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

February 2020

Abstract

This paper proposes a quasi maximum likelihood (QML) 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. Order conditions for identification of the number of interactive effects are established, and conditions are derived under which the parameters are almost surely locally identified. It is shown that global identification is possible only when the model does not contain lagged dependent variables. The QML 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 QML estimator has small bias and RMSE, and correct empirical size in most settings. The practical use of the QML approach is illustrated by means of two empirical applications 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, 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 22730178, 25780153, 17K03660) 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 generalized method of moments (GMM) and likelihood approaches have been advanced to estimate such panel data models. See, for example, Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover (1995), 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. 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 behavior of the fixed effects and GMM estimators, respectively.[1] 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 recent survey of short T panel data models with interactive effects can be found in Sarafidis and Wansbeek (2012).

This paper, building on the work of Hsiao et al. (2002), proposes a transformed QML approach applied to the short T dynamic panel data model after first-differencing. In addition to the standard individual and time fixed effects, we also allow for interactive 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 order conditions for identification of the number of interactive effects, and derive conditions under which the parameters are almost surely locally identified. It emerges that global identification is possible only when the model does not contain lagged dependent variables. These results can be useful for the development of QML theory in the case of more general models. The QML estimators are shown to be consistent and asymptotically normally distributed 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

[1] For 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).

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 QML estimator compares favourably to 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 QML approach is illustrated with two empirical applications from the literature, focussing on the importance of allowing for interactive effects in empirical analysis. The first application 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 application estimates growth regressions using the recent data analysed by Acemoglu et al. (2019). In the case of both applications 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 allow initial values, y_{i0}, $i = 1, 2, ..., N$, to depend on the fixed effects as well as on other parameters, which is in contrast to Bai's assumption that requires the initial values to be drawn independently of the fixed effects and other unknown parameters. Third, we provide a formal treatment of identification of short T dynamic panel data models with a multi-factor error structure, and propose a sequential multiple testing likelihood procedure to consistently estimate the number of factors, topics that are not addressed by Bai (2013).

The rest of this paper is organised as follows. Section 2 sets out the dynamic panel data model and its assumptions. Section 3 develops the quasi-likelihood approach and derives a solution using an eigenvalue approach. Identification of the number of factors and the parameters of the model are discussed in Section 4. Section 5 establishes the consistency of the QML estimator and derives its asymptotic distribution. Section 6 presents the sequential MTLR procedure for estimating the number of factors. Section 7 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 QML estimator. Empirical applications are provided in Section 8. 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 $\mathbf{w} = (w_1, w_2, ..., w_n)'$ 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 $\|\mathbf{w}\| = \left(\Sigma_{i=1}^{n} w_i^2\right)^{1/2}$ 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 C_0 such that $|y_n|/x_n \leq C_0$ for all n. $y_n = o(x_n)$ if $y_n/x_n \to 0$ as $n \to \infty$. If $\{y_n\}_{n=1}^{\infty}$ and $\{x_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $y_n = \ominus(x_n)$ 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. Small positive constants will be denoted by ϵ. $E_0(.)$ denotes expectations taken under the true probability measure. $\to_p$ and $\overset{a.s.}{\to}$ denote convergence in probability and almost sure (a.s.) convergence, respectively. $\to_d$ denotes convergence in distribution for fixed T and as $N \to \infty$.

2 A dynamic panel data model with interactive error components

We begin with the following standard dynamic panel data model with time and fixed effects

$$y_{it} = \gamma y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \delta_t + \zeta_{it}, \text{ for } t = 0, 1, 2, ..., T, \text{and } i = 1, 2, ..., N, \tag{1}$$

where $\mathbf{x}_{it}$ is a $k \times 1$ vector of regressors that vary both across i and t, $|\gamma| < K$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown coefficients, with $\|\boldsymbol{\beta}\| < K$, and K denotes a finite positive constant. α_i and δ_t denote unit-specific fixed effects and time effects, respectively. We consider T to be fixed, and allow $N \to \infty$, under which the unit root case where $|\gamma| = 1$ is also covered. It is assumed that the observations $\{y_{it}, \mathbf{x}_{it}$, for $t = 0, 1, ..., T; i = 1, 2, ..., N\}$ are available for estimation of γ and $\boldsymbol{\beta}$, which are the parameters of interest.

Specification (1) is the standard short T dynamic panel data model used extensively in the empirical literature assuming that the errors, ζ_{it}, are independently distributed across i and t. In this paper we contribute to this literature by allowing the errors to have the following multi-factor structure

$$\zeta_{it} = \sum_{j=1}^{m} \eta_{ij} f_{jt} + u_{it} = \boldsymbol{\eta}_i' \mathbf{f}_t + u_{it}, \tag{2}$$

where $\boldsymbol{\eta}_i'\mathbf{f}_t$ is an interactive effect with $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$, an $m \times 1$ vector of unobserved common factors, and $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, ..., \eta_{im})'$, an $m \times 1$ vector of associated factor loadings, with u_{it} denoting the remaining idiosyncratic error terms. The above specification includes a number of models considered in the literature and reviewed in Section 1 as special cases. It also provides a direct generalization of Hsiao and Tahmiscioglu (2008) who consider estimation of (1) with IID errors using the transformed MLE procedure. The model considered by Ahn et al. (2013) allows for errors to have the multi-factor structure as in (2) with additive individual and time effects subsumed therein for particular values of η_{ij} and f_{jt}. They treat the factor loadings as random and allow them to be correlated with the regressors.

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 $\boldsymbol{\beta}$. Initially, we develop our proposed estimation method assuming that m is known, and consider the problem of consistent estimation of m in Section 6.1.

We make the following assumptions:

Assumption 1 *The idiosyncratic errors, u_{it}, for $i = 1, 2, ..., N$ 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 $t = 1, 2, ..., T$, and the $m \times 1$ vector of factors $\mathbf{f}_t$, vary across t, so that $\Delta\delta_t \neq 0$ and $\mathbf{g}_t = \Delta\mathbf{f}_t \neq \mathbf{0}$ at least for some $t = 2, ..., T$, $m < T$, and $\sup_t \|\mathbf{g}_t\| < K$.*

Assumption 3 *The unobserved $m \times 1$ factor loadings, $\boldsymbol{\eta}_i$, for $i = 1, 2, ..., N$ 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 $i = 1, 2, ..., N$ 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 $i = 1, 2, ..., N$ follows the multi-factor model*

$$\Delta\mathbf{x}_{it} = \boldsymbol{\delta}_{x,t} + \mathbf{E}_{i,x}\mathbf{g}_{x,t} + \mathbf{v}_{it}, \text{ for all } t = ... - 2, -1, 0, 1, 2, ..., \tag{3}$$

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, $\mathbf{g}_{x,t} = (g_{x,1t}, g_{x,2t}, ..., g_{x,m_x t})'$ is a $m_x \times 1$ vector of common factors, $\mathbf{E}_{i,x} = \left(\boldsymbol{\eta}_{i1,x}, \boldsymbol{\eta}_{i2,x}, ..., \boldsymbol{\eta}_{i,m_x,x}\right)$ 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 $g_{x,jt}$, $\mathbf{\Psi}_j$ for $j = 0, 1, ...$ are $k \times k$ matrices of fixed constants such that $\sum_{j=0}^{\infty} \|\mathbf{\Psi}_j\| < K$, $\sup_t \|\boldsymbol{\delta}_{x,t}\| < K$, and $\sup_{j,t} |g_{x,jt}| < K$. Furthermore, $\mathbf{E}_{i,x}$ is distributed independently over i, and of $\boldsymbol{\eta}_i$ and $u_{it'}$ for all i, t, and t', $E(\boldsymbol{\eta}_{ij,x}) = \mathbf{0}$, $E\left(\boldsymbol{\eta}_{ij,x}\boldsymbol{\eta}'_{ij',x}\right) = \mathbf{V}_j$ if $j = j'$ and $E\left(\boldsymbol{\eta}_{ij,x}\boldsymbol{\eta}'_{ij',x}\right) = \mathbf{0}$ for all $j \neq j' = 1, 2, ..., m_x$, $\sup_{i,j} E\left\|\boldsymbol{\eta}_{ij,x}\right\|^{4+\epsilon} < K$, $\boldsymbol{\varepsilon}_{it} \sim IID(\mathbf{0}, \mathbf{I}_k)$ with $\sup_{i,t} E\|\boldsymbol{\varepsilon}_{it}\|^{4+\epsilon} < K$ for some small $\epsilon > 0$, and $\boldsymbol{\varepsilon}_{it}$ are distributed independently of u_{it} for all i, t and t'.*

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 $Var(u_{it}) = \sigma_t^2$, which is discussed further in Section 3. 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. Otherwise no restrictions are imposed on $\mathbf{g}_t$. 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. This assumption can be somewhat relaxed as we note below. 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 our model specification can be viewed as a direct generalization of the standard time and fixed effects models considered routinely in the empirical literature. Our specification also differs from the one considered by Bai (2013) who does not model the fixed effects explicitly but assumes 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 component of the individual specific effects, α_i, is given by

$$\eta_{i1} = \sum_{t=1}^{T} \mathbf{b}_t' \left[\mathbf{x}_{it} - E(\mathbf{x}_{it})\right] + \varepsilon_{\eta_{i1}}, \text{ for } i = 1, 2, ..., N, \tag{4}$$

where $(\mathbf{b}_1', \mathbf{b}_2', ..., \mathbf{b}_T')'$ 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 $\mathbf{x}_{it}$ is stationary and $E(\mathbf{x}_{it})$ being a constant is then absorbed in an intercept. But in our more general context where $\mathbf{x}_{it}$ could be non-stationary the use of the Mundlak scheme in (4) could be problematic.

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, $\mathbf{g}_{x,t}$ which could be correlated with $\mathbf{f}_t$, and a stationary component $\mathbf{v}_{it}$ which is assumed to be cross-sectionally independent. The first-difference formulation allows $\mathbf{x}_{it}$ to have unit roots as well as being stationary. The case of stationary $\mathbf{x}_{it}$ arises when $\boldsymbol{\delta}_{x,t} = \mathbf{0}$, $\sum_{j=0}^{\infty} \mathbf{\Psi}_j = \mathbf{0}$, and $\mathbf{f}_{x,t} = \sum_{\tau=0}^{t} \mathbf{g}_{x,\tau}$ is a stationary process, otherwise $\mathbf{x}_{it}$ will be non-stationary. The assumption that the factor loadings, $\boldsymbol{\eta}_{ij,x}$, $j = 1, 2, ..., m_x$ have zero mean and are uncorrelated over j is made for convenience, and can be relaxed without any consequences for the subsequent analysis. Assumption 5 also allows $\Delta\mathbf{x}_{it}$ to be correlated over time with the errors ζ_{it}, through possible non-zero correlations between $\mathbf{f}_t$ and $\mathbf{g}_{x,t}$, the common components of ζ_{it} and $\mathbf{x}_{it}$, but requires their idiosyncratic components, u_{it} and $\mathbf{v}_{it}$ and their factor loadings, $\boldsymbol{\eta}_i$ and $\mathbf{E}_{i,x}$, to be independently distributed over i. The key requirements are therefore cross-sectional independence of $\mathbf{E}_{i,x}$ and $\boldsymbol{\eta}_i$, and the independence of the idiosyncratic components of y_{it} and $\mathbf{x}_{it}$.

