$, $\xi_{i1} = \tilde{\mathbf{g}}'_1 \boldsymbol{\eta}_i + v_{i1}$, and $\|\tilde{\mathbf{g}}_1\| < K$ and $\sup_i E |v_{i1}|^{4+\epsilon} < K$ (see (26) and related results). Hence, $\left(E |\xi_{it}|^{4+\epsilon} \right)^{\frac{1}{4+\epsilon}} \leq K$, for $t = 1, 2, \dots, T$ and (A.1) follows as required. To establish condition (A.2), using (14) we first note that

$$\|\Delta \mathbf{x}_{it}\| \leq \|\boldsymbol{\delta}_{x,t}\| + \sum_{j=1}^{m_x} |g_{x,jt}| \|\boldsymbol{\eta}_{ij,x}\| + \sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_j\| \|\boldsymbol{\varepsilon}_{i,t-j}\|,$$

and by the Minkowski inequality for infinite sums we have

$$(E \|\Delta \mathbf{x}_{it}\|^p)^{1/p} \leq \|\boldsymbol{\delta}_{x,t}\| + \sum_{j=1}^{m_x} |g_{x,jt}| (E \|\boldsymbol{\eta}_{ij,x}\|^p)^{1/p} + \sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_j\| (E \|\boldsymbol{\varepsilon}_{i,t-j}\|^p)^{1/p},$$

for any $p \geq 1$. Set $p = 4 + \epsilon$, and note that under Assumption 5, $\sup_t \|\boldsymbol{\delta}_{x,t}\| < K$, $\sup_{j,t} |g_{x,jt}| < K$, $\sup_{i,j} E \|\boldsymbol{\eta}_{ij,x}\|^{4+\epsilon} < K$, $\sup_{i,t} E \|\boldsymbol{\varepsilon}_{it}\|^{4+\epsilon} < K$, and $\sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_j\| < K$. Therefore, $\left(E \|\Delta \mathbf{x}_{it}\|^{4+\epsilon} \right)^{1/(4+\epsilon)} \leq K$, and (A.2) follows as required. ■

Lemma 2 Consider the $T \times 1$ vector of composite errors $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{iT})'$, where ξ_{i1} is defined by (25) and ξ_{it} , for $t = 2, 3, \dots, T$ are defined by (17). Suppose that the conditions of Lemma 1 hold and T is fixed. Then

$$\sup_i E \|\boldsymbol{\xi}_i\|^4 < K < \infty, \quad (\text{A.3})$$

$$\sup_i E \|\mathbf{Z}_i\|^4 < K, \sup_i E \|\Delta \mathbf{y}_i\|^4 < K, \text{ and } \sup_i E \|\Delta \mathbf{W}_i\|^4 < K < \infty. \quad (\text{A.4})$$

²⁴It is worth emphasising that this and other lemmas are established for a finite T and conditional on given values of time effects, namely \mathbf{g}_t , $\boldsymbol{\delta}_t$, $\boldsymbol{\delta}_{x,t}$, and $\mathbf{g}_{x,t}$, for $t = 1, 2, \dots, T$.

Proof. To obtain (A.3) note that

$$\|\boldsymbol{\xi}_i\|^4 = \|\boldsymbol{\xi}_i \boldsymbol{\xi}_i'\|^2 = \text{tr}(\boldsymbol{\xi}_i \boldsymbol{\xi}_i' \boldsymbol{\xi}_i \boldsymbol{\xi}_i') = (\boldsymbol{\xi}_i' \boldsymbol{\xi}_i)^2 = \left(\sum_{t=1}^T \xi_{it}^2 \right)^2.$$

Then by Minkowski's inequality we have

$$E \|\boldsymbol{\xi}_i\|^4 = E \left(\sum_{t=1}^T \xi_{it}^2 \right)^2 \leq \left(\sum_{t=1}^T [E(\xi_{it}^4)]^{1/2} \right)^2,$$

and since $\sup_i E(|\xi_{it}|^{4+\epsilon}) < K$ for $t = 1, 2, \dots, T$ from result (A.1) of Lemma 1, result (A.3) follows noting that T is fixed. To establish (A.4), note that $\Delta \mathbf{W}_i = (\mathbf{I}_T, \Delta \mathbf{X}_i, \Delta \mathbf{y}_{i,-1}) = (\mathbf{I}_T, \Delta \mathbf{X}_i, \mathbf{L} \Delta \mathbf{y}_i)$, where $\Delta \mathbf{y}_{i,-1} = (0, \Delta y_{i1}, \dots, \Delta y_{i,T-1})'$, $\Delta \mathbf{X}_i$ and $\Delta \mathbf{y}_i$ are given by (43) and (42), and recall \mathbf{L} is the lag matrix operator which is given explicitly by

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & 0 \\ \vdots & 1 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad (\text{A.5})$$

with $\|\mathbf{L}\|^2 = T-1$. It is now easily seen that $\|\Delta \mathbf{W}_i\|^2 \leq T + \|\Delta \mathbf{X}_i\|^2 + (T-1) \|\Delta \mathbf{y}_i\|^2$, and by Minkowski's inequality we obtain

$$\left(E \|\Delta \mathbf{W}_i\|^4 \right)^{1/2} \leq T + \left(E \|\Delta \mathbf{X}_i\|^4 \right)^{1/2} + (T-1) \left(E \|\Delta \mathbf{y}_i\|^4 \right)^{1/2}.$$

Also $\|\Delta \mathbf{X}_i\|^2 = \|\Delta \mathbf{x}_{i1}\|^2 + 2 \sum_{t=2}^T \|\Delta \mathbf{x}_{it}\|^2$, and since by result (A.2) of Lemma 1 $\sup_{i,t} E \left(\|\Delta \mathbf{x}_{it}\|^{4+\epsilon} \right) < K$, it then follows that $\sup_i E \|\Delta \mathbf{X}_i\|^4 < K$. Similarly, using (42), we have

$$\|\Delta \mathbf{y}_i\| \leq \|\mathbf{a}\| + \|\mathbf{B}^{-1}(\gamma)\| \|\boldsymbol{\delta}\| \|\Delta \mathbf{X}_i\| + \|\mathbf{B}^{-1}(\gamma)\| \|\boldsymbol{\xi}_i\|,$$

and by assumption $\|\mathbf{a}\| < K$, $\|\boldsymbol{\delta}\| < K$, and $\|\mathbf{B}^{-1}(\gamma)\| < K$. Also by result (A.1) of Lemma 1 $\sup_{i,t} E |\xi_{it}|^{4+\epsilon} < K$, and it is already established that $\sup_i E \|\Delta \mathbf{X}_i\|^4 < K$. Hence,

$$\left(E \|\Delta \mathbf{y}_i\|^4 \right)^{1/4} \leq \|\mathbf{a}\| + \|\mathbf{B}^{-1}(\gamma)\| \|\boldsymbol{\delta}\| \left(E \|\Delta \mathbf{X}_i\|^4 \right)^{1/4} + \|\mathbf{B}^{-1}(\gamma)\| \left(E \|\boldsymbol{\xi}_i\|^4 \right)^{1/4},$$

and it follows that $\sup_i E \|\Delta \mathbf{y}_i\|^4 < K$, as required. ■

Lemma 3 Consider the model given by (28) and let

$$\boldsymbol{\xi}_i(\boldsymbol{\varphi}) = \Delta \mathbf{y}_i - \Delta \mathbf{W}_i \boldsymbol{\varphi}, \quad \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) = E [\boldsymbol{\xi}_i(\boldsymbol{\varphi}) \boldsymbol{\xi}_i'(\boldsymbol{\varphi})].$$

Define

$$\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0) = \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0), \quad (\text{A.6})$$

and suppose that Assumptions 1-7(ii),(iii) and 8, as well as the order condition (41) hold. Then

$$E_0 [\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)] = \mathbf{b}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0) = [\mathbf{0}, \mathbf{0}, -\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0)]', \quad (\text{A.7})$$

where

$$\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = \text{tr} \{ [\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) - \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)] \mathbf{C}(\boldsymbol{\psi}, \boldsymbol{\gamma}_0) \} \quad (\text{A.8})$$

and

$$\mathbf{C}(\psi, \gamma_0) = \Sigma_\xi(\psi)^{-1} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_0^{T-3} & \gamma_0^{T-4} & \cdots & 0 & 0 \\ \gamma_0^{T-2} & \gamma_0^{T-3} & \cdots & 1 & 0 \end{pmatrix}. \quad (\text{A.9})$$

Furthermore

$$E_0[\mathbf{d}_i(\psi_0, \varphi_0)] = \mathbf{0}, \text{ for } i = 1, 2, \dots, N, \quad (\text{A.10})$$

$$\mathbf{b}_N(\psi, \varphi_0) = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i(\psi, \varphi_0) \xrightarrow{a.s.} \mathbf{b}(\psi, \varphi_0) = [\mathbf{0}, \mathbf{0}, -\kappa(\psi, \psi_0)]', \quad (\text{A.11})$$

$$\mathbf{b}_N(\psi_0, \varphi_0) = \frac{1}{N} \sum_{i=1}^N \Delta \mathbf{W}_i' \Sigma_\xi(\psi_0)^{-1} \xi_i(\varphi_0) \xrightarrow{a.s.} \mathbf{0}, \quad (\text{A.12})$$

and

$$\Sigma_{N,\xi}(\psi_0) = \frac{1}{N} \sum_{i=1}^N \xi_i(\varphi_0) \xi_i(\varphi_0)' \xrightarrow{a.s.} \Sigma_\xi(\psi_0). \quad (\text{A.13})$$

Proof. First recall that $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\psi}')'$ with $\boldsymbol{\varphi} = (\boldsymbol{\delta}', \boldsymbol{\gamma}')'$, $\boldsymbol{\delta} = (\mathbf{d}', \boldsymbol{\phi}')' = (\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}')'$ where $\boldsymbol{\phi} = (\boldsymbol{\pi}', \boldsymbol{\beta}')'$ and $\boldsymbol{\psi} = (\omega, \sigma^2, \text{vec}(\mathbf{Q}))'$. Under (28),

$$\xi_i(\varphi_0) = \Delta \mathbf{y}_i - \Delta \mathbf{W}_i \varphi_0 = \mathbf{G}_0 \boldsymbol{\eta}_{0i} + \mathbf{r}_{0i}, \quad (\text{A.14})$$

where $\mathbf{G}_0, \boldsymbol{\eta}_{0i}$, and \mathbf{r}_{0i} denote the values of $\mathbf{G}, \boldsymbol{\eta}_i$ and \mathbf{r}_i evaluated at $\boldsymbol{\psi} = \psi_0$. It is now easily seen that $E_0[\xi_i(\varphi_0)] = \mathbf{0}$, and $\text{Var}[\xi_i(\varphi_0)] = E_0[\xi_i(\varphi_0) \xi_i'(\varphi_0)] = \Sigma_\xi(\psi_0)$. Also under Assumptions 1-6, $\xi_i(\varphi) = \mathbf{G} \boldsymbol{\eta}_i + \mathbf{r}_i$ are independently distributed over i for all values of $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\epsilon$, and $\Delta \mathbf{x}_{it}$ is independently distributed from u_{it} and $\boldsymbol{\eta}_i$. Partition $\Delta \mathbf{W}_i$ as $\Delta \mathbf{W}_i = (\mathbf{I}_T, \Delta \mathbf{X}_i, \Delta \mathbf{y}_{i,-1})$, where \mathbf{I}_T is the identity matrix of order T , $\Delta \mathbf{X}_i$ is given by (43) and $\Delta \mathbf{y}_{i,-1} = (0, \Delta y_{i1}, \dots, \Delta y_{i,T-1})' = \mathbf{L} \Delta \mathbf{y}_i$, where \mathbf{L} and $\Delta \mathbf{y}_i$ are given by (A.5) and (42). Also, using (42) and evaluating it at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we have

$$\Delta \mathbf{y}_i = \mathbf{B}(\gamma_0)^{-1} (\Delta \mathbf{X}_i \phi_0 + \mathbf{d}_0) + \mathbf{B}(\gamma_0)^{-1} \xi_i(\varphi_0), \quad (\text{A.15})$$

where $\mathbf{B}(\gamma)$ is defined by (37). Consider now (A.6), and note that

$$\mathbf{d}_i(\psi, \varphi_0) = \Delta \mathbf{W}_i' \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0) = \begin{pmatrix} \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0) \\ \Delta \mathbf{X}_i \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0) \\ \Delta \mathbf{y}_i' \mathbf{L}' \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0) \end{pmatrix} = \begin{pmatrix} \mathbf{d}_{1i}(\psi, \varphi_0) \\ \mathbf{d}_{2i}(\psi, \varphi_0) \\ d_{3i}(\psi, \varphi_0) \end{pmatrix}. \quad (\text{A.16})$$

Further, using (A.15), write $d_{3i}(\psi, \varphi_0)$ as

$$\begin{aligned} d_{3i}(\psi, \varphi_0) &= \left[\mathbf{B}(\gamma_0)^{-1} (\Delta \mathbf{X}_i \phi_0 + \mathbf{d}_0) + \mathbf{B}(\gamma_0)^{-1} \xi_i(\varphi_0) \right]' \mathbf{L}' \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0) \\ &= (\Delta \mathbf{X}_i \phi_0 + \mathbf{d}_0)' \mathbf{B}(\gamma_0)^{-1} \mathbf{L}' \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0) + \xi_i'(\varphi_0) \mathbf{B}(\gamma_0)^{-1} \mathbf{L}' \Sigma_\xi(\psi)^{-1} \xi_i(\varphi_0). \end{aligned} \quad (\text{A.17})$$

Also under Assumptions 1, 3, and 5, $\Delta \mathbf{X}_i$ and $\xi_i(\varphi_0)$ are cross-sectionally independently distributed, and $E_0[\xi_i(\varphi_0)] = \mathbf{0}$. Hence

$$E_0[\mathbf{d}_{1i}(\psi, \varphi_0)] = \mathbf{0}, \text{ and } E_0[\mathbf{d}_{2i}(\psi, \varphi_0)] = \mathbf{0}, \text{ for all } i, \quad (\text{A.18})$$

and

$$\begin{aligned}
E_0 [d_{3i}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)] &= E_0 \left[\boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0) \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \right] \\
&= \text{tr} \left\{ \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} E_0 [\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0)] \right\} \\
&= \text{tr} \left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \mathbf{L} \mathbf{B}(\gamma_0)^{-1} \right].
\end{aligned}$$

Also, using (38) and (A.5), we have

$$\boldsymbol{\Gamma}(\gamma_0) = \mathbf{L} \mathbf{B}(\gamma_0)^{-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_0^{T-3} & \gamma_0^{T-4} & \cdots & 0 & 0 \\ \gamma_0^{T-2} & \gamma_0^{T-3} & \cdots & 1 & 0 \end{pmatrix}.$$

Hence, $\text{tr} [\mathbf{L} \mathbf{B}(\gamma_0)^{-1}] = 0$, and $E_0 [d_{3i}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)]$ can be written as

$$E_0 [d_{3i}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)] = -\text{tr} \{ [\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) - \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)] \mathbf{C}(\boldsymbol{\psi}, \gamma_0) \} = -\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0), \quad (\text{A.19})$$

where $\mathbf{C}(\boldsymbol{\psi}, \gamma_0) = \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \mathbf{L} \mathbf{B}(\gamma_0)^{-1}$. Using (A.19) and (A.18) now yields (A.7), as required. Result (A.10) then follows immediately, noting that $E_0 [d_{3i}(\boldsymbol{\psi}_0, \boldsymbol{\varphi}_0)] = \text{tr} [\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \mathbf{L} \mathbf{B}(\gamma_0)^{-1}] = \text{tr} [\mathbf{L} \mathbf{B}(\gamma_0)^{-1}] = 0$. To establish (A.11), since $\Delta \mathbf{X}_i$ and $\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)$ are cross-sectionally independent for $i = 1, 2, \dots, N$ it follows that $\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)$ are also independently distributed across i . Hence to show that $\mathbf{b}_N(\boldsymbol{\psi}, \boldsymbol{\varphi}_0) = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)$ converges almost surely to $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E_0 [\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)]$, it is sufficient to show that $\sup_i E_0 \|\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)\|^2 < K$. Consider each of the three terms of $\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)$ in turn. First, from result (A.3) and Liapunov's inequality we have that $E \|\boldsymbol{\xi}_i\|^2 < K < \infty$ and noting that by assumption 7(ii) $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}$ is positive definite uniformly in $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$, then

$$\sup_i E_0 \|\mathbf{d}_{1i}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)\|^2 \leq \left\| \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right\|^2 \sup_i E_0 \|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\|^2 < K. \quad (\text{A.20})$$

Similarly, using in addition result (A.4) we have

$$\sup_i E_0 \|\mathbf{d}_{2i}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)\|^2 \leq \sup_i E \|\Delta \mathbf{X}_i\|^2 \left\| \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right\|^2 \sup_i E_0 \|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\|^2 < K. \quad (\text{A.21})$$

Finally, applying the Minkowski inequality to (A.17) we have

$$\begin{aligned}
\left[E_0 \|d_{3i}(\boldsymbol{\psi}, \boldsymbol{\varphi}_0)\|^2 \right]^{1/2} &\leq \left[E_0 \left\| (\Delta \mathbf{X}_i \boldsymbol{\phi}_0 + \mathbf{d}_0)' \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \right\|^2 \right]^{1/2} \\
&\quad + \left[E_0 \left\| \boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0) \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \right\|^2 \right]^{1/2},
\end{aligned}$$

$$\begin{aligned}
E_0 \left\| (\Delta \mathbf{X}_i \boldsymbol{\phi}_0 + \mathbf{d}_0)' \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \right\|^2 &\leq E_0 \|\Delta \mathbf{X}_i \boldsymbol{\phi}_0 + \mathbf{d}_0\|^2 \left\| \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right\|^2 \\
&\quad \times E_0 \|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\|^2, \\
E_0 \left\| \boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0) \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \right\|^2 &\leq \left\| \mathbf{B}(\gamma_0)'^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right\|^2 E_0 \|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\|^4.
\end{aligned}$$

But $\left\| \mathbf{B}(\gamma_0)^{-1} \mathbf{L}' \boldsymbol{\Sigma}_\xi(\psi)^{-1} \right\|^2 \leq \left\| \boldsymbol{\Sigma}_\xi(\psi)^{-1} \right\|^2 \left\| \mathbf{L} \right\|^2 \left\| \mathbf{B}(\gamma_0)^{-1} \right\|^2$, and it is easily seen that $\left\| \mathbf{L} \right\|^2 = T - 1$, and $\left\| \mathbf{B}(\gamma_0)^{-1} \right\| \leq \sum_{t=1}^T |\gamma_0|^{t-1} < K$. Also, by results of Lemma 2, $\sup_i E_0 \left\| \boldsymbol{\xi}_i(\varphi_0) \right\|^4 < K$, and $\left\| \boldsymbol{\Sigma}_\xi(\psi)^{-1} \right\| < K$, by assumption. Further, $E_0 \left\| \Delta \mathbf{X}_i \boldsymbol{\phi}_0 + \mathbf{d}_0 \right\|^2 \leq \left\| \boldsymbol{\phi}_0 \right\|^2 E \left\| \Delta \mathbf{X}_i \right\|^2 + \left\| \mathbf{d}_0 \right\|^2$ which is uniformly bounded under results (A.4) of Lemma 2, noting that $\boldsymbol{\phi}_0$ and \mathbf{d}_0 are defined on a compact set and are bounded as well. Therefore, $\sup_i E_0 \left\| d_{3i}(\psi, \varphi_0) \right\|^2 < K$. Now using this result together with (A.20) and (A.21) in (A.16) we have

$$\sup_i E_0 \left\| \mathbf{d}_i(\psi, \varphi_0) \right\|^2 = \sup_i E_0 \left\| \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\psi)^{-1} \boldsymbol{\xi}_i(\varphi_0) \right\|^2 < K,$$

which establishes that $\mathbf{d}_i(\psi, \varphi_0)$ is uniformly L_2 -bounded, besides being cross-sectionally independent. Hence,

$$\mathbf{b}_N(\psi, \varphi_0) = N^{-1} \sum_{i=1}^N \mathbf{d}_i(\psi, \varphi_0) \xrightarrow{a.s.} \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E_0 [\mathbf{d}_i(\psi, \varphi_0)] = [\mathbf{0}, \mathbf{0}, -\kappa(\psi, \psi_0)]',$$

which establishes (A.11). Result (A.12) follows from the above by setting $\psi = \psi_0$ and noting from (A.10) that $E_0 [\mathbf{d}_i(\psi_0, \varphi_0)] = \mathbf{0}$. Finally, since $\sup_i E_0 \left\| \boldsymbol{\xi}_i(\varphi_0) \boldsymbol{\xi}_i'(\varphi_0) \right\|^2 < K$, for a finite T (see result (A.3) of Lemma 2), and by assumption $\boldsymbol{\xi}_i(\varphi_0) \boldsymbol{\xi}_i'(\varphi_0)$ are distributed independently over i , then

$$\boldsymbol{\Sigma}_{N,\xi}(\psi_0) = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\xi}_i(\varphi_0) \boldsymbol{\xi}_i(\varphi_0)' \xrightarrow{a.s.} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E_0 [\boldsymbol{\xi}_i(\varphi_0) \boldsymbol{\xi}_i(\varphi_0)'],$$

and result (A.13) follows, since $E_0 [\boldsymbol{\xi}_i(\varphi_0) \boldsymbol{\xi}_i'(\varphi_0)] = \boldsymbol{\Sigma}_\xi(\psi_0)$. ■

Lemma 4 Consider the average log-likelihood function

$$\bar{\ell}_N(\boldsymbol{\theta}) = \bar{\ell}_N(\varphi, \psi) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\xi(\psi)| - \frac{1}{2N} \sum_{i=1}^N \boldsymbol{\xi}_i(\varphi)' \boldsymbol{\Sigma}_\xi(\psi)^{-1} \boldsymbol{\xi}_i(\varphi) \quad (\text{A.22})$$

$\bar{\ell}_N(\boldsymbol{\theta}) = N^{-1} \ell_N(\boldsymbol{\theta})$ and $\ell_N(\boldsymbol{\theta})$ is defined by (34). Then under Assumptions 1-7(ii),(iii) and 8, and the order condition (41), we have

$$\bar{\ell}_N(\boldsymbol{\theta}_0) \xrightarrow{a.s.} -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_\xi(\psi_0)| - \frac{T}{2}, \quad (\text{A.23})$$

and

$$\begin{aligned} \bar{\ell}_N(\boldsymbol{\theta}) &\xrightarrow{a.s.} -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\xi(\psi)| - \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}_\xi(\psi)^{-1} \boldsymbol{\Sigma}_\xi(\psi_0) \right] \\ &\quad - \frac{1}{2} (\varphi - \varphi_0)' \mathbf{A}(\psi) (\varphi - \varphi_0) - (\gamma - \gamma_0) \kappa(\psi, \psi_0), \end{aligned} \quad (\text{A.24})$$

where $\kappa(\psi, \psi_0)$ is defined by (A.8). Also

$$\bar{\ell}_N(\boldsymbol{\theta}_0) - \bar{\ell}_N(\boldsymbol{\theta}) \xrightarrow{a.s.} \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\boldsymbol{\theta}_0) - \bar{\ell}_N(\boldsymbol{\theta})] \geq 0, \quad (\text{A.25})$$

where

$$\begin{aligned} \lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\boldsymbol{\theta}_0) - \bar{\ell}_N(\boldsymbol{\theta})] &= \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}_\xi(\psi)^{-1} \boldsymbol{\Sigma}_\xi(\psi_0) \right] - \frac{1}{2} \log (|\boldsymbol{\Sigma}_\xi(\psi_0)| / |\boldsymbol{\Sigma}_\xi(\psi)|) - \frac{T}{2} \\ &\quad + \frac{1}{2} (\varphi - \varphi_0)' \mathbf{A}(\psi) (\varphi - \varphi_0) + (\gamma - \gamma_0) \kappa(\psi, \psi_0). \end{aligned} \quad (\text{A.26})$$

Proof. Result (A.23) follows by evaluating (A.22) under $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and using (A.13) from Lemma 3. To establish (A.24) we first note that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\epsilon$, $\boldsymbol{\xi}_i(\boldsymbol{\varphi}) = \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) - \Delta \mathbf{W}_i(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)$, and using this result in (A.22) we have

$$\begin{aligned} \bar{\ell}_N(\boldsymbol{\theta}) &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2N} \left[\sum_{i=1}^N [\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) - \Delta \mathbf{W}_i(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)]' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right. \\ &\quad \left. \times [\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) - \Delta \mathbf{W}_i(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)] \right] \\ &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2} \left[\text{tr} \left(\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \left[\frac{1}{N} \sum_{i=1}^N \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)' \right] \right) \right. \\ &\quad \left. - 2(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \mathbf{b}_N(\boldsymbol{\psi}, \boldsymbol{\varphi}_0) + (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \mathbf{A}_N(\boldsymbol{\psi}) (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0) \right], \end{aligned} \quad (\text{A.27})$$

where

$$\mathbf{A}_N(\boldsymbol{\psi}) = \frac{1}{N} \sum_{i=1}^N \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \Delta \mathbf{W}_i, \quad \mathbf{b}_N(\boldsymbol{\psi}, \boldsymbol{\varphi}_0) = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0), \quad (\text{A.28})$$

and $\mathbf{d}_i(\boldsymbol{\psi}, \boldsymbol{\varphi}_0) = \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)$, as defined by (A.6).

Next consider $\mathbf{A}_N(\boldsymbol{\psi}) = \frac{1}{N} \sum_{i=1}^N \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \Delta \mathbf{W}_i$ and note that

$$\sup_i E \left\| \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \Delta \mathbf{W}_i \right\|^2 < \left\| \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right\|^2 \sup_i E \left\| \Delta \mathbf{W}_i \right\|^4 < K,$$

where $\left\| \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \right\| < K$ under condition (ii) of Assumption 7, and $\sup_i E \left\| \Delta \mathbf{W}_i \right\|^4 < K$ by Lemma 2. Also under Assumptions 1, 3, and 5, $\Delta \mathbf{W}_i$ are cross-sectionally independent. This follows since $\Delta \mathbf{x}_i$ are independent across i by Assumption 5 (see also the expression for $\Delta \mathbf{x}_i$ given by (20)), and Δy_{it} being a function of $\Delta \mathbf{x}_{it}$ and ξ_{it} (see (42)) are also cross-sectionally independent noting that ξ_{it} are cross-sectionally independent under Assumptions 1 and 3. Hence, $\mathbf{A}_N(\boldsymbol{\psi}) \xrightarrow{a.s.} \mathbf{A}(\boldsymbol{\psi})$ for every $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$ (see, for example, Davidson (1994, Theorem 19.4)).

Result (A.24) then follows using (A.11) and (A.13) from Lemma 3 in (A.27) evaluated at $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$, respectively. Results (A.25) and (A.26) follow from the sure convergence property of (A.23) and (A.24). That $\lim_{N \rightarrow \infty} E_0 [\bar{\ell}_N(\boldsymbol{\theta}_0) - \bar{\ell}_N(\boldsymbol{\theta})] \geq 0$ follows from the Kullback–Leibler type information inequality and Jensen’s inequality (see for example Section 2.1 of Lee and Yu (2016)). ■

Lemma 5 Consider the average log-likelihood function defined by (53) and (35):

$$\begin{aligned} \bar{\ell}_N(\boldsymbol{\theta}) &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2N} \sum_{i=1}^N \boldsymbol{\xi}_i'(\boldsymbol{\varphi}) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}), \\ \boldsymbol{\xi}_i(\boldsymbol{\varphi}) &= \Delta \mathbf{y}_i - \Delta \mathbf{W}_i \boldsymbol{\varphi}, \end{aligned}$$

and suppose that Assumptions 1-7(ii),(iii) and 8, as well as the order condition (41), hold. Denote the average score function by $\bar{\mathbf{s}}_N(\boldsymbol{\theta}) = \partial \bar{\ell}_N(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. Then

$$\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbf{0}, \quad (\text{A.29})$$

$$\sqrt{N} \bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \rightarrow_d N[\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0)], \quad (\text{A.30})$$

where

$$\mathbf{J}(\boldsymbol{\theta}_0) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E [\boldsymbol{\omega}_i(\boldsymbol{\theta}_0) \boldsymbol{\omega}_i'(\boldsymbol{\theta}_0)], \quad (\text{A.31})$$

$$\boldsymbol{\omega}_i(\boldsymbol{\theta}_0) = \begin{pmatrix} \Delta \mathbf{W}_i' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \\ \boldsymbol{\nu}_i(\boldsymbol{\theta}_0) \end{pmatrix}, \quad (\text{A.32})$$

with the j^{th} element of $\boldsymbol{\nu}_i(\boldsymbol{\theta}_0)$ given by

$$\nu_{ij}(\boldsymbol{\theta}_0) = \frac{1}{2} \boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) - \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right]. \quad (\text{A.33})$$

A consistent estimator of $\mathbf{J}(\boldsymbol{\theta}_0)$ is given by

$$\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\omega}_i(\hat{\boldsymbol{\theta}}) \boldsymbol{\omega}'_i(\hat{\boldsymbol{\theta}}), \quad (\text{A.34})$$

where $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\epsilon} \bar{\ell}_N(\boldsymbol{\theta})$.

Proof. Let $\bar{\mathbf{s}}_N(\boldsymbol{\theta}) = \left(\bar{\mathbf{s}}'_{N,\varphi}(\boldsymbol{\theta}), \bar{\mathbf{s}}'_{N,\psi}(\boldsymbol{\theta}) \right)'$, $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_{n_\psi})'$, where $n_\psi = \dim(\boldsymbol{\psi}) = 1 + Tm - m(m-1)/2$, and note that

$$\begin{aligned} \bar{\mathbf{s}}_{N,\varphi}(\boldsymbol{\theta}) &= \frac{\partial \bar{\ell}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} = \frac{1}{N} \sum_{i=1}^N \Delta \mathbf{W}'_i \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}), \\ \bar{\mathbf{s}}_{N,\psi_j}(\boldsymbol{\theta}) &= \frac{\partial \bar{\ell}_N(\boldsymbol{\theta})}{\partial \psi_j} = -\frac{1}{2} \frac{\partial \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})|}{\partial \psi_j} + \frac{1}{2N} \sum_{i=1}^N \boldsymbol{\xi}'_i(\boldsymbol{\varphi}) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})}{\partial \psi_j} \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}), \end{aligned}$$

for $j = 1, 2, \dots, n_\psi$. Using (A.6), and result (A.12) of Lemma 3, it then readily follows that

$$\bar{\mathbf{s}}_{N,\varphi}(\boldsymbol{\theta}_0) = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i(\boldsymbol{\theta}_0) \xrightarrow{a.s} \mathbf{0}, \quad (\text{A.35})$$

Also

$$E_0 \left[\boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) \right] = \text{tr} \left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right],$$

and using well known results on the partial derivatives of the determinants, we have (see, for example, Magnus and Neudecker (1988, p.151)).

$$\frac{\partial \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)|}{\partial \psi_j} = \text{tr} \left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right],$$

and hence $\bar{\mathbf{s}}_{N,\psi}(\boldsymbol{\theta})$ can be written alternatively as

$$\bar{\mathbf{s}}_{N,\psi_j}(\boldsymbol{\theta}_0) = \frac{\partial \bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial \psi_j} = \frac{1}{N} \sum_{i=1}^N \nu_{ij}.$$

where

$$\nu_{ij}(\boldsymbol{\theta}_0) = \frac{1}{2} \boldsymbol{\xi}'_i(\boldsymbol{\varphi}_0) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}_0) - \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right]. \quad (\text{A.36})$$

Therefore,

$$\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) = \begin{pmatrix} \bar{\mathbf{s}}_{N,\varphi}(\boldsymbol{\theta}_0) \\ \bar{\mathbf{s}}_{N,\psi}(\boldsymbol{\theta}_0) \end{pmatrix} = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i(\boldsymbol{\theta}_0) \\ \frac{1}{N} \sum_{i=1}^N \boldsymbol{\nu}_i(\boldsymbol{\theta}_0) \end{pmatrix},$$

where $\boldsymbol{\nu}_i(\boldsymbol{\theta}_0) = (\nu_{i1}(\boldsymbol{\theta}_0), \nu_{i2}(\boldsymbol{\theta}_0), \dots, \nu_{i,n_\psi}(\boldsymbol{\theta}_0))'$.

$$\sup_i E \|\boldsymbol{\nu}_i(\boldsymbol{\theta}_0)\|^2 = \sup_i E (\boldsymbol{\nu}'_i(\boldsymbol{\theta}_0) \boldsymbol{\nu}_i(\boldsymbol{\theta}_0)) = \sum_{j=1}^{n_\psi} \sup_i E (\nu_{ij}^2(\boldsymbol{\theta}_0)) \leq n_\psi \sup_{i,j} E |\nu_{ij}(\boldsymbol{\theta}_0)|^2,$$

and application of Minkowski's inequality to (A.36) yields

$$\sup_i E |\nu_{ij}(\boldsymbol{\theta}_0)|^2 \leq \frac{1}{4} \left[\left\| \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \right\|^2 \left\| \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right\| \left(\sup_i E \|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\|^4 \right)^{1/2} + |C| \right]^2,$$

where $C = \text{tr} \left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right]$. But under Assumption 7(ii) and noting that n_ψ is finite, we also have $\left\| \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j} \right\| < K$ and $\left\| \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \right\| < K$, and from result (A.3) $\sup_i E \|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\|^4 < K$. Therefore, $\sup_i E \|\boldsymbol{\nu}_i(\boldsymbol{\theta}_0)\|^2 < K$. Also recall that $\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)$ are independently distributed over i , which implies that $\boldsymbol{\nu}_i$ are also independently distributed across i . Therefore, $\boldsymbol{\nu}_i$ have zero means (by construction), are independently distributed over i and have bounded second-order moments, which ensure that $\bar{\mathbf{s}}_{N,\psi}(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbf{0}$, and together with (A.35) yields $\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbf{0}$, as required. Consider now the limiting distribution of $\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0)$ and note that

$$\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) = \begin{pmatrix} \sqrt{N}\bar{\mathbf{s}}_{N,\varphi}(\boldsymbol{\theta}_0) \\ \sqrt{N}\bar{\mathbf{s}}_{N,\psi}(\boldsymbol{\theta}_0) \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{i=1}^N \mathbf{d}_i(\boldsymbol{\theta}_0) \\ \sum_{i=1}^N \boldsymbol{\nu}_i(\boldsymbol{\theta}_0) \end{pmatrix} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\omega}_i(\boldsymbol{\theta}_0),$$

where $\boldsymbol{\omega}_i(\boldsymbol{\theta}_0) = (\mathbf{d}'_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}'_i(\boldsymbol{\theta}_0))'$, and it is already established that $\boldsymbol{\omega}_i(\boldsymbol{\theta}_0)$ are independently distributed over i , have zero means and bounded second-order moments. Therefore, by the Liapounov central limit theorem and the Cramér-Wold device we have²⁵ $\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \rightarrow_d N[\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0)]$, where $\mathbf{J}(\boldsymbol{\theta}_0)$ is given by (A.31), as required. \sqrt{N} consistency of $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})$ for $\mathbf{J}(\boldsymbol{\theta}_0)$ follows from the local consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$ on Θ_ϵ , and the independence of $\boldsymbol{\omega}_i(\boldsymbol{\theta}_0)$ over i . ■

A.2 Proofs of Propositions and Theorems

Proof of Theorem 2. Firstly, under the assumptions of the theorem it suffices to show that $\bar{C}_N(\boldsymbol{\theta}) = -2\bar{\ell}_N(\boldsymbol{\theta}) \xrightarrow{a.s.} \bar{C}(\boldsymbol{\theta})$ uniformly on Θ_ϵ (see Section 6), which together with the result in Proposition 1 and that $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)$ and $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})$ commute deliver local consistency. From results in Lemma 4 (see (A.25) and (A.26)) it follows that $\bar{C}_N(\boldsymbol{\theta}) = -2\bar{\ell}_N(\boldsymbol{\theta}) \xrightarrow{a.s.} \bar{C}(\boldsymbol{\theta})$ for every $\boldsymbol{\theta} \in \Theta_\epsilon$, where

$$\bar{C}_N(\boldsymbol{\theta}) = \bar{C}_N(\boldsymbol{\varphi}, \boldsymbol{\psi}) = T \ln(2\pi) + \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| + \frac{1}{N} \sum_{i=1}^N \boldsymbol{\xi}_i(\boldsymbol{\varphi})' \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi})$$

and

$$\bar{C}(\boldsymbol{\theta}) = \bar{C}(\boldsymbol{\varphi}, \boldsymbol{\psi}) = \chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \mathbf{A}(\boldsymbol{\psi}) (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0) + 2(\gamma - \gamma_0) \kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + C(\boldsymbol{\psi}_0),$$

and the term C does not depend on $\boldsymbol{\theta}$. Since $\bar{\ell}_N(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ by assumption, this pointwise result holds uniformly on Θ_ϵ by the uniform law of large numbers, so long as the dominance condition

$$E \sup_{\boldsymbol{\theta} \in \Theta_\epsilon} \left| \boldsymbol{\xi}'_i(\boldsymbol{\varphi}) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}) + T \ln(2\pi) + \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| \right| < \infty$$

holds; see for example Pötscher and Prucha (2001, Theorem 23).

Since T is finite, it is sufficient to show that

$$E \sup_{\boldsymbol{\theta} \in \Theta_\epsilon} \left| \boldsymbol{\xi}'_i(\boldsymbol{\varphi}) \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}) + \ln |\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| \right| < \infty.$$

²⁵ See, for example, White (2001, Theorem 5.10).

We have that

$$E \sup_{\theta \in \Theta_\epsilon} \left| \xi'_i(\varphi) \Sigma_\xi(\psi)^{-1} \xi_i(\varphi) + \ln |\Sigma_\xi(\psi)| \right| \leq E \sup_{\theta \in \Theta_\epsilon} \left| \xi'_i(\varphi) \Sigma_\xi(\psi)^{-1} \xi_i(\varphi) \right| + \sup_{\psi \in \Theta_\psi} |\ln |\Sigma_\xi(\psi)||.$$

Starting with the second term and using Assumption 7(ii) and the property that for any positive definite real $n \times n$ matrix \mathbf{A} , $\ln |\mathbf{A}| \leq \text{tr}(\mathbf{A}) - n$,

$$\begin{aligned} \sup_{\psi \in \Theta_\psi} |\ln |\Sigma_\xi(\psi)|| &\leq \sup_{\psi \in \Theta_\psi} |\text{tr}[\Sigma_\xi(\psi)] - T| \\ &\leq \sup_{\psi \in \Theta_\psi} \left(\sum_{t=1}^T \lambda_t[\Sigma_\xi(\psi)] \right) + T \\ &\leq T \sup_{\psi \in \Theta_\psi} (\lambda_{\max}[\Sigma_\xi(\psi)]) - T \leq T(c_{\max} - 1) < \infty. \end{aligned}$$

For the first term, defining $\Theta_\varphi = \Theta_\delta \times \mathcal{N}_\epsilon(\varrho_0)$, we have

$$\begin{aligned} E \sup_{\theta \in \Theta_\epsilon} \left| \xi'_i(\varphi) \Sigma_\xi(\psi)^{-1} \xi_i(\varphi) \right| &\leq E \sup_{\theta \in \Theta_\epsilon} \left| \text{tr}[\xi_i(\varphi) \xi'_i(\varphi) \Sigma_\xi(\psi)^{-1}] \right| \\ &\leq E \sup_{\theta \in \Theta_\epsilon} \left\{ \lambda_{\max}[\Sigma_\xi(\psi)^{-1}] \|\xi_i(\varphi)\|^2 \right\} \\ &\leq E \sup_{\psi \in \Theta_\psi} \lambda_{\max}[\Sigma_\xi(\psi)^{-1}] E \sup_{\varphi \in \Theta_\varphi} \|\xi_i(\varphi)\|^2 \\ &\leq E \left(\inf_{\psi \in \Theta_\psi} \lambda_{\min}[\Sigma_\xi(\psi)] \right)^{-1} E \sup_{\varphi \in \Theta_\varphi} \|\xi_i(\varphi)\|^2 \\ &\leq \frac{1}{c_{\min}} E \sup_{\varphi \in \Theta_\varphi} \|\xi_i(\varphi)\|^2. \end{aligned}$$

Further

$$\begin{aligned} E \sup_{\varphi \in \Theta_\varphi} \|\xi_i(\varphi)\|^2 &= E \sup_{\varphi \in \Theta_\varphi} \|\Delta \mathbf{y}_i - \Delta \mathbf{W}_i \varphi\|^2 \\ &\leq E \|\Delta \mathbf{y}_i\|^2 + E \|\Delta \mathbf{W}_i\|^2 \sup_{\varphi \in \Theta_\varphi} \|\varphi\|^2. \end{aligned}$$

But given that Θ_ϵ is a compact set $\sup_{\varphi \in \Theta_\varphi} \|\varphi\|^2$ is bounded. Furthermore, from result (A.4) of Lemma 2 and Liapunov's inequality we have that $E \|\Delta \mathbf{y}_i\|^2 < K < \infty$ and $E \|\Delta \mathbf{W}_i\|^2 < K < \infty$. Since c_{\min}^{-1} is bounded by Assumption 7(ii) it follows that $E \sup_{\theta \in \Theta_\epsilon} \left| \xi'_i(\varphi) \Sigma_\xi(\psi)^{-1} \xi_i(\varphi) \right| < \infty$ and hence the dominance condition holds.

To establish asymptotic normality of $\hat{\boldsymbol{\theta}}$, by application of the mean value theorem to $\bar{\ell}_N(\boldsymbol{\theta})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we first note that

$$\bar{\ell}_N(\boldsymbol{\theta}) - \bar{\ell}_N(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{H}_N(\bar{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \quad (\text{A.37})$$

where $\bar{\mathbf{s}}_N(\boldsymbol{\theta}) = \partial \bar{\ell}_N(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, $\mathbf{H}_N(\boldsymbol{\theta}) = -\partial^2 \bar{\ell}_N(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$, and $\bar{\boldsymbol{\theta}}$ lies on a line segment joining $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. By result (A.29) of Lemma 5, and combining (54) and (55) we have

$$\begin{aligned} \bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) &\xrightarrow{a.s.} \mathbf{0}, \\ 2 [\bar{\ell}_N(\boldsymbol{\theta}_0) - \bar{\ell}_N(\boldsymbol{\theta})] &\xrightarrow{a.s.} \chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + 2(\gamma - \gamma_0) \kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \mathbf{A}(\boldsymbol{\psi}) (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0). \end{aligned}$$

Hence, in view of (A.37) we must also have

$$(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{H}_N(\bar{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \xrightarrow{a.s.} \chi(\boldsymbol{\psi}, \boldsymbol{\psi}_0) + (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \mathbf{A}(\boldsymbol{\psi}) (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0) + 2(\gamma - \gamma_0) \kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0). \quad (\text{A.38})$$

But by Proposition 1 and given that $\Sigma_\xi(\psi_0)$ and $\Sigma_\xi(\psi)$ commute, on Θ_ϵ the right hand side of (A.38) can be equal to zero if and only if $\theta = \theta_0$, and hence we must also have

$$\mathbf{H}_N(\bar{\theta}) \xrightarrow{a.s.} \mathbf{H}(\theta_0), \quad (\text{A.39})$$

where $\mathbf{H}(\theta_0)$ must be a positive definite matrix given by $\mathbf{H}(\theta_0) = \lim_{N \rightarrow \infty} E_0 [-\partial^2 \bar{\ell}_N(\theta_0) / \partial \theta \partial \theta']$. Applying the mean value theorem to $\bar{s}_N(\hat{\theta})$ around $\hat{\theta} = \theta_0$ we have

$$\mathbf{0} = \sqrt{N} \bar{s}_N(\hat{\theta}) = \sqrt{N} \bar{s}_N(\theta_0) - \mathbf{H}_N(\check{\theta}) \sqrt{N}(\hat{\theta} - \theta_0)$$

where $\check{\theta}$ lies on a line segment joining $\hat{\theta}$ and θ_0 . Then,

$$\sqrt{N}(\hat{\theta} - \theta_0) = \mathbf{H}_N^{-1}(\check{\theta}) \left[\sqrt{N} \bar{s}_N(\theta_0) \right].$$

Since $\check{\theta}$ lies between $\hat{\theta}$ and θ_0 and $\hat{\theta}$ is almost surely locally consistent for θ_0 on the set Θ_ϵ so is $\check{\theta}$, and as in (A.39) above $\mathbf{H}_N(\check{\theta}) \xrightarrow{a.s.} \mathbf{H}(\theta_0)$. In addition, using result (A.30) of Lemma 5, we have $\sqrt{N} \bar{s}_N(\theta_0) \rightarrow_d N[\mathbf{0}, \mathbf{J}(\theta_0)]$, where $\mathbf{J}(\theta_0)$ is given by (A.31). Hence

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow_d N(\mathbf{0}, \mathbf{V}_\theta).$$

where \mathbf{V}_θ has the familiar sandwich form

$$\mathbf{V}_\theta = \mathbf{H}^{-1}(\theta_0) \mathbf{J}(\theta_0) \mathbf{H}^{-1}(\theta_0).$$

■

Proof of Theorem 3. Denote the exactly identified estimator of θ (under H_1) by $\hat{\theta}_{m_{\max}}$ with its dimension $n_\theta^* = 3 + T(k+1) + k + (T-2)(T+3)/2$, and the constrained estimator of θ under $H_0 : m = m_0 < T-2$ by $\hat{\theta}_{m_0}$. The latter estimator is obtained under $\mathbf{r}(\theta_0) = \mathbf{0}$, where $\mathbf{r}(\theta_0)$ is the $r_0 \times 1$ vector of restrictions on $\ell_N(\theta)$, the log-likelihood function defined by (34), implied by setting $m = m_0$. Since $\hat{\theta}_{m_0}$ is the constrained estimator of θ under $H_0 : \mathbf{r}(\theta_0) = \mathbf{0}$, by using the results from constrained optimisation (see, for example, Davidson (2000, pp.289-290)), we have

$$\sqrt{N}(\hat{\theta}_{m_0} - \theta_0) \stackrel{a}{\sim} \mathbf{F}_0 \sqrt{N} \bar{s}_N(\theta_0) \quad (\text{A.40})$$

where \bar{s}_N is the score function in Lemma 5 which satisfies

$$\sqrt{N} \bar{s}_N(\theta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{J}_0) \quad (\text{A.41})$$

and

$$\mathbf{F}_0 = \mathbf{H}_0^{-1} - \mathbf{H}_0^{-1} \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1}. \quad (\text{A.42})$$

Also for the unconstrained estimator $\hat{\theta}_{m_{\max}}$, using result (65) in Section 6, we have

$$\sqrt{N}(\hat{\theta}_{m_{\max}} - \theta_0) \stackrel{a}{\sim} \mathbf{H}_0^{-1} \sqrt{N} \bar{s}_N(\theta_0) \quad (\text{A.43})$$

Consider now the mean value expansion of $\ell_N(\hat{\theta}_{m_0})$ around $\hat{\theta} = \hat{\theta}_{m_{\max}}$ given by

$$\begin{aligned} \ell_N(\hat{\theta}_{m_0}) &= \ell_N(\hat{\theta}_{m_{\max}}) + \frac{\partial \ell_N(\hat{\theta}_{m_{\max}})}{\partial \theta} (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}}) \\ &\quad + \frac{1}{2} (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}})' \left(\frac{\partial^2 \ell_N(\bar{\theta})}{\partial \theta \partial \theta'} \right) (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}}), \end{aligned}$$

where $\bar{\theta}$ lies on points between $\hat{\theta}_{m_0}$ and $\hat{\theta}_{m_{\max}}$. Since $\hat{\theta}_{m_{\max}}$ is the unconstrained ML estimator, we have $\partial \ell_N(\hat{\theta}_{m_{\max}})/\partial \theta = \mathbf{0}$, and

$$2 \left[\ell_N(\hat{\theta}_{m_{\max}}) - \ell_N(\hat{\theta}_{m_0}) \right] = \sqrt{N} (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}})' \left(\frac{-1}{N} \frac{\partial^2 \ell_N(\bar{\theta})}{\partial \theta \partial \theta'} \right) \sqrt{N} (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}}). \quad (\text{A.44})$$

Since $\hat{\theta}_{m_{\max}}$ and $\hat{\theta}_{m_0} \xrightarrow{p} \theta_0$ under $m = m_0$, we have $\bar{\theta} \xrightarrow{p} \theta_0$ and

$$2 \left[\ell_N(\hat{\theta}_{m_{\max}}) - \ell_N(\hat{\theta}_{m_0}) \right] \stackrel{a}{\sim} \sqrt{N} (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}})' \mathbf{H}_0 \sqrt{N} (\hat{\theta}_{m_0} - \hat{\theta}_{m_{\max}}). \quad (\text{A.45})$$

Using (A.40) and (A.43), we have the following result:

$$\sqrt{N} (\hat{\theta}_{m_{\max}} - \hat{\theta}_{m_0}) \stackrel{a}{\sim} (\mathbf{H}_0^{-1} - \mathbf{F}_0) \sqrt{N} \bar{\mathbf{s}}_N(\theta_0) = (\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{J}_0^{1/2} \mathbf{z}_n(\theta_0) \quad (\text{A.46})$$

where $\mathbf{z}(\theta_0) = \mathbf{J}_0^{-1/2} \sqrt{N} \bar{\mathbf{s}}_N(\theta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{n_\theta^*})$, which follows from (A.41). Then, using (A.46) in (A.44), we have

$$2 \left[\ell_N(\hat{\theta}_{m_{\max}}) - \ell_N(\hat{\theta}_{m_0}) \right] \stackrel{a}{\sim} \mathbf{z}(\theta_0)' \mathbf{A}_0 \mathbf{z}(\theta_0)$$

where

$$\mathbf{A}_0 = \mathbf{J}_0^{1/2} (\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{H}_0 (\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{J}_0^{1/2} = \mathbf{J}_0^{1/2} \mathbf{H}_0^{-1} \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0^{1/2}. \quad (\text{A.47})$$

Since $\mathbf{J}_0^{1/2} \mathbf{H}_0^{-1}$ is full rank under Theorem 2, then, $\text{rank}(\mathbf{A}_0) = \text{rank}(\mathbf{R}_0) = r_0$, and, hence, only r_0 eigenvalues of \mathbf{A}_0 are non-zero. Furthermore, since \mathbf{A}_0 is symmetric and positive semi-definite, the r_0 eigenvalues of \mathbf{A}_0 are positive, which are denoted by $w_1, w_2, \dots, w_{r_0} > 0$. Then, using the spectral decomposition of \mathbf{A}_0 , we obtain the following result

$$2 \left[\ell_N(\hat{\theta}_{m_{\max}}) - \ell_N(\hat{\theta}_{m_0}) \right] \stackrel{a}{\sim} \sum_{j=1}^{r_0} w_j z_j^2$$

where $z_j \sim \text{IIDN}(0, 1)$, as required for the first part of the theorem under the H_0 .

Consider now the asymptotic distribution of the log-likelihood ratio statistic under the η -local alternative $H_{1N} : \mathbf{r}(\theta_{1N}) = \mathbf{0}$, where $\theta_{1N} = \theta_0 + N^{-\eta/2} \boldsymbol{\kappa}$, with $\boldsymbol{\kappa}' \boldsymbol{\kappa} > 0$. With a slight abuse of notation we continue to denote by $\hat{\theta}_{m_0}$ the constrained estimator of θ now under H_{1N} , and by $\hat{\theta}_{m_{\max}}$ the unconstrained estimator of θ under H_{1N} . First note that (by the mean value theorem around θ_0)

$$\sqrt{N} \bar{\mathbf{s}}_N(\theta_1) = \sqrt{N} \bar{\mathbf{s}}_N(\theta_0) + \sqrt{N} \frac{\partial \bar{\mathbf{s}}_N(\theta_N^*)}{\partial \theta} (\theta_{1N} - \theta_0), \quad (\text{A.48})$$

where the rows of $\partial \bar{\mathbf{s}}_N(\theta)/\partial \theta$ are evaluated at θ^* , points between θ_0 , and θ_{1N} . Also using (A.40) and (A.43) under H_{1N} we have

$$\sqrt{N} (\hat{\theta}_{m_{\max}} - \hat{\theta}_{m_0}) \stackrel{a}{\sim} (\mathbf{H}_{1N}^{-1} - \mathbf{F}_{1N}) \sqrt{N} \bar{\mathbf{s}}_N(\theta_{1N})$$

where $\mathbf{H}_{1N} = \mathbf{H}(\theta_{1N})$ and \mathbf{F}_{1N} is defined analogously to \mathbf{F}_0 given above, namely

$$\mathbf{F}_{1N} = \mathbf{H}_{1N}^{-1} - \mathbf{H}_{1N}^{-1} \mathbf{R}_{1N}' (\mathbf{R}_{1N} \mathbf{H}_{1N}^{-1} \mathbf{R}_{1N}')^{-1} \mathbf{R}_{1N} \mathbf{H}_{1N}^{-1}, \quad (\text{A.49})$$

with $\mathbf{R}_{1N} = \mathbf{R}(\theta_{1N})$. Now using (A.48) we have

$$\begin{aligned} \sqrt{N} (\hat{\theta}_{m_{\max}} - \hat{\theta}_{m_0}) &\stackrel{a}{\sim} (\mathbf{H}_{1N}^{-1} - \mathbf{F}_{1N}) \left[\sqrt{N} \bar{\mathbf{s}}_N(\theta_0) + \sqrt{N} \frac{\partial \bar{\mathbf{s}}_N(\theta_N^*)}{\partial \theta} (\theta_{1N} - \theta_0) \right] \\ &= (\mathbf{H}_{1N}^{-1} - \mathbf{F}_{1N}) \left[\sqrt{N} \bar{\mathbf{s}}_N(\theta_0) + \sqrt{N} \frac{\partial \bar{\mathbf{s}}_N(\theta^*)}{\partial \theta} N^{-\eta/2} \boldsymbol{\kappa} \right]. \end{aligned}$$

Let $\mathbf{z}(\boldsymbol{\theta}_0) = \mathbf{J}_0^{-1/2} \sqrt{N} \bar{\mathbf{s}}_N(\boldsymbol{\theta}_0)$ and note that under H_{1N}

$$\frac{\partial \bar{\mathbf{s}}_N(\bar{\boldsymbol{\theta}}_N^*)}{\partial \boldsymbol{\theta}} = \frac{1}{N} \frac{\partial^2 \ell_N(\boldsymbol{\theta}_N^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \rightarrow_p -\mathbf{H}(\boldsymbol{\theta}_0) = -\mathbf{H}_0.$$

Then

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}}_{m_{\max}} - \hat{\boldsymbol{\theta}}_{m_0} \right) \stackrel{a}{\sim} (\mathbf{H}_{1N}^{-1} - \mathbf{F}_{1N}) \mathbf{J}_0^{1/2} \mathbf{z}(\boldsymbol{\theta}_0) - \sqrt{N} (\mathbf{H}_{1N}^{-1} - \mathbf{F}_{1N}) \mathbf{H}_0 N^{-\eta/2} \boldsymbol{\kappa}, \quad (\text{A.50})$$

where, as noted above, $\mathbf{z}(\boldsymbol{\theta}_0) \stackrel{a}{\sim} N(\mathbf{0}, \mathbf{I}_{n_\theta^*})$. The first component of (A.50) relates to the null hypothesis, whilst the second component relates to the "non-centrality" parameter which diverges since $\eta < 1$. Note also that, $\mathbf{H}_{1N} = \mathbf{H}(\boldsymbol{\theta}_0 + N^{-\eta/2} \boldsymbol{\kappa})$ and $\mathbf{R}_{1N} = \mathbf{R}(\boldsymbol{\theta}_0 + N^{-\eta/2} \boldsymbol{\kappa})$, and converge to \mathbf{H}_0 and \mathbf{R}_0 , respectively, which in view of (A.49), also implies that $\mathbf{F}_{1N} \rightarrow \mathbf{F}_0$, as $N \rightarrow \infty$. Using (A.50) in (A.44) we now have

$$\begin{aligned} \mathcal{LR}_N &= 2 \left[\ell_N(\hat{\boldsymbol{\theta}}_{m_{\max}}) - \ell_N(\hat{\boldsymbol{\theta}}_{m_0}) \right] \stackrel{a}{\sim} \\ &\left[(\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{J}_0^{1/2} \mathbf{z}(\boldsymbol{\theta}_0) - N^{\frac{(1-\eta)}{2}} (\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{H}_0 \boldsymbol{\kappa} \right]' \mathbf{H}_0 \\ &\times \left[(\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{J}_0^{1/2} \mathbf{z}(\boldsymbol{\theta}_0) - N^{\frac{(1-\eta)}{2}} (\mathbf{H}_0^{-1} - \mathbf{F}_0) \mathbf{H}_0 \boldsymbol{\kappa} \right]. \end{aligned}$$

Recalling that $(\mathbf{H}_0^{-1} - \mathbf{F}_0) = \mathbf{H}_0^{-1} \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1}$ (see (A.42)), \mathcal{LR}_N can also be written as

$$\mathcal{LR}_N \stackrel{a}{\sim} \mathbf{z}(\boldsymbol{\theta}_0)' \mathbf{S}_a \mathbf{z}(\boldsymbol{\theta}_0) - 2N^{\frac{(1-\eta)}{2}} \boldsymbol{\kappa}' \mathbf{S}_b' \mathbf{z}(\boldsymbol{\theta}_0) + N^{(1-\eta)} \boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa}, \quad (\text{A.51})$$

where

$$\begin{aligned} \mathbf{S}_a &= \mathbf{J}_0^{1/2} \mathbf{H}_0^{-1} \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0^{1/2}, \\ \mathbf{S}_b' &= \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0^{1/2}, \\ \mathbf{S}_c &= \mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0. \end{aligned}$$

Under the assumptions of the theorem, \mathbf{H}_0 is positive definite and \mathbf{R}_0 is full rank and so

$$\lambda_{\min} \left(\mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \right) > 0,$$

and since $\boldsymbol{\kappa}' \boldsymbol{\kappa} > 0$, then

$$\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa} > \boldsymbol{\kappa}' \boldsymbol{\kappa} \lambda_{\min} \left(\mathbf{R}_0' (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}_0')^{-1} \mathbf{R}_0 \right) > 0. \quad (\text{A.52})$$

Recall also that \mathbf{J}_0 is positive definite. Then \mathbf{S}_a is positive semi-definite with r_0 non-zero eigenvalues which we denote by w_i^* for $i = 1, 2, \dots, n$. It is clear that under $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, \mathbf{S}_a coincides with \mathbf{A}_0 given by (A.47) and $w_i^* = w_i$. In the present context it is still the case that $\mathbf{z}(\boldsymbol{\theta}_0)' \mathbf{S}_a \mathbf{z}(\boldsymbol{\theta}_0) = \sum_{j=1}^{r_0} w_j^* z_j^2$ which is a weighted average of chi-squared variates and is stochastically bounded, namely $\mathbf{z}(\boldsymbol{\theta}_0)' \mathbf{S}_a \mathbf{z}(\boldsymbol{\theta}_0) = O_p(1)$. Post-multiplying both sides of (A.51) by $N^{\frac{-(1-\eta)}{2}}$, and rearranging the terms we have

$$\begin{aligned} &N^{\frac{-(1-\eta)}{2}} \mathcal{LR}_N - N^{\frac{(1-\eta)}{2}} (\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa}) \stackrel{a}{\sim} -2\boldsymbol{\kappa}' \mathbf{S}_b' \mathbf{z}(\boldsymbol{\theta}_0) + N^{\frac{-(1-\eta)}{2}} \mathbf{z}(\boldsymbol{\theta}_0)' \mathbf{S}_a \mathbf{z}(\boldsymbol{\theta}_0) \\ &= -2\boldsymbol{\kappa}' \mathbf{S}_b' \mathbf{z}(\boldsymbol{\theta}_0) + o_p(1), \end{aligned}$$

since $\eta < 1$, and $N^{\frac{-(1-\eta)}{2}} \mathbf{z}(\boldsymbol{\theta}_0)' \mathbf{S}_a \mathbf{z}(\boldsymbol{\theta}_0) \rightarrow_p 0$, with $N \rightarrow \infty$. Furthermore, since $\mathbf{z}(\boldsymbol{\theta}_0) \sim N(\mathbf{0}, \mathbf{I}_{n_\theta^*})$, it then follows that

$$\frac{N^{\frac{-(1-\eta)}{2}} \mathcal{LR}_N - N^{\frac{(1-\eta)}{2}} (\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa})}{2\sqrt{\boldsymbol{\kappa}' \mathbf{S}_b \mathbf{S}_b' \boldsymbol{\kappa}}} \stackrel{a}{\sim} N(0, 1). \quad (\text{A.53})$$

Note also that

$$\mathbf{S}'_b \mathbf{S}_b = \mathbf{R}'_0 (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}'_0)^{-1} \mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{J}_0 \mathbf{H}_0^{-1} \mathbf{R}'_0 (\mathbf{R}_0 \mathbf{H}_0^{-1} \mathbf{R}'_0)^{-1} \mathbf{R}_0,$$

and $\mathbf{S}'_b \mathbf{S}_b$ is a positive definite matrix since by assumption $\text{rank}(\mathbf{R}_0) = r_0$, and \mathbf{H}_0 and \mathbf{J}_0 are positive definite matrices. Then

$$\boldsymbol{\kappa}' \mathbf{S}'_b \mathbf{S}_b \boldsymbol{\kappa} > \boldsymbol{\kappa}' \boldsymbol{\kappa} \lambda_{\min}(\mathbf{S}'_b \mathbf{S}_b) > 0. \quad (\text{A.54})$$

■

Proof of Corollary 1. The type II error probability of testing $H_0 : \mathbf{r}(\boldsymbol{\theta}_0) = 0$ against η -local alternatives, H_{1N} , is given by

$$\beta_N = \Pr [\mathcal{LR}_N \leq c_N^2(r_0) | H_{1N}],$$

which can be written equivalently as (recall that $\boldsymbol{\kappa}' \mathbf{S}_b \mathbf{S}'_b \boldsymbol{\kappa} > 0$)

$$\beta_N = \Pr \left[\frac{N^{-\frac{(1-\eta)}{2}} \mathcal{LR}_N - N^{\frac{(1-\eta)}{2}} (\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa})}{2\sqrt{\boldsymbol{\kappa}' \mathbf{S}_b \mathbf{S}'_b \boldsymbol{\kappa}}} \leq \frac{N^{-\frac{(1-\eta)}{2}} c_N^2(r_0) - N^{\frac{(1-\eta)}{2}} (\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa})}{2\sqrt{\boldsymbol{\kappa}' \mathbf{S}_b \mathbf{S}'_b \boldsymbol{\kappa}}} | H_{1N} \right]$$

Now using result (71) of Theorem 3 and taking limits as $N \rightarrow \infty$ we have (noting that $\eta < 1$)

$$\begin{aligned} \lim_{N \rightarrow \infty} \beta_N &= \lim_{N \rightarrow \infty} \Phi \left(\frac{-N^{\frac{(1-\eta)}{2}} \boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa} + N^{-\frac{(1-\eta)}{2}} c_N^2(r_0)}{2\sqrt{\boldsymbol{\kappa}' \mathbf{S}'_b \mathbf{S}_b \boldsymbol{\kappa}}} \right) \\ &= \lim_{N \rightarrow \infty} \Phi \left(-\frac{N^{\frac{(1-\eta)}{2}} \boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa} \left(1 - \frac{N^{-(1-\eta)} c_N^2(r_0)}{(\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa})} \right)}{2\sqrt{\boldsymbol{\kappa}' \mathbf{S}'_b \mathbf{S}_b \boldsymbol{\kappa}}} \right), \end{aligned}$$

where $\boldsymbol{\kappa}' \mathbf{S}_c \boldsymbol{\kappa} / \sqrt{\boldsymbol{\kappa}' \mathbf{S}'_b \mathbf{S}_b \boldsymbol{\kappa}} > 0$, which follows using (A.52) and (A.54) of Theorem 3. The desired result, $\lim_{N \rightarrow \infty} (\beta_N) = 0$, now follows since by assumption $\eta < 1$, and $N^{-(1-\eta)} c_N^2(r_0) \rightarrow 0$ as $N \rightarrow \infty$. ■

Proof of Proposition 2. Consider the type I error of the test and note that

$$\alpha_N = \Pr (\mathcal{LR}_N > c_N^2(r_0) | H_0) = \Pr \left(\sum_{i=1}^{r_0} w_i z_i^2 > c_N^2(r_0) \right),$$

where $z_i \sim \text{IIDN}(0, 1)$. Using Lemma A1 of the theory supplement to Chudik et al. (2018) we have that

$$\alpha_N = \Pr \left(\sum_{i=1}^{r_0} w_i z_i^2 > c_N^2(r_0) \right) \leq \sum_{i=1}^{r_0} \Pr (w_i z_i^2 > r_0^{-1} c_N^2(r_0)).$$

Therefore, since $w_i > 0$

$$\alpha_N \leq \sum_{i=1}^{r_0} \Pr (z_i^2 > (r_0 w_i)^{-1} c_N^2(r_0)) \leq r_0 \sup_i \Pr (z_i^2 > \theta_i^2 c_N^2(r_0)), \quad (\text{A.55})$$

where $\theta_i^2 = (r_0 w_i)^{-1} > 0$. But since $z_i \sim N(0, 1)$, then

$$\begin{aligned} \Pr (z_i^2 > \theta_i^2 c_N^2(r_0)) &= 1 - \Pr (-\theta_i |c_N(r_0)| \leq z_i \leq \theta_i |c_N(r_0)|) \\ &= 2\Phi(-\theta_i |c_N(r_0)|). \end{aligned}$$

Using this result in (A.55) we have

$$\alpha_N \leq 2r_0 \sup_i \Phi(-\theta_i |c_N(r_0)|) = 2r_0 \Phi(-\theta_{\min} |c_N(r_0)|) = 2h[1 - \Phi(\theta_{\min} |c_N(r_0)|)],$$

where $\theta_{\min}^2 = r_0^{-1} \inf_i w_i^{-1} = r_0^{-1} w_1^{-1} > 0$. Hence $\Phi(\theta_{\min} |c_N(r_0)|) \leq 1 - \alpha_N/2r_0$, and

$$\alpha_N \leq 2r_0 [1 - \Phi(\theta_{\min} |c_N(r_0)|)] = 2r_0 \Phi(-\theta_{\min} |c_N(r_0)|).$$

Since $\theta_{\min} |c_N(r_0)| > 0$, then by (A.1) in Lemma 1 of Bailey et al. (2019, BPS)

$$\Phi(-\theta_{\min} |c_N(r_0)|) \leq (1/2) \exp \left[-\frac{1}{2} \theta_{\min}^2 c_N^2(r_0) \right],$$

and hence

$$\alpha_N \leq r_0 \exp \left[-\frac{1}{2} \theta_{\min}^2 c_N^2(r_0) \right] = r_0 \exp \left[-\frac{c_N^2(r_0)}{2r_0 w_1} \right].$$

Since w_1 is bounded and strictly positive, it then follows that $\lim_{N \rightarrow \infty} \alpha_N = 0$, so long as $c_N^2(r_0) \rightarrow \infty$. Furthermore, due to the monotonicity property of $\Phi(\cdot)$ we have that (for α_N sufficiently small) $\theta_{\min} |c_N(r_0)| \leq \Phi^{-1} \left(1 - \frac{\alpha_N}{2r_0} \right)$, or $c_N^2(r_0) \leq \theta_{\min}^{-2} \left[\Phi^{-1} \left(1 - \frac{\alpha_N}{2r_0} \right) \right]^2$. By Lemma 3 of BPS, $\left[\Phi^{-1} \left(1 - \frac{\alpha_N}{2r_0} \right) \right]^2 \leq 2 \ln \left(\frac{r_0}{\alpha_N} \right)$, and hence it also follows that

$$c_N^2(r_0) \leq 2\theta_{\min}^{-2} \ln \left(\frac{r_0}{\alpha_N} \right) = 2w_1 r_0 \ln \left(\frac{r_0}{\alpha_N} \right). \quad (\text{A.56})$$

■

Proof of Theorem 4. To show that \hat{m} is almost surely (locally) consistent for the true number of factors m_0 on Θ_ϵ , we will show that $\lim_{N \rightarrow \infty} \Pr(\hat{m} = m_0) = 1$ on Θ_ϵ . Consider the event $\{\hat{m} > m_0\}$ on Θ_ϵ . For this event to be true it must be the case that for some $t \in \{1, 2, \dots, T-2\}$, at a certain stage in the sequential estimation, when testing $H_0: m = m_0 = t-1$ against $H_1: m = m_{\max} = T-2$, the null hypothesis of the true number of factors is rejected. That is,

$$\begin{aligned} \Pr(\hat{m} > m_0) &\leq P(\exists t, m_0 \text{ is rejected} | H_0) \\ &\leq \sum_{t=1}^{m_0+1} \Pr(\mathcal{LR}_N > c_N^2(r_0) | H_0), \end{aligned} \quad (\text{A.57})$$

where $c_N^2(r_0)$ denotes the critical value of the test recalling that r_0 is the number of over-identified restrictions imposed under the H_0 , given by (67). For any given t , using the result in Proposition 2 for $c_N^2(r_0) \rightarrow \infty$ as $N \rightarrow \infty$, we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \alpha_N &= \lim_{N \rightarrow \infty} \Pr(\mathcal{LR}_N > c_N^2(r_0) | H_0) \\ &= \lim_{N \rightarrow \infty} \Pr \left(\sum_{i=1}^{r_0} w_i z_i^2 > c_N^2(r_0) \right) = 0, \end{aligned} \quad (\text{A.58})$$

(recall that $z_i \sim IIDN(0, 1)$). Then, from (A.57) using (A.58) it follows that

$$\Pr(\hat{m} > m_0) \leq (m_0 + 1) \max_{1 \leq t \leq m_0+1} \Pr(\mathcal{LR}_N > c_N^2(r_0) | H_0) \rightarrow 0 \quad (\text{A.59})$$

as $N \rightarrow \infty$ on Θ_ϵ . Next consider the event $\{\hat{m} < m_0\}$ on Θ_ϵ , and note that

$$\begin{aligned} \Pr(\hat{m} < m_0) &= \Pr \left(\max_{1 \leq t \leq T-2} \mathcal{LR}_N \leq c_N^2(r_0) | H_0 \text{ is false} \right) \\ &\leq \sum_{t=1}^{T-2} \Pr(\mathcal{LR}_N \leq c_N^2(r_0) | H_0 \text{ is false}). \end{aligned} \quad (\text{A.60})$$

Using result (74) of Corollary 1, for $N^{-(1-\eta)}c_N^2(r_0) \rightarrow 0$ as $N \rightarrow \infty$ so long as $\eta < 1$, we have for the probability of type II error of the test that

$$\lim_{N \rightarrow \infty} \beta_N = \lim_{N \rightarrow \infty} \Pr(\mathcal{LR}_N \leq c_N^2(r_0) | H_0 \text{ is false}) = 0. \quad (\text{A.61})$$

Then similar to the $\{\hat{m} > m_0\}$ case, from (A.60) and using (A.61) it readily follows that $\lim_{N \rightarrow \infty} \Pr(\hat{m} < m_0) = 0$ on Θ_ϵ which together with (A.59) establishes the desired result. ■

References

- Acemoglu, D. and Naidu, S. and Restrepo, P. and Robinson, J.A. (2019). Democracy does cause growth. *Journal of Political Economy* 127, 47–100.
- Ahn, S.C. (2015). Comment on "IV estimation of panels with factor residuals" by D. Robertson and V. Sarafidis. *Journal of Econometrics* 185, 542–544.
- Ahn, S.C. and Horenstein, A.R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203–1227.
- Ahn, S.C. and Lee, H.Y. and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101, 219–255.
- (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174, 1–14.
- Ahn, S.C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5–27.
- Allen, D., and Chan, F., and McAleer, M. and Peiris, M.S. (2008). Finite sample properties of the QMLE for the log-ACD model: Application to Australian stocks. *Journal of Econometrics* 147, 163–185.
- Anderson, T.W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598–606.
- Arellano, M. and Bond, S. (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 Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29–51.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.
- (2013). Likelihood approach to dynamic panel models with interactive effects. mimeo.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- (2013). Principal components estimation and identification of static factors. *Journal of Econometrics* 176, 18–29.
- Bailey, N. and Pesaran, M.H. and Smith, L.V. (2019). A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics* 208, 507–534.

- Baltagi, B. (2006). Estimating an economic model of crime using panel data from north carolina. *Journal of Applied Econometrics* 21, 543–547.
- Barro, R.J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics* 106, 407–443.
- Bauer, P., Pötscher, B.M. and Hackl, P. (1988). Model selection by multiple test procedures. *Statistics* 19, 39–44.
- Bhargava, A. and Sargan, J. (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635–1659.
- Binder, M. and Hsiao, C. and Pesaran, M.H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795–837.
- Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.
- Bonferroni, C.E. (1936). Statistical theory of classes and calculation of probabilities. *Pub R Ist Superiore Sci Econ Commerc Firenze* 8, 36–62.
- Caselli, F. and Esquível, G. and Lefort, F. (1996). Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth* 1, 363–389.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.
- (1984). Panel data. In *Handbook of Econometrics*, Volume 2, Chapter 22, pp. 1248–1318. North-Holland.
- Chudik, A. and Kapetanios, G. and Pesaran, M.H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica* 86, 1479–1512.
- Chudik, A. and Pesaran, M.H. (2021). An augmented Anderson-Hsiao estimator for dynamic short-T panels. *Mimeo*.
- Chudik, A. and Pesaran, M.H. and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal* 14, C45–C90.
- Cornwell, C. and Trumbull, W.N. (1994). Estimating the economic model of crime with panel data. *The Review of Economics and Statistics* 76, 360–366.
- Cragg, J.G. and Donald, S.G. (1997). Inferring the rank of a matrix. *Journal of Econometrics* 76, 223–250.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.
- (2000). *Econometric Theory*. Wiley-Blackwell.
- Han, H. and Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Hayakawa, K. (2012). GMM estimation of a short dynamic panel data model with interactive fixed effects. *Journal of the Japan Statistical Society* 42, 109–123.