Remark 1 *Our assumptions allow for non-zero correlation between $\mathbf{x}_{it}$ and ζ_{it} through their dependence on common factors, but requires 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 (1) and (2) where $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ is modelled jointly as in Holtz-Eakin et al. (1988) and Binder et al. (2005). In addition, possible correlations between $\boldsymbol{\eta}_i$ and the regressors $\Delta\mathbf{x}_i$ can be dealt with using the Mundlak device as set out above in the case of the fixed effects. Though we do not pursue this here, we investigate the effect of such correlations on the proposed QML estimator in the subsequent Monte Carlo experiments.*

Finally, while the composite error term, ζ_{it}, in (1) is cross-sectionally heteroskedastic through the presence of the interactive effects, allowing explicitly for the same in the idiosyncratic error, u_{it}, of (2) 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.

Combining (1) and (2), and eliminating the individual effects by first-differencing we have

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \boldsymbol{\beta}' \Delta \mathbf{x}_{it} + d_t + \mathbf{g}'_t \boldsymbol{\eta}_i + \Delta u_{it}, \text{ for } t = 2, 3, ..., T;\ i = 1, 2, ..., N, \tag{5}$$

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

$$\xi_{it} = \mathbf{g}'_t \boldsymbol{\eta}_i + \Delta u_{it}, \text{ for } t = 2, 3, ..., T. \tag{6}$$

For the specification of Δy_{i1} we make the following assumption about the initialization of (5):

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*

$$E\left(\Delta y_{i,-S+1} \left| \Delta \mathbf{x}_i \right.\right) = a_S + \boldsymbol{\pi}'_S \Delta \mathbf{x}_i, \tag{7}$$

where $\Delta\mathbf{x}_i = (\Delta\mathbf{x}'_{i1}, \Delta\mathbf{x}'_{i2}, ..., \Delta\mathbf{x}'_{iT})'$ is the $kT \times 1$ vector of observations on the regressors, 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 $\varpi_i = \Delta y_{i,-S+1} - E\left(\Delta y_{i,-S+1} \left| \Delta\mathbf{x}_i \right.\right)$, and suppose that $\varpi_i \sim IID(0, \sigma^2_\varpi)$, $0 < \sigma^2_\varpi < K$, and $\sup_i E\left|\varpi_i\right|^{4+\epsilon} < K$.

This assumption is not that restrictive 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 (1). The main restriction here is the assumed linearity of (7).

Given the above assumptions, we can now derive an expression for Δy_{i1} that depends on the observables and the unknown parameters only. This is in contrast to Bai's initial value assumption that requires y_{i0} for $i = 1, 2, ..., N$ to be independent draws from a distribution with means and variances that do not depend on any of the model's unknown parameters, effectively conditioning the analysis on given values of y_{i0} (see Bai (2013, pp.5-7)).

Using (5), 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

$$\Delta y_{i1} = \gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} + \widetilde{d}_1 + \widetilde{\mathbf{g}}'_1 \boldsymbol{\eta}_i + \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j}, \tag{8}$$

where $\widetilde{d}_1 = \sum_{j=0}^{S-1} \gamma^j d_{1-j}$, and $\widetilde{\mathbf{g}}_1 = \sum_{j=0}^{S-1} \gamma^j \mathbf{g}_{1-j}$. 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 $\Delta\mathbf{x}^0_i = (\Delta\mathbf{x}'_{i0}, \Delta\mathbf{x}'_{i,-1}, ..., \Delta\mathbf{x}'_{i,-S+3})'$, 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

$$\Delta \mathbf{x}_i^0 = \boldsymbol{\delta}_x^0 + \sum_{j=1}^{m_x} \left(\mathbf{g}_{x,j}^0 \otimes \boldsymbol{\eta}_{ij,x} \right) + \mathbf{v}_i^0, \text{ and } \Delta \mathbf{x}_i = \boldsymbol{\delta}_x + \sum_{j=1}^{m_x} \left(\mathbf{g}_{x,j} \otimes \boldsymbol{\eta}_{ij,x} \right) + \mathbf{v}_i, \tag{9}$$

where $\boldsymbol{\delta}_x^0 = (\boldsymbol{\delta}'_{x,0}, \boldsymbol{\delta}'_{x,-1}, ..., \boldsymbol{\delta}'_{x,-S+3})'$, $\mathbf{g}_{x,j}^0 = (g_{x,j,0}, g_{x,j,-1}, ..., g_{x,j,-S+3})'$, and $\mathbf{v}_i^0 = \left(\mathbf{v}'_{i0}, \mathbf{v}'_{i,-1}, ..., \mathbf{v}'_{i,-S+3} \right)'$, and similarly $\boldsymbol{\delta}_x = (\boldsymbol{\delta}'_{x,1}, \boldsymbol{\delta}'_{x,2}, ..., \boldsymbol{\delta}'_{x,T})$, $\mathbf{g}_{x,j} = (g_{x,j1}, g_{x,j2}, ..., g_{x,jT})'$, and $\mathbf{v}_i = (\mathbf{v}'_{i1}, \mathbf{v}'_{i2}, ..., \mathbf{v}'_{iT})'$. Also $E\left(\Delta \mathbf{x}_i^0\right) = \boldsymbol{\delta}_x^0$, $E\left(\Delta \mathbf{x}_i\right) = \boldsymbol{\delta}_x$, and using linear projections, we have

$$E\left(\Delta \mathbf{x}_i^0 \left| \Delta \mathbf{x}_i \right.\right) = \boldsymbol{\delta}_x^0 + \boldsymbol{\Omega}_{01} \boldsymbol{\Omega}_{11}^{-1} \left(\Delta \mathbf{x}_i - \boldsymbol{\delta}_x \right) \tag{10}$$

where

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

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 $E\left(\mathbf{v}_i \mathbf{v}'_i\right) = \boldsymbol{\Omega}_{v,11}$ and $E\left(\mathbf{v}_i^0 \mathbf{v}'_i\right) = \boldsymbol{\Omega}_{v,01}$ that also do not depend on i. Now using (10) along with (7) we have

$$E\left(\gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} \left| \Delta \mathbf{x}_i \right. \right) = a + \boldsymbol{\pi}' \Delta \mathbf{x}_i, \tag{11}$$

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

$$\chi_i = \left(\gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} \right) - E\left(\gamma^S \Delta y_{i,-S+1} + \sum_{j=0}^{S-1} \gamma^j \boldsymbol{\beta}' \Delta \mathbf{x}_{i,1-j} \left| \Delta \mathbf{x}_i \right. \right), \tag{12}$$

and by construction χ_i is a martingale difference process. Also in view of Assumptions 5 and 6 and by application of the Minkowski inequality to both sides of χ_i we have $\sup_i \left|\chi_i\right|^{4+\epsilon} < K$.[2] Hence, using (11) and (12) in (8) we have

$$\Delta y_{i1} = d_1 + \boldsymbol{\pi}' \Delta \mathbf{x}_i + \xi_{i1}, \tag{13}$$

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

$$\xi_{i1} = \widetilde{\mathbf{g}}'_1 \boldsymbol{\eta}_i + v_{i1}, \tag{14}$$

and

$$v_{i1} = \sum_{j=0}^{S-1} \gamma^j \Delta u_{i,1-j} + \chi_i. \tag{15}$$

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 of $\Delta \mathbf{x}_i$ and $\boldsymbol{\eta}_i$. Further, by application of the Minkowski inequality to (15) we have $\sup_i \left|v_{i1}\right|^{4+\epsilon} < K$, and under Assumptions 5 and 6, $\sup_i Var\left(\chi_i\right) < K$; as a result $0 < \omega_{\min} < \omega < \omega_{\max} < \infty$, where $\omega_{\min}$ and $\omega_{\max}$ are fixed constants. Finally, using (15) we have

$$Cov\left(v_{i1}, \Delta u_{it}\right) = \begin{cases} -\sigma^2 & \text{for } t = 2 \\ 0 & \text{for } t = 3, 4, ..., T \end{cases}. \tag{16}$$

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

Remark 2 *As noted earlier, in the case where* $|\gamma| < 1$ *and* $S \to \infty$ *we have* $\Delta y_{i1} = d_1 + \boldsymbol{\pi}'\Delta\mathbf{x}_i + \xi_{i1}$*, where* ξ_{i1} *is defined by (14), with* v_{i1} *given by* $v_{i1} = \sum_{j=0}^{\infty}\gamma^j\Delta u_{i,1-j} + \chi_i$*, and*

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

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*

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

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

We can now combine the processes for Δy_{i1} and Δy_{it} conditional on $\Delta y_{i,t-1}$, for $t = 2, 3, ..., T$ to write down the quasi-likelihood function of the first-differenced model. Writing (5) and (13) in matrix notation we note that

$$\Delta\mathbf{y}_i = \Delta\mathbf{W}_i\boldsymbol{\varphi} + \boldsymbol{\xi}_i,\ \ \boldsymbol{\xi}_i = \mathbf{G}\boldsymbol{\eta}_i + \mathbf{r}_i, \tag{17}$$

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

$$\Delta\mathbf{W}_i = \begin{pmatrix} 1 & 0 & \ldots & 0 & \Delta\mathbf{x}_i' & 0 & 0 \\ 0 & 1 & \ldots & 0 & \mathbf{0} & \Delta\mathbf{x}_{i2}' & \Delta y_{i1} \\ \vdots & \vdots & \ldots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & \mathbf{0} & \Delta\mathbf{x}_{iT}' & \Delta y_{i,T-1} \end{pmatrix}, \tag{18}$$

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

In using the first-differenced specification (17), 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 $\Delta\mathbf{x}_i = (\Delta\mathbf{x}_{i1}', \Delta\mathbf{x}_{i2}', ..., \Delta\mathbf{x}_{iT}')$ are independently distributed over i. This follows since the cross sectional-variation of $\Delta\mathbf{x}_i$, given by (9), is governed by $\mathbf{v}_i$ and $\left\{\boldsymbol{\eta}_{ij,x}, \text{ for } j = 1, 2, ..., m_x\right\}$ that are assumed to be distributed independently of $\boldsymbol{\eta}_i$ and Δu_{it} for all i and t (see Assumption 5). Furthermore, conditional on the common factors, $\mathbf{g}_{x,t}$, $\Delta\mathbf{x}_i$ are cross-sectionally independent which allows us to apply the law of large numbers to averages of $\Delta\mathbf{x}_i$ and quadratic forms in $\Delta\mathbf{x}_i$.