- (2016). Identification problem of GMM estimators for short panel data models with interactive fixed effects. *Economics Letters* 139, 22–26.
- Hayakawa, K. and Pesaran, M.H. (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics* 188, 111–134.
- Hayashi, K. and Bentler, P.M. and Yuan, K-H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal* 14, 505–526.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- Holtz-Eakin, D. and Newey, W.K. and Rosen, H.S. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56, 1371–1395.
- Hsiao, C. and Pesaran, M.H. and Tahmiscioglu, K.A. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107–150.
- Hsiao, C. and Tahmiscioglu, K.A. (2008). Estimation of dynamic panel data models with both individual and time specific effects. *Journal of Statistical Planning and Inference* 138, 2698–2721.
- Islam, N. (1995). Growth empirics: A panel data approach. *The Quarterly Journal of Economics* 110, 1127–1170.
- Jacob, J. and Osang, T. (2018). Democracy and growth: A dynamic panel data study. *The Singapore Economic Review*, 1–40.
- Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics* 28, 397–409.
- Kapetanios, G. and Pesaran, M.H. and Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics* 160, 326–348.
- Kristensen, D. and Rahbek, A.C. (2010). Likelihood-based inference for cointegration with nonlinear error-correction. *Journal of Econometrics* 158, 78–94.
- Lee, K. and Pesaran, M.H. and Smith, R.P. (1997). Growth and convergence in a multi-country empirical stochastic solow model. *Journal of Applied Econometrics* 12, 358–392.
- Lee, K. and Pesaran, M.H. and Smith, R. (1998). Growth empirics: A panel data approach—a comment. *The Quarterly Journal of Economics* 113, 319–323.
- Lee, L.-F. and Yu, J. (2016). Identification of spatial durbin panel models. *Journal of Applied Econometrics* 31, 133–162.
- Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.
- Lütkepohl, H. (1996). *Handbook of matrices*. New York: John Wiley.
- Magnus, J.R. and Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Oxford: John Wiley & Sons.
- Mankiw, N.G. and Romer, D. and Weil, D.N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics* 107, 407–437.

- Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics* 31, 451–472.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* 46, 69–85.
- Nauges, C. and Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique* 70, 53–75.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92, 1004–1016.
- Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with cross section dependence. *Econometrica* 74, 967–1012.
- Pesaran, M.H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182–202.
- Phillips, P.C.B. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 127, 162–188.
- Pötscher, B.M. and Prucha, I.R. (2001). Basic elements of asymptotic theory. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, pp. 201–229. Oxford: Blackwell.
- Robertson, D. and Sarafidis, V. (2015). IV estimation of panels with factor residuals. *Journal of Econometrics* 185, 526–541.
- Sala-i-Martin, X.X. (1996). The classical approach to convergence analysis. *The Economic Journal* 106, 1019–1036.
- Sarafidis, V. and Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal* 12, 62–81.
- Sarafidis, V. and Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31, 483–531.
- Somé, J. and Pasali, S. and Kaboine, M. (2019). Exploring the impact of healthcare on economic growth in Africa. *Applied Economics and Finance* 6, 45–57.
- White, H. (2001). *Asymptotic Theory for Econometricians*. San Diego: Academic Press.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25–51.

Online Supplement for
Short T Dynamic Panel Data Models with Individual, Time and Interactive Effects

Kazuhiko Hayakawa
Hiroshima University

M. Hashem Pesaran
Department of Economics
University of Southern California, USA, and Trinity College, Cambridge

L. Vanessa Smith
University of York

July 2021

S.1 Introduction

This supplement is organised as follows: Section S.2 provides the derivations for the rank conditions associated with the quasi-differenced GMM estimators given in the related literature section of the paper. Section S.3 outlines the eigenvalue approach used for computing the TQML estimator. Section S.4 gives the derivations of the initial values used for the Monte Carlo (MC) analysis. Sections S.5 and S.6 provide details for the computation of the Bai-QML and GMM estimators, respectively. Sections S.7 and S.8 give additional MC results for the stationary and unit root cases, respectively. To save space the results for the ARX(1) model are given only for the case where $\sigma_v^2 = 1$. The results for other values, $\sigma_v^2 = \{0.5, 1.5\}$, are very similar and are available upon request.

Section S.9 gives the details of the MC experiments we carried out for the robustness analysis and the associated results, covering the effects of initial values deviating from the steady state distribution (applicable only for the stationary case), the use of alternative p-values ($p = 0.01$, $p = 0.10$) in implementing the MTLR test, allowing for non-zero correlation of the factor loadings and the regressors, and for weakly cross-correlated factor loadings. The last three experiments are presented for the stationary case. Qualitatively similar results were obtained for the unit root case and are available upon request. All results are given for $\beta_0 = 1$ and are based on 2000 replications. Also, all MC results are obtained using the Multiple Testing Likelihood Ratio (MTLR) test for selecting the number of factors with $p = 0.05$ unless otherwise stated. Lastly, Section S.10 discusses the case of time series heteroskedasticity in the idiosyncratic errors.

S.2 Rank conditions for quasi-differenced GMM estimators

Here we consider the rank conditions with respect to the moment conditions $E[\mathbf{m}_N(\boldsymbol{\theta}_0)] = 0$, defined by (7) in the paper, where

$$\mathbf{m}_N(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^N \mathbf{z}_i \nu_{i3}(\boldsymbol{\theta}),$$

with $\mathbf{z}_i = (\mathbf{w}_i', \mathbf{x}_i')'$, $\mathbf{w}_i = (y_{i0}, y_{i1})'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})'$, and

$$\nu_{i3}(\boldsymbol{\theta}) = y_{i3} - (b_3 + \gamma) y_{i2} + b_3 \gamma y_{i1} - \beta x_{i3} + b_3 \beta x_{i2}.$$

To simplify the notations we denote b_3 as b , so that $\boldsymbol{\theta} = (\gamma, b, \beta)'$. Following standard results from the GMM literature (see, for example, Chapter 10 of Pesaran (2015)) for identification it is required that \mathbf{S}_N (5×5 matrix) and \mathbf{D}_N (5×3 matrix) defined by

$$\nu_{i3}(\boldsymbol{\theta}) = y_{i3} - (b_3 + \gamma) y_{i2} + b_3 \gamma y_{i1} - \beta x_{i3} + b_3 \beta x_{i2},$$

$$\mathbf{D}_N(\boldsymbol{\theta}) = \frac{\partial \mathbf{m}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \text{ and } \mathbf{S}_N(\boldsymbol{\theta}) = N \mathbf{m}_N(\boldsymbol{\theta}) \mathbf{m}_N'(\boldsymbol{\theta}),$$

are full rank matrices and that $\mathbf{S} = \lim_{N \rightarrow \infty} E_0[\mathbf{S}_N(\boldsymbol{\theta}_0)]$ is positive definite, and $\mathbf{D} = \lim_{N \rightarrow \infty} E_0[\mathbf{D}_N(\boldsymbol{\theta}_0)]$ has full column rank. To derive \mathbf{S} and \mathbf{D} note that

$$\begin{aligned} \mathbf{D}_N(\boldsymbol{\theta}) &= N^{-1} \sum_{i=1}^N \mathbf{z}_i \frac{\partial \nu_{i3}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &= - \left(N^{-1} \sum_{i=1}^N \mathbf{z}_i (y_{i2} - b y_{i1}), \quad N^{-1} \sum_{i=1}^N \mathbf{z}_i (y_{i2} - \gamma y_{i1} - \beta x_{i2}), \quad N^{-1} \sum_{i=1}^N \mathbf{z}_i (x_{i3} - b x_{i2} \beta) \right) \end{aligned}$$

and

$$\mathbf{S}_N(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^N \sum_{j=1}^N \nu_{i3}(\boldsymbol{\theta}) \nu_{j3}(\boldsymbol{\theta}) \mathbf{z}_i \mathbf{z}_j'. \quad (\text{S.2})$$

Consider first the limit of $\mathbf{S}_N(\boldsymbol{\theta})$, and note that under the assumption of conditional cross sectional independence we have

$$\begin{aligned}\mathbf{S}_N(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{S} &= \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E(\mathbf{z}_i \mathbf{z}_i') E(\varepsilon_{i3} - b\varepsilon_{i2})^2 \\ \mathbf{S} &= (1 + b^2) \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{z}_i \mathbf{z}_i').\end{aligned}$$

Also

$$E(\mathbf{z}_i \mathbf{z}_i') = \begin{pmatrix} E(\mathbf{w}_i \mathbf{w}_i') & E(\mathbf{w}_i \mathbf{x}_i') \\ E(\mathbf{x}_i \mathbf{w}_i') & E(\mathbf{x}_i \mathbf{x}_i') \end{pmatrix},$$

where

$$\begin{aligned}\mathbf{w}_i &= \begin{pmatrix} y_{i0} \\ y_{i1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + \varepsilon_{i0} \\ \boldsymbol{\pi}'_1 \mathbf{x}_i + \lambda_i(\gamma f_0 + f_1) + \gamma \varepsilon_{i0} + \varepsilon_{i1} \end{pmatrix}, \\ \boldsymbol{\pi}_1 &= \gamma \boldsymbol{\pi}_0 + \beta \mathbf{e}_1,\end{aligned}\tag{S.3}$$

and \mathbf{e}_s is a 3×1 vector of zeros except for its s^{th} element which is unity.

$$E(\mathbf{w}_i \mathbf{w}_i') = \begin{pmatrix} \boldsymbol{\pi}'_0 E(\mathbf{x}_i \mathbf{x}_i') \boldsymbol{\pi}_0 + f_0^2 \lambda_i^2 + \sigma_i^2 & \boldsymbol{\pi}'_0 E(\mathbf{x}_i \mathbf{x}_i') \boldsymbol{\pi}_1 + f_0(\gamma f_0 + f_1) \lambda_i^2 + \gamma \sigma_i^2 \\ \boldsymbol{\pi}'_1 E(\mathbf{x}_i \mathbf{x}_i') \boldsymbol{\pi}_1 + (\gamma f_0 + f_1)^2 \lambda_i^2 + \sigma_i^2(1 + \gamma^2) \end{pmatrix}.$$

$$\begin{aligned}E(\mathbf{w}_i \mathbf{x}_i') &= \begin{pmatrix} E(\boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + \varepsilon_{i0}) \mathbf{x}_i' \\ E[\boldsymbol{\pi}'_1 \mathbf{x}_i + \lambda_i(\gamma f_0 + f_1) + \gamma \varepsilon_{i0} + \varepsilon_{i1}] \mathbf{x}_i' \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\pi}'_0 E(\mathbf{x}_i \mathbf{x}_i') \\ \boldsymbol{\pi}'_1 E(\mathbf{x}_i \mathbf{x}_i') \end{pmatrix}.\end{aligned}$$

Let

$$\mathbf{A} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 (\mathbf{x}_i \mathbf{x}_i'), \quad d_{\lambda\sigma} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 \lambda_i^2 \geq 0, \quad d_{\sigma\sigma} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 > 0,$$

$$N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{w}_i \mathbf{w}_i') = \lim_{N \rightarrow \infty} \begin{pmatrix} \boldsymbol{\pi}'_0 \mathbf{A} \boldsymbol{\pi}_0 + f_0^2 d_{\lambda\sigma} + d_{\sigma\sigma} & \boldsymbol{\pi}'_0 \mathbf{A} \boldsymbol{\pi}_1 + f_0(\gamma f_0 + f_1) d_{\lambda\sigma} + \gamma d_{\sigma\sigma} \\ \boldsymbol{\pi}'_1 \mathbf{A} \boldsymbol{\pi}_1 + f_0(\gamma f_0 + f_1) d_{\lambda\sigma} + \gamma d_{\sigma\sigma} & \boldsymbol{\pi}'_1 \mathbf{A} \boldsymbol{\pi}_1 + (\gamma f_0 + f_1)^2 d_{\lambda\sigma} + (1 + \gamma^2) d_{\sigma\sigma} \end{pmatrix},$$

$$N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{w}_i \mathbf{x}_i') = \begin{pmatrix} \boldsymbol{\pi}'_0 \mathbf{A} \\ \boldsymbol{\pi}'_1 \mathbf{A} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}'_0 \mathbf{A} \\ \gamma \boldsymbol{\pi}'_0 \mathbf{A} + \beta \mathbf{e}'_1 \mathbf{A} \end{pmatrix}$$

and note that

$$\begin{aligned}\mathbf{S} &= (1 + b^2) \lim_{N \rightarrow \infty} \begin{pmatrix} N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{w}_i \mathbf{w}_i') & N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{w}_i \mathbf{x}_i') \\ N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{x}_i \mathbf{w}_i') & N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{x}_i \mathbf{x}_i') \end{pmatrix} \\ &= (1 + b^2) \begin{pmatrix} \boldsymbol{\pi}'_0 \mathbf{A} \boldsymbol{\pi}_0 + f_0^2 d_{\lambda\sigma} + d_{\sigma\sigma} & \boldsymbol{\pi}'_0 \mathbf{A} \boldsymbol{\pi}_1 + f_0(\gamma f_0 + f_1) d_{\lambda\sigma} + \gamma d_{\sigma\sigma} & \boldsymbol{\pi}'_0 \mathbf{A} \\ \boldsymbol{\pi}'_1 \mathbf{A} \boldsymbol{\pi}_0 + f_0(\gamma f_0 + f_1) d_{\lambda\sigma} + \gamma d_{\sigma\sigma} & \boldsymbol{\pi}'_1 \mathbf{A} \boldsymbol{\pi}_1 + (\gamma f_0 + f_1)^2 d_{\lambda\sigma} + (1 + \gamma^2) d_{\sigma\sigma} & \gamma \boldsymbol{\pi}'_0 \mathbf{A} + \beta \mathbf{e}'_1 \mathbf{A} \\ \mathbf{A} \boldsymbol{\pi}_0 & \gamma \mathbf{A} \boldsymbol{\pi}_0 + \beta \mathbf{A} \mathbf{e}_1 & \mathbf{A} \end{pmatrix}.\end{aligned}$$

It is clear that in general for \mathbf{S} to be positive definite it is necessary that \mathbf{A} is positive definite. Since $\mathbf{A} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 (\mathbf{x}_i \mathbf{x}_i') \geq \inf_i (\sigma_i^2) \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i'$, and by assumption $\inf_i (\sigma_i^2) >$

$c_{\min} > 0$, then it is sufficient if $E(\mathbf{x}_i \mathbf{x}_i')$ is a positive definite matrix, which is likely to be so if x_{it} varies sufficiently across $t = 1, 2, 3$. Note that even if $\boldsymbol{\pi}_0 = 0$ and $d_{\lambda\sigma} = 0$ (cases to be considered below) then

$$\mathbf{S} = (1 + b^2) \begin{pmatrix} d_{\sigma\sigma} & \gamma d_{\sigma\sigma} & \mathbf{0} \\ \gamma d_{\sigma\sigma} & \beta^2 \mathbf{e}_1' \mathbf{A} \mathbf{e}_1 + (1 + \gamma^2) d_{\sigma\sigma} & \beta \mathbf{e}_1' \mathbf{A} \\ \mathbf{0} & \beta \mathbf{A} \mathbf{e}_1 & \mathbf{A} \end{pmatrix},$$

which is a positive definite matrix so long as $\mathbf{A} > \mathbf{0}$ and $d_{\sigma\sigma} > 0$. This result holds even if $\beta = 0$.

Now consider \mathbf{D}_N defined by (S.1), and since $y_{i2} - \gamma y_{i1} - \beta x_{i2} = \lambda_i f_2 + \varepsilon_{i2}$, then \mathbf{D}_N can be written equivalently as

$$\mathbf{D}_N = - \begin{pmatrix} N^{-1} \sum_{i=1}^N \mathbf{z}_i (y_{i2} - b y_{i1}), & N^{-1} \sum_{i=1}^N \mathbf{z}_i (\lambda_i f_2 + \varepsilon_{i2}), & N^{-1} \sum_{i=1}^N \mathbf{z}_i (x_{i3} - b x_{i2}) \end{pmatrix}. \quad (\text{S.4})$$

First we note that

$$\mathbf{z}_i = \begin{pmatrix} y_{i0} \\ y_{i1} \\ x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + \varepsilon_{i0} \\ \boldsymbol{\pi}'_1 \mathbf{x}_i + \lambda_i (\gamma f_0 + f_1) + \gamma \varepsilon_{i0} + \varepsilon_{i1} \\ x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix},$$

where $\boldsymbol{\pi}_1 = \gamma \boldsymbol{\pi}_0 + \beta \mathbf{e}_1$. Also

$$y_{i2} = \boldsymbol{\pi}'_2 \mathbf{x}_i + \lambda_i (\gamma^2 f_0 + \gamma f_1 + f_2) + \gamma^2 \varepsilon_{i0} + \gamma \varepsilon_{i1} + \varepsilon_{i2}, \quad (\text{S.5})$$

with

$$\boldsymbol{\pi}_2 = \gamma^2 \boldsymbol{\pi}_0 + (\gamma \mathbf{e}_1 + \mathbf{e}_2) \beta. \quad (\text{S.6})$$

Furthermore, to simplify the exposition we assume \mathbf{x}_i have zero means and are uncorrelated with the loadings, namely $E(\lambda_i x_{it}) = 0$. Then it is easily established that

$$\begin{aligned} N^{-1} \sum_{i=1}^N \mathbf{z}_i (\lambda_i f_2 + \varepsilon_{i2}) &= \begin{pmatrix} N^{-1} \sum_{i=1}^N y_{i0} (\lambda_i f_2 + \varepsilon_{i2}) \\ N^{-1} \sum_{i=1}^N y_{i1} (\lambda_i f_2 + \varepsilon_{i2}) \\ N^{-1} \sum_{i=1}^N \mathbf{x}_i (\lambda_i f_2 + \varepsilon_{i2}) \end{pmatrix} \\ &= \begin{pmatrix} N^{-1} \sum_{i=1}^N (\boldsymbol{\pi}'_0 \mathbf{x}_i + \lambda_i f_0 + \varepsilon_{i0}) (\lambda_i f_2 + \varepsilon_{i2}) \\ N^{-1} \sum_{i=1}^N (\boldsymbol{\pi}'_1 \mathbf{x}_i + \lambda_i (\gamma f_0 + f_1) + \gamma \varepsilon_{i0} + \varepsilon_{i1}) (\lambda_i f_2 + \varepsilon_{i2}) \\ N^{-1} \sum_{i=1}^N \mathbf{x}_i (\lambda_i f_2 + \varepsilon_{i2}) \end{pmatrix} \\ &\rightarrow_p \begin{pmatrix} f_0 f_2 \bar{d}(\boldsymbol{\lambda}) \\ (\gamma f_0 + f_1) f_2 \bar{d}(\boldsymbol{\lambda}) \\ \mathbf{0} \end{pmatrix}, \end{aligned}$$

where $\bar{d}(\boldsymbol{\lambda}) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \lambda_i^2$. Similarly,

$$N^{-1} \sum_{i=1}^N \mathbf{z}_i (x_{i3} - b x_{i2}) \rightarrow_p \begin{pmatrix} \boldsymbol{\pi}'_0 \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \\ \boldsymbol{\pi}'_1 \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \\ \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \end{pmatrix}.$$

where $\boldsymbol{\Sigma}_{xx} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i'$. Finally, to obtain the limit of the first column of \mathbf{D}_N , using (2) and (S.5) we first note that

$$y_{i2} - b y_{i1} = [(\gamma - b) \boldsymbol{\pi}_1 + \beta \mathbf{e}_2]' \mathbf{x}_i + [(\gamma - b) (\gamma f_0 + f_1) + f_2] \lambda_i + (\gamma - b) \gamma \varepsilon_{i0} + (\gamma - b) \varepsilon_{i1} + \varepsilon_{i2}.$$

Then it follows that

$$y_{i2} - b y_{i1} = [(\gamma - b) \boldsymbol{\pi}_1 + \beta \mathbf{e}_2] + \lambda_i [(\gamma - b) (\gamma f_0 + f_1) + f_2]$$

$$N^{-1} \sum_{i=1}^N y_{i0} (y_{i2} - by_{i1}) \rightarrow_p q_1, \quad N^{-1} \sum_{i=1}^N y_{i1} (y_{i2} - by_{i1}) \rightarrow_p q_2,$$

where

$$q_1 = \boldsymbol{\pi}'_0 \boldsymbol{\Sigma}_{xx} [(\gamma - b)\boldsymbol{\pi}_1 + \beta \mathbf{e}_2] + f_0 [(\gamma - b)(\gamma f_0 + f_1) + f_2] \bar{d}(\boldsymbol{\lambda}) + \gamma(\gamma - b) \bar{\sigma}^2, \quad (\text{S.7})$$

$$q_2 = \boldsymbol{\pi}'_1 \boldsymbol{\Sigma}_{xx} [(\gamma - b)\boldsymbol{\pi}_1 + \beta \mathbf{e}_2] + (\gamma f_0 + f_1) [(\gamma - b)(\gamma f_0 + f_1) + f_2] \bar{d}(\boldsymbol{\lambda}) + (\gamma - b)(1 + \gamma^2) \bar{\sigma}^2 \quad (\text{S.8})$$

and $\bar{\sigma}^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$. Similarly

$$N^{-1} \sum_{i=1}^N \mathbf{x}_i (y_{i2} - by_{i1}) \rightarrow_p \boldsymbol{\Sigma}_{xx} [(\gamma - b)\boldsymbol{\pi}_1 + \beta \mathbf{e}_2].$$

Collecting the above results in (S.4), we have

$$\mathbf{D}_N \rightarrow_p \mathbf{D} = - \begin{pmatrix} q_1 & f_0 f_2 \bar{d}(\boldsymbol{\lambda}) & \boldsymbol{\pi}'_0 \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \\ q_2 & (\gamma f_0 + f_1) f_2 \bar{d}(\boldsymbol{\lambda}) & \boldsymbol{\pi}'_1 \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \\ \boldsymbol{\Sigma}_{xx} [(\gamma - b)\boldsymbol{\pi}_1 + \beta \mathbf{e}_2] & \mathbf{0} & \boldsymbol{\Sigma}_{xx} (\mathbf{e}_3 - b \mathbf{e}_2) \end{pmatrix}. \quad (\text{S.9})$$

The rank of \mathbf{D} depends on $\boldsymbol{\theta}$, as well as the parameters of the \mathbf{x}_i process, and the strength of the common factor, as measured by $\bar{d}(\boldsymbol{\lambda})$. It is not possible to be sure that \mathbf{D} will be full rank for all values of $\boldsymbol{\theta}$; the rank could become deficient due to the particular values that the incidental parameters, such as b and $\bar{d}(\boldsymbol{\lambda})$ could take.

S.3 An eigenvalue approach for computing the TQML estimator

Consider the log-likelihood given in (34) without any restrictions on \mathbf{Q} , which can be further written as

$$\begin{aligned} \ell_N(\boldsymbol{\theta}) &= \ell_N(\boldsymbol{\varphi}, \boldsymbol{\psi}) = -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln(\sigma^2) \\ &\quad - \frac{N}{2} \ln |\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'| - \frac{1}{2\sigma^2} \sum_{i=1}^N \boldsymbol{\xi}'_i(\boldsymbol{\varphi}) (\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}')^{-1} \boldsymbol{\xi}_i(\boldsymbol{\varphi}). \end{aligned} \quad (\text{S.10})$$

To compute the TQML estimator consider (S.10) and note that since $\boldsymbol{\Omega}$ is a positive definite matrix and $\mathbf{Q}\mathbf{Q}'$ is rank deficient (recall that $m < T$), we have $|\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}'| = |\boldsymbol{\Omega}| |\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}|$, and using the Woodbury matrix identity

$$\begin{aligned} (\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}')^{-1} &= \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1} \mathbf{Q} (\mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q})^{-1} \mathbf{Q}' \boldsymbol{\Omega}^{-1} \\ &= \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1} \mathbf{Q} \mathbf{A}^{-1} \mathbf{Q}' \boldsymbol{\Omega}^{-1}, \end{aligned} \quad (\text{S.11})$$

where \mathbf{A} is a non-singular matrix defined by

$$\mathbf{A} = \mathbf{I}_m + \mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q}. \quad (\text{S.12})$$

Using the above results in (S.10), and after some simplification the quasi-log-likelihood function can be written as

$$N^{-1} \ell_N(\boldsymbol{\theta}) \propto -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \ln |\boldsymbol{\Omega}| - \frac{1}{2} \ln |\mathbf{A}| - \frac{1}{2\sigma^2} [\text{tr}(\mathbf{B}_N \boldsymbol{\Omega}^{-1}) - \text{tr}(\mathbf{B}_N \boldsymbol{\Omega}^{-1} \mathbf{Q} \mathbf{A}^{-1} \mathbf{Q}' \boldsymbol{\Omega}^{-1})], \quad (\text{S.13})$$

where $|\boldsymbol{\Omega}| = 1 + T(\omega - 1)$, and

$$\mathbf{B}_N(\boldsymbol{\varphi}) = N^{-1} \sum_{i=1}^N \boldsymbol{\xi}_i(\boldsymbol{\varphi}) \boldsymbol{\xi}'_i(\boldsymbol{\varphi}). \quad (\text{S.14})$$

For analytical convenience we further define $\mathbf{P} = \mathbf{\Omega}^{-1/2} \mathbf{Q} \mathbf{A}^{-1/2}$. Note that since \mathbf{A} and $\mathbf{\Omega}$ are non-singular matrices, then $\text{rank}(\mathbf{P}) = m$, as well. Further, it is easily seen that

$$\mathbf{I}_m - \mathbf{P}'\mathbf{P} = \mathbf{I}_m - \mathbf{A}^{-1/2} \mathbf{Q}' \mathbf{\Omega}^{-1} \mathbf{Q} \mathbf{A}^{-1/2},$$

and using $\mathbf{Q}' \mathbf{\Omega}^{-1} \mathbf{Q} = \mathbf{A} - \mathbf{I}_m$ from (S.12), we have

$$\mathbf{A}^{-1} = \mathbf{I}_m - \mathbf{P}'\mathbf{P}. \quad (\text{S.15})$$

Similarly,

$$\text{tr}(\mathbf{B}_N \mathbf{\Omega}^{-1} \mathbf{Q} \mathbf{A}^{-1} \mathbf{Q}' \mathbf{\Omega}^{-1}) = \sigma^2 \text{tr}[\mathbf{P}' \mathbf{C}_N(\boldsymbol{\theta}_c) \mathbf{P}],$$

where

$$\mathbf{C}_N(\boldsymbol{\theta}_c) = \sigma^{-2} \mathbf{\Omega}^{-1/2} \mathbf{B}_N(\boldsymbol{\varphi}) \mathbf{\Omega}^{-1/2}, \quad (\text{S.16})$$

and $\boldsymbol{\theta}_c = (\boldsymbol{\varphi}', \omega, \sigma^2)'$ where subscript c refers to $\boldsymbol{\theta}_c$ being the concentrated parameter vector.

Using the above results, the quasi-log-likelihood function given by (S.13) can now be written as

$$N^{-1} \ell_N(\boldsymbol{\theta}_c, \mathbf{P}) \propto -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \ln[1 + T(\omega - 1)] + \frac{1}{2} \ln |\mathbf{I}_m - \mathbf{P}'\mathbf{P}| - \frac{1}{2} \{ \text{tr}[\mathbf{C}_N(\boldsymbol{\theta}_c)] - \text{tr}[\mathbf{P}' \mathbf{C}_N(\boldsymbol{\theta}_c) \mathbf{P}] \}. \quad (\text{S.17})$$

In line with the discussion in Section 4, \mathbf{P} is not identified without additional restrictions. It is easily seen that the value of $\ell_N(\boldsymbol{\theta}_c, \mathbf{P})$ is invariant to the orthonormal transformation of \mathbf{P} . To see this consider the transformation $\tilde{\mathbf{P}} = \mathbf{P} \mathbf{\Xi}$, where $\mathbf{\Xi}$ is an $m \times m$ orthonormal matrix such that $\mathbf{\Xi}' \mathbf{\Xi} = \mathbf{I}_m$. Then it is readily verified that $N^{-1} \ell_N(\boldsymbol{\theta}_c, \mathbf{P}) = N^{-1} \ell_N(\boldsymbol{\theta}_c, \tilde{\mathbf{P}})$. Let $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m)$, where \mathbf{p}_t is the t^{th} column of \mathbf{P} , and \mathbf{p}_t is a $T \times 1$ vector of unknown parameters. Since $\text{rank}(\mathbf{P}) = m$, then $\mathbf{P}'\mathbf{P}$ can be diagonalised by an orthonormal transformation, and without loss of generality we impose the following $m(m-1)/2$ orthogonality conditions

$$\mathbf{p}_t' \mathbf{p}_s = 0, \text{ for all } s \neq t = 1, 2, \dots, m. \quad (\text{S.18})$$

Under these restrictions the quasi-log-likelihood function, (S.17), simplifies to

$$N^{-1} \ell_N(\boldsymbol{\theta}_c, \mathbf{P}) \propto -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \ln[1 + T(\omega - 1)] + \frac{1}{2} \sum_{t=1}^m \ln(1 - \mathbf{p}_t' \mathbf{p}_t) + \frac{1}{2} \sum_{t=1}^m \mathbf{p}_t' \mathbf{C}_N(\boldsymbol{\theta}_c) \mathbf{p}_t - \frac{1}{2} \text{tr}[\mathbf{C}_N(\boldsymbol{\theta}_c)]. \quad (\text{S.19})$$

Taking first derivatives with respect to \mathbf{p}_t and setting these derivatives to zero now yields

$$\mathbf{C}_N(\boldsymbol{\theta}_c) \hat{\mathbf{p}}_t - \left(\frac{1}{1 - \hat{\mathbf{p}}_t' \hat{\mathbf{p}}_t} \right) \hat{\mathbf{p}}_t = \mathbf{0}, \quad \text{for } t = 1, 2, \dots, m, \quad (\text{S.20})$$

where $\hat{\mathbf{p}}_t$ is the quasi-maximum likelihood estimator of \mathbf{p}_t (in terms of $\boldsymbol{\theta}_c$). Therefore, $\hat{\mathbf{p}}_t$ is the eigenvector of $\mathbf{C}_N(\boldsymbol{\theta}_c)$ associated with the first m largest non-zero eigenvalues of $\mathbf{C}_N(\boldsymbol{\theta}_c)$, which we denote by $\lambda_1(\boldsymbol{\theta}_c) > \lambda_2(\boldsymbol{\theta}_c) > \dots > \lambda_m(\boldsymbol{\theta}_c) > 0$. Note that $\mathbf{C}_N(\boldsymbol{\theta}_c)$ is a symmetric positive definite matrix with all real eigenvalues $\lambda_t(\boldsymbol{\theta}_c) > 0$, for $t = 1, 2, \dots, T$. We also have

$$\lambda_t(\boldsymbol{\theta}_c) = \frac{1}{1 - \hat{\mathbf{p}}_t' \hat{\mathbf{p}}_t}, \quad \text{and} \quad \hat{\mathbf{p}}_t' \mathbf{C}_N(\boldsymbol{\theta}_c) \hat{\mathbf{p}}_t = \lambda_t(\boldsymbol{\theta}_c) - 1.$$

Hence, the concentrated quasi log-likelihood function in terms of $\boldsymbol{\theta}_c$ can be written as

$$N^{-1} \ell_N(\boldsymbol{\theta}_c; m) \propto -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \ln[1 + T(\omega - 1)] - \frac{1}{2} \sum_{t=1}^m \ln[\lambda_t(\boldsymbol{\theta}_c)] + \frac{1}{2} \sum_{t=1}^m [\lambda_t(\boldsymbol{\theta}_c) - 1] - \frac{1}{2} \sum_{t=1}^T \lambda_t(\boldsymbol{\theta}_c), \quad (\text{S.21})$$

where $\lambda_t(\boldsymbol{\theta}_c)$ is the t^{th} eigenvalue of $\mathbf{C}_N(\boldsymbol{\theta}_c)$, given by (S.16). This concentrated quasi log-likelihood function can now be maximised with respect to $\boldsymbol{\theta}_c = (\boldsymbol{\varphi}', \omega, \sigma^2)'$. The TQML estimators, $\hat{\lambda}_t(\boldsymbol{\theta}_c)$, can then be computed using the TQML estimator of $\boldsymbol{\theta}_c$ and their corresponding variance covariance matrix can be computed using the delta method. Due to the possibility of local maxima, in maximising (S.21) we initialise the optimisation process with a number of starting values, randomly selected from the uniform distribution, specifically

$$\gamma_{ini} \sim U(-0.999, 0.999), \sigma_{ini}^2 \sim U(0.1, 1), \omega_{ini} \sim U(1, 2),$$

with the initial values for the remaining parameters, namely $\boldsymbol{\delta} = (\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}')'$ generated from a $U(-1, 1)$.

With regard to the computation of $\hat{\mathbf{p}}_t$ it is important to bear in mind that standard eigenvector routines provide eigenvectors that are typically orthonormalised. Whilst in the above analysis, $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \dots, \hat{\mathbf{p}}_m$ are orthogonal to each other, their length is not unity and is given by $\hat{\mathbf{p}}_t' \hat{\mathbf{p}}_t = 1 - 1/\lambda_t(\boldsymbol{\theta}_c)$.

S.4 Steady state distribution of y_{it} in the stationary case

Consider the panel data model

$$y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}, \quad |\gamma| < 1,$$

where

$$\zeta_{it} = \sum_{\ell=1}^m \eta_{\ell i} f_{\ell t} + u_{it}, \quad (\text{S.22})$$

$$x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} + v_{it}, \quad (\text{S.23})$$

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

$$v_{it} = \rho_x v_{i,t-1} + (1 - \rho_x^2)^{1/2} \varepsilon_{it}, \quad |\rho_x| < 1, \text{ for } t = 1, \dots, T, \quad (\text{S.24})$$

$$\varepsilon_{it} \sim IIDN(0, \sigma_{iv}^2), \text{ and } v_{i0} \sim IIDN(0, \sigma_{iv}^2),$$

which ensures that $Var(v_{it}) = \sigma_{iv}^2$. Further,

$$f_{\ell t} = \rho_{\ell f} f_{\ell,t-1} + (1 - \rho_{\ell f}^2)^{1/2} \varepsilon_{f\ell t}, \quad \varepsilon_{f\ell t} \sim IIDN(0, 1),$$

with $f_{\ell,0} = 0$, for $\ell = 1, 2, \dots, m$, and $t = 1, \dots, T$. Also to simplify the derivations we set $\rho_{\ell f} = \rho_f$ for all ℓ . From the above specifications of v_{it} and \mathbf{f}_t it readily follows that

$$E(v_{it}) = 0, \quad E(\mathbf{f}_t) = 0, \quad Cov(v_{i,t-j}, v_{i,t-j'}) = \sigma_{iv}^2 \rho_x^{|j-j'|} \text{ and } Cov(\mathbf{f}_{t-j}, \mathbf{f}_{t-j'}) = \rho_f^{|j-j'|} \mathbf{I}_m. \quad (\text{S.25})$$

Due to the dependence of x_{it} and ζ_{it} on the same unobserved factors, the regressors and the errors of the above regression are correlated. Following Pesaran and Smith (1994) we base the derivation of the steady state distribution of y_{it} on the following reduced form regressions

$$y_{it} = \tilde{\alpha}_i + \delta_t + \gamma y_{i,t-1} + \beta v_{it} + \mathbf{c}_i' \mathbf{f}_t + u_{it}, \quad (\text{S.26})$$

where

$$\tilde{\alpha}_i = \alpha_i + \beta \alpha_{xi}, \quad (\text{S.27})$$

$$\mathbf{c}_i' \mathbf{f}_t = \sum_{\ell=1}^m \eta_{\ell i} f_{\ell t} + \beta \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} = \sum_{\ell=1}^{\max(m, m_x)} c_{\ell i} f_{\ell t}, \quad (\text{S.28})$$

where $c_{\ell i}$ for all i and $\ell = 1, 2, \dots, \max(m, m_x)$, are defined implicitly. Using the results in (S.25), and noting that $\mathbf{f}_t, u_{it'}$ and \mathbf{v}_{is} are mutually uncorrelated for all values of t, t' and s , it then follows, conditional on $\tilde{\alpha}_i$ and \mathbf{c}_i , that (without loss of generality we set $\delta_t = 0$)

$$E(y_{it}|\tilde{\alpha}_i, \mathbf{c}_i) = \gamma E(y_{i,t-1}|\tilde{\alpha}_i, \mathbf{c}_i) + \tilde{\alpha}_i \quad (\text{S.29})$$

$$\begin{aligned} \text{Var}(y_{it}|\tilde{\alpha}_i, \mathbf{c}_i) &= \gamma^2 \text{Var}(y_{i,t-1}|\tilde{\alpha}_i, \mathbf{c}_i) + \beta^2 \text{Var}(\mathbf{v}_{it}|\tilde{\alpha}_i, \mathbf{c}_i) + \mathbf{c}_i' \text{Cov}(\mathbf{f}_t \mathbf{f}_t') \mathbf{c}_i + \sigma^2 \\ &\quad + 2\gamma \text{Cov}(y_{i,t-1}, \mathbf{c}_i' \mathbf{f}_t|\tilde{\alpha}_i, \mathbf{c}_i) + 2\gamma\beta \text{Cov}(y_{i,t-1}, \mathbf{v}_{it}|\tilde{\alpha}_i, \mathbf{c}_i). \end{aligned} \quad (\text{S.30})$$

Also, the steady state values of the covariances in the above expression are given by (upon using (S.25))

$$\begin{aligned} \text{Cov}(y_{i,t-1}, \mathbf{c}_i' \mathbf{f}_t|\tilde{\alpha}_i, \mathbf{c}_i) &= \sum_{j=0}^{\infty} \gamma^j \mathbf{c}_i' E(\mathbf{f}_{t-j-1} \mathbf{f}_t') \mathbf{c}_i = (\mathbf{c}_i' \mathbf{c}_i) \sum_{j=0}^{\infty} \rho_f^{j+1} \gamma^j = \frac{(\mathbf{c}_i' \mathbf{c}_i) \rho_f}{1 - \gamma \rho_f}, \\ \text{Cov}(y_{i,t-1}, \mathbf{v}_{it}|\tilde{\alpha}_i, \mathbf{c}_i) &= \beta \sigma_{iv}^2 \sum_{j=0}^{\infty} \gamma^j E(\mathbf{v}_{i,t-j-1} \mathbf{v}_{it}) = \beta \sigma_{iv}^2 \sum_{j=0}^{\infty} \rho_x^{j+1} \gamma^j = \frac{\beta \rho_x \sigma_{iv}^2}{1 - \gamma \rho_x}. \end{aligned}$$

Using the above results in (S.30) and noting that in steady state $E(y_{it}|\tilde{\alpha}_i, \mathbf{c}_i) = E(y_{i0}|\tilde{\alpha}_i, \mathbf{c}_i)$ and $\text{Var}(y_{it}|\tilde{\alpha}_i, \mathbf{c}_i) = \text{Var}(y_{i0}|\tilde{\alpha}_i, \mathbf{c}_i)$ we have

$$E(y_{it}|\tilde{\alpha}_i, \mathbf{c}_i) = \mu_{i0} = \frac{\alpha_i + \beta \alpha_{xi}}{1 - \gamma}, \quad (\text{S.31})$$

$$\text{Var}(y_{it}|\tilde{\alpha}_i, \mathbf{c}_i) = \sigma_{i0}^2 = \frac{\sigma^2 + \mathbf{a}_x \beta^2 \sigma_{iv}^2 + \mathbf{a}_f \mathbf{a}_i}{1 - \gamma^2}, \quad (\text{S.32})$$

where

$$\mathbf{a}_i = \mathbf{c}_i' \mathbf{c}_i = \sum_{\ell=1}^m \eta_{\ell i}^2 + \beta^2 \sum_{\ell=1}^{m_x} \vartheta_{\ell i}^2 + 2\beta \sum_{\ell=1}^{\min(m, m_x)} \eta_{\ell i} \vartheta_{\ell i}, \quad (\text{S.33})$$

and

$$\mathbf{a}_x = \left(\frac{1 + \gamma \rho_x}{1 - \gamma \rho_x} \right), \text{ and } \mathbf{a}_f = \left(\frac{1 + \gamma \rho_f}{1 - \gamma \rho_f} \right). \quad (\text{S.34})$$

S.5 Quasi-log-likelihood function of Bai (2013)

Consider

$$y_{it} = \gamma y_{i,t-1} + \delta_t + \mathbf{x}_{it}' \boldsymbol{\beta} + \mathbf{f}_t' \boldsymbol{\lambda}_i + \varepsilon_{it}, \text{ for } t = 1, 2, 3, \dots, T, i = 1, 2, \dots, N, \quad (\text{S.35})$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$, $\mathbf{x}_i \boldsymbol{\beta} = (\mathbf{I}_T \otimes \boldsymbol{\beta}') \text{vec}(\mathbf{x}_i') = (\mathbf{I}_T \otimes \boldsymbol{\beta}') \mathbf{w}_i$, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{i, \tilde{m}})' = (\alpha_i, \lambda_{i2}, \dots, \lambda_{i, \tilde{m}})'$, $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{\tilde{m}, t})'$ with $\tilde{m} = m + 1$, and

$$y_{i0} = \delta_0^* + \sum_{s=1}^T \mathbf{x}_{is}' \boldsymbol{\pi}_s + \mathbf{f}_0^{*'} \boldsymbol{\lambda}_i + \varepsilon_{i0}^*, \quad (\text{S.36})$$

with $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \dots, \boldsymbol{\pi}_T')'$.