3 Quasi Maximum Likelihood Estimation

Consider the panel data model given by (17) and note that under Assumption 1, and using (14) and (16), we have (recall also that $v_{i1} \sim IID(0, \omega\sigma^2)$)

$$E(\mathbf{r}_i\mathbf{r}_i') = \sigma^2\boldsymbol{\Omega}, \tag{19}$$

where

$$E(\mathbf{r}_i\mathbf{r}_i') = \sigma^2 \begin{pmatrix} \omega & -1 & & & 0 \\ -1 & 2 & \ddots & & 0 \\ & & \ddots & & \\ & & \ddots & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix} = \sigma^2\boldsymbol{\Omega}, \tag{20}$$

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, we have

$$Var(\boldsymbol{\xi}_i) = \mathbf{\Sigma}_\xi(\boldsymbol{\psi}) = \sigma^2\mathbf{\Omega} + \mathbf{G}\mathbf{\Omega}_\eta\mathbf{G}' = \sigma^2\left(\mathbf{\Omega} + \mathbf{Q}\mathbf{Q}'\right) \tag{21}$$

where $\mathbf{Q} = (1/\sigma)\mathbf{G}\mathbf{\Omega}_\eta^{1/2}$, $rank(\mathbf{Q}) = m$, and $\boldsymbol{\psi} = \left(\omega, \sigma^2, vec(\mathbf{Q})'\right)'$. With this normalisation, the quasi-log-likelihood of the transformed model (17) is given by

$$\begin{aligned}\ell_N(\boldsymbol{\theta}) &= \ell_N(\boldsymbol{\varphi}, \boldsymbol{\psi}) = -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\mathbf{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2}\sum_{i=1}^{N}\boldsymbol{\xi}_i'(\boldsymbol{\varphi})\mathbf{\Sigma}_\xi(\boldsymbol{\psi})^{-1}\boldsymbol{\xi}_i(\boldsymbol{\varphi}) &(22)\\ &= -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma^2) - \frac{N}{2}\ln\left|\mathbf{\Omega} + \mathbf{Q}\mathbf{Q}'\right| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\boldsymbol{\xi}_i'(\boldsymbol{\varphi})\left(\mathbf{\Omega} + \mathbf{Q}\mathbf{Q}'\right)^{-1}\boldsymbol{\xi}_i(\boldsymbol{\varphi}), &(23)\end{aligned}$$

where

$$\boldsymbol{\xi}_i(\boldsymbol{\varphi}) = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\varphi}, \tag{24}$$

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

The above log-likelihood function can be readily modified to allow for time series heteroskedasticity, so that $Var(u_{it}) = \sigma_t^2$, for $t = 1, 2, ..., T$. We only need to replace $\sigma^2\mathbf{\Omega}$ by

$$E(\mathbf{r}_i\mathbf{r}_i') = \begin{pmatrix} \omega\sigma_1^2 & -\sigma_1^2 & 0 & \cdots & 0 & 0 & 0 \\ -\sigma_1^2 & \sigma_1^2 + \sigma_2^2 & -\sigma_2^2 & \ddots & \vdots & 0 & 0 \\ 0 & -\sigma_2^2 & \sigma_2^2 + \sigma_3^2 & \ddots & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{T-2}^2 & \sigma_{T-2}^2 + \sigma_{T-1}^2 & -\sigma_{T-1}^2 \\ 0 & 0 & 0 & \cdots & 0 & -\sigma_{T-1}^2 & \sigma_{T-1}^2 + \sigma_T^2 \end{pmatrix}, \tag{25}$$

with the resultant log-likelihood maximized with respect to $\omega, \sigma_1^2, \sigma_2^2, ..., \sigma_T^2$ and the remaining parameters. This extension does not pose additional difficulties noting that T is fixed as $N \to \infty$. However, it does impact the identification conditions to be addressed below.

4 Identification conditions

We shall first derive necessary order conditions on m and T for identification, and then subject to these order conditions we derive additional conditions under which the parameters are locally identified, and show that global identification of short T panels with an error multi-factor structure is possible only in the case of panels without a lagged dependent variable.

We begin our investigation by considering the order condition for identification of the panel AR(1) model. Using (5) and (13), we note that in this case

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

[3] In the Monte Carlo and empirical applications that follow the QML estimates are obtained by maximizing a concentrated version of the likelihood function in (23). This is derived using an eigenvalue approach which greatly simplifies the computations. For details see Section I of the online supplement.

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 $i = 1, 2, ..., N$, where $\mathbf{d} = (d_1, ..., d_T)'$, $\Delta\mathbf{y}_i$ and $\boldsymbol{\xi}_i$ are as defined above, and

$$\mathbf{B}(\gamma) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\gamma & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\gamma & 1 \end{pmatrix}. \tag{26}$$

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

$$\mathbf{B}^{-1}(\gamma) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \gamma & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \gamma^{T-1} & \cdots & \gamma & 1 \end{pmatrix}, \tag{27}$$

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

$$\mathbf{a} = \mathbf{B}^{-1}(\gamma)\mathbf{d} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \gamma & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \gamma^{T-1} & \cdots & \gamma & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_T \end{pmatrix} = \begin{pmatrix} d_1 \\ \gamma d_1 + d_2 \\ \vdots \\ \gamma^{T-1}d_1 + \gamma^{T-2}d_2 + ... + \gamma d_{T-1} + d_T \end{pmatrix}. \tag{28}$$

Since $\mathbf{d}$ is a $T \times 1$ unrestricted parameter vector, then $\mathbf{a}$ is also unrestricted, namely knowing $\mathbf{a}$ does not help to identify γ. Therefore, γ can only be identified from the $T(T+1)/2$ distinct elements of $Var(\Delta\mathbf{y}_i)$ which is given by

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

where $\boldsymbol{\varrho} = \left(\gamma, \omega, \sigma^2\right)'$. But since $\mathbf{Q}$ enters $\boldsymbol{\Sigma}(\boldsymbol{\varrho}, \mathbf{Q})$ 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 $rank(\mathbf{A}) = T$, then $\boldsymbol{\varrho}$ is not identified. Hence, for identification of $\boldsymbol{\varrho}$, we must have $rank(\mathbf{A}) = rank(\mathbf{Q}) = m < T$. When $rank(\mathbf{Q}) = m$, $\mathbf{Q}$ is identified only up to an $m \times m$ non-singular transformation. However, the number of non-redundant parameters of $\mathbf{Q}$ is given by $mT - m(m-1)/2$ (see Hayashi et al. (2007, p.507)).[4] Hence, the order condition for identification of $\boldsymbol{\varrho}$ and the non-redundant elements of $\mathbf{Q}$ is given by

$$T(T+1)/2 \geq 3 + Tm - m(m-1)/2. \tag{29}$$

This order condition is satisfied if $T \geq 3$, for $m = 0, 1, 2, .., m_{\max}$ where $m_{\max}$ is the largest value of m that satisfies (29), that is $m_{\max} = T - 2$. It is easily seen that the above condition is not satisfied if $m = T - 1$. The maximized log-likelihood values for the rank deficient cases, $m = 0, 1, 2..., m_{\max}$ can be computed using (S.11) in the supplement. In the case where the errors, u_{it}, are heteroskedastic over time there are an additional $T - 1$ new error variances to estimate and the above order condition 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 example for $m = 1$ we need $T \geq 4$, and for $m = 2$ we need $T \geq 6$.

[4] Note that $m(m+1)/2$ restrictions are imposed by re-writting $\mathbf{G}\boldsymbol{\Omega}_\eta\mathbf{G}'$ as $\mathbf{Q}\mathbf{Q}'$. To achieve the usual m^2 restrictions imposed on $Var(\mathbf{G}\boldsymbol{\eta}_i)$ in traditional factor analysis an additional $m(m-1)/2$ restrictions need to be imposed on $\mathbf{Q}$. In the concentrated version of the loglikelihood of (17) used to obtain all Monte Carlo and empirical results that follow, these restrictions are imposed on $\mathbf{Q}$. See also Section I of the online supplement.

Consider now the more general case where the panel AR(1) model also contains exogenous regressors, but to simplify the exposition we continue to assume that $\sigma_t^2 = \sigma^2$ for all t. For this case note that the system of equations (17) can be written equivalently as

$$\Delta \mathbf{y}_i = \mathbf{a} + \widetilde{\mathbf{Z}}_i\left(\gamma\right)\boldsymbol{\delta} + \mathbf{B}^{-1}\left(\gamma\right)\boldsymbol{\xi}_i, \tag{30}$$

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

$$\mathbf{Z}_i = \begin{pmatrix} \Delta \mathbf{x}_i' & \mathbf{0} \\ \mathbf{0} & \Delta \mathbf{x}_{i2}' \\ \vdots & \vdots \\ \mathbf{0} & \Delta \mathbf{x}_{iT}' \end{pmatrix}. \tag{31}$$

It is clear from (30) that $\mathbf{a}$ and $\boldsymbol{\delta}$, and hence $\mathbf{d}$ and $\boldsymbol{\delta}$, are uniquely identified for a given value of γ. But it is already established that γ is identified from the covariance of $\mathbf{B}^{-1}(\gamma)\boldsymbol{\xi}_i$, given by $\boldsymbol{\Sigma}(\boldsymbol{\varrho}, \mathbf{Q}) = \sigma^2\mathbf{B}(\gamma)^{-1}(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}')\mathbf{B}'(\gamma)^{-1}$, if the order condition (29) is met. Note that $\boldsymbol{\Sigma}(\boldsymbol{\varrho}, \mathbf{Q})$ does not depend on $\mathbf{d}$ and $\boldsymbol{\delta}$, and hence knowing $\mathbf{d}$ and $\boldsymbol{\delta}$ will not help identification of γ. As a result, the order condition (29) continues to be sufficient for identification of the parameters of the panel ARX(1) model.

To investigate necessary and sufficient conditions for identification of the parameters, we consider the average log-likelihood function defined by (22) which we write as

$$\bar{\ell}_N(\boldsymbol{\theta}) = N^{-1}\ell_N(\boldsymbol{\varphi}, \boldsymbol{\psi}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2N}\sum_{i=1}^{N}\boldsymbol{\xi}_i'(\boldsymbol{\varphi})\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}\boldsymbol{\xi}_i(\boldsymbol{\varphi}), \tag{32}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\varphi}', \boldsymbol{\psi}'\right)'$, $\boldsymbol{\varphi} = \left(\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}', \gamma\right)' = \left(\boldsymbol{\lambda}', \gamma\right)'$ with $\boldsymbol{\lambda} = \left(\mathbf{d}', \boldsymbol{\pi}', \boldsymbol{\beta}'\right)'$, $\boldsymbol{\psi} = \left(\omega, \sigma^2, \mathbf{q}'\right)'$, and $\mathbf{q}$ refers to the $[Tm - m(m-1)/2] \times 1$ vector containing the non-reduntant elements of $\mathbf{Q}$. Suppose that $\boldsymbol{\lambda} \in \boldsymbol{\Theta}_\lambda$, $\gamma \in \boldsymbol{\Theta}_\gamma$, and $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$, and denote the true values of $\boldsymbol{\lambda}$, γ and $\boldsymbol{\psi}$ by $\boldsymbol{\lambda}_0, \gamma_0$, and $\boldsymbol{\psi}_0$, respectively. Consider the set $\mathcal{N}_\epsilon(\gamma_0)$ defined as follows:

Definition 1 *Let $\mathcal{N}_\epsilon(\gamma_0)$ be a set in the closed neighbourhood of γ_0 defined by*

$$\mathcal{N}_\epsilon(\gamma_0) = \{\gamma \in \boldsymbol{\Theta}_\gamma,\ |\gamma - \gamma_0| \leq \epsilon\},$$

for some small $\epsilon > 0$, where $\boldsymbol{\Theta}_\gamma$ is a compact subset of $\mathbb{R}$.

We now show that $\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}_0', \boldsymbol{\psi}_0')' = \left(\boldsymbol{\lambda}_0', \gamma_0, \boldsymbol{\psi}_0'\right)'$ is identified on $\boldsymbol{\Theta}_\epsilon = \boldsymbol{\Theta}_\lambda \times \mathcal{N}_\epsilon(\gamma_0) \times \boldsymbol{\Theta}_\psi$. For this purpose, we require the following additional assumption.

Assumption 7 *(i) $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\epsilon = \boldsymbol{\Theta}_\lambda \times \mathcal{N}_\epsilon(\gamma_0) \times \boldsymbol{\Theta}_\psi$, where $\boldsymbol{\Theta}_{\boldsymbol{\lambda}} = \boldsymbol{\Theta}_d \times \boldsymbol{\Theta}_\pi \times \boldsymbol{\Theta}_\beta$ and $\boldsymbol{\Theta}_\psi = \boldsymbol{\Theta}_\omega \times \boldsymbol{\Theta}_\sigma \times \boldsymbol{\Theta}_q$, $\boldsymbol{\Theta}_d$, $\boldsymbol{\Theta}_\pi$, $\boldsymbol{\Theta}_\beta$ and $\boldsymbol{\Theta}_q$ are compact subsets of $\mathbb{R}^{n_d}$, $\mathbb{R}^{n_\pi}$, $\mathbb{R}^{n_\beta}$, and $\mathbb{R}^{n_q}$, respectively; $\boldsymbol{\Theta}_\omega$ and $\boldsymbol{\Theta}_\sigma$ are compact subsets of $\mathbb{R}$, where $n_d = T$, $n_\pi = kT$, $n_\beta = k$, and $n_q = Tm - m(m-1)/2$; $\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}_0', \boldsymbol{\psi}_0')' = \left(\boldsymbol{\lambda}_0', \gamma_0, \boldsymbol{\psi}_0'\right)'$ lies in the interior of $\boldsymbol{\Theta}_\epsilon$ (ii) $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) = \sigma^2(\boldsymbol{\Omega} + \mathbf{Q}\mathbf{Q}')$, and for some $c_{\max} > c_{\min} > 0$, $c_{\min} \leq \inf_{\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi}\lambda_{\min}[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})] < \sup_{\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi}\lambda_{\max}[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})] \leq c_{\max}$, and (iii) as $N \to \infty$*

$$\mathbf{A}_N(\boldsymbol{\psi}) = \frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}\Delta\mathbf{W}_i \overset{a.s.}{\to} \mathbf{A}(\boldsymbol{\psi}) \ \textit{uniformly in } \boldsymbol{\Theta}_\psi, \tag{33}$$

where $\mathbf{A}(\boldsymbol{\psi}) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left(\Delta\mathbf{W}_i'\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}\Delta\mathbf{W}_i\right)$ is positive definite for all values of $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$.

The first part of this assumption is standard and rules out parameter values on the boundary of the parameter space, and since $\mathcal{N}_\epsilon(\gamma_0)$ is a subset of $\mathbf{\Theta}_\gamma$ which is compact, it also follows that $\mathbf{\Theta}_\epsilon$ being the Cartesian product of compact sets, is itself compact, namely $\mathbf{\Theta}_\epsilon \in \mathbb{R}^{n_\theta}$, where $n_\theta = 3 + T(k+1) + k + Tm - m(m-1)/2$. Note also that order condition (29) is taken into account in setting n_θ. The eigenvalue conditions on $\mathbf{\Sigma}_\xi(\boldsymbol{\psi})$ in the second part of the assumption are required for the proof of the consistency results. This part of the assumption also holds when the order condition is met and $\omega > 1 - \frac{1}{T}$. Recall that under the latter $\mathbf{\Omega}$ is a positive definite matrix and $\mathbf{Q}$ is rank deficient, and under Assumption 1, $0 < \sigma^2 < K$. For γ we need to distinguish between the case where S is fixed (namely initialization is from a finite past) and when $S \to \infty$. Under the former, it is only required that $|\gamma| < K$, which includes the unit root case ($|\gamma| = 1$). Under the latter (when $S \to \infty$), we must have $|\gamma| < 1$. Consider now the third part of Assumption 7, and note that

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

where $\left\|\mathbf{\Sigma}_\xi(\boldsymbol{\psi})^{-1}\right\| < K$ 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 conditional on $\mathbf{g}_{x,t}$, $\Delta\mathbf{x}_i$ are independent across i by Assumption 5 (see also the expression for $\Delta\mathbf{x}_i$ given by (9)), and Δy_{it} being a function of $\Delta\mathbf{x}_{it}$ and ξ_{it} (see (30)) 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 $\boldsymbol{\psi} \in \mathbf{\Theta}_\psi$ (see, for example, Davidson (1994, Theorem 19.4)). Under condition (ii) of Assumption 7 it is trivial to see that this result also holds uniformly in $\mathbf{\Theta}_\psi$. Finally, the condition that $\mathbf{A}(\boldsymbol{\psi})$ is a positive definite matrix is needed for identification of $\boldsymbol{\varphi}$.

The main identification result is set out in the following proposition:

Proposition 1 *Consider the model given by (1) and (2), with the associated log-likelihood function for first-differences given by (22). Suppose that Assumptions 1-7, and the order condition (29) hold. Then $\boldsymbol{\theta}_0$ is almost surely locally identified for values of γ sufficiently close to γ_0, as formalised by Definition 1.*

Remark 3 *In the absence of lagged dependent variables in (1), $\boldsymbol{\theta}_0$ is almost surely globally identified even if $m > 0$. This can be easily seen from the proof of Proposition 1 in the Appendix. Similarly γ and $\boldsymbol{\beta}$ are globally identified when $m = 0$. Therefore, lack of global identification in short T panels arises when the panel data model contains both dynamics and latent factors.*

5 Asymptotic properties of the QML estimator

The analysis of consistency and asymptotic normality of the QML estimator, $\widehat{\boldsymbol{\theta}} = \arg\max_{\theta\in\mathbf{\Theta}_\epsilon} \bar{\ell}_N(\boldsymbol{\theta})$, now follows by application of standard results from the literature. Almost sure local consistency of $\widehat{\boldsymbol{\theta}}$ follows, for example, from a straightforward adaptation of Theorem 9.3.1 of Davidson (2000). Specifically: (i) $\mathbf{\Theta}_\epsilon$ as a subset of $\mathbf{\Theta}$ is compact, (ii) setting $\bar{C}_N(\boldsymbol{\theta}) = -2\bar{\ell}_N(\boldsymbol{\theta})$, and $\bar{C}(\boldsymbol{\theta}) = E_0[\bar{C}_N(\boldsymbol{\theta})]$, under Assumptions 1-7, and using (A.37) and (A.38) in the Appendix we have that $\bar{C}_N(\boldsymbol{\theta}) \xrightarrow{a.s.} \bar{C}(\boldsymbol{\theta})$ uniformly on $\mathbf{\Theta}_\epsilon$ and (iii) $\boldsymbol{\theta}_0$ is the unique minimum of $\bar{C}(\boldsymbol{\theta})$ on $\mathbf{\Theta}_\epsilon$, and is an interior point of $\mathbf{\Theta}_\epsilon$, by assumption. Condition (iii) follows directly from condition (ii) and Proposition 1 (see Davidson (2000, Theorem 9.3.4)). Therefore, all three conditions of Theorem 9.3.1 of Davidson are satisfied and $\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$ on the set $\mathbf{\Theta}_\epsilon$.

The asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ is derived by taking a Taylor expansion of $\frac{\partial\bar{\ell}_N(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} = \mathbf{0}$ at $\boldsymbol{\theta}_0$ and checking the asymptotic behaviour of the score function, $\bar{\mathbf{s}}_N(\boldsymbol{\theta}) = \frac{\partial\bar{\ell}_N(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$, and Hessian matrix, $\mathbf{H}_N(\boldsymbol{\theta}) = -\frac{\partial^2\bar{\ell}_N(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$. If $E_0\left[\frac{\bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right] = \mathbf{0}$ and $\mathbf{H}_N(\breve{\boldsymbol{\theta}}) \xrightarrow{a.s.} \mathbf{H}(\boldsymbol{\theta}_0)$, the asymptotic normality of the QMLE will follow from the mean value theorem:

$$\mathbf{0} = \sqrt{N}\bar{\mathbf{s}}_N(\widehat{\boldsymbol{\theta}}) = \sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) - \mathbf{H}_N(\breve{\boldsymbol{\theta}})\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{34}$$

where $\breve{\boldsymbol{\theta}}$ lies between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. The resultant asymptotic distribution is summarized in the following theorem:

Theorem 1 *Consider the dynamic panel data model given by (1) with interactive effects as in (2). Suppose that Assumptions 1 to 7, the order condition (29) and Proposition 1 hold. Denote the QML estimator of $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}} = \arg\max_{\theta \in \boldsymbol{\Theta}_\epsilon} \bar{\ell}_N(\boldsymbol{\theta})$, where $\bar{\ell}_N(\boldsymbol{\theta})$ is given by (32). Then, $\widehat{\boldsymbol{\theta}}$ is almost surely locally consistent for $\boldsymbol{\theta}_0$ on $\boldsymbol{\Theta}_\epsilon$ for values of γ sufficiently close to γ_0 as formalised by Definition 1, and*

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow_d N\left[\mathbf{0}, \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\theta}_0)\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right], \tag{35}$$

where $\mathbf{H}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} E_0\left[-\frac{\partial^2 \bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]$ and $\mathbf{J}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} E_0\left[N\frac{\partial\bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]$, both assumed to exist.

When $\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)$ is Gaussian $\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow_d N\left[\mathbf{0}, \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right]$. A consistent estimator for the variance in (35) can be obtained by substituting $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$ in the expressions for $\mathbf{J}(\boldsymbol{\theta}_0)$ and $\mathbf{H}(\boldsymbol{\theta}_0)$.

6 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 et al. (1988) and Cragg and Donald (1997), Ahn et al. (2013) choose the significance level b_N such that $b_N \to 0$ and $-\ln(b_N)/N \to 0$ as $N \to \infty$. 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 individual effects 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 critical values of the tests to take account of the multiple testing nature of the procedure in terms of T, as well as adjusting the critical values 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.

6.1 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 (29), 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 maximized log-likelihood function, $\ell_N\left(\widehat{\boldsymbol{\theta}}_{m_0}\right)$, is computed by maximizing (32) subject to r_0 over-identifying restrictions given by

$$r_0 = T(T+1)/2 - 3 - \left[Tm_0 - m_0(m_0-1)/2\right]. \tag{36}$$

The LR statistic for testing H_0: $m = m_0$ against H_1: $m = m_{\max} = T - 2$, is then given by

$$\mathcal{LR}_N(m_{\max}, m_0) = 2\left[\ell_N\left(\widehat{\boldsymbol{\theta}}_{m_{\max}}\right) - \ell_N\left(\widehat{\boldsymbol{\theta}}_{m_0}\right)\right], \text{ for } m_0 = 0, 1, 2, .., T-3, \tag{37}$$

where $\widehat{\boldsymbol{\theta}}_{m_{\max}}$and $\widehat{\boldsymbol{\theta}}_{m_0}$ are the QML estimators of $\boldsymbol{\theta}$ under $m = m_0$ and $m = T-2$, respectively, and $\ell_N(\boldsymbol{\theta})$ is given by (22). The following theorem provides the asymptotic distribution of $\mathcal{LR}_N(m_{\max}, m_0)$.