To account for the correlation of the factor loading corresponding to the individual effects with the regressors \mathbf{x}_{it} , the Mundlak-Chamberlain projection is applied to the factor loadings $\boldsymbol{\lambda}_i$.

Projecting $\boldsymbol{\lambda}_i$ on $\mathbf{w}_i = \text{vec}(\mathbf{x}_i')$

$$\boldsymbol{\lambda}_i = \boldsymbol{\lambda} + \phi_1 \mathbf{x}_{i1} + \dots \phi_T \mathbf{x}_{iT} + \boldsymbol{\eta}_i,$$

or

$$\boldsymbol{\lambda}_i = \boldsymbol{\lambda} + \phi \mathbf{w}_i + \boldsymbol{\eta}_i, \quad (\text{S.37})$$

where λ is the intercept, η_i is the projection residual, ϕ_1, \dots, ϕ_T are matrices ($\tilde{m} \times k$) of projection coefficients. By definition $E(\eta_i) = \mathbf{0}$, and the regressors are uncorrelated with the projection residual for all t .

Substituting (S.37) into (S.35) and absorbing $\mathbf{f}_t' \lambda_i$ into δ_t , for $t \geq 1$,

$$y_{it} = \gamma y_{i,t-1} + \delta_t + \mathbf{x}_{it}' \beta + \mathbf{f}_t' \phi \mathbf{w}_i + \mathbf{f}_t' \eta_i + \varepsilon_{it}, \text{ for } t = 1, 2, 3, \dots, T, i = 1, 2, \dots, N, \quad (\text{S.38})$$

and

$$y_{i0} = \delta_0^* + \mathbf{w}_i' \pi + \mathbf{f}_0^{*'} \eta_i + \varepsilon_{i0}^*. \quad (\text{S.39})$$

Stacking the system of $T + 1$ equations given by (S.38) and (S.39) yields

$$\mathbf{B}^+ \mathbf{y}_i^+ = \mathbf{C} \mathbf{w}_i + \boldsymbol{\delta}^+ + \mathbf{F}^+ \eta_i + \boldsymbol{\varepsilon}_i^+,$$

where

$$\mathbf{y}_i^+ = \begin{bmatrix} y_{i0} \\ \mathbf{y}_i \end{bmatrix}, \quad \boldsymbol{\delta}^+ = \begin{bmatrix} \delta_0^* \\ \boldsymbol{\delta} \end{bmatrix}, \quad \mathbf{F}^+ = \begin{bmatrix} \mathbf{f}_0^{*'} \\ \mathbf{F} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i^+ = \begin{bmatrix} \varepsilon_{i0}^* \\ \boldsymbol{\varepsilon}_i \end{bmatrix},$$

with

$$\mathbf{B}^+ = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -\gamma & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -\gamma & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \pi' \\ \mathbf{I}_T \otimes \beta' + \mathbf{F} \phi \end{bmatrix}$$

and $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$, $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_T)'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$.

Let $\boldsymbol{\Sigma}^+ = \mathbf{F}^+ \boldsymbol{\Sigma}_\eta \mathbf{F}^{+'} + \boldsymbol{\Sigma}_\varepsilon$, where $\boldsymbol{\Sigma}_\eta = E(\eta_i \eta_i')$ and $\boldsymbol{\Sigma}_\varepsilon = E(\varepsilon_i^+ \varepsilon_i^{+'}) = \text{diag}(\sigma_0^2, \sigma^2 \mathbf{I}_T)$. Furthermore, let $\mathbf{u}_i^+ = \mathbf{B}^+ \mathbf{y}_i^+ - \mathbf{C} \mathbf{w}_i - \boldsymbol{\delta}^+$.

Following Bai (2013) we consider the following normalisation

$$\mathbf{F}^+ = (\mathbf{I}_{\tilde{m}}, \mathbf{F}_2')'. \quad (\text{S.40})$$

The quasi-log-likelihood function for $(y_{i0}, y_{i1}, \dots, y_{iT})$, conditional on \mathbf{w}_i , is then given by

$$\ell_N \propto -\frac{N}{2} \ln |\boldsymbol{\Sigma}^+| - \frac{1}{2} \sum_{i=1}^N \mathbf{u}_i^{+'} (\boldsymbol{\Sigma}^+)^{-1} \mathbf{u}_i^+,$$

where a number of random initial values are considered in maximising the above likelihood.

S.6 The GMM approach

Let us consider a GMM approach to estimate the dynamic panel data model with interactive effects:

$$y_{it} = \alpha_i + \mathbf{w}_{it}' \boldsymbol{\delta} + \boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{it}, \quad (i = 1, 2, \dots, N; t = 1, 2, \dots, T) \quad (\text{S.41})$$

where $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}_{it}')'$, $\boldsymbol{\delta} = (\gamma, \beta')'$, $\boldsymbol{\lambda}_i = (\lambda_{1i}, \dots, \lambda_{mi})'$ and $\mathbf{f}_t = (f_{1t}, \dots, f_{mt})'$ are $(m \times 1)$ vectors 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. A similar model is considered in Ahn et al. (2013), but there are two differences. The first is that the model under consideration is a dynamic model whereas Ahn et al. (2013) consider a static model. This difference does not cause a serious problem in implementing GMM estimation: minor corrections when selecting the instruments suffice. The second difference is that the current model contains time-invariant fixed effects α_i whereas the model considered in Ahn et al. (2013) does not. Thus the method by Ahn et al. (2013) cannot be applied directly in this case. Hence, we consider two approaches to use the method proposed by Ahn et al. (2013). The first approach is to regard the time-invariant fixed effects as an additional factor to be estimated. The second approach is to take the first-differences prior to applying the quasi-difference approach by Ahn et al. (2013), which is similar to Nauges and Thomas (2003). In the following, we describe each approach.

Approach 1: Quasi-differencing

By incorporating α_i into $\lambda_i' \mathbf{f}_t$ in (S.41), we have the following alternative expression

$$y_{it} = \mathbf{w}_{it}' \boldsymbol{\delta} + \tilde{\lambda}_i' \tilde{\mathbf{f}}_t + \varepsilon_{it},$$

where $\tilde{\lambda}_i = (\alpha_i, \lambda_{1i}, \dots, \lambda_{mi})'$ and $\tilde{\mathbf{f}}_t = (1, f_{1t}, \dots, f_{mt})'$. The model in matrix notation can be written as

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\delta} + \tilde{\mathbf{F}} \tilde{\lambda}_i + \boldsymbol{\varepsilon}_i, \quad (\text{S.42})$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$, $\mathbf{W}_i = (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iT})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$ and $\tilde{\mathbf{F}} = (\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_T)'$ is a $T \times \tilde{m}$ matrix. Define $\tilde{\boldsymbol{\Psi}} = \tilde{\mathbf{F}} \tilde{\mathbf{F}}^{-1}$ where $\tilde{\mathbf{F}} = (\tilde{\mathbf{f}}_{T-\tilde{m}+1}, \dots, \tilde{\mathbf{f}}_T)'$. To separately identify $\tilde{\mathbf{F}}$ from $\tilde{\lambda}_i$, \tilde{m}^2 restrictions are imposed on the factors 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 obtained as the first $T-\tilde{m}$ rows of $\tilde{\boldsymbol{\Psi}}$. Let $\mathbf{H}_Q = (\mathbf{I}_{T-\tilde{m}}, -\boldsymbol{\Psi})'$, so that $\mathbf{H}_Q' \tilde{\mathbf{F}} = (\mathbf{I}_{T-\tilde{m}}, -\boldsymbol{\Psi})(\boldsymbol{\Psi}', \mathbf{I}_{\tilde{m}})' = \mathbf{0}_{(T-\tilde{m}) \times \tilde{m}}$. Then, pre-multiplying equation (S.42) by \mathbf{H}_Q' removes the unobservable effects so that

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

or

$$\begin{aligned} \dot{\mathbf{y}}_i &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + \boldsymbol{\Psi} \ddot{\mathbf{y}}_i - \boldsymbol{\Psi} \ddot{\mathbf{W}}_i \boldsymbol{\delta} + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi} \ddot{\boldsymbol{\varepsilon}}_i \\ &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + (\mathbf{I}_{T-\tilde{m}} \otimes \ddot{\mathbf{y}}_i) \text{vec}(\boldsymbol{\Psi}) - \left(\text{vec}(\ddot{\mathbf{W}}_i)' \otimes \mathbf{I}_{T-\tilde{m}} \right) \text{vec}(\boldsymbol{\delta}' \otimes \boldsymbol{\Psi}) + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi} \ddot{\boldsymbol{\varepsilon}}_i, \end{aligned} \quad (\text{S.43})$$

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

The t^{th} equation is given by

$$y_{it} = \boldsymbol{\delta}' \mathbf{w}_{it} + \boldsymbol{\psi}_t' \ddot{\mathbf{y}}_i - \boldsymbol{\psi}_t' \ddot{\mathbf{W}}_i \boldsymbol{\delta} + v_{it}, \quad (i = 1, \dots, N; t = 1, \dots, T - \tilde{m}), \quad (\text{S.44})$$

where $v_{it} = (\varepsilon_{it} - \boldsymbol{\theta}_t' \ddot{\boldsymbol{\varepsilon}}_i)$. Since \mathbf{x}_{it} is strictly exogenous, a large number of moment conditions are available. However, as using many instruments causes a large finite sample bias, we consider $(k+1)(T-\tilde{m})(T-\tilde{m}+1)/2 + k(T-\tilde{m})\tilde{m}$ moment conditions given by $E[\mathbf{z}_{it} v_{it}] = \mathbf{0}$, for $t = 1, \dots, T - \tilde{m}$, where $\mathbf{z}_{it} = (y_{i0}, \dots, y_{i,t-1}, \mathbf{x}_{i1}', \dots, \mathbf{x}_{it}', \mathbf{x}_{i,T-\tilde{m}+1}', \dots, \mathbf{x}_{iT}')'$. In addition to the commonly used instruments $(y_{i0}, \dots, y_{i,t-1}, \mathbf{x}_{i1}', \dots, \mathbf{x}_{it}')$, we also use $\mathbf{x}_{i,T-\tilde{m}+1}', \dots, \mathbf{x}_{iT}'$ as instruments since they are included in the regressor $\ddot{\mathbf{W}}$. In matrix notation the moment conditions can be written as $E[\mathbf{Z}_i \mathbf{v}_i(\boldsymbol{\theta})] = \mathbf{0}$, where $\mathbf{Z}_i = \text{diag}(\mathbf{z}_{i1}', \dots, \mathbf{z}_{i,T-\tilde{m}}')$, $\mathbf{v}_i(\boldsymbol{\theta}) = (v_{i1}, \dots, v_{i,T-\tilde{m}})'$ and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\psi}')'$ with $\boldsymbol{\psi} = \text{vec}(\boldsymbol{\Psi})$.

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

$$\hat{\boldsymbol{\theta}}_{QD1} = \arg \min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{v}_i(\boldsymbol{\theta}) \right), \quad (\text{S.45})$$

and

$$\hat{\boldsymbol{\theta}}_{QD2} = \arg \min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{v}_i(\hat{\boldsymbol{\theta}}_{QD1}) \mathbf{v}_i(\hat{\boldsymbol{\theta}}_{QD1})' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{v}_i(\boldsymbol{\theta}) \right). \quad (\text{S.46})$$

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

$$\text{Var}(\hat{\boldsymbol{\theta}}_{QD1}) = N^{-1} \left(\hat{\mathbf{G}}_{QD1}' \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{QD1} \right)^{-1} \hat{\mathbf{G}}_{QD1}' \hat{\mathbf{W}}^{-1} \hat{\boldsymbol{\Omega}}_{QD1} \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{QD1} \left(\hat{\mathbf{G}}_{QD1}' \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{QD1} \right)^{-1} \quad (\text{S.47})$$

$$\text{Var}(\hat{\boldsymbol{\theta}}_{QD2}) = N^{-1} \left(\hat{\mathbf{G}}_{QD2}' \hat{\boldsymbol{\Omega}}_{QD2}^{-1} \hat{\mathbf{G}}_{QD2} \right)^{-1}, \quad (\text{S.48})$$

where $\hat{\mathbf{G}}_j = \partial \bar{\mathbf{g}}(\hat{\boldsymbol{\theta}}_j) / \partial \boldsymbol{\theta}'$ for $j = QD1, QD2$, 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.

Approach 2: Quasi-differencing after first-differencing

Taking the first-differences of model (S.41) to remove α_i we have

$$\Delta y_{it} = \Delta \mathbf{w}'_{it} \boldsymbol{\delta} + \boldsymbol{\lambda}'_i \Delta \mathbf{f}_t + \Delta \varepsilon_{it}, \quad (i = 1, 2, \dots, N; t = 2, 3, \dots, T)$$

where $\Delta \mathbf{w}_{it} = (\Delta y_{i,t-1}, \Delta \mathbf{x}'_{it})'$, $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$, and $\Delta \mathbf{f}_t = \mathbf{f}_t - \mathbf{f}_{t-1}$. The model in notation can be written as

$$\Delta \mathbf{y}_i = \Delta \mathbf{W}_i \boldsymbol{\delta} + \Delta \mathbf{F} \boldsymbol{\lambda}_i + \Delta \boldsymbol{\varepsilon}_i, \quad (\text{S.49})$$

where $\Delta \mathbf{y}_i = (\Delta y_{i2}, \dots, \Delta y_{iT})'$, $\Delta \mathbf{W}_i = (\Delta \mathbf{w}_{i2}, \dots, \Delta \mathbf{w}_{iT})'$, $\Delta \boldsymbol{\varepsilon}_i = (\Delta \varepsilon_{i2}, \dots, \Delta \varepsilon_{iT})'$ and $\Delta \mathbf{F} = (\Delta \mathbf{f}_2, \dots, \Delta \mathbf{f}_T)'$ is a $(T-1) \times m$ matrix. Define $\tilde{\boldsymbol{\Phi}} = \Delta \mathbf{F} (\overline{\Delta \mathbf{F}})^{-1}$ where $\overline{\Delta \mathbf{F}} = (\Delta \mathbf{f}_{T-m+1}, \dots, \Delta \mathbf{f}_T)'$. To separately identify $\Delta \mathbf{F}$ from $\boldsymbol{\lambda}_i$, m^2 restrictions are imposed on the factors such that $\Delta \mathbf{F} = (\boldsymbol{\Phi}', \mathbf{I}_m)'$ where $\boldsymbol{\Phi}$ is a $(T-1-m) \times m$ matrix of unrestricted parameters obtained as the first $T-1-m$ rows of $\tilde{\boldsymbol{\Phi}}$. Let $\mathbf{H}_D = (\mathbf{I}_{T-1-m}, -\boldsymbol{\Phi})'$, so that $\mathbf{H}'_D \Delta \mathbf{F} = (\mathbf{I}_{T-1-m}, -\boldsymbol{\Phi})(\boldsymbol{\Phi}', \mathbf{I}_m)' = \mathbf{0}_{(T-1-m) \times m}$. Then, pre-multiplying equation (S.49) by \mathbf{H}'_D removes the unobservable effects so that

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

or

$$\begin{aligned} \Delta \dot{\mathbf{y}}_i &= \Delta \dot{\mathbf{W}}_i \boldsymbol{\delta} + \boldsymbol{\Phi} \Delta \ddot{\mathbf{y}}_i - \boldsymbol{\Phi} \Delta \ddot{\mathbf{W}}_i \boldsymbol{\delta} + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Phi} \Delta \ddot{\boldsymbol{\varepsilon}}_i \\ &= \Delta \dot{\mathbf{W}}_i \boldsymbol{\delta} + (\mathbf{I}_{T-1-m} \otimes \Delta \ddot{\mathbf{y}}'_i) \text{vec}(\boldsymbol{\Phi}) - \left(\text{vec}(\Delta \ddot{\mathbf{W}}_i) \otimes \mathbf{I}_{T-1-m} \right) \text{vec}(\boldsymbol{\delta}' \otimes \boldsymbol{\Phi}) + \Delta \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Phi} \Delta \ddot{\boldsymbol{\varepsilon}}_i, \end{aligned}$$

where $\Delta \dot{\mathbf{y}}_i = (\Delta y_{i2}, \dots, \Delta y_{i,T-m})'$, $\Delta \ddot{\mathbf{y}}_i = (\Delta y_{i,T-m+1}, \dots, \Delta y_{iT})'$, $\Delta \dot{\mathbf{W}}_i = (\Delta \mathbf{w}_{i2}, \dots, \Delta \mathbf{w}_{i,T-m})'$, $\Delta \ddot{\mathbf{W}}_i = (\Delta \mathbf{w}_{i,T-m+1}, \dots, \Delta \mathbf{w}_{iT})'$, $\boldsymbol{\Phi}' = (\phi_2, \dots, \phi_{T-m})$, $\Delta \dot{\boldsymbol{\varepsilon}}_i = (\Delta \varepsilon_{i2}, \dots, \Delta \varepsilon_{i,T-m})'$, and $\Delta \ddot{\boldsymbol{\varepsilon}}_i = (\Delta \varepsilon_{i,T-m+1}, \dots, \Delta \varepsilon_{iT})'$.

The t^{th} equation is given by

$$\Delta y_{it} = \boldsymbol{\delta}' \Delta \mathbf{w}_{it} + \phi'_t \Delta \ddot{y}_i - \phi'_t \Delta \ddot{\mathbf{W}}_i \boldsymbol{\delta} + \Delta v_{it}, \quad (i = 1, \dots, N; t = 2, \dots, T-m), \quad (\text{S.50})$$

where $\Delta v_{it} = (\Delta \varepsilon_{it} - \phi'_t \Delta \ddot{\boldsymbol{\varepsilon}}_i)$. Since \mathbf{x}_{it} is strictly exogenous, a large number of moment conditions are available. However, since using many instruments causes a large finite sample bias, we consider $(k+1)(T-1-m)(T-m)/2 + k(T-1-m)m + k(T-1-m)$ moment conditions given by $E[\mathbf{z}_{it} \Delta v_{it}] = \mathbf{0}$, for $t = 2, \dots, T-m$, where $\mathbf{z}_{it} = (y_{i0}, \dots, y_{i,t-1}, \mathbf{x}'_{i0}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{it}, \mathbf{x}'_{i,T-m+1}, \dots, \mathbf{x}'_{iT})'$. In addition to the commonly used instruments $(y_{i0}, \dots, y_{i,t-1}, \mathbf{x}'_{i0}, \dots, \mathbf{x}'_{it})$, we also use $\mathbf{x}'_{i,T-m+1}, \dots, \mathbf{x}'_{iT}$ as instruments since they are included in the regressor $\Delta \ddot{\mathbf{W}}$. Also, compared to the quasi-difference approach, we additionally use \mathbf{x}_{i0} as instruments. This is because without \mathbf{x}_{i0} , the local identification assumption is not satisfied for the linear GMM estimator which is used as the starting value to obtain nonlinear GMM estimators. In matrix notation the moment conditions can be written as $E[\mathbf{Z}'_i \Delta \mathbf{v}_i(\boldsymbol{\theta})] = \mathbf{0}$, where $\mathbf{Z}_i = \text{diag}(\mathbf{z}'_{i2}, \dots, \mathbf{z}'_{i,T-m})$, $\Delta \mathbf{v}_i(\boldsymbol{\theta}) = (\Delta v_{i2}, \dots, \Delta v_{i,T-m})'$ and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\phi}')'$ with $\boldsymbol{\phi} = \text{vec}(\boldsymbol{\Phi})$.

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

$$\hat{\boldsymbol{\theta}}_{FD1} = \arg \min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^N \Delta \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{v}_i(\boldsymbol{\theta}) \right), \quad (\text{S.51})$$

and

$$\hat{\boldsymbol{\theta}}_{FD2} = \arg \min_{\boldsymbol{\theta}} \left(\frac{1}{N} \sum_{i=1}^N \Delta \mathbf{v}_i(\boldsymbol{\theta})' \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{v}_i(\hat{\boldsymbol{\theta}}_{FD1}) \Delta \mathbf{v}_i(\hat{\boldsymbol{\theta}}_{FD1})' \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{v}_i(\boldsymbol{\theta}) \right). \quad (\text{S.52})$$

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

$$Var(\hat{\theta}_{FD1}) = N^{-1} \left(\hat{\mathbf{G}}'_{FD1} \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{FD1} \right)^{-1} \hat{\mathbf{G}}'_{FD1} \hat{\mathbf{W}}^{-1} \hat{\boldsymbol{\Omega}}_{FD1} \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{FD1} \left(\hat{\mathbf{G}}'_{FD1} \hat{\mathbf{W}}^{-1} \hat{\mathbf{G}}_{FD1} \right)^{-1} \quad (\text{S.53})$$

$$Var(\hat{\theta}_{FD2}) = N^{-1} \left(\hat{\mathbf{G}}'_{FD2} \hat{\boldsymbol{\Omega}}_{FD2}^{-1} \hat{\mathbf{G}}_{FD2} \right)^{-1}, \quad (\text{S.54})$$

where $\hat{\mathbf{G}}_j = \partial \bar{\mathbf{g}}(\hat{\theta}_j) / \partial \theta'$ for $j = FD1, FD2$, with $\mathbf{g}_i(\hat{\theta}_j) = \mathbf{Z}'_i \Delta \mathbf{v}_i(\hat{\theta}_j)$ and $\bar{\mathbf{g}}(\hat{\theta}_j) = N^{-1} \sum_{i=1}^N \mathbf{g}_i(\hat{\theta}_j)$, $\hat{\mathbf{W}} = N^{-1} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{Z}_i$, and $\hat{\boldsymbol{\Omega}}_j = N^{-1} \sum_{i=1}^N \mathbf{g}_i(\hat{\theta}_j) \mathbf{g}_i(\hat{\theta}_j)'$. The derivatives involved in $\hat{\mathbf{G}}_j$ are computed numerically.

Starting values

For the computation of the above nonlinear GMM estimators, starting values are required. When the number of moment conditions is greater than the unknown reduced form parameters we use the linear GMM estimator by Hayakawa (2012) as the starting value. This can reduce the computational time compared to employing several random starting values which we use in the alternative case.

To define the linear GMM estimator, let us define $L_1 = L_2 = 1$ for $\tilde{m} = 1$, and $\mathbf{L}_1 = (\mathbf{I}_{\tilde{m}}, \mathbf{0}_{\tilde{m}})$ and $\mathbf{L}_2 = (\mathbf{0}_{\tilde{m}}, \mathbf{I}_{\tilde{m}})$ for $\tilde{m} > 1$. Also, define $\tilde{\mathbf{y}}_i = (y_{i,T-\tilde{m}}, y_{i,T-\tilde{m}+1}, \dots, y_{iT})' = (y_{i,T-\tilde{m}}, \tilde{\mathbf{y}}'_i)'$. Then, noting that $\ddot{\mathbf{W}}_i = (\ddot{\mathbf{y}}_{i,-1}, \ddot{\mathbf{X}}_{it})$ where $\ddot{\mathbf{y}}_{i,-1} = (y_{i,T-\tilde{m}}, y_{i,T-\tilde{m}+1}, \dots, y_{iT-1})'$, $\ddot{\mathbf{y}}_i = \mathbf{L}_2 \tilde{\mathbf{y}}_i$ and $\ddot{\mathbf{y}}_{i,-1} = \mathbf{L}_1 \tilde{\mathbf{y}}_i$, we have

$$\begin{aligned} \dot{\mathbf{y}}_i &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + \Psi \ddot{\mathbf{y}}_i - \Psi \ddot{\mathbf{W}}_i \boldsymbol{\delta} + \dot{\boldsymbol{\epsilon}}_i - \Psi \ddot{\boldsymbol{\epsilon}}_i \\ &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + \Psi \mathbf{L}_2 \tilde{\mathbf{y}}_i - \Psi \left(\gamma \mathbf{L}_1 \tilde{\mathbf{y}}_i + \ddot{\mathbf{X}}_i \boldsymbol{\beta} \right) + \dot{\boldsymbol{\epsilon}}_i - \Psi \ddot{\boldsymbol{\epsilon}}_i \\ &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + \Psi (\mathbf{L}_2 - \gamma \mathbf{L}_1) \tilde{\mathbf{y}}_i - \Psi \ddot{\mathbf{X}}_i \boldsymbol{\beta} + \mathbf{v}_i \\ &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + \Upsilon \tilde{\mathbf{y}}_i - \Psi \ddot{\mathbf{X}}_i \boldsymbol{\beta} + \mathbf{v}_i \\ &= \dot{\mathbf{W}}_i \boldsymbol{\delta} + (\mathbf{I}_{T-\tilde{m}} \otimes \tilde{\mathbf{y}}'_i) \text{vec}(\Upsilon') - \left(\text{vec}(\ddot{\mathbf{X}}_i)' \otimes \mathbf{I}_{T-\tilde{m}} \right) \text{vec}(\boldsymbol{\beta}' \otimes \Psi) + \mathbf{v}_i \\ &= \tilde{\mathbf{X}}_i \boldsymbol{\pi} + \mathbf{v}_i \end{aligned}$$

where $\Upsilon = \Psi (\mathbf{L}_2 - \gamma \mathbf{L}_1)$, $\mathbf{X}_i = \left(\dot{\mathbf{W}}_i, (\mathbf{I}_{T-\tilde{m}} \otimes \tilde{\mathbf{y}}'_i), - \left(\text{vec}(\ddot{\mathbf{X}}_i)' \otimes \mathbf{I}_{T-\tilde{m}} \right) \right)$ and $\boldsymbol{\pi} = (\boldsymbol{\delta}', \text{vec}(\Upsilon'), \text{vec}(\boldsymbol{\beta}' \otimes \Psi'))' = (\boldsymbol{\pi}'_1, \boldsymbol{\pi}'_2, \boldsymbol{\pi}'_3)'$ with $\boldsymbol{\pi}_1 = \boldsymbol{\delta}$, $\boldsymbol{\pi}_2 = \text{vec}(\Upsilon')$, $\boldsymbol{\pi}_3 = \text{vec}(\boldsymbol{\beta}' \otimes \Psi)$. We consider this particular model rather than the original model (S.43) because perfect multicollinearity between $\tilde{\mathbf{y}}_i$ and $\ddot{\mathbf{W}}_i$ occurs in (S.43) when $\tilde{m} > 1$. Since this is a linear model in $\boldsymbol{\pi}$ with moment conditions $E[\mathbf{Z}'_i \mathbf{v}_i(\boldsymbol{\pi})] = \mathbf{0}$, a closed form solution is obtained as

$$\begin{aligned} \hat{\boldsymbol{\pi}} &= \left[\left(\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{X}}'_i \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}'_i \tilde{\mathbf{X}}_i \right) \right]^{-1} \\ &\quad \times \left[\left(\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{X}}'_i \mathbf{Z}_i \right) \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{Z}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}'_i \dot{\mathbf{y}}_i \right) \right]. \end{aligned}$$

Hence, $\hat{\boldsymbol{\pi}}_1$ and $\hat{\boldsymbol{\pi}}_2$ are consistent estimates of $\boldsymbol{\delta}$ and $\text{vec}(\Upsilon')$, respectively. To recover Ψ from the estimate of Υ , since

$$\text{vec}(\Upsilon') = \text{vec}((\mathbf{L}_2 - \gamma \mathbf{L}_1)' \Psi') = (\mathbf{I}_{T-\tilde{m}} \otimes (\mathbf{L}_2 - \gamma \mathbf{L}_1)') \text{vec}(\Psi') = \mathbf{A} \text{vec}(\Psi'),$$

$\text{vec}(\Psi')$ is obtained as $\text{vec}(\Psi') = (\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' \text{vec}(\Upsilon')$. In the computation of the nonlinear GMM estimators, estimates of $\boldsymbol{\delta}$ and $\text{vec}(\Psi')$ are obtained from $\hat{\boldsymbol{\pi}}_1$ and $\hat{\boldsymbol{\pi}}_2$ and are used as the starting values

of the numerical optimization. For those cases where random starting values are used γ is generated as $U(-0.999, 0.999)$, β as $U(-1, 1)$ and ψ_j as $\psi_{j0} \times U(0.9, 1.1)$ where ψ_{j0} denotes the true value of ψ_j , j th element of $vec(\Psi')$.

The same procedure can be used in approach 2 by replacing the \mathbf{y}_i 's and \mathbf{W}_i 's with their first differences.

The AR(1) panel data model

Estimation of the AR(1) model is exactly the same as above after removing all \mathbf{x} 's from both the model and instruments. However, for the starting value, we cannot use the linear estimator since the number of moment conditions is always smaller than that of the unknown reduced form parameters. Hence in the Monte Carlo simulations for this case we use random starting values. Specifically, we use

$$\gamma_{ini} \sim U(-0.999, 0.999), \psi_{j,ini} \sim \psi_{j,0} \times U(-0.5, 0.5), \quad (j = 1, \dots, (T - \tilde{m})\tilde{m})$$

for approach 1 and

$$\gamma_{ini} \sim U(-0.999, 0.999), \psi_{j,ini} \sim \psi_{j,0} \times U(-0.5, 0.5), \quad (j = 1, \dots, (T - 1 - m)m)$$

for approach 2 where $\psi_{j,0}$ is the true value of ψ_j .

S.7 Monte Carlo Results for the Stationary Case

A1: Selecting the number of factors

Table A1(i): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure in the case of the AR(1) panel data model

$T = 5$												
κ^2	0.25			0.5			1			2		
m_0	0	1	2	0	1	2	0	1	2	0	1	2
$\gamma_0 = 0.4$												
100	99.4	25.5	0.9	99.4	88.2	17.1	99.4	99.7	88.9	99.4	99.7	99.9
300	99.8	93.7	16.5	99.8	100.0	95.4	99.8	100.0	100.0	99.8	100.0	100.0
500	99.9	100.0	56.1	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0
1000	99.9	100.0	99.2	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0
$\gamma_0 = 0.8$												
100	99.2	53.4	1.5	99.2	98.7	28.7	99.2	99.8	96.3	99.2	99.7	100.0
300	99.8	99.6	23.3	99.8	100.0	98.9	99.8	100.0	100.0	99.8	100.0	100.0
500	99.9	100.0	65.2	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0
1000	99.9	100.0	99.7	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0
$T = 10$												
κ^2	0.25			0.5			1			2		
m_0	0	1	2	0	1	2	0	1	2	0	1	2
$\gamma_0 = 0.4$												
100	99.5	97.1	13.2	99.5	99.6	90.8	99.5	99.6	99.7	99.5	99.6	99.7
300	99.8	100.0	95.4	99.8	100.0	100.0	99.8	100.0	100.0	99.8	100.0	100.0
500	99.9	100.0	99.9	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0
1000	99.7	100.0	100.0	99.7	100.0	100.0	99.7	100.0	100.0	99.7	100.0	100.0
$\gamma_0 = 0.8$												
100	99.7	96.6	15.1	99.7	99.5	93.5	99.7	99.5	99.7	99.7	99.6	99.7
300	99.8	100.0	96.7	99.8	100.0	100.0	99.8	100.0	100.0	99.8	100.0	99.9
500	99.9	99.9	100.0	99.9	99.9	100.0	99.9	99.9	100.0	99.9	99.9	100.0
1000	99.6	100.0	100.0	99.6	100.0	100.0	99.6	100.0	100.0	99.6	100.0	100.0

Note: \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$. See also the note to Table 1.

Table A1(ii): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure in the case of the ARX(1) panel data model

$T = 5$																		
$\kappa^2 = 0.25$									$\kappa^2 = 0.5$									
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
N	$\gamma_0 = 0.4$																	
100	99.7	99.7	99.8	46.3	51.5	52.6	1.1	1.2	1.2	99.2	99.3	99.3	97.9	98.1	98.1	17.7	18.3	18.5
300	99.9	100.0	100.0	99.7	99.9	100.0	21.9	23.5	23.3	99.4	100.0	100.0	100.0	100.0	100.0	97.2	97.6	97.7
500	99.8	99.9	99.9	99.9	99.9	99.9	67.4	69.0	69.1	99.6	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	99.9	99.9	99.9	99.9	99.9	99.9	99.6	99.7	99.7	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\gamma_0 = 0.8$																		
100	99.7	99.6	99.6	56.2	56.9	57.4	1.4	1.6	1.7	99.4	99.4	99.4	97.9	98.0	98.0	19.2	18.9	19.0
300	99.9	99.9	99.9	100.0	100.0	100.0	24.8	24.7	24.5	100.0	100.0	100.0	100.0	100.0	100.0	98.2	98.1	98.1
500	99.9	99.9	99.9	99.9	99.9	99.9	71.1	71.1	71.1	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	99.9	99.9	99.9	99.9	99.9	99.9	99.8	99.8	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\kappa^2 = 1$									$\kappa^2 = 2$									
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
N	$\gamma_0 = 0.4$																	
100	99.7	99.7	99.8	97.8	98.7	99.0	29.4	31.0	31.0	99.2	99.3	99.3	99.5	99.6	99.6	93.5	94.2	94.4
300	99.9	100.0	100.0	100.0	100.0	100.0	98.9	99.5	99.4	99.4	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9
500	99.8	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0	99.6	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	99.9	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\gamma_0 = 0.8$																		
100	99.7	99.6	99.6	99.1	99.2	99.3	32.6	33.0	33.1	99.4	99.4	99.4	99.5	99.6	99.6	94.4	94.7	94.4
300	99.9	99.9	99.9	100.0	100.0	100.0	99.5	99.5	99.5	100.0	100.0	100.0	100.0	99.9	99.9	99.8	99.8	99.8
500	99.9	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	99.9	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$T = 10$																		
$\kappa^2 = 0.25$									$\kappa^2 = 0.5$									
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
N	$\gamma_0 = 0.4$																	
100	99.2	99.3	99.3	97.9	98.1	98.1	17.7	18.3	18.5	99.2	99.3	99.3	99.5	99.6	99.6	93.5	94.2	94.4
300	99.4	100.0	100.0	100.0	100.0	100.0	97.2	97.6	97.7	99.4	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9
500	99.6	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.6	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\gamma_0 = 0.8$																		
100	99.4	99.4	99.4	97.9	98.0	98.0	19.2	18.9	19.0	99.4	99.4	99.4	99.5	99.6	99.6	94.4	94.7	94.4
300	100.0	100.0	100.0	100.0	100.0	100.0	98.2	98.1	98.1	100.0	100.0	100.0	100.0	99.9	99.9	99.8	99.8	99.8
500	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\kappa^2 = 1$									$\kappa^2 = 2$									
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
N	$\gamma_0 = 0.4$																	
100	99.2	99.3	99.3	99.5	99.6	99.7	99.8	99.7	99.7	99.2	99.3	99.3	99.7	99.6	99.6	99.7	99.7	99.7
300	99.4	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	99.4	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9
500	99.6	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.6	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\gamma_0 = 0.8$																		
100	99.4	99.4	99.4	99.5	99.6	99.6	99.7	99.7	99.7	99.4	99.4	99.4	99.5	99.6	99.6	99.7	99.7	99.7
300	100.0	100.0	100.0	100.0	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0	100.0	99.9	99.9	99.9	99.9	99.9
500	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

See the note to Table A1(i).

A2: Bias, RMSE and Size

Table A2(i): Bias($\times 100$) and RMSE($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m}

$T = 5, \gamma_0 = 0.4$										$T = 5, \gamma_0 = 0.8$							
		Bias($\times 100$)				RMSE($\times 100$)				Bias($\times 100$)				RMSE($\times 100$)			
κ^2		0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
N	$m_0 = 0$																
100		0.42	0.42	0.42	0.42	8.69	8.69	8.69	8.69	0.65	0.65	0.65	0.65	12.29	12.29	12.29	12.29
300		-0.03	-0.03	-0.03	-0.03	4.26	4.26	4.26	4.26	1.42	1.42	1.42	1.42	9.26	9.26	9.26	9.26
500		0.03	0.03	0.03	0.03	3.22	3.22	3.22	3.22	1.46	1.46	1.46	1.46	7.80	7.80	7.80	7.80
1000		0.00	0.00	0.00	0.00	2.29	2.29	2.29	2.29	1.02	1.02	1.02	1.02	6.07	6.07	6.07	6.07
	$m_0 = 1$																
100		24.98	5.19	0.41	0.23	33.05	18.36	9.39	7.79	7.22	1.11	1.42	1.38	15.51	13.99	12.99	11.19
300		1.96	-0.05	-0.09	-0.11	11.04	5.64	4.99	4.17	1.20	1.28	1.00	0.46	11.06	10.41	9.04	6.86
500		0.15	0.10	0.05	0.01	4.53	4.17	3.68	3.07	1.68	1.46	0.96	0.40	9.48	8.64	7.12	5.09
1000		0.05	0.05	0.04	0.03	3.25	3.02	2.67	2.22	1.43	1.13	0.61	0.27	7.70	6.77	5.08	3.56
	$m_0 = 2$																
100		6.61	13.75	4.09	0.34	13.61	25.13	16.38	7.89	7.09	5.07	1.82	1.50	14.00	15.66	16.38	11.31
300		5.43	1.25	0.20	0.13	10.92	8.49	4.99	4.14	6.76	1.81	1.38	0.81	13.80	10.54	4.99	6.82
500		3.12	0.08	0.05	0.04	8.58	4.36	3.81	3.16	4.31	1.50	0.98	0.49	11.71	8.74	3.81	5.12
1000		0.12	0.04	0.02	0.01	3.38	2.98	2.62	2.18	1.23	0.89	0.45	0.19	7.43	6.34	2.62	3.45
$T = 10, \gamma_0 = 0.4$										$T = 10, \gamma_0 = 0.8$							
		Bias($\times 100$)				RMSE($\times 100$)				Bias($\times 100$)				RMSE($\times 100$)			
κ^2		0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
N	$m_0 = 0$																
100		-0.03	-0.03	-0.03	-0.03	3.76	3.76	3.76	3.76	1.94	1.94	1.94	1.94	7.90	7.90	7.90	7.90
300		-0.04	-0.04	-0.04	-0.04	2.18	2.18	2.18	2.18	0.68	0.68	0.68	0.68	4.62	4.62	4.62	4.62
500		-0.01	-0.01	-0.01	-0.01	1.70	1.70	1.70	1.70	0.26	0.26	0.26	0.26	3.09	3.09	3.09	3.09
1000		-0.01	-0.01	-0.01	-0.01	1.22	1.22	1.22	1.22	0.18	0.18	0.18	0.18	2.24	2.24	2.24	2.24
	$m_0 = 1$																
100		0.11	-0.04	-0.05	-0.06	4.87	4.52	4.20	3.75	1.08	0.50	0.23	0.08	7.05	5.83	4.64	3.48
300		0.03	0.02	0.02	0.01	2.67	2.55	2.38	2.13	0.24	0.15	0.08	0.04	3.53	2.98	2.41	1.89
500		-0.05	-0.06	-0.06	-0.05	2.11	2.03	1.90	1.70	0.07	0.04	0.01	-0.01	2.58	2.28	1.88	1.49
1000		-0.03	-0.02	-0.01	-0.01	1.48	1.42	1.32	1.17	0.00	0.00	0.00	0.00	1.74	1.55	1.30	1.03
	$m_0 = 2$																
100		5.48	0.66	-0.08	-0.05	8.23	6.57	5.12	4.48	7.57	1.11	0.19	0.04	11.64	7.58	5.32	3.93
300		0.26	0.02	0.04	0.05	3.58	3.07	2.81	2.46	0.51	0.16	0.08	0.06	4.62	3.44	2.66	2.06
500		-0.12	-0.11	-0.10	-0.09	2.50	2.35	2.16	1.90	-0.06	-0.08	-0.09	-0.08	2.98	2.51	2.06	1.61
1000		-0.02	-0.01	0.00	0.00	1.84	1.74	1.59	1.39	0.03	0.03	0.01	0.00	2.02	1.75	1.44	1.11

Note: γ is the coefficient of the lagged dependent variable given in (1) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table 1.

Table A2(ii): Size($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m}

$T = 5, \gamma_0 = 0.4$					$T = 5, \gamma_0 = 0.8$				$T = 10, \gamma_0 = 0.4$				$T = 10, \gamma_0 = 0.8$			
κ^2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
N	$m_0 = 0$															
100	6.2	6.2	6.2	6.2	21.3	21.3	21.3	21.3	6.5	6.5	6.5	6.5	16.4	16.4	16.4	16.4
300	5.4	5.4	5.4	5.4	19.2	19.2	19.2	19.2	5.1	5.1	5.1	5.1	8.7	8.7	8.7	8.7
500	4.8	4.8	4.8	4.8	14.6	14.6	14.6	14.6	5.9	5.9	5.9	5.9	6.7	6.7	6.7	6.7
1000	4.5	4.5	4.5	4.5	12.1	12.1	12.1	12.1	5.4	5.4	5.4	5.4	5.7	5.7	5.7	5.7
	$m_0 = 1$															
100	52.6	15.7	5.1	6.2	54.3	21.6	19.6	12.6	6.9	6.0	6.1	5.7	12.1	7.6	4.9	4.9
300	9.3	3.8	5.1	5.9	16.9	17.0	11.9	6.7	4.0	4.3	4.5	5.1	4.3	4.3	4.7	5.2
500	2.6	3.3	3.9	4.5	12.7	12.3	7.1	4.5	5.4	5.7	6.0	6.1	4.5	5.1	5.4	5.5
1000	3.2	4.2	4.7	5.2	10.0	8.1	4.7	4.5	4.7	4.9	4.9	5.0	4.5	4.6	4.2	4.1
	$m_0 = 2$															
100	8.6	26.2	11.5	4.7	42.2	43.0	19.8	11.4	33.6	9.6	5.8	6.3	37.9	10.2	5.3	6.2
300	23.2	6.1	3.9	4.5	49.3	15.9	10.3	5.4	5.8	4.4	4.6	5.0	4.8	3.3	4.0	4.5
500	24.6	2.5	3.1	3.8	31.2	11.4	6.3	3.3	3.4	4.2	4.9	4.9	3.1	4.1	4.7	5.3
1000	2.6	2.6	3.3	3.8	7.8	6.6	4.4	3.9	3.4	4.0	4.7	4.9	3.6	4.0	4.0	4.3

See the note to Table A2(i).

Table A2(iii): Bias($\times 100$) and RMSE($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\sigma_v^2 = 1$)

$T=5, \gamma_0=0.4$									$T=5, \gamma_0=0.8$								
κ^2	Bias($\times 100$)				RMSE($\times 100$)				Bias($\times 100$)				RMSE($\times 100$)				
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	
γ																	
N	$m_0=0$																
100	-0.15	-0.15	-0.15	-0.15	3.45	3.45	3.45	3.45	-0.07	-0.07	-0.07	-0.07	3.02	3.02	3.02	3.02	
300	-0.04	-0.04	-0.04	-0.04	1.97	1.97	1.97	1.97	-0.05	-0.05	-0.05	-0.05	1.71	1.71	1.71	1.71	
500	0.02	0.02	0.02	0.02	1.47	1.47	1.47	1.47	0.00	0.00	0.00	0.00	1.27	1.27	1.27	1.27	
1000	-0.05	-0.05	-0.05	-0.05	1.08	1.08	1.08	1.08	-0.03	-0.03	-0.03	-0.03	0.93	0.93	0.93	0.93	
$m_0=1$																	
100	2.45	0.21	0.09	0.08	5.82	4.49	4.30	4.10	0.96	0.28	0.23	0.21	4.91	4.91	4.74	4.53	
300	-0.03	-0.04	-0.05	-0.06	2.45	2.42	2.39	2.31	-0.02	-0.02	-0.02	-0.03	2.64	2.60	2.56	2.47	
500	0.02	0.02	0.01	0.01	1.86	1.86	1.83	1.75	0.01	0.02	0.02	0.02	1.98	1.96	1.92	1.88	
1000	-0.05	-0.05	-0.04	-0.04	1.37	1.37	1.35	1.29	-0.02	-0.02	-0.02	-0.02	1.44	1.43	1.41	1.36	
$m_0=2$																	
100	1.29	1.49	0.37	0.21	4.21	5.39	4.70	4.27	0.57	0.69	0.47	0.36	3.78	4.90	4.99	4.60	
300	0.78	0.03	0.03	0.04	2.87	2.51	2.46	2.35	0.24	0.07	0.07	0.08	2.60	2.70	2.63	2.52	
500	0.31	0.07	0.07	0.07	2.16	1.96	1.94	1.87	0.11	0.09	0.10	0.10	2.12	2.13	2.10	2.03	
1000	0.06	0.05	0.05	0.05	1.41	1.41	1.39	1.33	0.05	0.05	0.05	0.05	1.51	1.49	1.47	1.41	
β																	
N	$m_0=0$																
100	-0.06	-0.06	-0.06	-0.06	4.44	4.44	4.44	4.44	-0.06	-0.06	-0.06	-0.06	4.55	4.55	4.55	4.55	
300	0.02	0.02	0.02	0.02	2.53	2.53	2.53	2.53	0.01	0.01	0.01	0.01	2.58	2.58	2.58	2.58	
500	0.04	0.04	0.04	0.04	1.92	1.92	1.92	1.92	0.04	0.04	0.04	0.04	1.97	1.97	1.97	1.97	
1000	0.00	0.00	0.00	0.00	1.38	1.38	1.38	1.38	0.00	0.00	0.00	0.00	1.40	1.40	1.40	1.40	
$m_0=1$																	
100	0.39	0.01	-0.01	-0.01	5.48	5.69	5.99	6.19	0.33	0.07	0.06	0.04	5.67	5.90	6.16	6.33	
300	-0.10	-0.13	-0.15	-0.16	3.00	3.20	3.39	3.52	-0.10	-0.12	-0.14	-0.15	3.11	3.29	3.46	3.57	
500	0.09	0.09	0.09	0.08	2.35	2.51	2.65	2.75	0.10	0.10	0.09	0.08	2.43	2.58	2.70	2.79	
1000	0.04	0.04	0.05	0.06	1.66	1.77	1.88	1.95	0.04	0.05	0.06	0.07	1.71	1.82	1.91	1.97	
$m_0=2$																	
100	0.27	0.29	0.27	0.33	5.73	6.85	8.33	10.58	0.28	0.38	0.41	0.44	5.88	7.11	8.56	10.75	
300	0.22	0.15	0.18	0.20	3.23	3.75	4.62	5.89	0.22	0.18	0.20	0.23	3.32	3.84	4.67	5.91	
500	0.10	0.09	0.11	0.14	2.49	2.90	3.55	4.51	0.11	0.12	0.14	0.17	2.60	3.00	3.63	4.57	
1000	-0.03	-0.04	-0.06	-0.09	1.77	2.05	2.51	3.18	-0.02	-0.03	-0.05	-0.07	1.83	2.11	2.55	3.21	
$T=10, \gamma_0=0.4$									$T=10, \gamma_0=0.8$								
κ^2	Bias($\times 100$)				RMSE($\times 100$)				Bias($\times 100$)				RMSE($\times 100$)				
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	
γ																	
N	$m_0=0$																
100	-0.06	-0.06	-0.06	-0.06	1.95	1.95	1.95	1.95	-0.03	-0.03	-0.03	-0.03	1.37	1.37	1.37	1.37	
300	0.08	0.08	0.08	0.08	1.14	1.14	1.14	1.14	0.04	0.04	0.04	0.04	0.77	0.77	0.77	0.77	
500	-0.01	-0.01	-0.01	-0.01	0.86	0.86	0.86	0.86	0.00	0.00	0.00	0.00	0.58	0.58	0.58	0.58	
1000	0.00	0.00	0.00	0.00	0.62	0.62	0.62	0.62	0.00	0.00	0.00	0.00	0.42	0.42	0.42	0.42	
$m_0=1$																	
100	-0.07	-0.10	-0.10	-0.11	2.23	2.19	2.15	2.09	-0.06	-0.07	-0.07	-0.07	1.60	1.57	1.54	1.49	
300	0.03	0.03	0.03	0.03	1.23	1.22	1.20	1.16	0.02	0.02	0.02	0.02	0.85	0.84	0.83	0.79	
500	-0.02	-0.02	-0.02	-0.02	0.94	0.93	0.92	0.90	-0.01	-0.01	-0.01	-0.01	0.67	0.66	0.65	0.63	
1000	0.01	0.01	0.01	0.01	0.68	0.68	0.67	0.65	0.00	0.00	0.00	0.01	0.47	0.47	0.46	0.44	
$m_0=2$																	
100	1.17	0.02	-0.09	-0.08	2.81	2.43	2.33	2.27	0.31	-0.02	-0.05	-0.05	1.68	1.63	1.59	1.53	
300	-0.04	-0.07	-0.06	-0.06	1.37	1.35	1.33	1.29	-0.02	-0.03	-0.02	-0.02	0.94	0.93	0.91	0.88	
500	-0.03	-0.03	-0.03	-0.03	1.00	1.00	0.98	0.96	-0.01	-0.01	-0.01	-0.02	0.71	0.70	0.69	0.67	
1000	0.02	0.02	0.02	0.02	0.71	0.71	0.70	0.69	0.01	0.01	0.01	0.01	0.49	0.49	0.48	0.47	
β																	
N	$m_0=0$																
100	-0.01	-0.01	-0.01	-0.01	3.04	3.04	3.04	3.04	-0.02	-0.02	-0.02	-0.02	3.02	3.02	3.02	3.02	
300	-0.05	-0.05	-0.05	-0.05	1.73	1.73	1.73	1.73	-0.03	-0.03	-0.03	-0.03	1.71	1.71	1.71	1.71	
500	0.00	0.00	0.00	0.00	1.34	1.34	1.34	1.34	0.00	0.00	0.00	0.00	1.33	1.33	1.33	1.33	
1000	0.01	0.01	0.01	0.01	0.96	0.96	0.96	0.96	0.01	0.01	0.01	0.01	0.95	0.95	0.95	0.95	
$m_0=1$																	
100	0.09	0.09	0.09	0.10	3.73	3.87	3.98	4.04	0.07	0.08	0.07	0.08	3.73	3.87	3.98	4.04	
300	0.01	0.01	0.01	0.01	2.15	2.24	2.29	2.32	0.02	0.02	0.02	0.02	2.14	2.22	2.28	2.31	
500	0.01	0.01	0.00	0.00	1.61	1.69	1.74	1.78	0.01	0.00	0.00	-0.01	1.59	1.66	1.72	1.76	
1000	0.03	0.03	0.03	0.03	1.13	1.18	1.21	1.23	0.03	0.03	0.04	0.04	1.12	1.17	1.20	1.22	
$m_0=2$																	
100	-0.20	0.09	0.15	0.15	4.55	5.25	6.27	7.46	0.12	0.13	0.13	0.15	4.51	5.23	6.24	7.42	
300	0.10	0.10	0.09	0.08	2.55	3.02	3.63	4.29	0.10	0.09	0.08	0.07	2.53	3.00	3.61	4.27	
500	0.01	0.01	0.02	0.03	1.99	2.36	2.85	3.36	0.01	0.01	0.01	0.03	1.98	2.36	2.84	3.35	
1000	0.01	0.02	0.04	0.06	1.38	1.63	1.96	2.32	0.02	0.03	0.05	0.07	1.37	1.63	1.95	2.31	

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (1). See also the note to Table A2(i).

Table A2(iv): Size($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\sigma_v^2 = 1$)

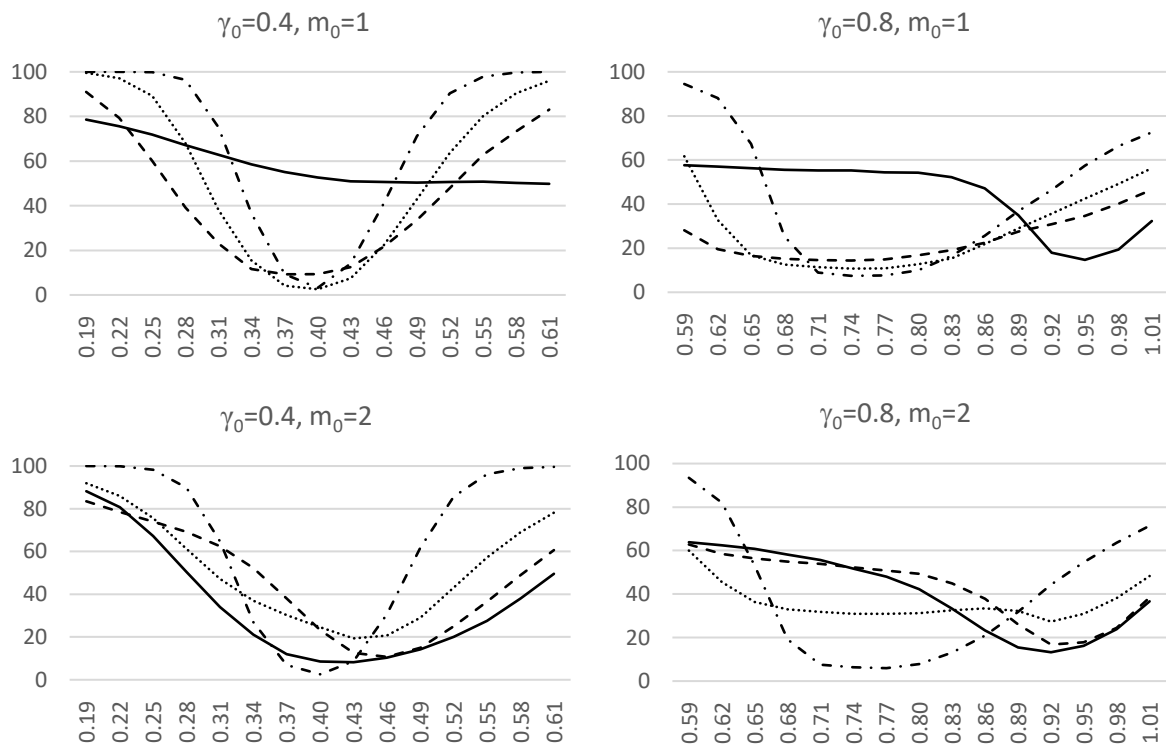
κ^2	$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$				$T = 10, \gamma_0 = 0.4$				$T = 10, \gamma_0 = 0.8$			
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
γ																
N	$m_0 = 0$															
100	5.9	5.9	5.9	5.9	6.6	6.6	6.6	6.6	5.4	5.4	5.4	5.4	5.8	5.8	5.8	5.8
300	5.6	5.6	5.6	5.6	6.1	6.1	6.1	6.1	5.3	5.3	5.3	5.3	5.1	5.1	5.1	5.1
500	5.1	5.1	5.1	5.1	4.4	4.4	4.4	4.4	4.5	4.5	4.5	4.5	4.3	4.3	4.3	4.3
1000	5.1	5.1	5.1	5.1	5.8	5.8	5.8	5.8	4.9	4.9	4.9	4.9	5.8	5.8	5.8	5.8
	$m_0 = 1$															
100	14.8	4.6	5.1	5.7	5.4	4.4	5.2	5.8	5.8	5.7	6.0	6.1	5.8	6.3	6.5	6.6
300	3.0	3.8	4.4	4.9	3.2	4.4	5.1	5.4	5.4	5.4	5.2	5.6	3.7	4.2	4.0	4.0
500	2.3	3.0	3.8	3.9	2.4	3.4	3.9	4.1	5.3	5.4	5.5	5.3	4.8	5.0	5.1	5.4
1000	3.2	4.1	4.5	5.0	3.5	4.1	4.5	4.8	5.1	5.2	5.4	5.2	5.0	5.3	5.4	5.4
	$m_0 = 2$															
100	7.5	8.8	5.8	5.7	6.2	4.5	4.7	5.1	11.1	5.3	5.8	6.5	6.6	5.3	5.9	6.3
300	8.0	3.3	4.1	4.4	4.4	3.7	4.8	5.3	4.0	5.1	5.4	5.5	3.4	4.4	4.8	4.8
500	5.6	2.9	3.6	4.3	3.0	3.3	4.6	5.1	3.4	3.8	4.3	4.9	3.7	4.4	4.7	5.0
1000	2.6	3.0	3.6	4.3	2.6	3.6	4.2	4.4	3.7	4.1	4.3	4.5	3.4	3.8	4.1	4.4
β																
N	$m_0 = 0$															
100	5.6	5.6	5.6	5.6	5.4	5.4	5.4	5.4	6.5	6.5	6.5	6.5	6.6	6.6	6.6	6.6
300	5.7	5.7	5.7	5.7	5.8	5.8	5.8	5.8	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
500	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.7	5.7	5.7	5.7	5.6	5.6	5.6	5.6
1000	5.0	5.0	5.0	5.0	4.9	4.9	4.9	4.9	5.6	5.6	5.6	5.6	5.8	5.8	5.8	5.8
	$m_0 = 1$															
100	4.8	5.1	5.6	5.6	4.9	5.3	5.5	5.6	6.2	6.4	6.3	6.1	5.9	6.3	6.2	6.4
300	4.8	4.4	4.9	5.0	4.6	4.8	4.9	5.2	6.4	6.5	6.0	5.6	5.9	6.1	5.6	5.4
500	5.2	5.7	5.5	5.4	4.9	5.1	5.3	5.3	4.9	5.0	5.2	5.4	5.2	5.2	5.2	5.4
1000	5.1	5.6	5.5	5.8	5.2	5.4	5.7	5.6	4.4	4.5	4.4	4.4	4.6	4.7	4.7	4.6
	$m_0 = 2$															
100	6.4	6.1	6.5	6.8	6.5	6.2	5.8	6.7	5.1	4.3	4.9	5.8	5.0	4.1	5.0	5.7
300	4.5	4.9	5.2	5.4	4.5	5.5	5.3	5.2	4.4	5.1	5.3	5.7	4.7	5.4	5.4	5.6
500	4.0	4.6	5.0	5.2	4.5	4.9	5.0	5.3	5.7	5.9	5.7	5.6	5.8	6.1	5.9	5.5
1000	5.4	5.3	4.9	4.9	4.8	5.1	5.2	4.8	5.9	5.7	5.3	4.9	6.2	6.0	5.3	5.0

See the note to Table A2(i).

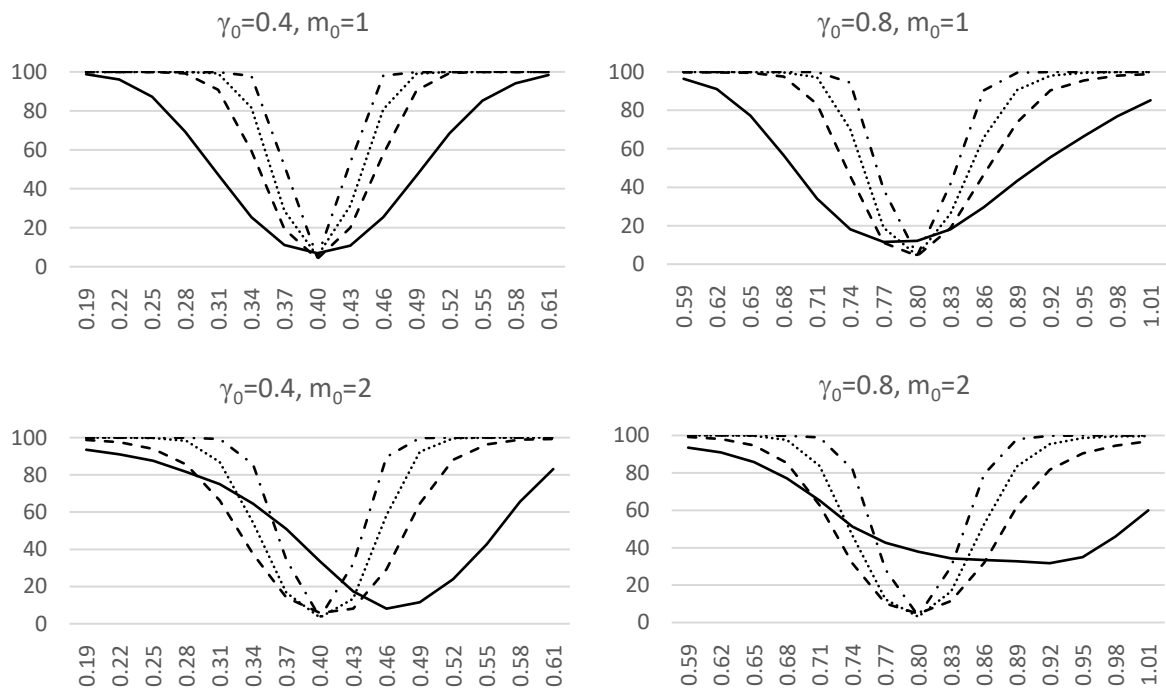
A3: Power Functions

Figure A3(i): Power functions for γ in the case of the AR(1) panel data model with different values of m and N ($\kappa^2=0.25$)

Panel A: $T=5$



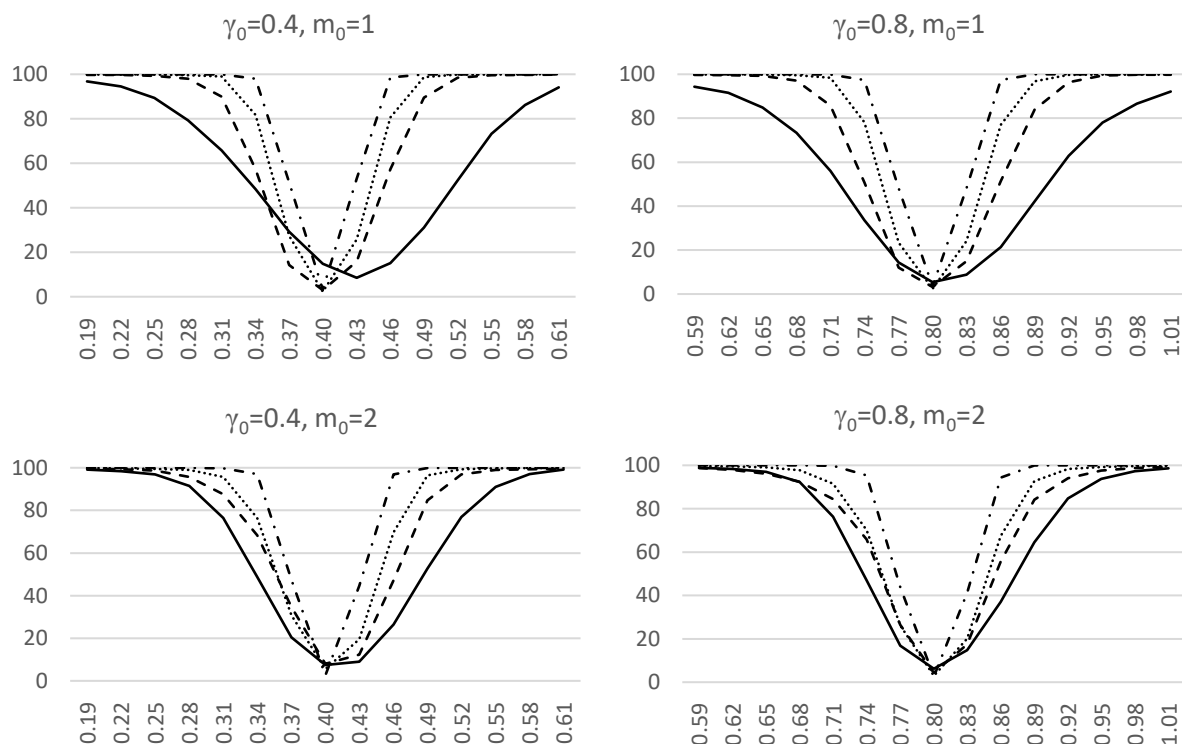
Panel B: $T=10$



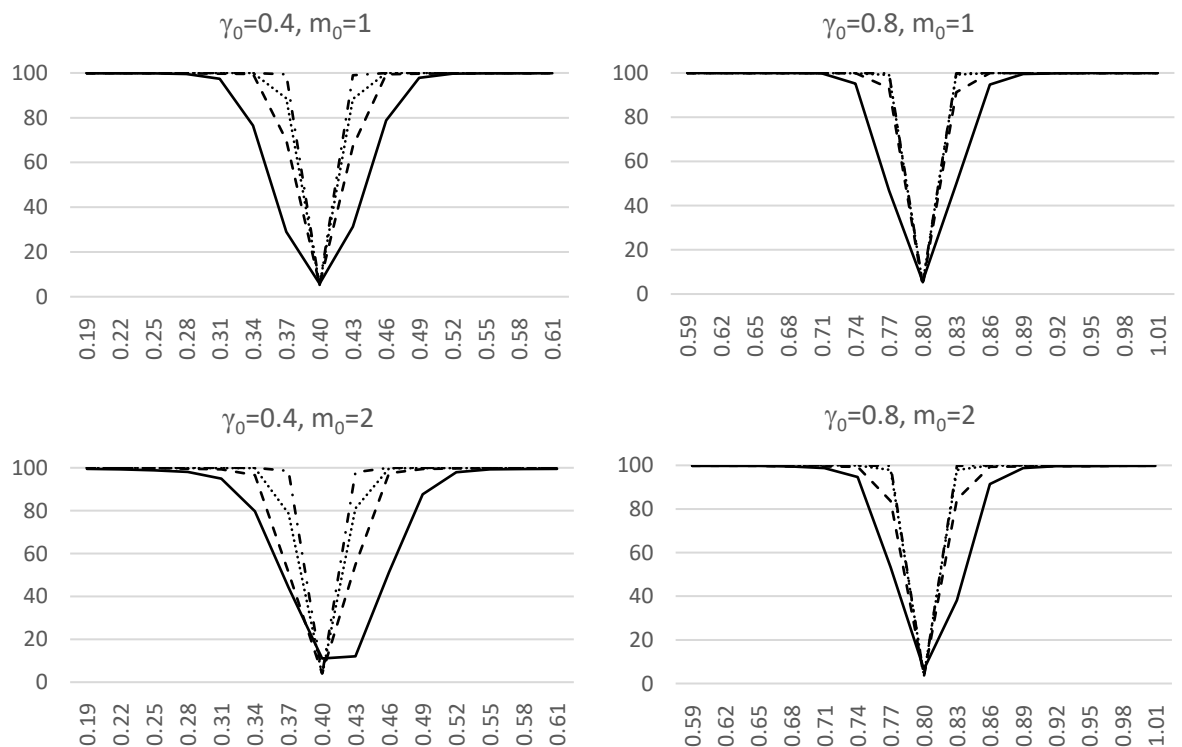
Note: — $N=100$ - - - $N=300$ $N=500$ - . - $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table 1.

Figure A3(ii): Power functions for γ in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.25$)

Panel A: $T=5$



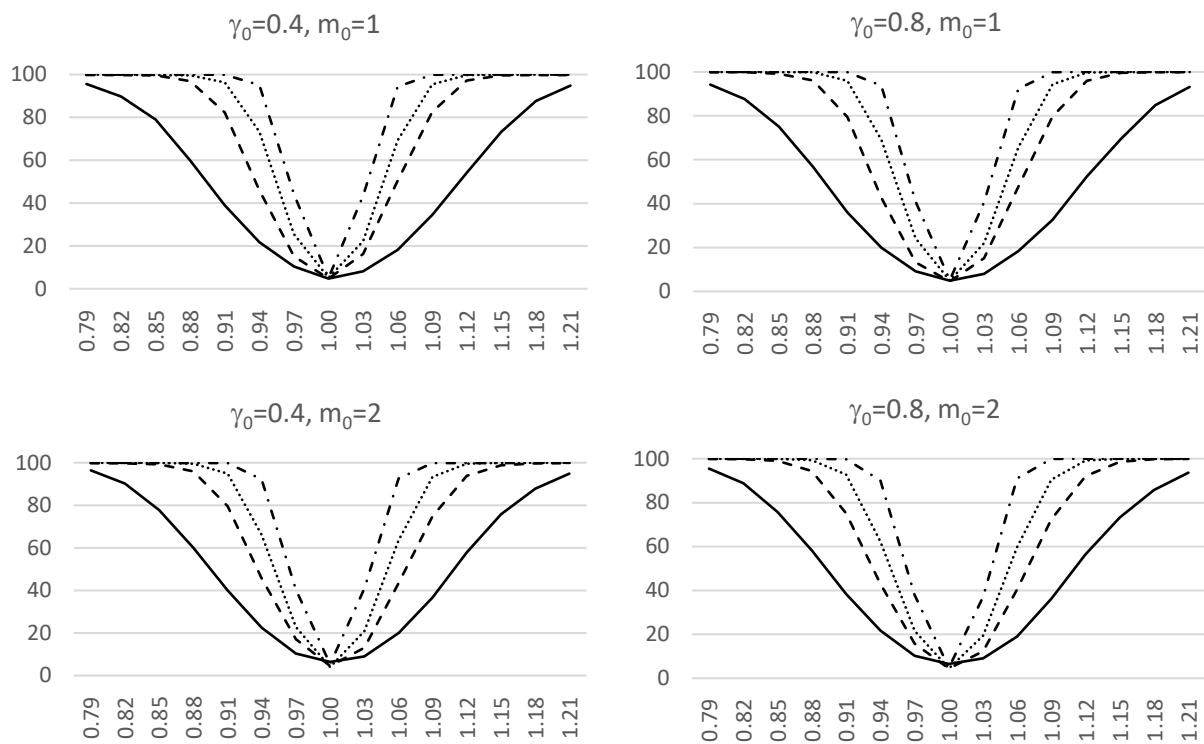
Panel B: $T=10$



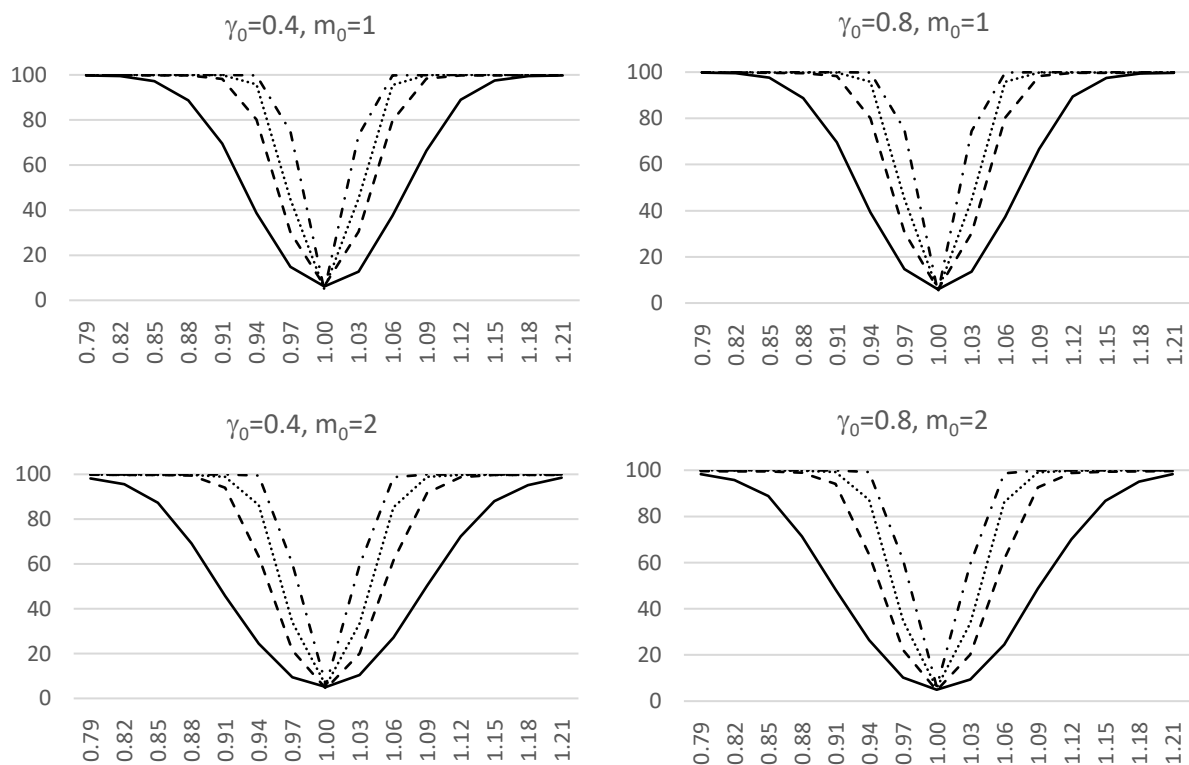
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13). See also the note to Table 1.