Theorem 2 *Consider the dynamic panel data model given by (1), and suppose that Assumptions 1 to 7, the order condition (29) and Proposition 1 hold. Denote the constrained QMLE of $\boldsymbol{\theta}$ obtained under $H_0 : m = m_0$ by $\widehat{\boldsymbol{\theta}}_{m_0}$ and its unconstrained estimator by $\widehat{\boldsymbol{\theta}}_{m_{\max}}$, where $m_{\max} = T - 2$. Also let the restrictions imposed under H_0 be given by $\mathbf{q}_0(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{q}_0(\boldsymbol{\theta})$ is the $r_0 \times 1$ vector function of $\boldsymbol{\theta}$ implied by setting $m = m_0$ where $r_0 = T(T+1)/2 - 3 - [Tm_0 - m_0(m_0-1)/2]$. Then, as $N \to \infty$ for a fixed T, $\mathcal{LR}_N(m_{\max}, m_0)$ defined by (37) has the following asymptotic distribution*

$$\mathcal{LR}_N(m_{\max}, m_0) \to_d \sum_{j=1}^{r_0} w_j z_j^2,$$

where $z_j \sim IIDN(0,1)$, $w_1, w_2, ..., w_{r_0}$ are the non-zero (positive) eigenvalues of the symmetric matrix

$$\mathbf{A}_0 = \mathbf{J}_0^{1/2}\mathbf{H}_0^{-1}\mathbf{Q}_0'\left(\mathbf{Q}_0\mathbf{H}_0^{-1}\mathbf{Q}_0'\right)^{-1}\mathbf{Q}_0\mathbf{H}_0^{-1}\mathbf{J}_0^{1/2},$$

with $\mathbf{J}_0 = \mathbf{J}(\boldsymbol{\theta}_0)$, $\mathbf{H}_0 = \mathbf{H}(\boldsymbol{\theta}_0)$, $\mathbf{Q}_0 = \mathbf{Q}(\boldsymbol{\theta}_0)$, and $\mathbf{Q}_0(\boldsymbol{\theta}) = \partial \mathbf{q}_0(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ of dimension $(r_0 \times n_\theta^*)$, with $n_\theta^* = 3 + T(k+1) + k + (T-2)(T+3)/2$.

Remark 4 *Note that the non-zero eigenvalues of $\mathbf{A}_0$ are also the eigenvalues of $\left(\mathbf{Q}_0\mathbf{H}_0^{-1}\mathbf{Q}_0'\right)^{-1}\mathbf{Q}_0\left(\mathbf{H}_0^{-1}\mathbf{J}_0\mathbf{H}_0^{-1}\right)\mathbf{Q}_0'$. Hence, if $\mathbf{J}_0 = \mathbf{H}_0$, this matrix becomes equal to $\mathbf{I}_{r_0}$ and we have $w_i = 1, (i = 1, 2, ..., r_0)$, which yields the familiar result*

$$\mathcal{LR}_N(m_{\max}, m_0) \to_d \chi^2_{r_0}.$$

This theorem shows that the use of LR tests in the non-Gaussian setting is non-standard and requires an explicit derivation of $\mathbf{q}_0(\boldsymbol{\theta}) = \mathbf{0}$. Furthermore, even in the standard case the use of the sequential LR procedure for 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 propose a novel approach for dealing 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.

Theorem 3 *Suppose under the null hypothesis H_0, the LR test statistic $\mathcal{LR}_N$ is distributed as $\sum_{i=1}^h w_i\chi_i^2(1)$, where the weights $w_1 \geq w_2 \geq ... \geq w_h > 0$ are finite constants, and $\chi_i^2(1)$ for $i = 1, 2, ..., h$ are independently distributed central chi-squared variates with 1 degree of freedom. Further suppose that under the alternative hypothesis H_1, $\mathcal{LR}_N$ is distributed as $\sum_{i=1}^h w_i\chi_i^2(1, \mu_{i,N}^2)$, where $\chi_i^2(1, \mu_{i,N}^2)$ for $i = 1, 2, ..., h$ are independently distributed non-central chi-squared variates with 1 degree of freedom and non-centrality parameter, $\mu_{i,N}^2$, $i = 1, 2, ..., h$. Denote the non-centrality parameter of the test under H_1 by $\mu_N^2 = \sum_{i=1}^h \mu_{i,N}^2$. Suppose h is a finite integer, and $\mu_N^2 = O(N)$. Denote type I and II errors of the test by α_N and β_N, respectively, and the critical value of the test by $c_N^2(h)$. Under Assumptions 1-7, if $c_N^2(h) \to \infty$ and $\mu_N^2 \to \infty$ as $N \to \infty$ such that $c_N^2(h)/\mu_N^2 \to 0$, then both α_N and $\beta_N \to 0$.*

Remark 5 *Clearly, the conditions of Theorem 3 are met if $\alpha_N = p/N^\delta$, with δ a finite positive constant. Further, using (A.50) in the Appendix we have*

$$\frac{c_N^2(h)}{\mu_N^2} \leq \frac{2\theta_{\min}^{-2}\ln\left(\frac{h}{\alpha_N}\right)}{\mu_N^2} = \frac{2w_1 h \ln\left(\frac{hN^\delta}{p}\right)}{\mu_N^2} = O\left(\frac{\delta\ln(N)}{\mu_N^2}\right), \tag{38}$$

and since by assumption $\mu_N^2 = O(N)$ it follows that $c_N^2(h)/\mu_N^2 \to 0$ as required.

Remark 6 *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. Then $\beta_N \to 0$ if $\ln N/\mu_N^2 \to 0$, which is satisfied in the standard case where $\mu_N^2 = O(N)$. 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 $\Pr(H_0)\alpha_N + \Pr(H_1)\beta_N \to 0$ as $N \to \infty$, 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 7 *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.*

Theorems 2 and 3 can now be used to develop a sequential approach for estimating (selecting) m. As the true number of factors, 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 null hypotheses of interest be $H_{T-2,0}, H_{T-2,1}, ..., H_{T-2,T-3}$, and write the $T-2$ LR tests as

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

where $\mathcal{LR}_N\left(m_{\max}, m_0\right)$ is given by (37), $CV_{N,T-2,t-1}$ is the critical value for the test of $H_{T-2,t-1}$, and $p_{N,T-2,t-1}$ is the realized p-value for $H_{T-2,t-1}$. The overall size of the test is now given by the family-wise error rate (FWER) defined by

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

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 Boole's inequality (also known as the union bound) we have

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

Hence, to obtain $FWER_N \leq p$, it is sufficient to set $p_{N,T-2,t-1} \leq p/(T-2)$. The individual critical values, $CV_{N,T-2,t-1}$ are based on the asymptotic critical values (as $N \to \infty$) of the χ^2 distribution, namely $\chi^2_{r_0}\left[p/\left(T-2\right)\right]$, where $p/(T-2)$ is the right-tail probability of the individual tests.

Utilizing the above results we propose the following sequential testing procedure to estimate m:

$$\widehat{m} = 0, \text{ if } \mathcal{LR}_N\left(m_{\max} = T-2, m_0 = 0\right) < CV_{N,T-2,0}$$

$$\widehat{m} = 1, \mathcal{LR}_N\left(m_{\max} = T-2, m_0 = 0\right) \geq CV_{N,T-2,0} \text{ and } \mathcal{LR}_N\left(m_{\max} = T-2, m_0 = 1\right) < CV_{N,T-2,1}$$

$$\begin{aligned} \widehat{m} \; = \; & 2, \mathcal{LR}_N\left(m_{\max} = T-2, m_0 = 0\right) \geq CV_{N,T-2,0}; \mathcal{LR}_N\left(m_{\max} = T-2, m_0 = 1\right) \geq CV_{N,T-2,1} \\ & \text{and } \mathcal{LR}_N\left(m_{\max} = T-2, m_0 = 2\right) < CV_{N,T-2,2}. \end{aligned}$$

This sequential procedure is continued until $t = T-3$. The consistency of $\widehat{m}$ for m_0 is established in the following theorem:

Theorem 4 *Let $\widehat{m}$ be the number of factors obtained using the sequential likelihood ratio procedure based on the statistic $\mathcal{LR}_N\left(m_{\max}, m_0\right)$ given by (37) for which Theorem 3 holds. Then* $\Pr(\widehat{m} = m_0) \rightarrow 1$.

By Theorems 3 and Theorem 4 it follows that $\widehat{m}$ obtained using the sequential MTLR procedure described above is a consistent estimator of the true number of factors m_0. In line with the above, in the ensuing Monte Carlo results when performing the sequential MTLR procedure we set $\alpha_N = \frac{p}{N(T-2)}$.

7 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.

7.1 Monte Carlo design

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

$$y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}, \tag{39a}$$

$$\zeta_{it} = \sum_{\ell=1}^{m_0} \eta_{\ell i} f_{\ell t} + u_{it} = \boldsymbol{\eta}_i' \mathbf{f}_t + u_{it}, \tag{39b}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$. 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 $\alpha_i \sim IID\mathcal{N}(0, 1)$. The factor loadings, $\boldsymbol{\eta}_i = (\eta_{1i}, \eta_{2i}, ..., \eta_{m_0,i})'$ are generated as

$$\eta_{\ell i} \sim IID\mathcal{N}\left(0, \frac{\kappa^2}{m_0}\right), \ \ell = 1, 2, ..., m_0. \tag{40}$$

We have scaled the variance of $\eta_{\ell i}$, $\sigma^2_{\eta_\ell}$, 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 $Var\left(\eta_{\ell i}\right) = 0$ for all ℓ. The strength of the factors is controlled by the parameter κ^2.

The idiosyncratic errors, u_{it}, for $t = 0, 1, ..., T$ and $i = 1, 2, ..., N$ are generated as $u_{it} \sim IID \frac{\sigma}{\sqrt{12}}(\chi^2_6 - 6)$ where χ^2_6 is a chi-square variate with six degrees of freedom. The regressors, x_{it}, for $i = 1, 2, ..., N$ are generated as

$$x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{i\ell} f_{\ell t} + \mathrm{v}_{it}, \ \mathrm{v}_{it} = \rho_x \mathrm{v}_{i,t-1} + \left(1 - \rho_x^2\right)^{1/2} \varepsilon_{it}, \quad \text{for } t = 1, 2, ..., T, \tag{41}$$

with $\rho_x = 0.95$, and $\varepsilon_{it} \sim IID\mathcal{N}(0, \sigma^2_{\mathrm{v}i})$. 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 $\mathrm{v}_{i0} \sim IID\mathcal{N}(0, \sigma^2_{\mathrm{v}i})$, for $i = 1, 2, ..., N$. This in turn ensures that $Var\left(\mathrm{v}_{it}\right) = \sigma^2_{\mathrm{v}i}$. These error variances are drawn as $\sigma^2_{\mathrm{v}i} \sim IID$ $\frac{1}{4}(\chi^2_2 + 2)\sigma^2_{\mathrm{v}}$, thus ensuring that $E\left(\sigma^2_{\mathrm{v}i}\right) = \sigma^2_{\mathrm{v}}$. The factor loadings in the x_{it} equations, $\vartheta_{i\ell}$, are generated as $\vartheta_{i\ell} \sim IID\mathcal{N}(0, \sigma^2_{\vartheta\ell})$, for $\ell = 1, 2, ..., m_x$. 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 $\sigma^2_{\vartheta\ell} = \sigma^2_{\mathrm{v}}/m_x$, for all ℓ. Finally, we set $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IID\mathcal{N}(0, 1)$, for all i. This specification ensures that the fixed effects are correlated with the regressors.