Figure A3(iii): Power functions for β in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.25$)

Panel A: $T=5$



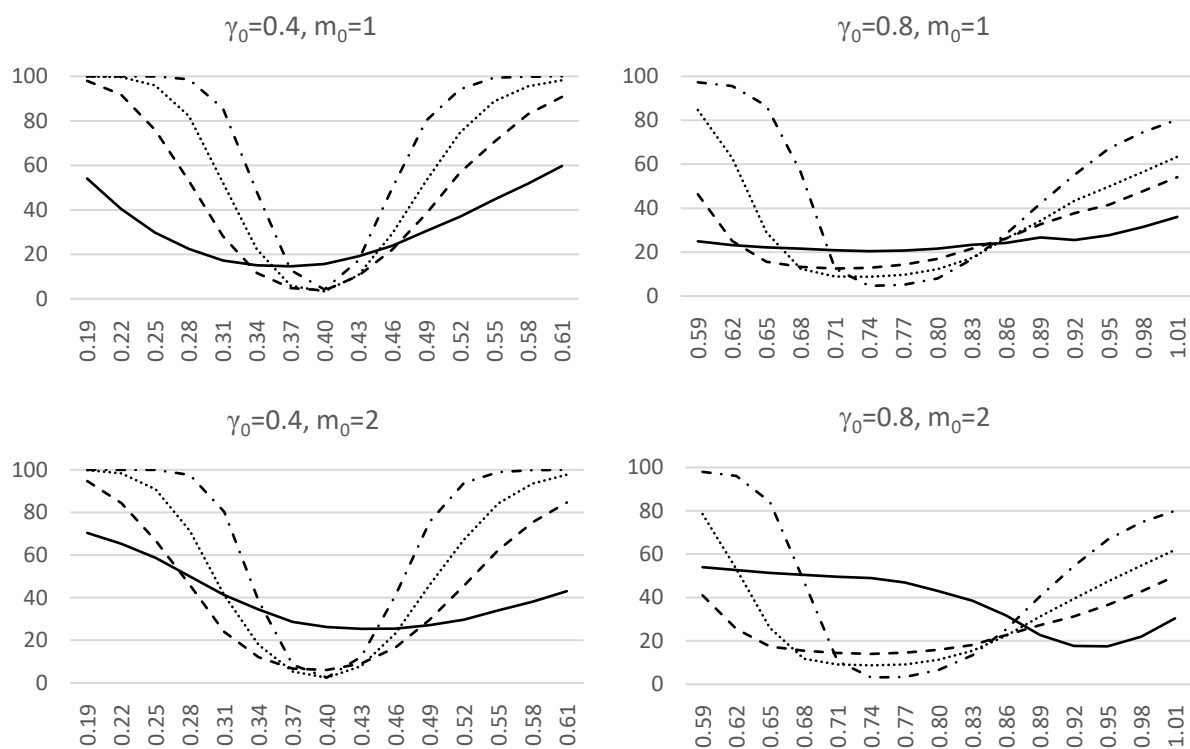
Panel B: $T=10$



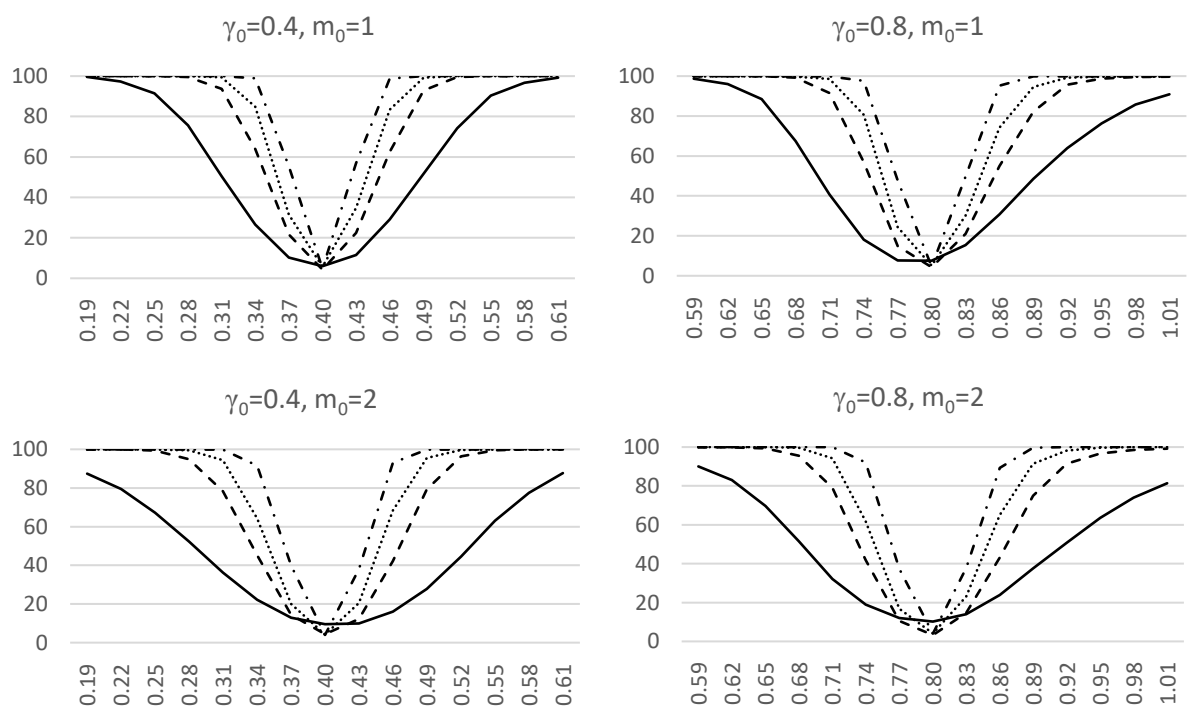
Note: — $N=100$ - - - $N=300$ $N=500$ - . - $N=1000$. β is the coefficient of the \mathbf{x}_{it} regressors in (13). See also the note to Figure A3(ii).

Figure A3(iv): Power functions for γ in the case of the AR(1) panel data model with different values of m and N ($\kappa^2=0.5$)

Panel A: $T=5$



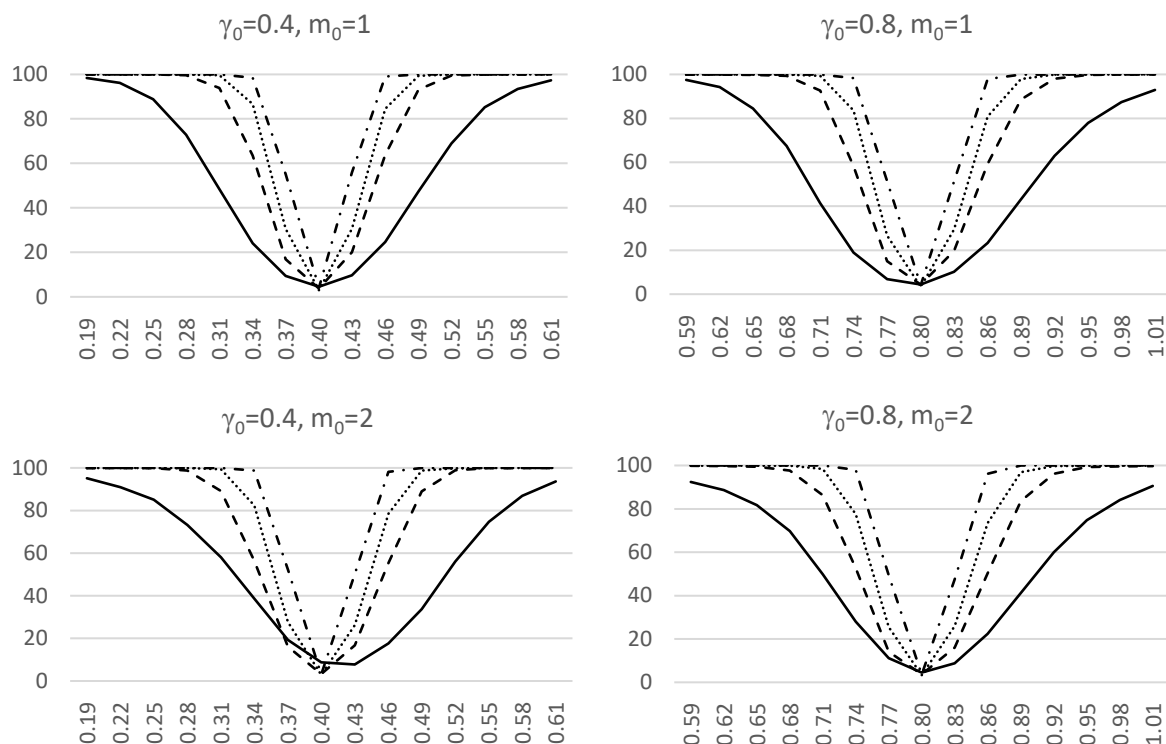
Panel B: $T=10$



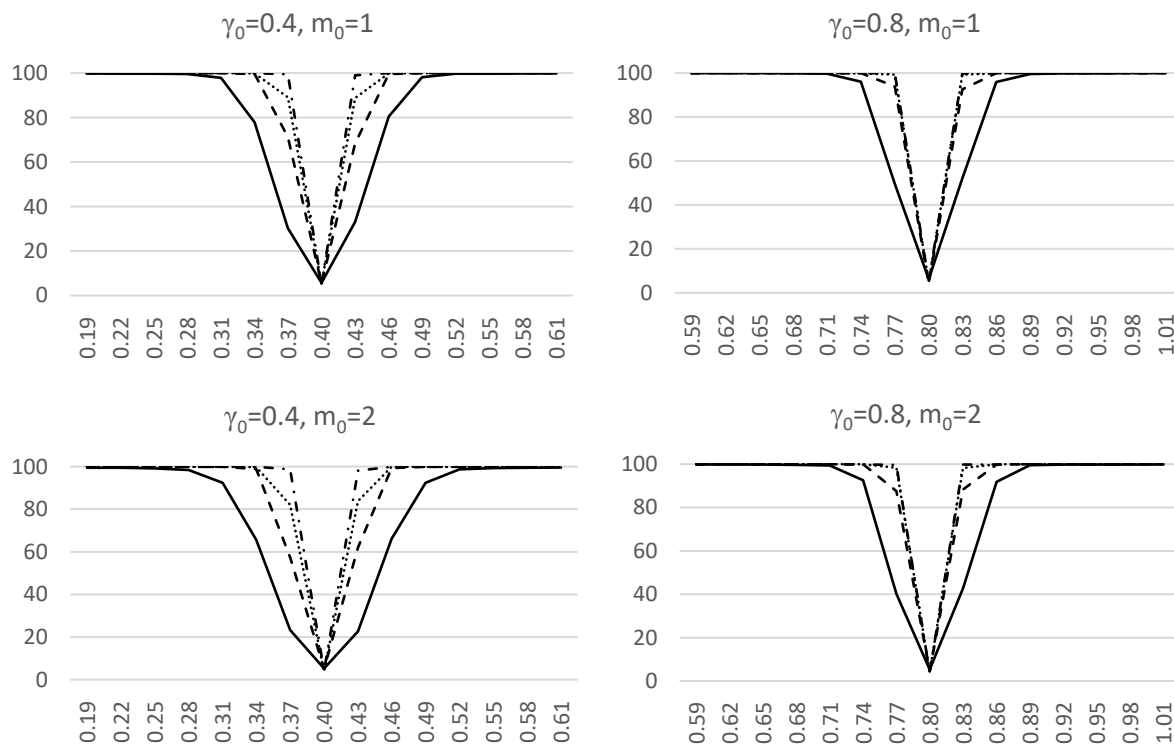
Note: — $N=100$ - - - $N=300$ $N=500$ - . - . $N=1000$. See also the note to Table A3(i).

Figure A3(v): Power functions for γ in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.5$)

Panel A: $T=5$



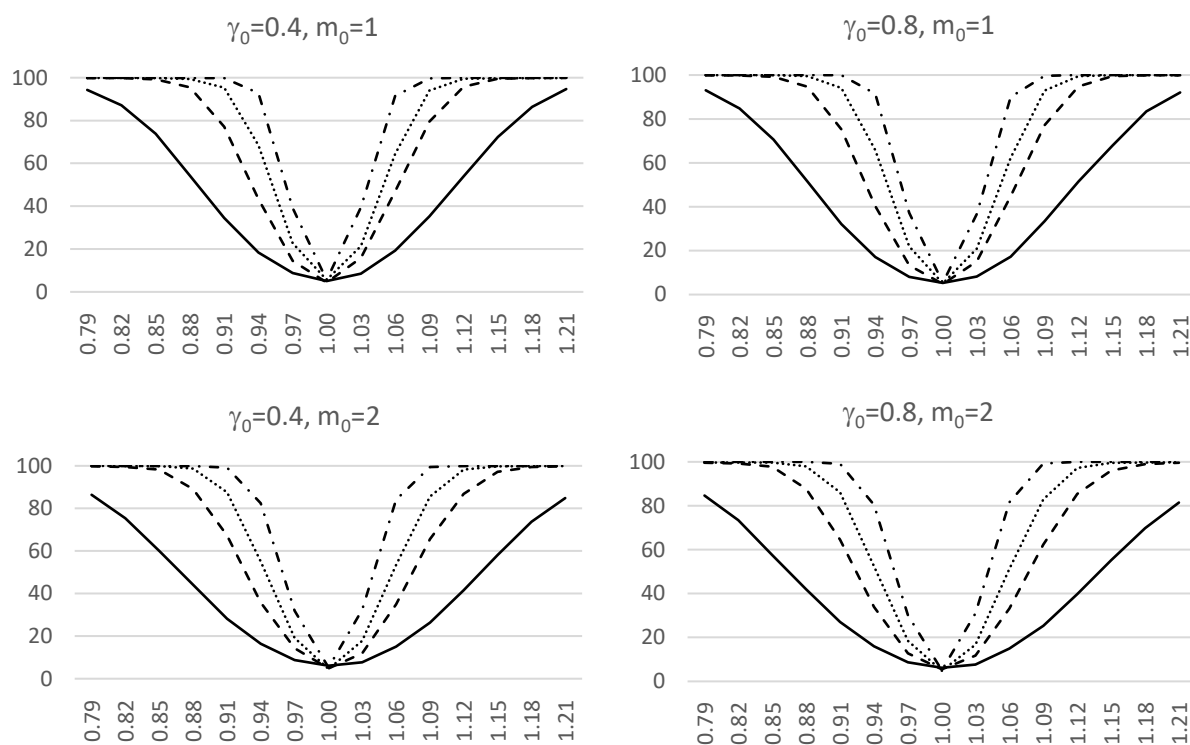
Panel B: $T=10$



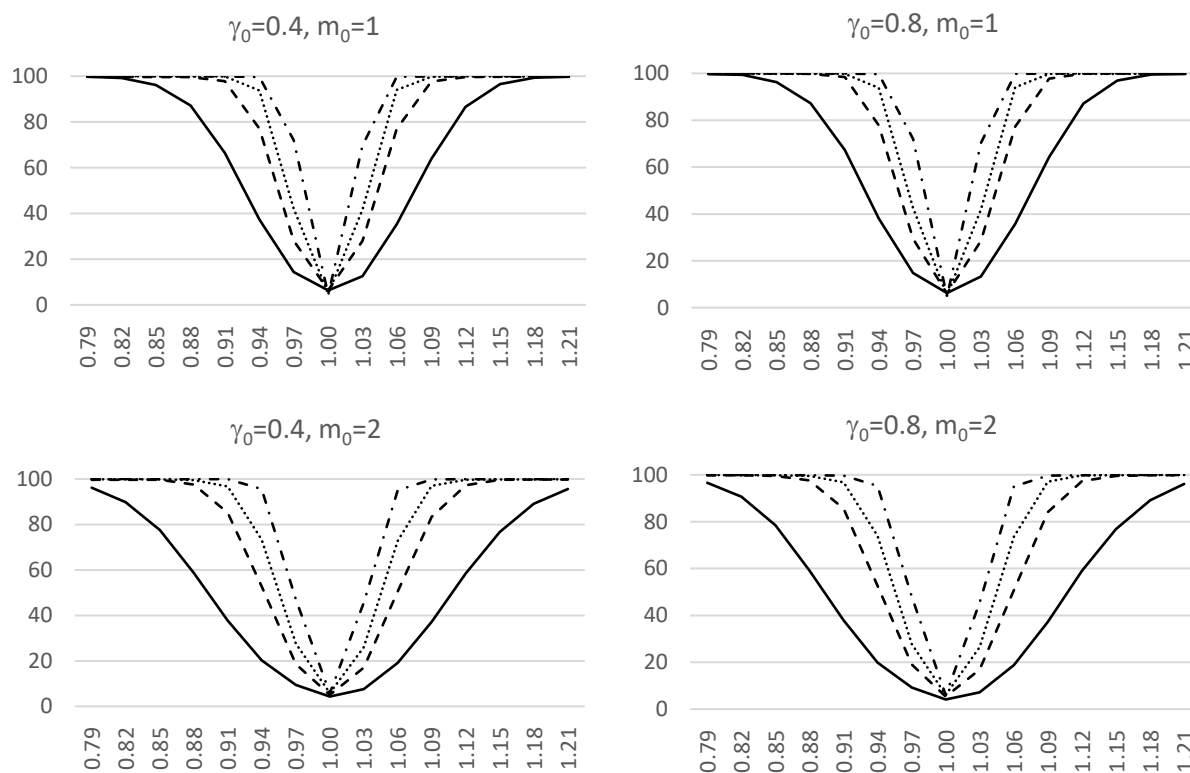
Note: — $N=100$ - - - $N=300$ $N=500$ - . - . $N=1000$. See also the note to Table A3(ii).

Figure A3(vi): Power functions for β in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.5$)

Panel A: $T=5$



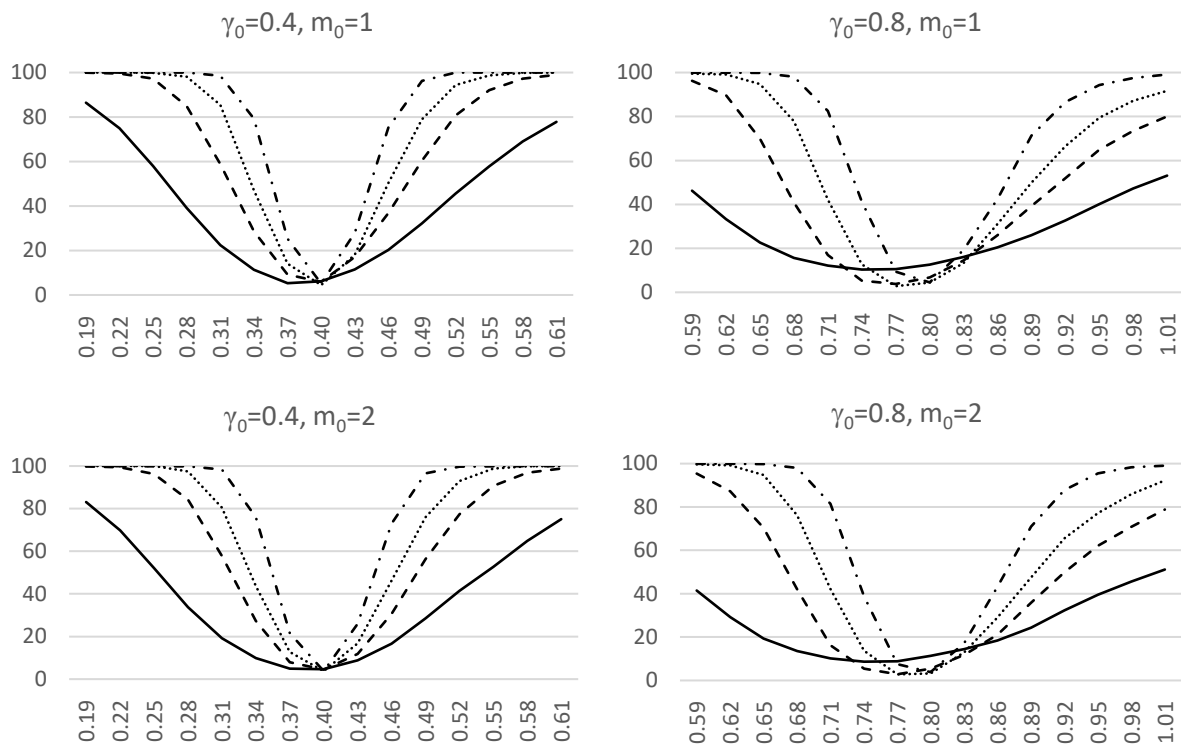
Panel B: $T=10$



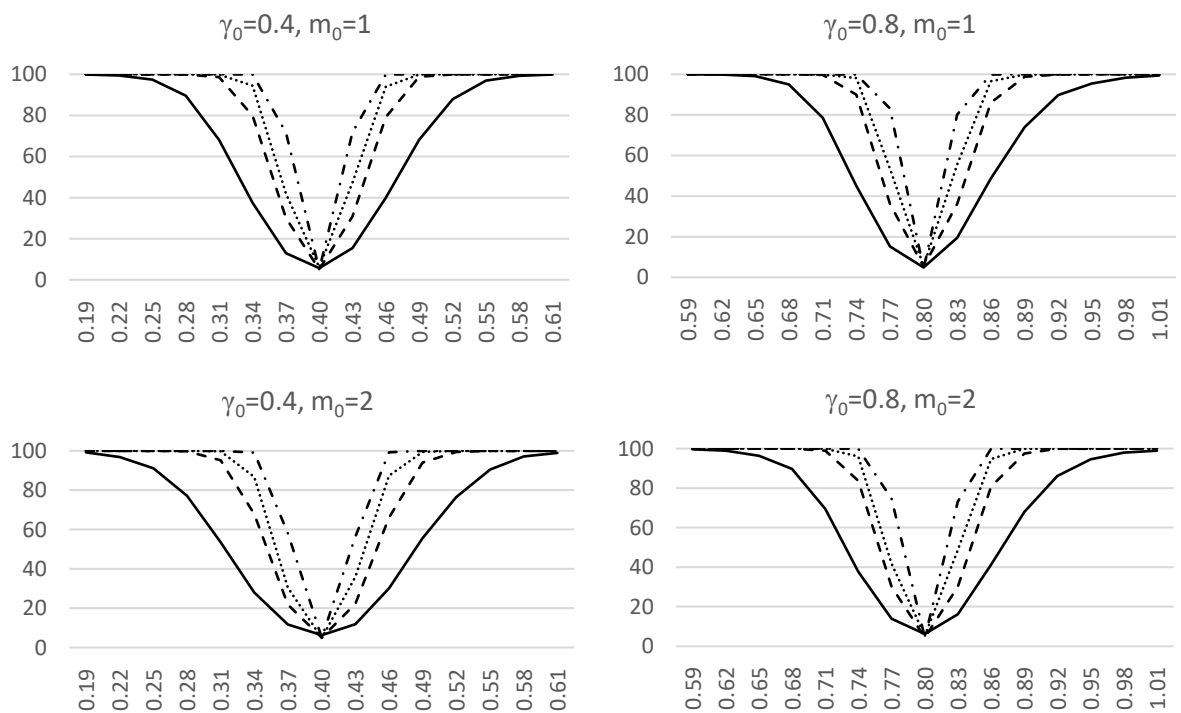
Note: — $N=100$ --- $N=300$ $N=500$ - - - - $N=1000$. See also note to Figure A3(v).

Figure A3(vii): Power functions for γ in the case of the AR(1) panel data model with different values of m and N ($\kappa^2=2$)

Panel A: $T=5$



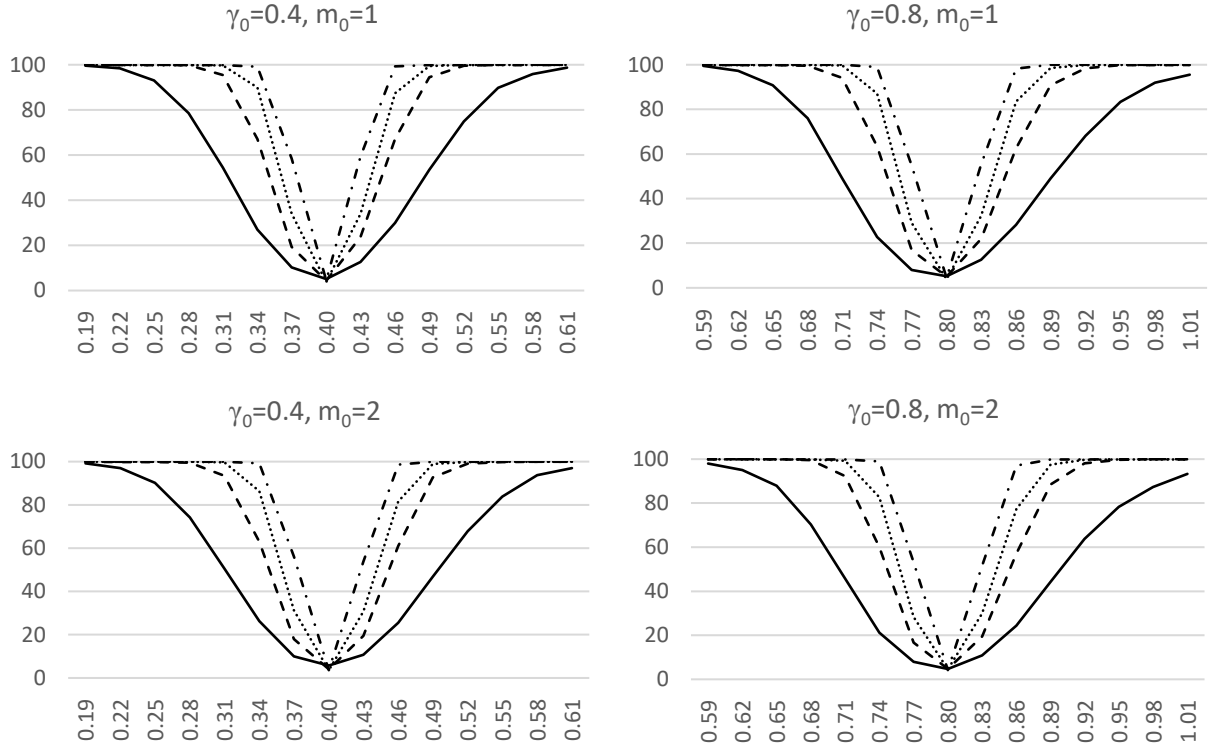
Panel B: $T=10$



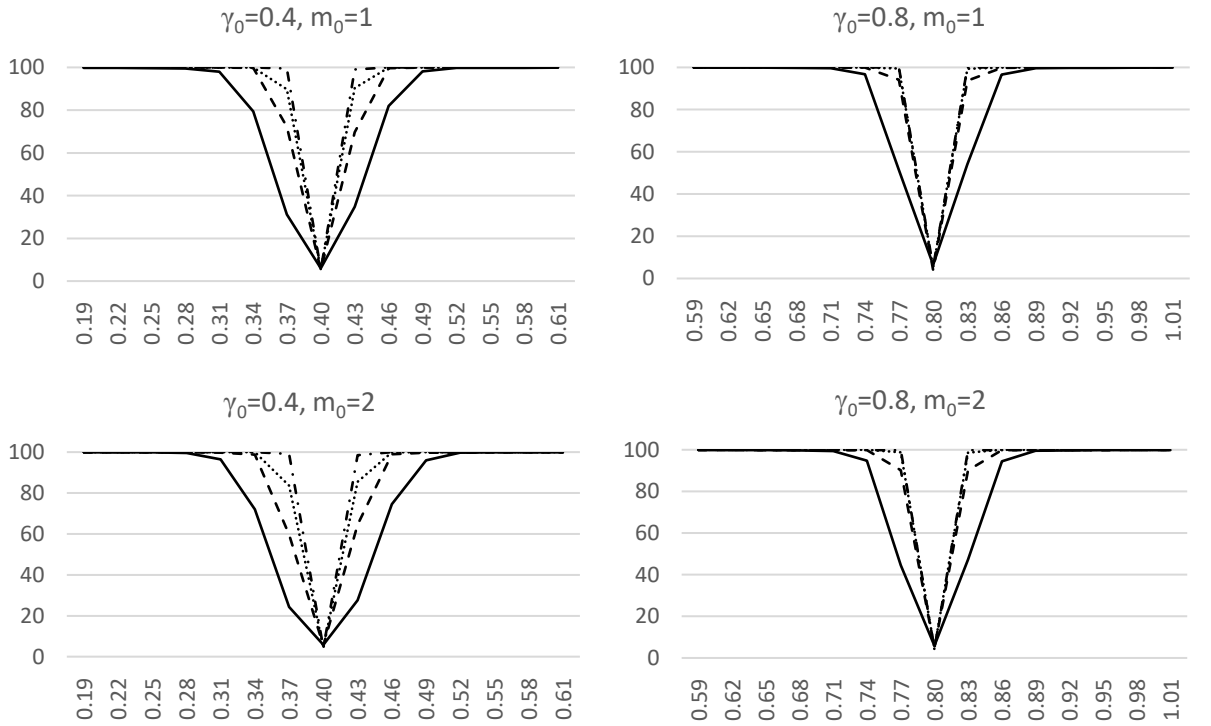
Note: — $N=100$ - - - $N=300$ $N=500$ - . - $N=1000$. See also the note to Table A3(i).

Figure A3(viii): Power functions for γ in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=2$)

Panel A: $T=5$



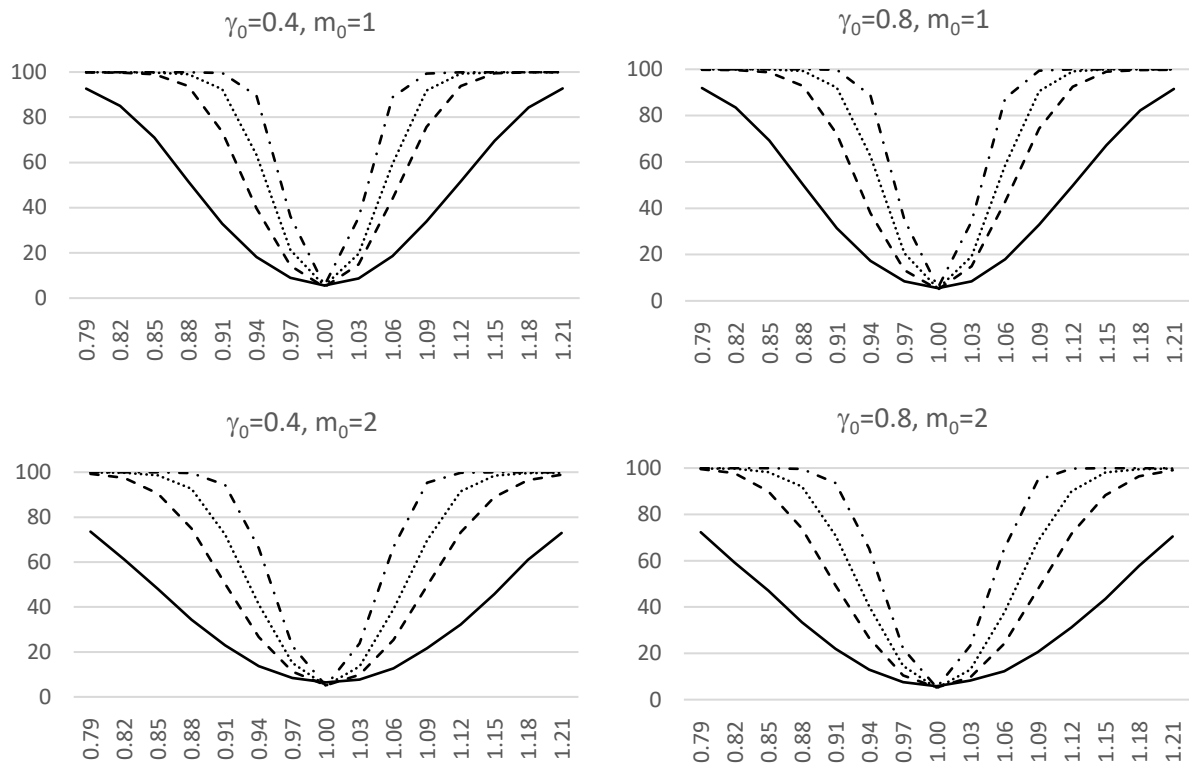
Panel B: $T=10$



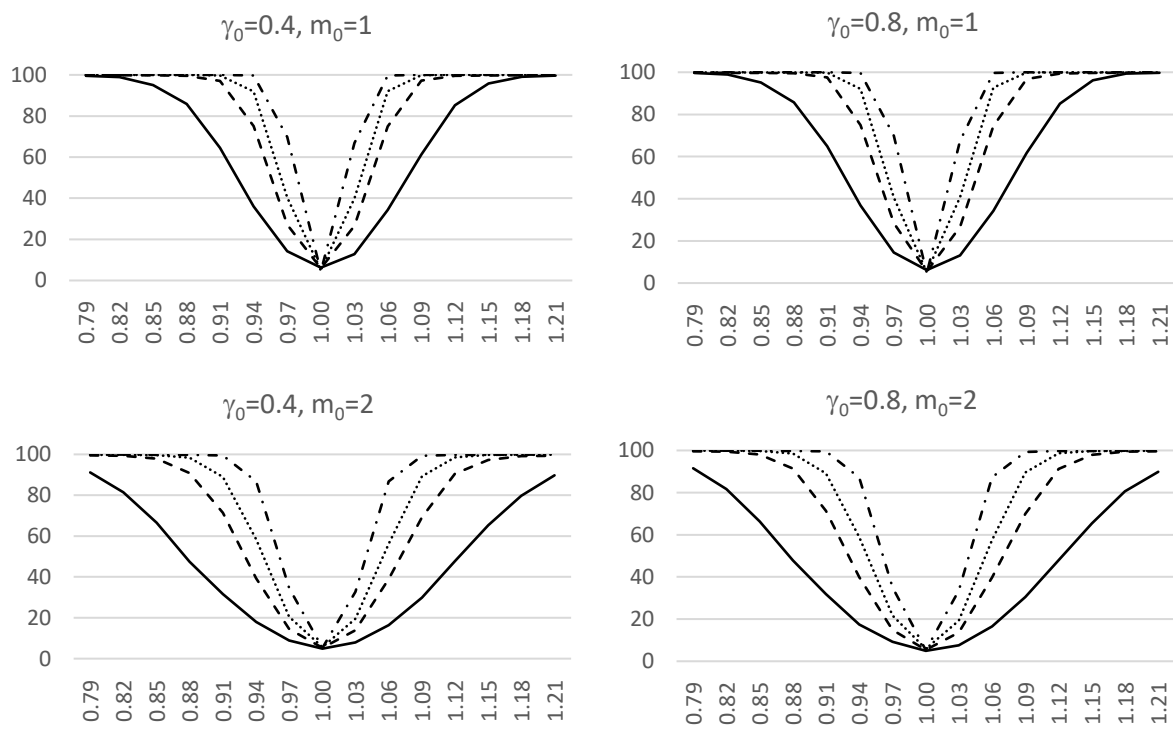
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. See also the note to Table A3(ii).

Figure A3(ix): Power functions for β in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=2$)

Panel A: $T=5$



Panel B: $T=10$



Note: — $N=100$ - - - $N=300$ $N=500$ - . - $N=1000$. See also the note to Figure A3(viii).

S.8 Unit Root Case ($\gamma_0 = 1$)

B1: Selecting the number of factors

Table B1(i): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure in the case of the AR(1) panel data model

$T = 5$													
κ^2		0.25			0.5			1			2		
$N \setminus m_0$	0	1	2	0	1	2	0	1	2	0	1	2	
100	99.5	58.8	1.4	99.5	98.8	32.1	99.5	99.6	96.5	99.5	99.6	100.0	
300	99.8	100.0	29.7	99.8	99.9	98.9	99.8	99.9	100.0	99.8	99.9	100.0	
500	99.8	100.0	74.7	99.8	100.0	100.0	99.8	100.0	100.0	99.8	100.0	100.0	
1000	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	
$T = 10$													
κ^2		0.25			0.5			1			2		
$N \setminus m_0$	0	1	2	0	1	2	0	1	2	0	1	2	
100	99.5	97.6	18.7	99.5	99.6	94.8	99.5	99.6	99.6	99.5	99.6	99.6	
300	100.0	99.9	97.8	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	
500	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	
1000	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	

Note: \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$. See also the note to Table 12.

Table B1(ii): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure in the case of the ARX(1) panel data model

$T = 5$																		
N			$\kappa^2 = 0.25$						$\kappa^2 = 0.5$									
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
100	99.5	99.6	99.6	57.8	57.7	57.6	1.3	1.3	1.2	99.5	99.6	99.6	99.2	99.3	99.2	32.5	32.3	32.3
300	100.0	100.0	100.0	100.0	100.0	100.0	26.3	26.4	26.4	100.0	100.0	100.0	100.0	100.0	100.0	99.5	99.5	99.5
500	99.9	99.9	99.9	100.0	100.0	100.0	71.3	71.5	71.5	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	99.8	99.8	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$\kappa^2 = 1$																		
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
100	99.5	99.6	99.6	99.9	99.9	99.9	97.3	97.2	97.3	99.5	99.6	99.6	99.9	99.9	99.9	100.0	100.0	100.0
300	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
500	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$T = 10$																		
N			$\kappa^2 = 0.25$						$\kappa^2 = 0.5$									
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
100	99.3	99.3	99.3	98.1	98.2	98.2	20.1	19.95	19.7	99.3	99.3	99.3	99.7	99.7	99.7	95.05	94.9	94.9
300	100.0	100.0	100.0	100.0	100.0	100.0	98.3	98.3	98.3	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0
$\kappa^2 = 1$																		
m_0	0			1			2			0			1			2		
σ_v^2	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
100	99.3	99.3	99.3	99.7	99.7	99.7	100.0	99.8	99.8	99.3	99.3	99.3	99.7	99.7	99.7	99.6	99.6	99.7
300	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

See the note to Table B1(i).

B2: Bias, RMSE and Size

Table B2(i): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m}

$T = 5$												
κ^2	Bias($\times 100$)				RMSE($\times 100$)				Size($\times 100$)			
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
N	$m_0 = 0$											
100	-1.49	-1.49	-1.49	-1.49	2.74	2.74	2.74	2.74	3.8	3.8	3.8	3.8
300	-0.89	-0.89	-0.89	-0.89	1.69	1.69	1.69	1.69	3.1	3.1	3.1	3.1
500	-0.67	-0.67	-0.67	-0.67	1.08	1.08	1.08	1.08	2.6	2.6	2.6	2.6
1000	-0.53	-0.53	-0.53	-0.53	1.25	1.25	1.25	1.25	2.4	2.4	2.4	2.4
	$m_0 = 1$											
100	-2.81	-3.04	-2.99	-2.97	5.44	5.80	5.70	5.66	4.3	4.4	5.4	6.0
300	-1.87	-1.84	-1.83	-1.82	3.48	3.45	3.43	3.42	2.8	4.0	4.9	5.2
500	-1.38	-1.35	-1.34	-1.34	2.34	2.27	2.25	2.24	2.8	3.4	3.7	3.9
1000	-0.99	-0.98	-0.97	-0.97	1.67	1.65	1.64	1.64	2.2	3.3	3.4	3.9
	$m_0 = 2$											
100	-2.01	-2.93	-3.00	-2.91	3.64	5.57	5.09	4.90	4.2	3.5	5.1	5.9
300	-1.65	-1.75	-1.70	-1.68	3.39	3.05	2.93	2.88	2.3	3.0	3.9	4.5
500	-1.43	-1.39	-1.37	-1.36	2.53	2.34	2.30	2.28	1.1	2.3	3.2	3.9
1000	-1.01	-0.99	-0.99	-0.98	1.70	1.66	1.65	1.65	1.4	2.5	3.3	3.7
$T = 10$												
κ^2	Bias($\times 100$)				RMSE($\times 100$)				Size($\times 100$)			
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
N	$m_0 = 0$											
100	-0.53	-0.53	-0.53	-0.53	1.24	1.24	1.24	1.24	3.3	3.3	3.3	3.3
300	-0.33	-0.33	-0.33	-0.33	0.50	0.50	0.50	0.50	4.2	4.2	4.2	4.2
500	-0.26	-0.26	-0.26	-0.26	0.37	0.37	0.37	0.37	2.5	2.5	2.5	2.5
1000	-0.20	-0.20	-0.20	-0.20	0.33	0.33	0.33	0.33	3.0	3.0	3.0	3.0
	$m_0 = 1$											
100	-0.63	-0.62	-0.61	-0.61	1.03	1.01	1.01	1.00	2.3	2.7	3.0	3.2
300	-0.40	-0.40	-0.39	-0.39	0.99	0.96	0.95	0.95	2.4	2.7	2.8	2.8
500	-0.31	-0.31	-0.31	-0.31	0.46	0.46	0.46	0.46	2.1	2.7	2.9	3.1
1000	-0.24	-0.24	-0.24	-0.24	0.33	0.33	0.33	0.33	2.2	2.3	2.4	2.6
	$m_0 = 2$											
100	-0.67	-0.68	-0.65	-0.65	1.43	1.41	1.11	1.10	3.2	3.3	3.8	4.0
300	-0.39	-0.38	-0.39	-0.38	0.61	0.60	0.59	0.59	1.5	1.9	2.3	2.8
500	-0.32	-0.32	-0.31	-0.32	0.48	0.48	0.48	0.48	1.8	2.2	2.4	2.8
1000	-0.24	-0.24	-0.24	-0.24	0.33	0.33	0.33	0.33	1.4	1.8	2.1	2.2

Note: γ is the coefficient of the lagged dependent variable given in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table B1(i).

Table B2(ii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\sigma_v^2 = 1$)

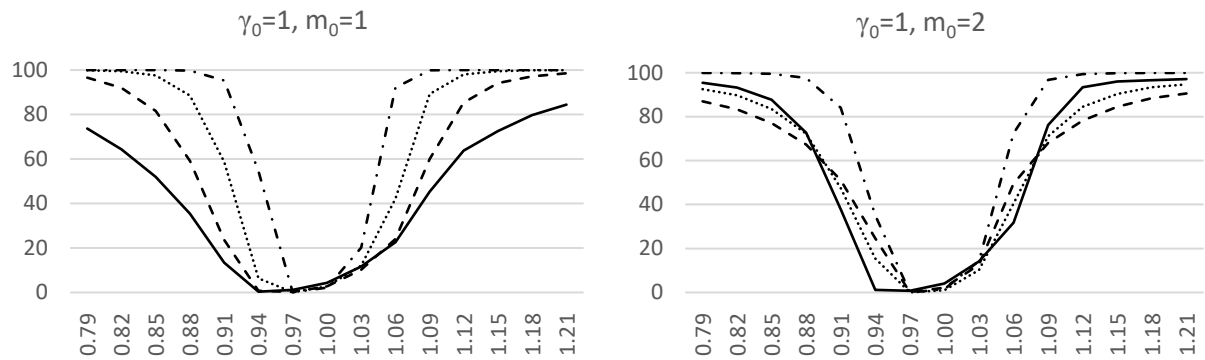
$T = 5$												
κ^2	Bias($\times 100$)				RMSE($\times 100$)				Size($\times 100$)			
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
γ												
N	$m_0 = 0$											
100	-1.28	-1.28	-1.28	-1.28	2.17	2.17	2.17	2.17	3.7	3.7	3.7	3.7
300	-0.77	-0.77	-0.77	-0.77	1.27	1.27	1.27	1.27	3.4	3.4	3.4	3.4
500	-0.58	-0.58	-0.58	-0.58	0.94	0.94	0.94	0.94	3.2	3.2	3.2	3.2
1000	-0.46	-0.46	-0.46	-0.46	0.70	0.70	0.70	0.70	3.3	3.3	3.3	3.3
N	$m_0 = 1$											
100	-1.84	-1.98	-2.00	-2.02	3.16	3.42	3.46	3.49	2.9	2.9	3.9	4.5
300	-1.19	-1.22	-1.24	-1.26	1.97	2.01	2.05	2.08	1.8	2.3	2.3	2.9
500	-0.93	-0.95	-0.97	-0.98	1.54	1.58	1.61	1.63	2.3	2.6	2.3	3.0
1000	-0.70	-0.73	-0.75	-0.76	1.15	1.19	1.23	1.25	2.7	3.3	3.5	3.7
N	$m_0 = 2$											
100	-1.56	-1.96	-2.02	-2.07	2.68	3.38	3.52	3.59	4.2	3.2	3.5	4.2
300	-1.06	-1.16	-1.19	-1.22	1.81	2.01	2.06	2.11	1.8	2.6	3.0	3.5
500	-0.90	-0.94	-0.97	-1.00	1.51	1.56	1.61	1.66	1.3	1.9	2.5	2.7
1000	-0.66	-0.69	-0.71	-0.73	1.08	1.12	1.16	1.20	1.9	2.4	2.8	3.1
β												
N	$m_0 = 0$											
100	-0.58	-0.58	-0.58	-0.58	4.47	4.47	4.47	4.47	5.5	5.5	5.5	5.5
300	-0.30	-0.30	-0.30	-0.30	2.55	2.55	2.55	2.55	5.0	5.0	5.0	5.0
500	-0.21	-0.21	-0.21	-0.21	1.94	1.94	1.94	1.94	4.0	4.0	4.0	4.0
1000	-0.18	-0.18	-0.18	-0.18	1.39	1.39	1.39	1.39	4.4	4.4	4.4	4.4
N	$m_0 = 1$											
100	-0.84	-0.95	-0.97	-0.99	5.44	5.68	5.95	6.15	4.2	4.1	4.5	4.8
300	-0.62	-0.66	-0.69	-0.72	3.04	3.21	3.38	3.50	3.8	4.0	4.2	3.9
500	-0.32	-0.34	-0.36	-0.38	2.36	2.49	2.62	2.71	4.7	4.9	4.5	4.3
1000	-0.26	-0.27	-0.27	-0.27	1.68	1.78	1.87	1.94	3.9	4.1	4.4	4.5
N	$m_0 = 2$											
100	-0.61	-0.69	-0.59	-0.47	5.70	6.84	8.26	10.46	5.8	5.1	5.1	6.3
300	-0.30	-0.32	-0.29	-0.23	3.25	3.77	4.61	5.86	3.7	4.0	4.5	4.6
500	-0.30	-0.29	-0.27	-0.21	2.51	2.91	3.56	4.50	3.1	3.4	3.9	4.3
1000	-0.31	-0.33	-0.34	-0.35	1.81	2.09	2.54	3.20	4.2	4.5	4.6	4.3
$T = 10$												
κ^2	Bias($\times 100$)				RMSE($\times 100$)				Size($\times 100$)			
	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
γ												
N	$m_0 = 0$											
100	-0.43	-0.43	-0.43	-0.43	0.67	0.67	0.67	0.67	3.3	3.3	3.3	3.3
300	-0.26	-0.26	-0.26	-0.26	0.37	0.37	0.37	0.37	2.1	2.1	2.1	2.1
500	-0.22	-0.22	-0.22	-0.22	0.30	0.30	0.30	0.30	2.5	2.5	2.5	2.5
1000	-0.18	-0.18	-0.18	-0.18	0.23	0.23	0.23	0.23	2.9	2.9	2.9	2.9
N	$m_0 = 1$											
100	-0.53	-0.53	-0.53	-0.53	0.84	0.84	0.84	0.84	3.0	3.6	3.6	3.6
300	-0.30	-0.30	-0.31	-0.31	0.45	0.45	0.46	0.46	1.9	2.0	2.3	2.1
500	-0.26	-0.26	-0.26	-0.26	0.37	0.37	0.37	0.37	2.0	2.5	2.8	2.5
1000	-0.20	-0.20	-0.20	-0.20	0.26	0.26	0.26	0.26	1.9	2.2	2.2	2.3
N	$m_0 = 2$											
100	-0.50	-0.49	-0.50	-0.50	0.79	0.79	0.80	0.81	2.7	2.0	2.4	2.8
300	-0.31	-0.31	-0.32	-0.32	0.46	0.47	0.47	0.48	2.0	2.0	2.1	1.9
500	-0.26	-0.26	-0.27	-0.27	0.37	0.38	0.39	0.39	2.3	2.4	2.5	2.8
1000	-0.19	-0.20	-0.20	-0.20	0.25	0.26	0.26	0.27	1.5	1.7	2.0	2.0
β												
N	$m_0 = 0$											
100	-0.13	-0.13	-0.13	-0.13	3.01	3.01	3.01	3.01	6.2	6.2	6.2	6.2
300	-0.09	-0.09	-0.09	-0.09	1.72	1.72	1.72	1.72	5.6	5.6	5.6	5.6
500	-0.05	-0.05	-0.05	-0.05	1.33	1.33	1.33	1.33	5.3	5.3	5.3	5.3
1000	-0.03	-0.03	-0.03	-0.03	0.95	0.95	0.95	0.95	4.8	4.8	4.8	4.8
N	$m_0 = 1$											
100	-0.04	-0.02	-0.02	-0.02	3.70	3.84	3.95	4.01	5.6	5.9	6.0	6.1
300	-0.05	-0.04	-0.04	-0.04	2.13	2.22	2.27	2.31	5.5	5.8	5.3	5.0
500	-0.04	-0.05	-0.05	-0.05	1.59	1.66	1.72	1.75	4.7	4.8	4.5	4.7
1000	-0.01	0.00	0.00	0.00	1.12	1.17	1.20	1.22	4.3	4.1	3.8	4.2
N	$m_0 = 2$											
100	0.00	0.14	0.28	0.42	4.51	5.22	6.25	7.44	4.6	4.2	5.2	5.1
300	0.07	0.11	0.17	0.24	2.52	2.99	3.60	4.27	4.4	4.9	5.0	5.0
500	-0.01	0.03	0.09	0.18	1.98	2.35	2.83	3.35	5.1	5.6	5.8	5.0
1000	0.00	0.05	0.11	0.18	1.37	1.63	1.95	2.31	5.5	4.9	4.7	4.0

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (13). See also the note to Table B1(i).

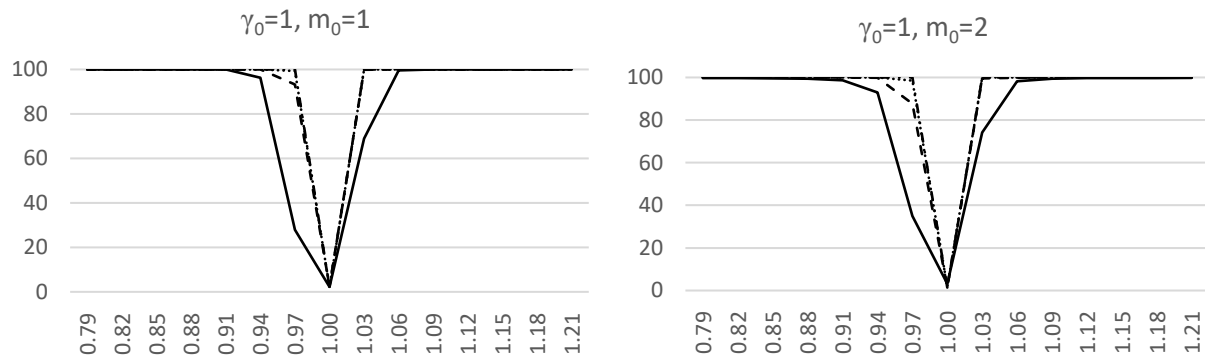
B3: Power Functions

Figure B3(i): Power functions for γ in the case of the AR(1) panel data model with different values of m and N ($\kappa^2=0.25$)

Panel A: $T=5$



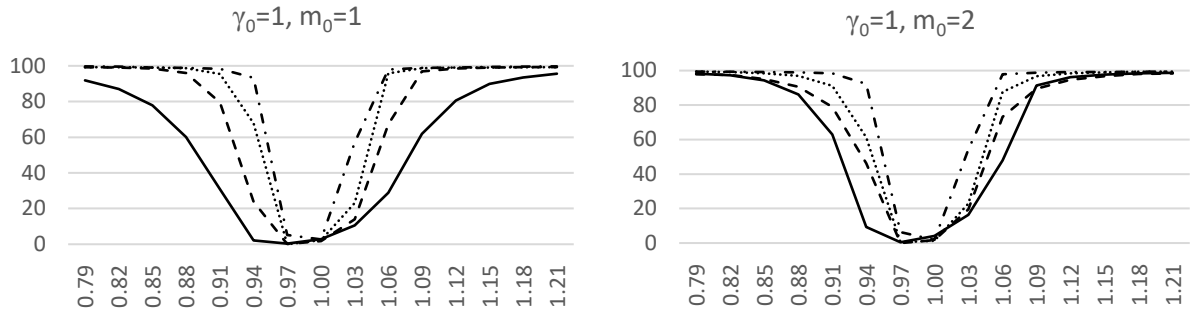
Panel B: $T=10$



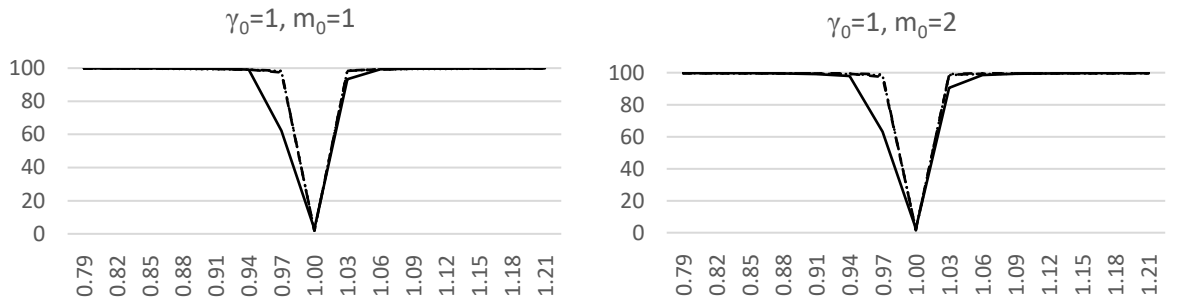
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/N(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table 4.

Figure B3(ii): Power functions for γ in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.25$)

Panel A: $T=5$



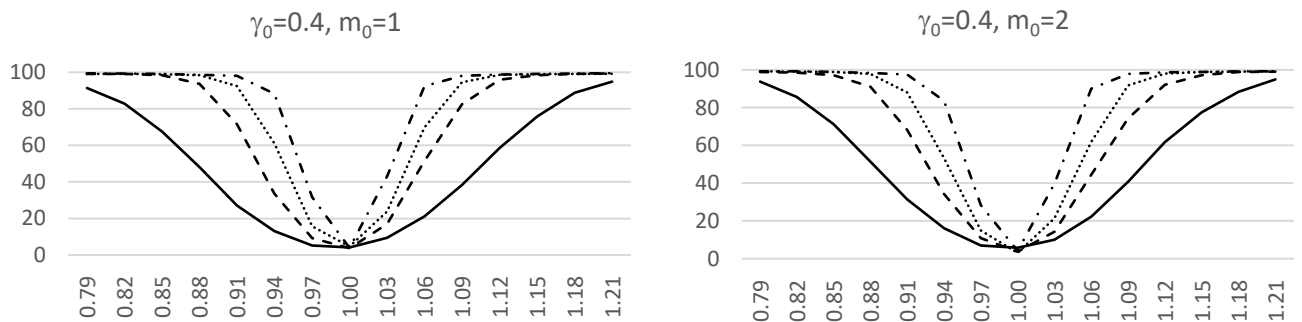
Panel B: $T=10$



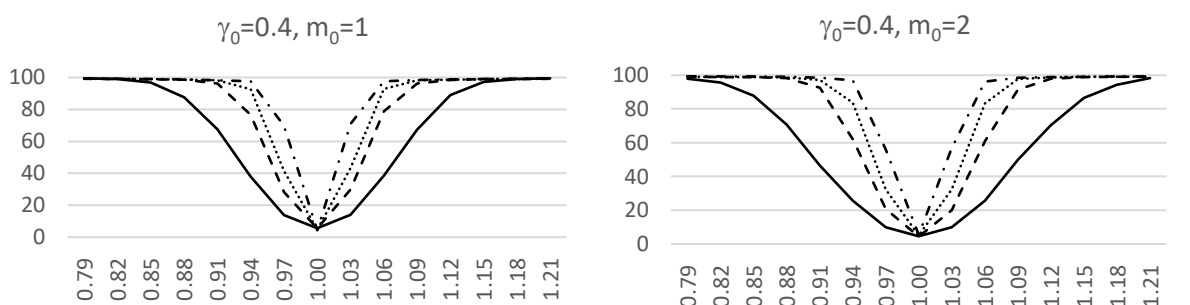
Note: — $N=100$ --- $N=300$ $N=500$ - - - $N=1000$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N=p/(T-2)$ and $p=0.05$; γ is the coefficient of the lagged dependent variable in (13). See also the note to Table 4.

Figure B3(iii): Power functions for β in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.25$)

Panel A: $T=5$



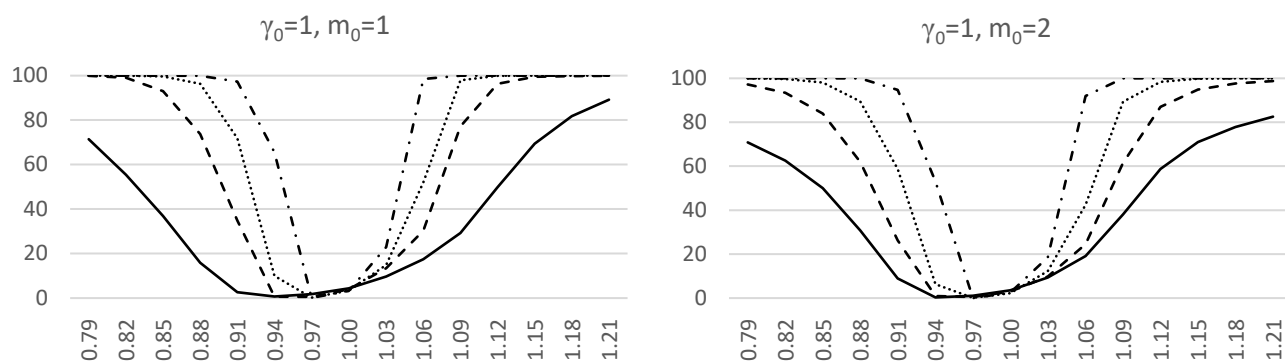
Panel B: $T=10$



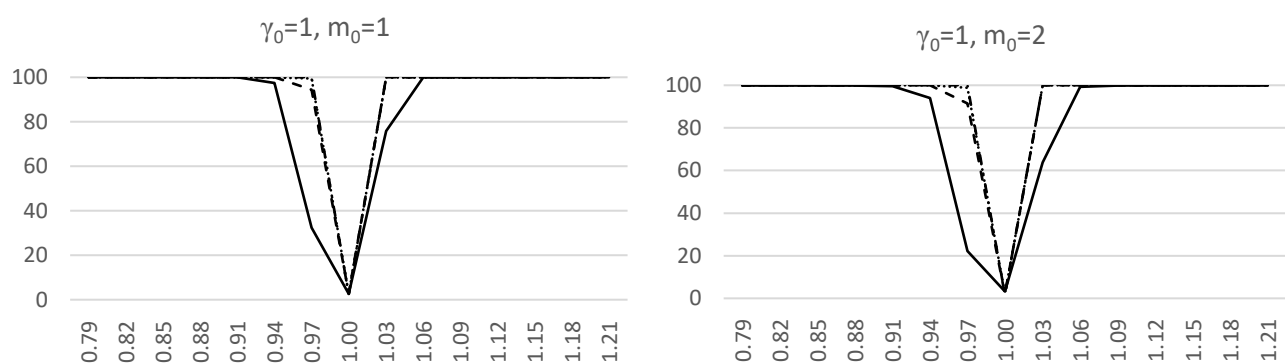
Note: — $N=100$ --- $N=300$ $N=500$ - - - $N=1000$. β is the coefficient of the \mathbf{x}_{it} regressors in (13). See also the note to Figure B3(ii).

Figure B3(iv): Power functions for γ in the case of the AR(1) panel data model with different values of m and N ($\kappa^2=0.5$)

Panel A: $T=5$



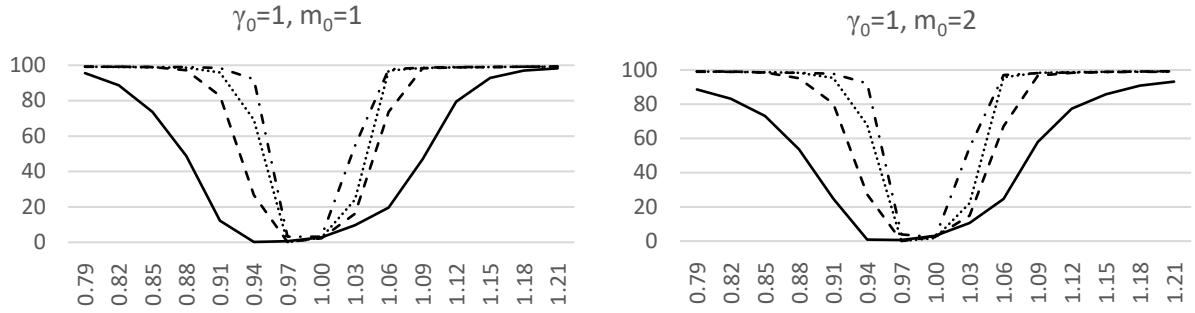
Panel B: $T=10$



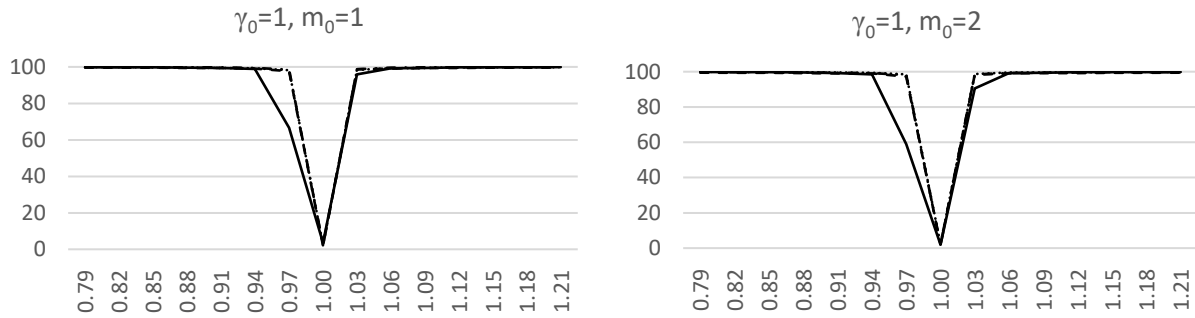
Note: — $N=100$ - - - $N=300$ $N=500$ - . - . $N=1000$. See also the note to Figure B3(i).

Figure B3(v): Power functions for γ in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.5$)

Panel A: $T=5$



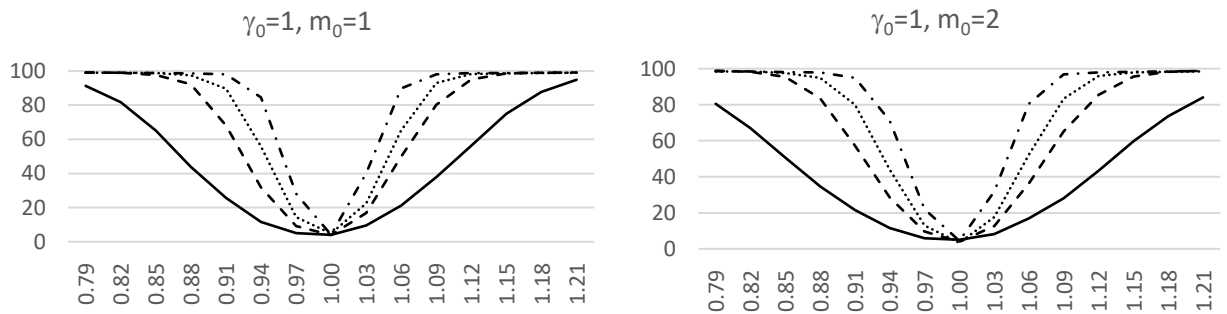
Panel B: $T=10$



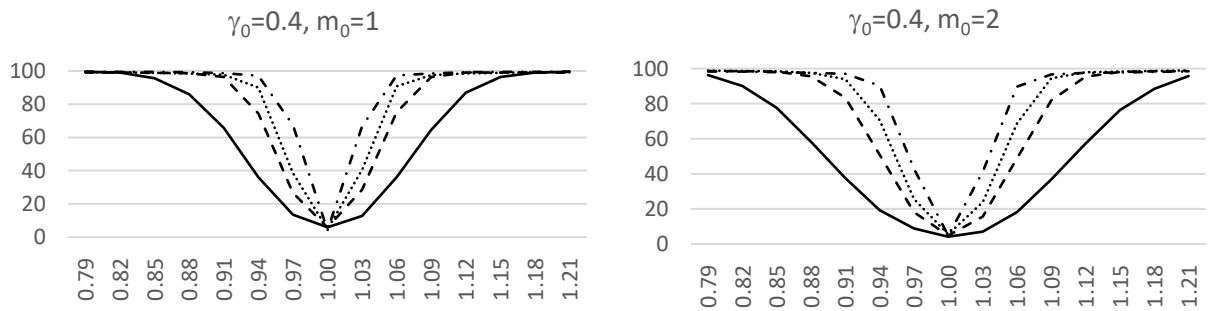
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. See also note to Figure B3(ii).

Figure B3(vi): Power functions for β in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=0.5$)

Panel A: $T=5$



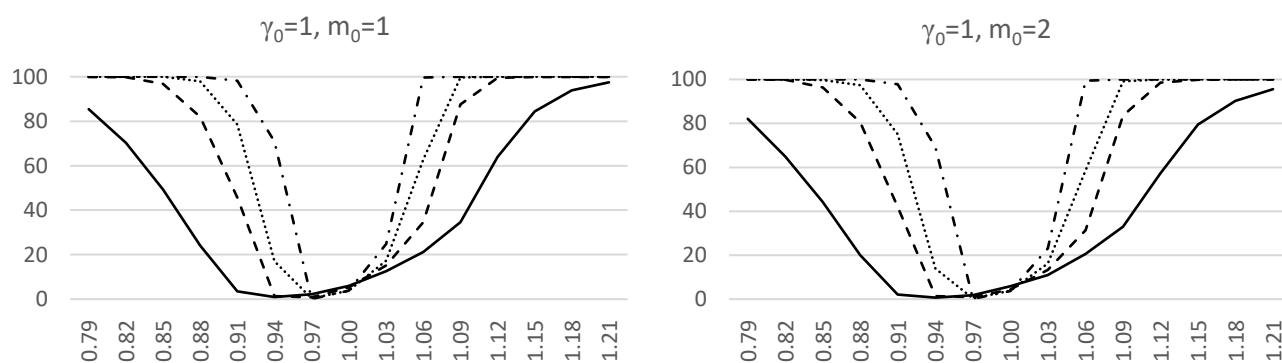
Panel B: $T=10$



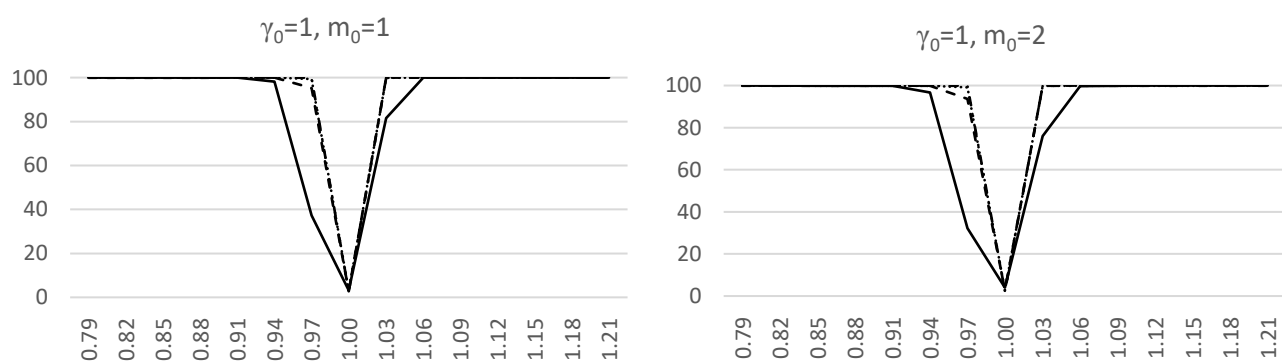
Note: — $N=100$ --- $N=300$ $N=500$ -.-.- $N=1000$. See also the note to Figure B3(v).

Figure B3(vii): Power functions for γ in the case of the AR(1) panel data model with different values of m and N ($\kappa^2=2$)

Panel A: $T=5$



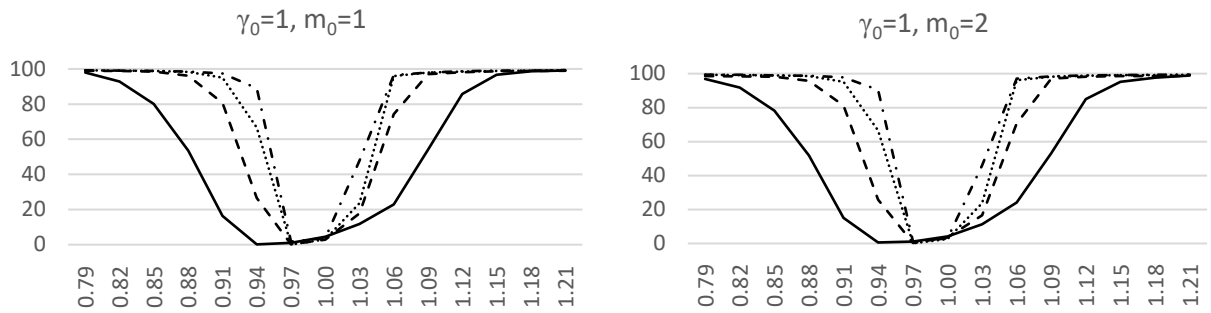
Panel B: $T=10$



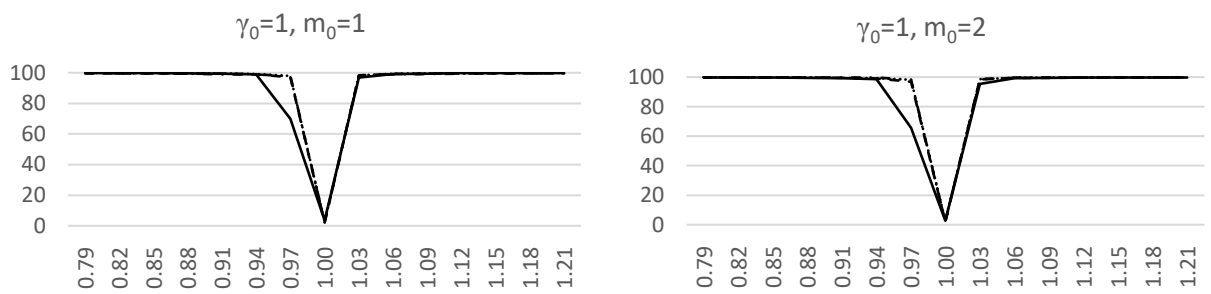
Note: — $N=100$ - - - $N=300$ $N=500$ - . - . $N=1000$. See also the note to Figure B3(i).