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

$$f_{\ell t} = \rho_{\ell f} f_{\ell, t-1} + \left(1 - \rho^2_{f\ell}\right)^{1/2} \varepsilon_{f\ell t}, \ \varepsilon_{f\ell t} \sim IID\mathcal{N}(0, 1), \text{ for } \ell = 1, 2, ..., m_0, \text{ and } t = 1, 2, ..., T, \tag{42}$$

with $\rho_{f\ell} = \rho_f = 0.5$, and $f_{\ell,0} = 0$ for $\ell = 1, 2, ..., m_0$. 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

$$\lambda_{f,NT} = \frac{N^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}'_i\left(T^{-1}\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}'_t\right)\boldsymbol{\eta}_i}{N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}u_{it}^2}, \tag{43}$$

$$\lambda_{x,NT} = \frac{N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(x_{it}-\alpha_{xi}\right)^2}{N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\zeta_{it}^2}, \tag{44}$$

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

$$T^{-1}\sum_{t=1}^{T}f_{\ell t} = 0,\; T^{-1}\sum_{t=1}^{T}f_{\ell t}^2 = 1,\; T^{-1}\sum_{t=1}^{T}f_{\ell t}f_{\ell' t} = 0, \text{ for all } \ell \text{ and } \ell \neq \ell'. \tag{45}$$

Under the above scaling and using (40) we have (for any finite T) and as $N \to \infty$

$$\lambda_f = \lim_{N\to\infty}\kappa_{f,NT} = \frac{E\left(\boldsymbol{\eta}'_i\boldsymbol{\eta}_i\right)}{\sigma^2} = \frac{\kappa^2}{\sigma^2}. \tag{46}$$

Similarly, using (41) and (39b) we have

$$\begin{aligned}\lambda_x &= \frac{\lim_{N\to\infty}\left[N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\sum_{\ell=1}^{m_x}\vartheta_{i\ell}f_{\ell t}+\mathrm{v}_{it}\right)^2\right]}{\lim_{N\to\infty}\left[N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{\eta}'_i\mathbf{f}_t+u_{it}\right)^2\right]} \\ &= \frac{2\sigma_{\mathrm{v}}^2}{\kappa^2+\sigma^2} = \frac{2\sigma_{\mathrm{v}}^2/\sigma^2}{1+\kappa^2/\sigma^2}.\end{aligned} \tag{47}$$

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 $\sigma_{\mathrm{v}}^2 = \{1/2, 1, 3/2\}$. 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, $\{y_{i0}; i = 1, 2, ..., N\}$, 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 set[5]

$$y_{i0} = \mu_{i0} + \sigma_{i0}\left(u_{i0}/\sigma\right), \text{ for } i = 1, 2, ..., N \tag{48}$$

where

$$\mu_{i0} = \frac{\alpha_i + \beta\alpha_{xi}}{1-\gamma},\; \sigma_{i0}^2 = \frac{\sigma^2 + \mathrm{a}_x\beta^2\sigma_{i\mathrm{v}}^2 + \mathrm{a}_f\mathrm{a}_i}{1-\gamma^2}, \tag{49}$$

$$\mathrm{a}_x = \frac{1+\gamma\rho_x}{1-\gamma\rho_x},\; \mathrm{a}_f = \frac{1+\gamma\rho_f}{1-\gamma\rho_f}, \tag{50}$$

$$\mathrm{a}_i = \sum_{\ell=1}^{m_0}\eta_{\ell i}^2 + \beta^2\sum_{\ell=1}^{m_x}\vartheta_{\ell i}^2 + 2\beta\sum_{\ell=1}^{\min(m_0,m_x)}\eta_{\ell i}\vartheta_{\ell i}, \tag{51}$$

[5]For the derivation of μ_{i0} and σ_{i0} see Section S.3 of the online supplement.

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 (48) 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

$$\Delta y_{i1} = \Delta\delta_1 + \beta\Delta x_{i1} + \Delta\zeta_{i1}, \tag{52}$$

$$\Delta y_{it} = \Delta\delta_t + \gamma\Delta y_{i,t-1} + \beta\Delta x_{it} + \Delta\zeta_{it},\ t = 2, 3, ..., T, \tag{53}$$

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

7.2 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 QML estimators 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 QML estimator with the GMM quasi-difference (QD) and first-difference (FD) estimators proposed by Ahn et al. (2013) (ALS), assuming m_0 is known.[6] 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 $\kappa^2 = \sigma_{\text{v}}^2 = 1$; results for other values of $\kappa^2 = \{1/4, 1/2, 2\}$ and $\sigma_{\text{v}}^2 = \{1/2, 3/2\}$ are provided in the online supplement and are discussed only briefly to save space. Further, we only resport the results for non-Gaussian errors. The results for the case of Gaussian errors are available upon request.

Unless otherwise stated, the sequential MTLR procedure is implemented using the $\mathcal{LR}_N(m_{\max}, m_0)$ statistic for testing $m = m_0 = \{0, 1, 2, .., T-3\}$ against $m = m_{\max} = T - 2$, with significance level $\alpha_N = \frac{p}{(T-2)N}$ and $p = 0.05$, using the critical values of the chi-square distribution with degrees of freedom as given by (36). The standard errors used for inference are based on equation (35) with all derivatives computed numerically. All tests are carried out at the 5% significance level and all experiments are replicated 2000 times.

7.2.1 Selecting the number of factors

Table 1 reports the number of times (in %) that the estimated number of factors, $\widehat{m}$, is equal to the true number of factors, m_0, following the sequential MTLR procedure outlined in Section 6.1. The results refer to the baseline case where $\kappa^2 = \sigma_{\text{v}}^2 = 1$ and show that $\widehat{m}$ 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) 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 $\widehat{m}$ is not that much affected as other values of σ_{v}^2 are considered.

[6] The estimation of the number of factors for the GMM type estimators turned out to be time consuming, and to simplify the comparisons we thought it more instructive to base our comparisons assuming that m is known. Also, as will be seen, under our approach m is generally well estimated.

7.2.2 Performance of the QML estimator

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

AR(1) For this model, bias, RMSE, and empirical size for the QML 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 QML 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 $\sigma_{\text{v}}^2 = \{1/2, 1, 3/2\}$ 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 QML estimators 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 is increased. 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) 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 $\sigma_{\text{v}}^2 = 1$, and are available upon request.

7.2.3 Comparison of QML and GMM estimators

Next we present simulation results comparing the QML estimator with the GMM estimators proposed by ALS. For this set of experiments the number of factors during estimation is set to the true number of factors. The GMM estimators include the quasi-differenced and first-differenced ALS one step and two step estimators denoted by QD1, QD2, FD1 and FD2, respectively. The details of how these estimators are computed are set out in Section II of the online supplement. Results for the ARX(1) model are reported in Table 4, but to save space the results for the AR(1) model are provided in the online supplement.

For the GMM estimation results, deviation from cross section averages is taken prior to estimation to remove the time effects. Since the individual effects are the loadings of a constant factor, the number of factors used for the QD estimators is $\widetilde{m} = m + 1$, while m factors are used for the FD estimators.

Table 4 reports the bias and RMSE of γ and β for the QML and GMM estimators, and shows that the QML 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 summarized in Table 5. 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 QML 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.[7]

For the AR(1) model the results are summarized in Table A2(v) of the online supplement, and again show that the QML estimator performs substantially better than the GMM estimators in terms of bias and RMSE.[8] When $\gamma_0 = 0.8$, the GMM estimators, especially FD1 and FD2, perform very poorly due to weak instruments whereas the QML estimator has small bias and RMSE. With regard to size shown in Table A2(vi), the GMM estimators display substantial size distortions while the QML estimator has empirical size close to the nominal value, except for the case where $\gamma_0 = 0.8$ and $N = 100$, for which minor distortion is observed.

7.2.4 The unit root case

The results for the unit root case are very similar to those already reported for the stationary case. Table 6 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 QML estimator when $\gamma_0 = 1$ are summarized in Tables 7 and 8 for the AR(1) and ARX(1) models respectively. As can be seen, 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 3 and 4 for the AR(1) and ARX(1) 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).

7.2.5 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 (48) 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.7 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) 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.

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

[8] The case of $T = 5$ is not reported for the AR(1) model because the number of unknown parameters exceeds that of the moment conditions.

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.

8 Empirical applications

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 application 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 application, 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 panel regressions are estimated with fixed and time effects, and the presence of interactive effects is investigated by first estimating m, the number of unobserved factors, subject to $m_{\max} = T - 2$.

8.1 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 ($WTUC_{it}$).[9] 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 6.1). We obtain $\widehat{m} = 3$ 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 ($\widehat{m} = 3$) if we set $p = 10\%$, although setting $p = 1\%$ yields $\widehat{m} = 2$. In Table 9 we report the results for $\widehat{m} = 3$, 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

[9] Cornwell and Trumbull (1994) and Baltagi (2006) also consider wage rates in other industries, as well as a number other variables, which we exclude to simplify the exposition and to avoid possible endogeneity of the included regressors.

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 $\widehat{m} = 3$.

Table 9: 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 ($y_{i,t-1}, \mathbf{x}_{it}$)	$\widehat{m} = 3$	$m = 0$
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 QML estimation after first differencing. $\widehat{m}$ is the latent factors estimated using the sequential MTLR procedure described in Section 6.1 with $m_{\max} = T - 2 = 4$ and $\alpha_N = 0.05/(N(T-2))$. The standard errors are computed according to equation (35). Figures in parentheses are standard errors. ***,** ,* denote significance at the 1%, 5% and 10% levels, respectively.

8.2 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.[10]

For this application the number of latent factors (m) was estimated to be $\widehat{m} = 2$, 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 $\widehat{m} = 2$ and $m = 0$ are summarized in Table 10. 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

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

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 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 $\widehat{m} = 2$.

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

Table 10: Dynamic panel regressions for cross country log per capita output equations (y_{it}) (1981-2005, five yearly $T = 5, N = 82$)

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

Note: $\widehat{m}$ is the estimated number of factors using the sequential MTLR procedure described in Section 6.1 with $m_{\max} = T - 2 = 3$ and $\alpha_N = 0.05/(N(T-2))$. See also the note to Table 9.

9 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 QML 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 QML estimator in terms of bias, RMSE and empirical size, and power. Empirical applications to cross county 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

Abdel-Aty, S. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2 distribution. *Biometrika 41*, 538–540.

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (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 A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

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

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

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*, 5–27.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*, 598–606.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29–51.

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

Bai, J. (2013). Likelihood approach to dynamic panel models with interactive effects. mimeo.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bailey, N., M. H. Pesaran, and L. V. Smith (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., B. M. Pötscher, and P. Hackl (1988). Model selection by multiple test procedures. *Statistics 19*, 39–44.

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

Blundell, R. and S. Bond (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., G. Esquivel, and F. Lefort (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.

Chamberlain, G. (1984). Panel data. In *Handbook of Econometrics*, Volume 2, Chapter 22, pp. 1248–1318. North-Holland.

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

Cornwell, C. and W. N. Trumbull (1994). Estimating the economic model of crime with panel data. *The Review of Economics and Statistics 76*, 360–366.

Cragg, J. G. and S. G. Donald (1997). Inferring the rank of a matrix. *Journal of Econometrics 76*, 223–250.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

Davidson, J. (2000). *Econometric theory.* Oxford: Wiley-Blackwell.

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.

Hayakawa, K. (2016). Identification problem of GMM estimators for short panel data models with interactive fixed effects. *Economics Letters 139*, 22–26.

Hayakawa, K. and M. H. Pesaran (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., P. M. Bentler, and K.-H. Yuan (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., W. K. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395.

Hsiao, C., M. H. Pesaran, and K. A. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*, 107–150.

Hsiao, C. and A. K. Tahmiscioglu (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 T. Osang (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., M. H. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics 160*, 326–348.

Lee, K., M. H. Pesaran, and R. Smith (1998). Growth empirics: a panel data approach–a comment. *The Quarterly Journal of Economics 113*, 319–323.