Figure B3(viii): Power functions for γ in the case of the ARX(1) panel data model with different values of m and N ($\kappa^2=2$)

Panel A: $T=5$



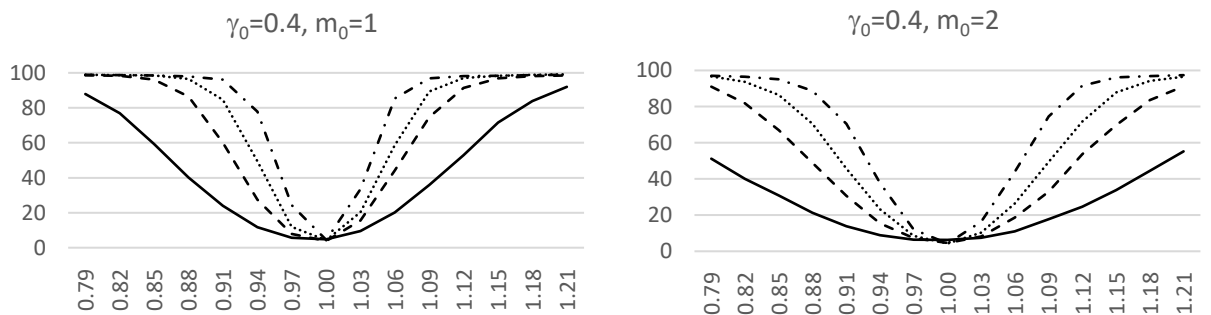
Panel B: $T=10$



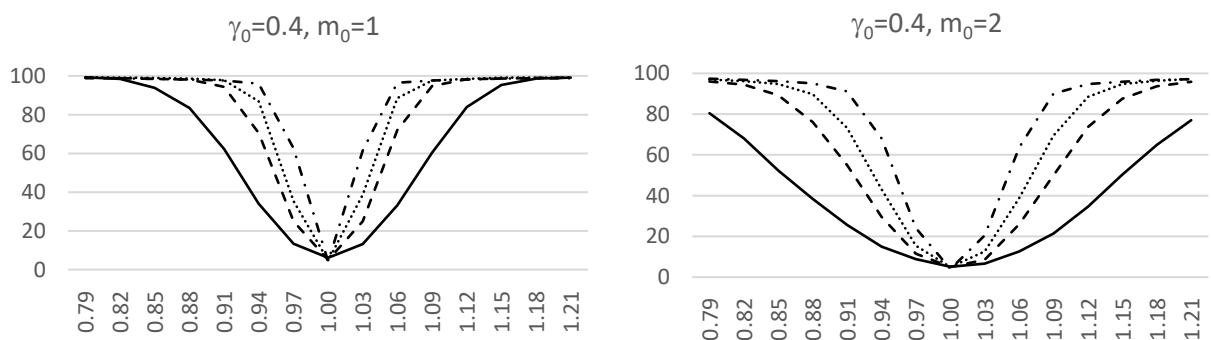
Note: — $N=100$ - - - $N=300$ $N=500$ - . - . $N=1000$. See also the note to Figure B3(ii).

Figure B3(ix): Power functions for β in the ARX(1) panel data model with different values of m and N ($\kappa^2=2$)

Panel A: $T=5$



Panel B: $T=10$



Note: — $N=100$ - - - $N=300$ $N=500$ - . - . $N=1000$. See also the note to Figure B3(viii).

S.9 Monte Carlo experiments for the robustness analysis

C1: Initial values deviating from the steady state distribution

Table C1(i): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure ($\sigma_v^2 = 1, \kappa^2 = 1$)

m_0	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	0	1	2	0	1	2	0	1	2	0	1	2
N	AR(1)											
100	99.4	99.7	87.8	99.2	99.7	96.2	99.7	99.5	99.7	99.6	99.5	99.7
300	99.7	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0
500	99.9	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0	99.8	100.0	100.0
1000	99.9	100.0	100.0	99.8	100.0	100.0	99.9	100.0	100.0	99.5	100.0	100.0
	ARX(1)											
100	99.7	100.0	96.5	99.5	99.9	96.8	99.4	99.6	99.7	99.6	99.6	99.8
300	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	99.8
500	99.9	99.9	100.0	99.9	99.9	100.0	99.9	100.0	100.0	99.9	100.0	100.0
1000	99.9	99.9	100.0	99.8	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: y_{it} is generated as $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}$, $\zeta_{it} = \eta'_i \mathbf{f}_t + u_{it}$ for $i = 1, 2, \dots, N; t = 1, \dots, T$ with $y_{i0} = \kappa_1 \mu_{i0} + \kappa_2 \sigma_{i0} (u_{i0}/\sigma)$ and $\kappa_1, \kappa_2 = 1.2, 0.8$. Under $m_0 = 0$, $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + u_{it}$. In the case of the AR(1) panel data model, $\beta = 0$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$. See also the note to Table 1.

Table C1(ii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m} ($\kappa^2 = 1$)

N	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
$m_0 = 0$												
100	0.56	9.26	6.4	0.74	12.46	22.0	-0.02	3.83	6.1	1.91	7.86	15.4
300	-0.01	4.47	5.5	1.17	9.10	19.2	-0.05	2.22	5.3	0.71	4.73	8.3
500	0.02	3.36	4.7	1.39	7.73	15.4	-0.01	1.72	5.8	0.24	3.03	6.4
1000	0.01	2.41	4.7	1.04	6.07	11.2	-0.01	1.25	5.6	0.20	2.39	6.0
$m_0 = 1$												
100	0.73	11.21	5.7	1.27	13.68	24.2	-0.04	4.52	5.9	0.38	5.37	6.6
300	-0.08	5.71	5.0	1.16	9.98	16.7	0.01	2.55	4.8	0.08	2.73	4.9
500	0.09	4.19	3.7	1.35	8.22	11.5	-0.06	2.06	6.4	0.02	2.15	5.4
1000	0.04	3.07	5.2	0.91	6.22	7.8	-0.03	1.42	4.8	-0.02	1.46	4.9
$m_0 = 2$												
100	4.81	17.79	14.6	1.69	14.06	23.8	-0.13	5.57	5.1	0.34	6.25	7.0
300	0.28	5.72	3.2	1.63	9.90	14.2	0.02	3.07	4.8	0.09	3.16	3.7
500	0.08	4.36	2.9	1.34	8.16	9.6	-0.10	2.35	4.6	-0.08	2.36	4.3
1000	0.03	2.99	3.6	0.75	5.82	5.8	0.00	1.75	4.7	0.03	1.65	4.4

Note: γ is the coefficient of the lagged dependent variable given in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table C1(i).

Table C1(iii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\sigma_v^2 = 1, \kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$				
Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)		Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)		Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)		Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
γ														
N	$m_0 = 0$													
100	-0.12	3.64	5.7	-0.05	3.15	6.8	-0.06	1.99	5.7	-0.03	1.41	6.8		
300	-0.04	2.08	6.1	-0.06	1.79	5.7	0.07	1.17	6.1	0.03	0.80	5.7		
500	0.02	1.55	5.3	0.01	1.34	5.1	-0.01	0.88	5.3	0.00	0.60	5.1		
1000	-0.05	1.14	5.6	-0.04	0.98	5.6	0.00	0.64	5.6	0.00	0.44	5.6		
$m_0 = 1$														
100	0.12	4.60	5.4	0.26	4.98	5.4	-0.10	2.22	6.1	-0.07	1.60	6.4		
300	-0.04	2.56	4.6	0.01	2.68	4.2	0.03	1.25	5.7	0.03	0.87	4.7		
500	0.02	1.97	4.0	0.03	2.03	3.7	-0.02	0.96	5.1	-0.02	0.69	5.5		
1000	-0.06	1.44	5.0	-0.04	1.48	4.7	0.01	0.69	5.5	0.00	0.49	5.4		
$m_0 = 2$														
100	0.41	5.09	6.1	0.52	5.27	4.9	-0.10	2.42	6.0	-0.06	1.66	5.4		
300	0.04	2.64	4.1	0.08	2.78	4.0	-0.06	1.38	5.4	-0.02	0.96	4.9		
500	0.07	2.09	4.6	0.10	2.22	4.9	-0.03	1.02	4.2	-0.01	0.73	4.5		
1000	0.05	1.49	4.0	0.05	1.54	4.5	0.02	0.73	4.4	0.01	0.51	4.4		
β														
$m_0 = 0$														
100	-0.05	4.45	5.8	-0.04	4.57	5.8	-0.02	3.03	5.8	-0.02	3.02	5.8		
300	0.02	2.53	5.7	0.00	2.58	5.6	-0.05	1.73	5.7	-0.03	1.71	5.6		
500	0.04	1.92	5.1	0.04	1.97	4.8	0.00	1.34	5.1	0.00	1.33	4.8		
1000	0.00	1.38	5.1	0.00	1.41	5.1	0.01	0.96	5.1	0.01	0.95	5.1		
$m_0 = 1$														
100	0.01	6.02	5.7	0.08	6.19	5.5	0.09	3.98	6.2	0.08	3.98	6.2		
300	-0.14	3.41	4.9	-0.12	3.48	5.1	0.01	2.29	5.8	0.02	2.28	5.5		
500	0.09	2.67	5.4	0.10	2.73	5.2	0.00	1.74	5.1	0.00	1.72	5.1		
1000	0.04	1.88	5.8	0.05	1.92	5.5	0.03	1.21	4.3	0.04	1.20	4.7		
$m_0 = 2$														
100	0.28	8.34	6.3	0.43	8.59	5.9	0.14	6.26	5.2	0.15	6.24	5.2		
300	0.18	4.62	5.3	0.21	4.68	5.3	0.09	3.63	5.4	0.08	3.61	5.5		
500	0.12	3.56	5.1	0.15	3.64	5.2	0.02	2.84	5.9	0.01	2.84	5.8		
1000	-0.06	2.51	4.7	-0.05	2.55	5.0	0.04	1.96	5.3	0.05	1.95	5.4		

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (13). See also the note to Table C1(i).

C2: Alternative p-values ($p = 0.01, p = 0.10$) for implementing the MTLR test

► Results for $p = 0.01$

Table C2(i): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure ($\sigma_v^2 = 1, \kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
m_0	0	1	2	0	1	2	0	1	2	0	1	2
N												
100	99.7	99.9	80.4	99.7	99.9	93.1	99.9	99.8	99.9	100.0	99.8	99.9
300	99.9	100.0	100.0	99.9	100.0	100.0	99.8	100.0	100.0	99.9	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	99.8	100.0	100.0
ARX(1)												
100	100.0	100.0	93.3	100.0	100.0	94.3	99.8	99.7	99.9	99.8	99.7	99.9
300	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
500	100.0	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.01$. See also the note to Table 1.

Table C2(ii): Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m} ($\kappa^2 = 1$)

	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
N	$m_0 = 0$											
100	0.44	8.64	6.1	0.73	12.11	21.2	-0.02	3.75	6.4	1.96	7.89	16.3
300	-0.03	4.26	5.4	1.41	9.26	19.2	-0.04	2.18	5.1	0.69	4.61	8.7
500	0.03	3.22	4.8	1.48	7.77	14.5	-0.01	1.70	5.9	0.26	3.09	6.7
1000	0.00	2.29	4.5	1.02	6.07	12.1	-0.01	1.22	5.4	0.22	2.37	5.8
	$m_0 = 1$											
100	0.45	9.32	5.1	1.43	13.00	19.6	-0.04	4.19	6.1	0.25	4.61	4.9
300	-0.10	4.98	5.1	0.99	9.04	11.9	0.02	2.38	4.5	0.08	2.41	4.7
500	0.05	3.68	3.9	0.96	7.12	7.1	-0.05	1.91	6.0	0.01	1.88	5.4
1000	0.04	2.67	4.7	0.61	5.08	4.7	-0.01	1.32	4.9	0.00	1.30	4.2
	$m_0 = 2$											
100	6.94	20.36	17.9	1.93	13.52	20.9	-0.09	5.13	5.9	0.19	5.32	5.3
300	0.20	4.99	3.9	1.38	8.97	10.3	0.04	2.81	4.6	0.08	2.66	4.0
500	0.05	3.81	3.1	0.98	7.06	6.3	-0.10	2.16	4.9	-0.09	2.06	4.7
1000	0.02	2.62	3.3	0.45	4.81	4.4	0.00	1.59	4.7	0.01	1.44	4.0

Note: γ is the coefficient of the lagged dependent variable given in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table C2(i).

Table C2(iii): Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of $\hat{\gamma}$ and $\hat{\beta}$ for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\sigma_v^2 = 1, \kappa^2 = 1$)

	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
γ												
N	$m_0 = 0$											
100	-0.14	3.45	5.9	-0.07	2.98	6.6	-0.05	1.94	5.4	-0.03	1.36	5.9
300	-0.04	1.97	5.6	-0.05	1.70	6.0	0.08	1.14	5.3	0.04	0.77	5.0
500	0.02	1.47	5.1	0.00	1.27	4.4	-0.01	0.86	4.5	0.00	0.58	4.3
1000	-0.05	1.08	5.2	-0.03	0.93	5.8	0.00	0.62	4.9	0.00	0.42	5.8
	$m_0 = 1$											
100	0.09	4.28	5.1	0.23	4.74	5.2	-0.10	2.15	6.0	-0.07	1.54	6.5
300	-0.05	2.39	4.4	-0.02	2.56	5.1	0.03	1.20	5.2	0.02	0.82	4.0
500	0.01	1.83	3.8	0.03	1.91	3.9	-0.02	0.92	5.5	-0.01	0.65	5.1
1000	-0.04	1.35	4.5	-0.02	1.41	4.5	0.01	0.67	5.4	0.00	0.46	5.4
	$m_0 = 2$											
100	0.46	4.84	6.3	0.48	4.99	4.6	-0.09	2.33	5.8	-0.05	1.59	5.9
300	0.03	2.46	4.1	0.07	2.63	4.8	-0.06	1.33	5.4	-0.02	0.91	4.8
500	0.07	1.94	3.6	0.10	2.10	4.6	-0.03	0.98	4.3	-0.01	0.69	4.7
1000	0.05	1.39	3.6	0.05	1.47	4.2	0.02	0.70	4.3	0.01	0.48	4.1
β												
	$m_0 = 0$											
100	-0.06	4.44	5.6	-0.06	4.55	5.4	-0.01	3.04	6.5	-0.02	3.02	6.6
300	0.02	2.53	5.7	0.00	2.58	5.8	-0.05	1.73	6.0	-0.03	1.71	6.0
500	0.04	1.92	5.2	0.04	1.97	5.2	0.00	1.34	5.7	0.00	1.33	5.6
1000	0.00	1.38	5.0	0.00	1.40	4.9	0.01	0.96	5.6	0.01	0.95	5.8
	$m_0 = 1$											
100	-0.01	5.98	5.6	0.05	6.16	5.5	0.09	3.98	6.3	0.07	3.98	6.2
300	-0.15	3.39	4.9	-0.14	3.46	4.9	0.01	2.29	6.0	0.02	2.28	5.6
500	0.09	2.65	5.5	0.10	2.70	5.3	0.00	1.74	5.2	0.00	1.72	5.2
1000	0.05	1.87	5.5	0.06	1.91	5.7	0.03	1.21	4.4	0.04	1.20	4.7
	$m_0 = 2$											
100	0.27	8.35	6.4	0.41	8.57	5.9	0.15	6.27	4.9	0.13	6.24	5.0
300	0.18	4.62	5.2	0.20	4.67	5.3	0.09	3.63	5.3	0.08	3.61	5.4
500	0.11	3.55	5.0	0.14	3.63	5.0	0.02	2.85	5.7	0.01	2.84	5.9
1000	-0.06	2.51	4.9	-0.05	2.55	5.2	0.04	1.96	5.3	0.05	1.95	5.3

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (13). See also the note to Table C2(i).

► Results for $p = 0.10$

Table C2(iv): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure ($\sigma_v^2 = 1, \kappa^2 = 1$)

m_0	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	0	1	2	0	1	2	0	1	2	0	1	2
N	AR(1)											
100	99.4	99.5	91.7	99.0	99.5	97.5	99.3	99.4	99.4	99.3	99.5	99.4
300	99.7	99.9	100.0	99.7	100.0	100.0	99.7	99.9	99.9	99.7	100.0	99.9
500	99.9	100.0	100.0	99.6	100.0	100.0	99.8	99.9	100.0	99.9	99.9	100.0
1000	99.9	100.0	100.0	99.8	100.0	100.0	99.6	100.0	100.0	99.5	100.0	100.0
N	ARX(1)											
100	99.5	99.8	97.6	99.4	99.7	98.0	99.2	99.4	99.6	99.1	99.4	99.6
300	99.8	100.0	100.0	99.7	100.0	100.0	100.0	99.9	99.7	100.0	99.9	99.8
500	99.8	99.9	100.0	99.9	99.9	100.0	99.9	100.0	100.0	99.9	100.0	100.0
1000	99.9	99.9	100.0	99.8	99.9	100.0	99.9	100.0	100.0	99.9	100.0	100.0

Note: \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.10$. See also the note to Table 1.

Table C2(v): Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, \hat{m} ($\kappa^2 = 1$)

N	$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
$m_0 = 0$												
100	0.40	8.71	6.2	0.59	12.38	21.3	-0.03	3.77	6.4	1.94	7.91	16.4
300	-0.02	4.26	5.4	1.39	9.32	19.2	-0.04	2.18	5.1	0.67	4.60	8.7
500	0.03	3.22	4.8	1.42	7.85	14.6	-0.01	1.70	5.9	0.26	3.09	6.7
1000	0.00	2.29	4.5	1.00	6.08	12.1	-0.01	1.22	5.4	0.18	2.24	5.7
$m_0 = 1$												
100	0.41	9.41	5.1	1.34	13.23	19.7	-0.05	4.21	9.6	0.23	4.64	19.3
300	-0.08	5.02	5.1	1.00	9.04	11.9	0.02	2.38	3.9	0.08	2.41	10.3
500	0.05	3.68	3.9	0.94	7.16	7.1	-0.05	1.91	3.1	0.01	1.88	6.3
1000	0.04	2.67	4.7	0.61	5.08	4.7	-0.01	1.32	3.3	0.00	1.30	4.4
$m_0 = 2$												
100	3.15	14.91	6.1	1.76	13.30	4.9	-0.08	5.13	5.9	0.18	5.33	5.3
300	0.20	4.99	4.5	1.38	8.97	4.7	0.04	2.81	4.6	0.08	2.66	4.0
500	0.05	3.81	6.0	0.98	7.06	5.4	-0.10	2.16	4.9	-0.09	2.06	4.7
1000	0.02	2.62	4.9	0.45	4.81	4.2	0.00	1.59	4.7	0.01	1.44	4.0

Note: γ is the coefficient of the lagged dependent variable given in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table C2(iv).

Table C2(vi): Bias($\times 100$), RMSE($\times 100$) and Size ($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, \hat{m} ($\sigma_v^2 = 1$, $\kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
Bias ($\times 100$)		RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
γ												
$m_0 = 0$												
100	-0.14	3.45	5.9	-0.07	3.03	6.6	-0.06	1.95	5.4	-0.03	1.37	5.8
300	-0.04	1.97	5.6	-0.05	1.74	6.1	0.08	1.14	5.3	0.04	0.77	5.1
500	0.01	1.47	5.1	0.00	1.27	4.4	-0.01	0.86	4.5	0.00	0.58	4.3
1000	-0.05	1.08	5.1	-0.03	0.93	5.8	0.00	0.62	4.9	0.00	0.42	5.8
$m_0 = 1$												
100	0.10	4.30	5.1	0.23	4.76	5.3	-0.10	2.15	6.0	-0.07	1.54	6.5
300	-0.05	2.39	4.4	-0.02	2.56	5.1	0.03	1.20	5.2	0.02	0.83	4.0
500	0.01	1.83	3.8	0.02	1.92	3.9	-0.02	0.92	5.5	-0.01	0.65	5.1
1000	-0.04	1.35	4.5	-0.02	1.41	4.5	0.01	0.67	5.4	0.00	0.46	5.4
$m_0 = 2$												
100	0.34	4.68	5.7	0.45	4.97	4.7	-0.08	2.33	5.8	-0.05	1.59	5.9
300	0.03	2.46	4.1	0.07	2.63	4.8	-0.06	1.33	5.4	-0.02	0.91	4.8
500	0.07	1.94	3.6	0.10	2.10	4.6	-0.03	0.98	4.3	-0.01	0.69	4.7
1000	0.05	1.39	3.6	0.05	1.47	4.2	0.02	0.70	4.3	0.01	0.48	4.1
β												
$m_0 = 0$												
100	-0.05	4.44	5.6	-0.06	4.55	5.4	-0.01	3.04	6.5	-0.02	3.02	6.6
300	0.02	2.53	5.7	0.00	2.58	5.9	-0.05	1.73	6.0	-0.03	1.71	6.0
500	0.04	1.92	5.2	0.04	1.97	5.2	0.00	1.34	5.7	0.00	1.33	5.6
1000	0.00	1.38	5.0	0.00	1.40	4.9	0.01	0.96	5.6	0.01	0.95	5.8
$m_0 = 1$												
100	-0.01	5.99	5.6	0.06	6.16	5.5	0.09	3.98	6.3	0.07	3.98	6.2
300	-0.15	3.39	4.9	-0.14	3.46	4.9	0.01	2.29	6.0	0.02	2.28	5.6
500	0.09	2.65	5.5	0.09	2.70	5.3	0.00	1.74	5.2	0.00	1.72	5.2
1000	0.05	1.88	5.5	0.06	1.91	5.7	0.03	1.21	4.4	0.04	1.20	4.7
$m_0 = 2$												
100	0.27	8.33	6.4	0.41	8.55	5.8	0.15	6.27	4.9	0.13	6.24	5.0
300	0.18	4.62	5.2	0.20	4.67	5.3	0.09	3.63	5.3	0.08	3.61	5.4
500	0.11	3.55	5.0	0.14	3.63	5.0	0.02	2.85	5.7	0.01	2.84	5.9
1000	-0.06	2.51	4.9	-0.05	2.55	5.2	0.04	1.96	5.3	0.05	1.95	5.3

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (13). See also the note to Table C2(iv).

C3: Correlation of factor loadings and regressors

In this experiment we allow the factor loadings η_i in the Monte Carlo design outlined in Section 8.1 to be correlated with the regressors x_{it} according to

$$\eta_{i\ell} = \kappa \sqrt{\frac{2}{m_0}} \left[\left(\sqrt{T} \bar{v}_i / \sigma_v \right) + v_{i\ell} \right], \text{ for } \ell = 1, 2, \dots, m_0 \quad (\text{S.55})$$

where $\bar{v}_i = T^{-1} \sum_{t=1}^T v_{it}$, with v_{it} representing the idiosyncratic component of x_{it} , defined by (78), and $v_{i\ell} \sim IIDN(0, 1)$, for $\ell = 1, 2, \dots, m_0$. The above formulation ensures that $Var(\eta_{i\ell}) = \frac{\kappa^2}{m_0}$, as in the baseline case where the loadings are uncorrelated with the regressors. The rest of the parameters are as described in Section 8.1.

Table C3(i): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure ($\sigma_v^2 = 1, \kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$		$T = 10, \gamma_0 = 0.4$		$T = 10, \gamma_0 = 0.8$	
m_0	1	2	1	2	1	2	1	2
N	AR(1)							
100	99.7	100.0	99.6	100.0	99.6	99.8	99.5	99.7
300	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ARX(1)							
100	99.9	100.0	99.9	100.0	99.6	99.7	99.6	99.7
300	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.8
500	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0
1000	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0

Note: y_{it} is generated as $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}$, $\zeta_{it} = \boldsymbol{\eta}'_i \mathbf{f}_t + u_{it}$ for $i = 1, 2, \dots, N; t = 1, \dots, T$ with $y_{i0} = \mu_{i0} + \sigma_{i0} (u_{i0}/\sigma)$. The factor loadings are generated as $\eta_{i\ell} = \kappa \sqrt{\frac{2}{m_0}} \left[\left(\sqrt{T} \bar{v}_i / \sigma_v \right) + v_{i\ell} \right]$, for $\ell = 1, 2, \dots, m_0$ where $\bar{v}_i = T^{-1} \sum_{t=1}^T v_{it}$, and $v_{i\ell} \sim IIDN(0, 1)$, for $\ell = 1, 2, \dots, m_0$. In the case of the AR(1) panel data model, $\beta = 0$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$. See also the note to Table 1.

Table C3(ii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ for the AR(1) panel data model, using the estimated number of factors, m , and the true number, m_0 ($\kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
N	$m_0 = 1$											
100	0.20	7.36	6.4	1.21	10.81	10.5	-0.05	3.68	5.6	0.08	3.41	5.5
300	-0.13	3.96	5.6	0.30	6.30	6.4	0.01	2.09	4.9	0.04	1.83	5.0
500	0.02	2.91	4.7	0.34	4.67	4.1	-0.06	1.67	5.8	-0.01	1.43	5.3
1000	0.04	2.11	5.5	0.24	3.27	4.8	-0.01	1.16	5.3	0.00	1.00	4.6
	$m_0 = 2$											
100	0.23	7.37	5.1	1.33	10.68	9.3	-0.05	4.40	6.5	0.02	3.79	6.3
300	0.12	3.91	4.3	0.67	6.22	5.0	0.05	2.43	5.3	0.06	2.00	4.5
500	0.03	3.01	4.2	0.38	4.65	3.6	-0.09	1.87	4.8	-0.08	1.56	5.4
1000	0.01	2.06	4.2	0.17	3.15	3.7	-0.01	1.36	5.0	0.00	1.07	4.0

Note: γ is the coefficient of the lagged dependent variable given in (1) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table C3(i).

Table C3(iii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, m , and the true number, m_0 ($\sigma_v^2 = 1$, $\kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
Bias	RMSE	Size		Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
($\times 100$)	($\times 100$)	($\times 100$)		($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)
γ												
N	$m_0 = 1$											
100	0.02	4.01	5.9	0.10	4.49	5.8	-0.08	2.09	6.4	-0.05	1.51	6.3
300	-0.12	2.28	5.6	-0.14	2.48	6.4	0.04	1.16	5.4	0.04	0.80	4.2
500	-0.06	1.74	4.2	-0.11	1.85	4.2	-0.01	0.90	5.4	0.01	0.63	5.5
1000	-0.11	1.28	4.8	-0.15	1.36	5.2	0.02	0.65	5.5	0.03	0.45	5.5
$m_0 = 2$												
100	0.07	4.27	5.8	0.19	4.73	5.7	-0.07	2.28	6.6	-0.02	1.57	6.2
300	-0.04	2.33	4.4	-0.06	2.55	5.0	-0.06	1.30	5.9	0.00	0.90	4.9
500	-0.04	1.84	4.4	-0.08	2.04	5.8	-0.03	0.96	5.1	0.01	0.68	5.5
1000	-0.06	1.32	4.6	-0.12	1.44	5.2	0.02	0.69	4.7	0.03	0.48	4.2
β												
N	$m_0 = 1$											
100	0.01	6.20	5.3	0.04	6.34	5.5	0.07	4.05	6.2	0.06	4.06	6.3
300	-0.15	3.53	5.1	-0.18	3.59	5.2	-0.01	2.33	5.6	0.00	2.32	5.5
500	0.07	2.77	5.6	0.04	2.81	5.3	-0.02	1.78	5.5	-0.02	1.76	5.4
1000	0.09	1.97	5.9	0.06	1.99	5.4	0.01	1.23	4.3	0.02	1.22	4.6
$m_0 = 2$												
100	0.49	11.19	6.8	0.56	11.35	6.5	-0.18	7.56	5.5	-0.17	7.53	5.6
300	0.38	6.24	5.5	0.37	6.27	5.0	-0.27	4.37	5.4	-0.28	4.35	5.3
500	0.28	4.74	5.0	0.26	4.80	5.5	-0.30	3.43	6.0	-0.31	3.43	5.8
1000	0.02	3.35	5.1	-0.02	3.38	5.3	-0.26	2.36	4.7	-0.26	2.34	4.7

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (13). See also the note to Table C3(i).

C4: Weakly cross-correlated factor loadings

Here we generate the factor loadings, $\eta_{i\ell}$, in the Monte Carlo design outlined in Section 8.1 to follow a first-order spatial autoregressive process defined by

$$\boldsymbol{\eta}_\ell = a\mathbf{W}\boldsymbol{\eta}_\ell + \mathbf{e}_\ell, \quad \ell = 1, 2, \dots, m_0, \quad (\text{S.56})$$

where $\boldsymbol{\eta}_\ell = (\eta_{1\ell}, \eta_{2\ell}, \dots, \eta_{N\ell})'$,

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 1/2 & 0 & 1/2 & 0 & & 0 \\ 0 & 1/2 & 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \ddots & 1/2 & 0 \\ \vdots & & & & 1/2 & 0 & 1/2 \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}, \quad (\text{S.57})$$

and $\mathbf{e}_\ell = (e_{1\ell}, e_{2\ell}, \dots, e_{N\ell})'$. For each i and ℓ , $e_{i\ell}$ are drawn as $IIDN(0, \sigma_{e\ell}^2)$. To ensure $N^{-1} \sum_{i=1}^N \text{Var}(\eta_{i\ell}) = \frac{\kappa^2}{m_0}$, for $\ell = 1, 2, \dots, m_0$ (which corresponds to the case of cross-sectionally independent factor loadings) we set

$$\sigma_{e\ell}^2 = \left(\frac{\kappa^2}{m_0} \right) \left\{ \frac{N}{\text{tr}[(\mathbf{I}_N - a\mathbf{W})^{-1}(\mathbf{I}_N - a\mathbf{W}')^{-1}]} \right\}. \quad (\text{S.58})$$

The rest of the parameters are as described in Section 8.1.

Table C4(i): Empirical frequency of correctly selecting the true number of factors, m_0 , using the sequential MTLR procedure ($\sigma_v^2 = 1, \kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$			$T = 5, \gamma_0 = 0.8$		$T = 10, \gamma_0 = 0.4$		$T = 10, \gamma_0 = 0.8$	
m_0	1	2	1	2	1	2	1	2
N	AR(1)							
100	99.6	86.3	99.8	95.6	99.6	99.7	99.5	99.8
300	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ARX(1)							
100	99.9	95.6	99.9	96.6	99.6	99.8	99.5	99.8
300	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9
500	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0
1000	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0

Note: y_{it} is generated as $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}$, $\zeta_{it} = \eta_i' \mathbf{f}_t + u_{it}$, for $i = 1, 2, \dots, N; t = 1, \dots, T$ with $y_{i0} = \mu_{i0} + \sigma_{i0}(u_{i0}/\sigma)$. The factor loadings $\eta_\ell = (\eta_{1\ell}, \eta_{2\ell}, \dots, \eta_{N\ell})'$ are generated as $\eta_\ell = a\mathbf{W}\eta_\ell + \mathbf{e}_\ell$, for $\ell = 1, 2, \dots, m_0$, where $\mathbf{e}_\ell = (e_{1\ell}, e_{2\ell}, \dots, e_{N\ell})'$, with $a = 0.4$ and \mathbf{W} is specified as in equation (S.57). For each i and ℓ , $e_{i\ell}$ are drawn as $IIDN(0, \sigma_{e\ell}^2)$ with $\sigma_{e\ell}^2 = \left(\frac{\kappa^2}{m_0}\right) \left\{ N / \text{tr} \left[(\mathbf{I}_N - a\mathbf{W})^{-1} (\mathbf{I}_N - a\mathbf{W}')^{-1} \right] \right\}$. In the case of the AR(1) panel data model, $\beta = 0$. \hat{m} is estimated using the sequential MTLR procedure described in Section 7.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$. See also the note to Table 1.

Table C4(ii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ for the AR(1) model, using the estimated number of factors, m , and the true number, m_0 ($\kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$				$T = 10, \gamma_0 = 0.4$				$T = 10, \gamma_0 = 0.8$			
Bias	RMSE	Size		Bias	RMSE	Size		Bias	RMSE	Size		Bias	RMSE	Size	
($\times 100$)	($\times 100$)	($\times 100$)		($\times 100$)	($\times 100$)	($\times 100$)		($\times 100$)	($\times 100$)	($\times 100$)		($\times 100$)	($\times 100$)	($\times 100$)	
N	$m_0 = 1$														
100	0.43	9.46	5.1	1.35	12.86	18.9		-0.06	4.22	5.8		0.23	4.70	5.1	
300	-0.08	4.99	5.4	1.03	9.07	11.6		0.03	2.39	4.5		0.09	2.43	4.9	
500	0.05	3.68	3.8	0.97	7.16	6.8		-0.06	1.90	5.5		0.01	1.88	5.5	
1000	0.03	2.67	4.8	0.61	5.09	4.7		-0.02	1.32	5.3		0.00	1.30	4.5	
N	$m_0 = 2$														
100	5.11	17.99	13.7	1.99	13.35	19.6		-0.09	5.10	6.0		0.20	5.24	5.1	
300	0.30	5.00	3.4	1.73	9.31	10.7		0.01	2.84	5.2		0.04	2.68	4.1	
500	-0.01	3.85	3.8	0.89	7.17	7.0		-0.07	2.15	4.3		-0.06	2.05	4.3	
1000	0.02	2.62	3.7	0.44	4.76	4.6		0.00	1.59	4.8		0.02	1.44	4.5	

Note: γ is the coefficient of the lagged dependent variable given in (13) in the absence of the \mathbf{x}_{it} regressors. See also the note to Table C4(i).

Table C4(iii): Bias($\times 100$), RMSE($\times 100$) and Size($\times 100$) of γ and β for the ARX(1) panel data model, using the estimated number of factors, m , and the true number, m_0 ($\sigma_v^2 = 1$, $\kappa^2 = 1$)

$T = 5, \gamma_0 = 0.4$				$T = 5, \gamma_0 = 0.8$			$T = 10, \gamma_0 = 0.4$			$T = 10, \gamma_0 = 0.8$		
Bias	RMSE	Size		Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
($\times 100$)	($\times 100$)	($\times 100$)		($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)	($\times 100$)
γ												
N												
$m_0 = 1$												
100	0.09	4.30	5.0	0.22	4.73	5.6	-0.10	2.15	6.4	-0.07	1.54	6.5
300	-0.05	2.39	4.4	-0.01	2.56	5.1	0.03	1.20	5.3	0.02	0.82	3.9
500	0.01	1.84	3.5	0.02	1.93	3.8	-0.02	0.92	5.5	-0.01	0.65	5.1
1000	-0.04	1.35	4.5	-0.02	1.40	4.4	0.01	0.67	5.3	0.00	0.46	5.4
$m_0 = 2$												
100	0.35	4.77	5.7	0.43	4.98	4.4	-0.08	2.31	5.5	-0.05	1.58	5.2
300	0.01	2.41	3.4	0.05	2.59	4.2	-0.08	1.33	5.3	-0.04	0.91	4.6
500	0.06	1.94	3.9	0.09	2.11	4.3	-0.03	0.97	4.6	-0.01	0.69	4.2
1000	0.06	1.36	3.2	0.06	1.45	3.8	0.02	0.70	4.7	0.01	0.48	4.1
β												
N												
$m_0 = 1$												
100	0.00	6.01	5.5	0.06	6.18	5.3	0.09	3.97	6.3	0.08	3.98	6.0
300	-0.15	3.37	4.9	-0.14	3.44	5.2	0.01	2.29	5.4	0.02	2.28	5.7
500	0.09	2.66	5.7	0.09	2.71	5.4	0.00	1.74	5.0	0.00	1.72	4.8
1000	0.06	1.88	5.7	0.06	1.92	5.5	0.03	1.21	4.5	0.04	1.20	4.5
$m_0 = 2$												
100	0.08	8.17	5.8	0.21	8.37	6.1	0.01	6.35	5.8	0.01	6.33	5.9
300	0.13	4.65	5.6	0.15	4.74	5.9	0.14	3.66	5.4	0.13	3.64	5.8
500	0.04	3.47	4.8	0.06	3.55	4.7	0.03	2.80	5.7	0.03	2.78	5.6
1000	-0.01	2.48	4.8	0.00	2.52	4.7	-0.04	1.99	5.2	-0.03	1.98	5.2

Note: γ and β are the coefficients of the lagged dependent variable and the \mathbf{x}_{it} regressor given in (13). See also the note to Table C4(i).

S.10 The case of heteroskedastic errors

The log-likelihood function in (34) can be modified to allow for time series heteroskedasticity. This involves replacing $\sigma^2 \mathbf{\Omega}$ by

$$E(\mathbf{r}_i \mathbf{r}_i') = \begin{pmatrix} \omega \sigma_1^2 & -\sigma_1^2 & 0 & \cdots & 0 & 0 & 0 \\ -\sigma_1^2 & \sigma_1^2 + \sigma_2^2 & -\sigma_2^2 & \ddots & \vdots & 0 & 0 \\ 0 & -\sigma_2^2 & \sigma_2^2 + \sigma_3^2 & \ddots & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{T-2}^2 & \sigma_{T-2}^2 + \sigma_{T-1}^2 & -\sigma_{T-1}^2 \\ 0 & 0 & 0 & \cdots & 0 & -\sigma_{T-1}^2 & \sigma_{T-1}^2 + \sigma_T^2 \end{pmatrix},$$

with the resultant log-likelihood maximised with respect to $\omega, \sigma_1^2, \sigma_2^2, \dots, \sigma_T^2$ and the remaining parameters. This extension does not pose additional difficulties, however it does impact the order conditions for identification. There are an additional $T - 1$ new error variances to estimate and the order condition in the case of an AR(1) model, for example, becomes $T(T + 1)/2 - (T + 2) \geq Tm - m(m - 1)/2$, and a larger T is required for identification when $m > 0$. For instance for $m = 1$ we need $T \geq 4$, and for $m = 2$ we need $T \geq 6$.

References

- Ahn, S.C. and Lee, H.Y. and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174, 1-14.
- Bai, J. (2013). Likelihood approach to dynamic panel models with interactive effects. Mimeo.
- Hayakawa, K. (2012). GMM estimation of a short dynamic panel data model with interactive fixed effects. *Journal of the Japan Statistical Society* 42, 109-123.
- Nauges, C. and Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique* 70, 53-75.
- Pesaran, M.H. (2015). *Time series and panel data econometrics*. Oxford University Press.
- Pesaran, M.H. and Smith, R.J. (1994). A generalized R^2 criterion for regression models estimated by the instrumental variables method. *Econometrica* 62, 705.