Lee, K., M. H. Pesaran, and R. P. Smith (1997). Growth and convergence in a multi-country empirical stochastic solow model. *Journal of Applied Econometrics 12*, 358–392.

Magnus, J. R. and H. Neudecker (1988). *Matrix differential calculus with applications in statistics and econometrics*. Oxford: John Wiley & Sons.

Mankiw, N. G., D. Romer, and D. N. Weil (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 A. Thomas (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.

Patnaik, P. B. (1949). The non-central χ^2- and f-distribution and their applications. *Biometrika 36*, 202–232.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with cross section dependence. *Econometrica 74*, 967–1012.

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

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

Phillips, P. C. B. and D. Sul (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics 127*, 162–188.

Pötscher, B. M. and I. R. Prucha (2001). Basic elements of asymptotic theory. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, pp. 201–229. Oxford: Blackwell.

Robertson, D. and V. Sarafidis (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.

Sankaran, M. (1959). On the non-central chi-square distribution. *Biometrika 46*, 235–237.

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

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

Somé, J., S. Pasali, and M. Kaboine (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.

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 ($\kappa^2 = \sigma_{\mathrm{v}}^2 = 1$)

	$T=5$						$T=10$					
m_0	0	1	2	0	1	2	0	1	2	0	1	2
	$\gamma_0=0.4$			$\gamma_0=0.8$			$\gamma_0=0.4$			$\gamma_0=0.8$		
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 $i = 1, 2, ..., N; t = 1, ..., T$, with $y_{i0} = \mu_{i0} + \sigma_{i0}(u_{i0}/\sigma)$ where $\mu_{i0} = (\alpha_i + \beta \alpha_{xi})/(1-\gamma)$ and $\sigma_{i0}^2 = (\sigma^2 + \mathrm{a}_x \beta^2 \sigma_{i\mathrm{v}}^2 + \mathrm{a}_f \mathrm{a}_i)/(1-\gamma^2)$. In addition, $\mathrm{a}_x = (1+\gamma\rho_x)/(1-\gamma\rho_x)$, $\mathrm{a}_f = (1+\gamma\rho_f)/(1-\gamma\rho_f)$ and $\mathrm{a}_i = \sum_{\ell=1}^{m_0} \eta_{\ell i}^2 + \beta^2 \sum_{\ell=1}^{m_x} \vartheta_{\ell i}^2 + 2\beta \sum_{\ell=1}^{\min(m_0, m_x)} \eta_{\ell i} \vartheta_{\ell i}$, where $\eta_{\ell i} \sim IID\mathcal{N}\left(0, \frac{\kappa^2}{m_0}\right)$, $\ell = 1, 2, ..., m_0$, $\vartheta_{i\ell} \sim IID\mathcal{N}(0, \sigma_{\vartheta\ell}^2)$, for $\ell = 1, 2, ..., m_x$, with $\sigma_{\vartheta\ell}^2 = \sigma_{\mathrm{v}}^2/m_x$, 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 $i = 1, 2, ..., N; t = 0, 1, ..., T$ where χ_6^2 is a chi-square variate with 6 degrees of freedom and $\sigma^2 = 1$. The fixed effects are generated as $\alpha_i \sim IID\mathcal{N}(0,1)$. The regressors, x_{it}, for $i = 1, 2, ..., N$ are generated as $x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{i\ell} f_{\ell t} + \mathrm{v}_{it}$, with $\mathrm{v}_{it} = \rho_x \mathrm{v}_{i,t-1} + \left(1-\rho_x^2\right)^{1/2} \varepsilon_{it}$, for $t = 1, 2, ..., T$, $\varepsilon_{it} \sim IID\mathcal{N}(0, \sigma_{\mathrm{v}i}^2)$, $\mathrm{v}_{i0} \sim IID\mathcal{N}(0, \sigma_{\mathrm{v}i}^2)$, for $i = 1, 2, ..., N$, with $\sigma_{\mathrm{v}i}^2 \sim IID\ \frac{1}{4}(\chi_2^2 + 2)\sigma_{\mathrm{v}}^2$ and $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IID\mathcal{N}(0,1)$, 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) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0=0.4$			$\gamma_0=0.8$			$\gamma_0=0.4$			$\gamma_0=0.8$		
N	$m_0=0$											
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) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = \sigma_v^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
	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(×100) and RMSE(×100) of γ and β for QML and GMM estimators in the case of the ARX(1) model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

	Bias(×100)					RMSE(×100)					Bias(×100)					RMSE(×100)				
	QML	GMM				QML	GMM				QML	GMM				QML	GMM			
		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2
	$\gamma_0 = 0.4$										$\gamma_0 = 0.8$									
										$T = 5$										
	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
										$T = 10$										
	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 supplementary material. "-" 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 5: Size($\times$100) of γ and β for the QML and GMM estimators in the case of the ARX(1) model, using the true number of factors, m_0 ($\kappa^2 = \sigma_v^2 = 1$)

	$T=5$										$T=10$									
	QML	GMM				QML	GMM				QML	GMM				QML	GMM			
		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2
	$\gamma_0=0.4$					$\gamma_0=0.8$					$\gamma_0=0.4$					$\gamma_0=0.8$				
	γ																			
N	$m_0=1$																			
100	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 4.

Table 6: 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	99.9	100.0	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 $\Delta y_{it} = \Delta\delta_t + \gamma\Delta y_{i,t-1} + \beta\Delta x_{it} + \Delta\zeta_{it}$, $t = 2, 3, ..., T$, with $\Delta\zeta_{it} = \sum_{\ell=1}^{m_0} \eta_{\ell i}\Delta f_{\ell t} + \Delta u_{it} = \boldsymbol{\eta}_i'\Delta\mathbf{f}_t + \Delta u_{it}$, $\Delta y_{i1} = \Delta\delta_1 + \beta\Delta x_{i1} + \Delta\zeta_{i1}$ and $\Delta y_{i0} = 0$, for $i = 1, 2, ..., N$, 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 $u_{it} \sim IID\frac{\sigma}{\sqrt{12}}(\chi_6^2 - 6)$ for $i = 1, 2, ..., N; t = 0, 1, ..., T$ where χ_6^2 is a chi-square variate with 6 degrees of freedom and $\sigma^2 = 1$. The fixed effects are generated as $\alpha_i \sim IID\mathcal{N}(0, 1)$ and the factor loadings are specified as $\eta_{\ell i} \sim IID\mathcal{N}\left(0, \frac{\kappa^2}{m_0}\right)$, $\ell = 1, 2, ..., m_0$. The regressors, x_{it}, for $i = 1, 2, ..., N$ are generated as $x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{i\ell} f_{\ell t} + \mathrm{v}_{it}$, $\mathrm{v}_{it} = \rho_x \mathrm{v}_{i,t-1} + \left(1 - \rho_x^2\right)^{1/2} \varepsilon_{it}$, for $t = 1, 2, ..., T$, with $\rho_x = 0.95$, $m_x = 2$, $\vartheta_{i\ell} \sim IID\mathcal{N}(0, \sigma_{\vartheta\ell}^2)$, for $\ell = 1, 2, ..., m_x$, and $\sigma_{\vartheta\ell}^2 = \sigma_v^2/m_x$ for all $\ell, \varepsilon_{it} \sim IID\mathcal{N}(0, \sigma_{vi}^2)$, $\mathrm{v}_{i0} \sim IID\mathcal{N}(0, \sigma_{vi}^2)$, for $i = 1, 2, ..., N$, with $\sigma_{vi}^2 \sim IID\ \frac{1}{4}(\chi_2^2 + 2)\sigma_v^2$ and $\alpha_{xi} = \alpha_i + v_i$, where $v_i \sim IID\mathcal{N}(0, 1)$, for all i. The remaining parameters are generated as described in Section 7.1. 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 7: Bias($\times$100), RMSE($\times$100) and Size ($\times$100) of γ for the AR(1) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = 1$)

	$T=5$			$T=10$		
	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 6.

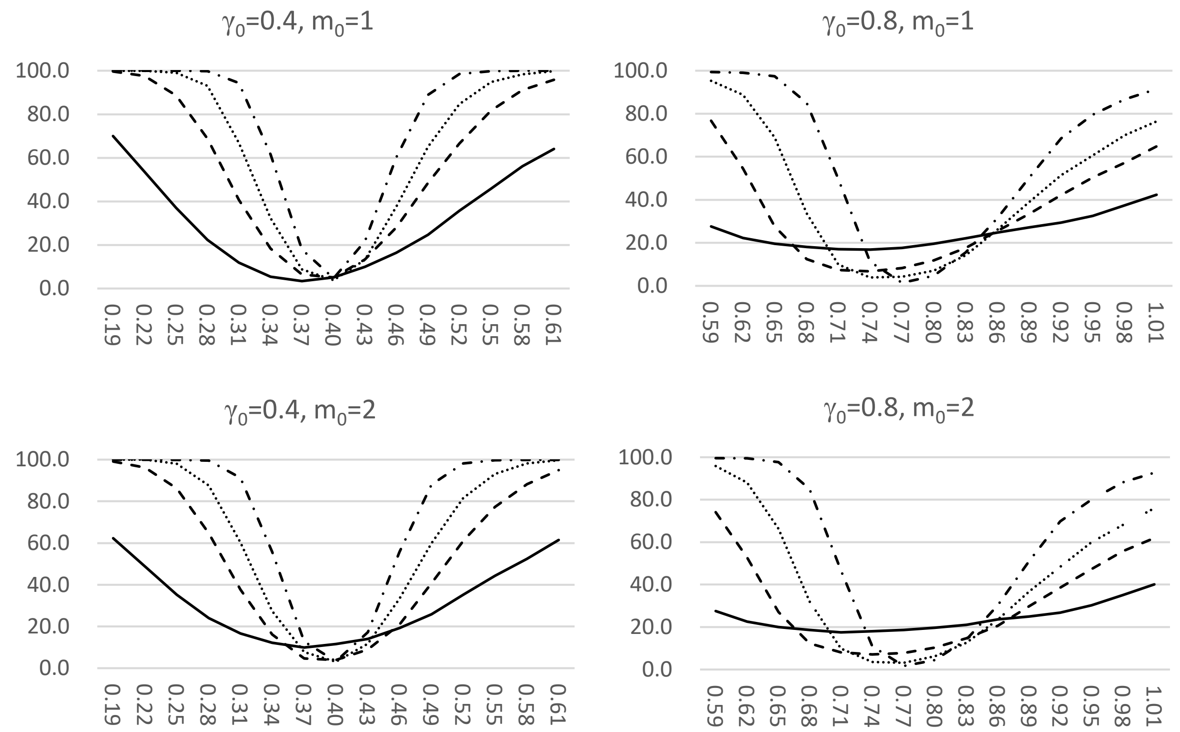
Table 8: Bias($\times$100), RMSE($\times$100) and Size($\times$100) of γ and β for the ARX(1) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = \sigma_{\mathrm{v}}^2 = 1$)

	$T=5$			$T=10$		
	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

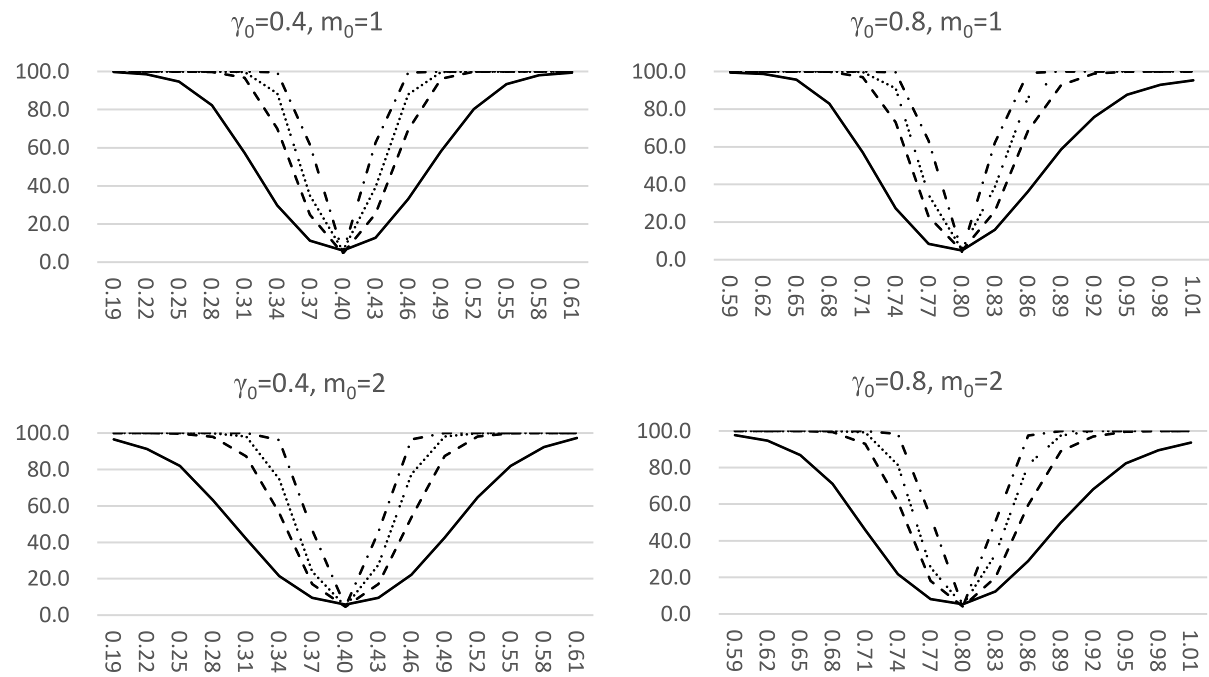
See the note to Table 6.

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

Panel A: T=5



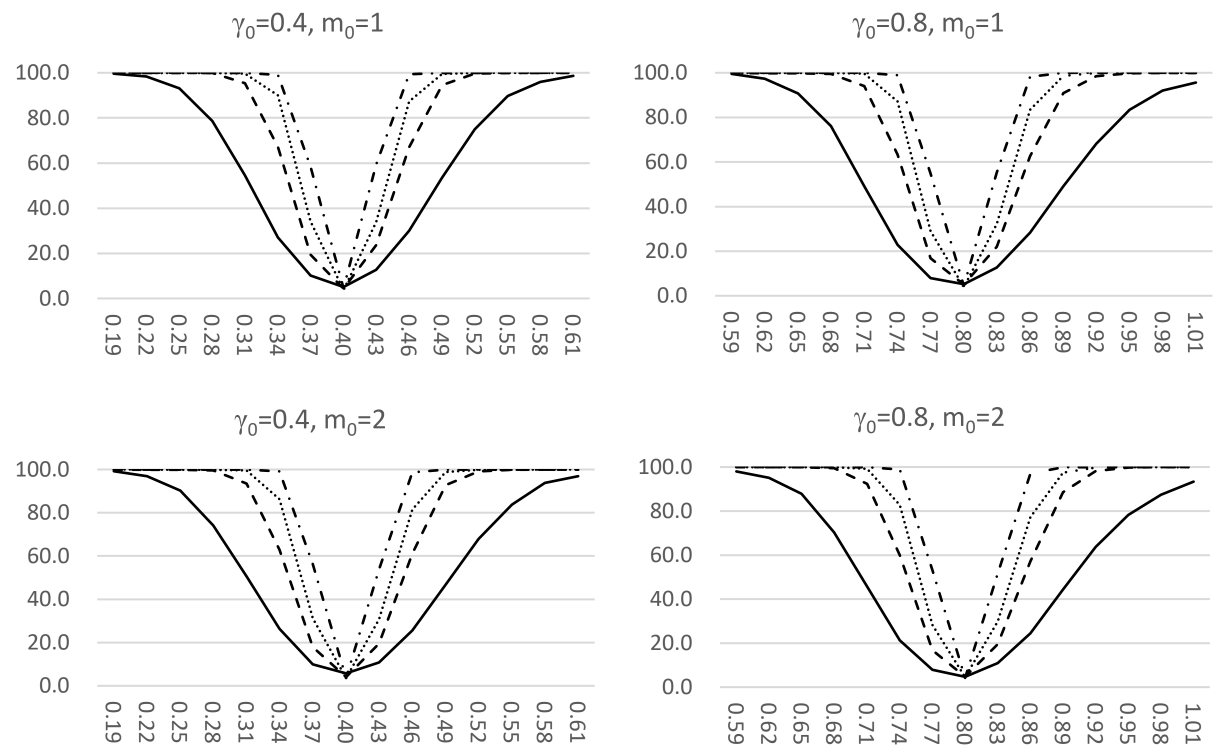
Panel B: T=10



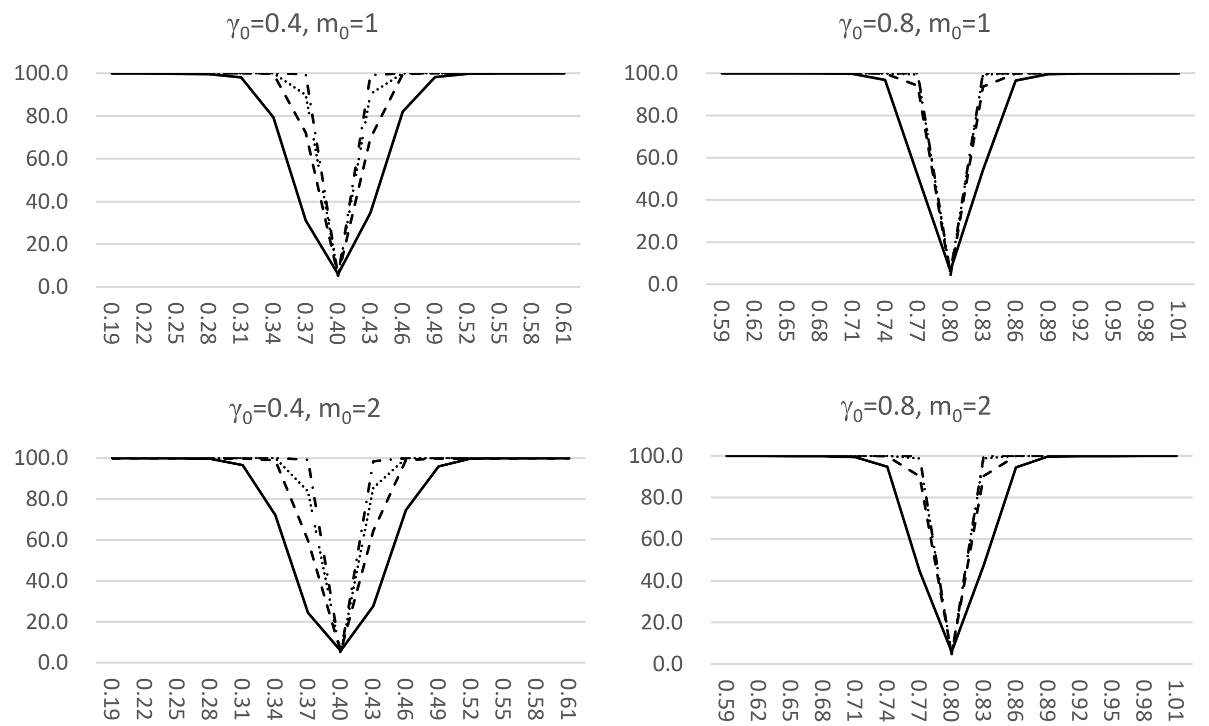
Note: —— N=100 - - - - N=300 ········· N=500 -·-·- N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with $\alpha_N=p/N(T-2)$ and p=0.05; γ is the coefficient of the lagged dependent variable given in (1) in the absence of the $\mathbf{x}_{it}$ regressors. See also the note to Table 1.

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

Panel A: T=5



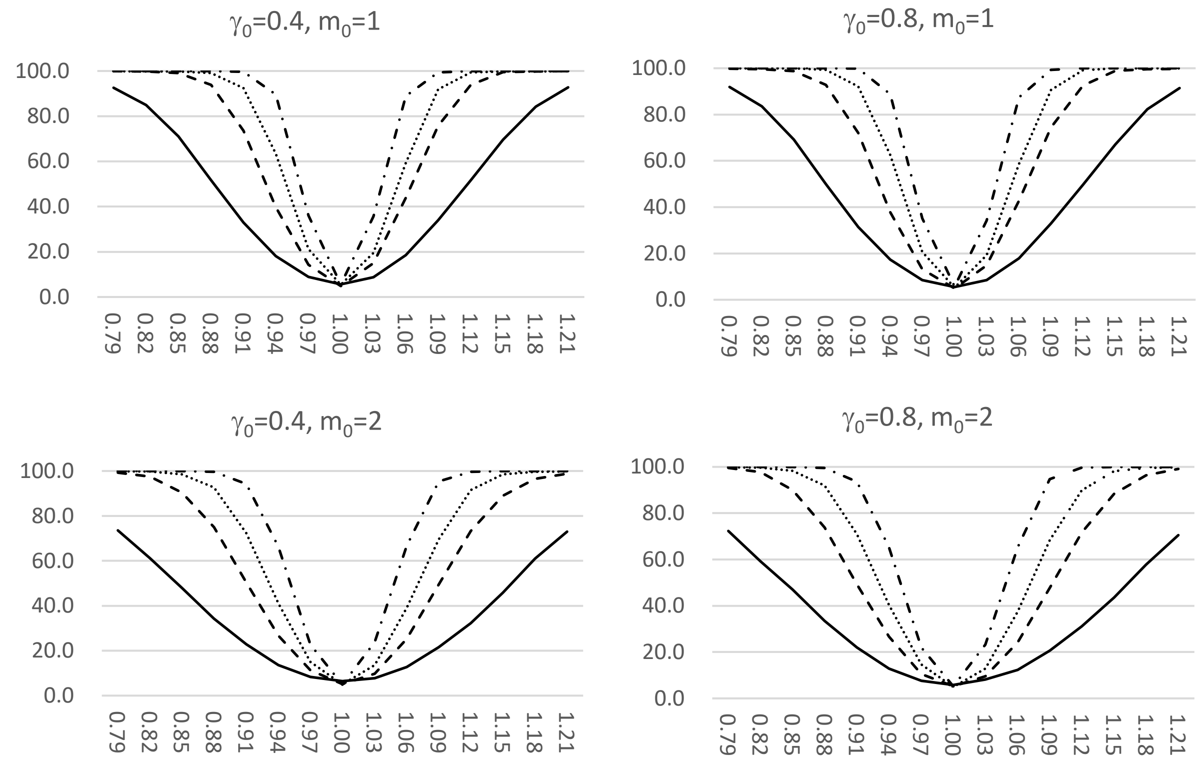
Panel B: T=10



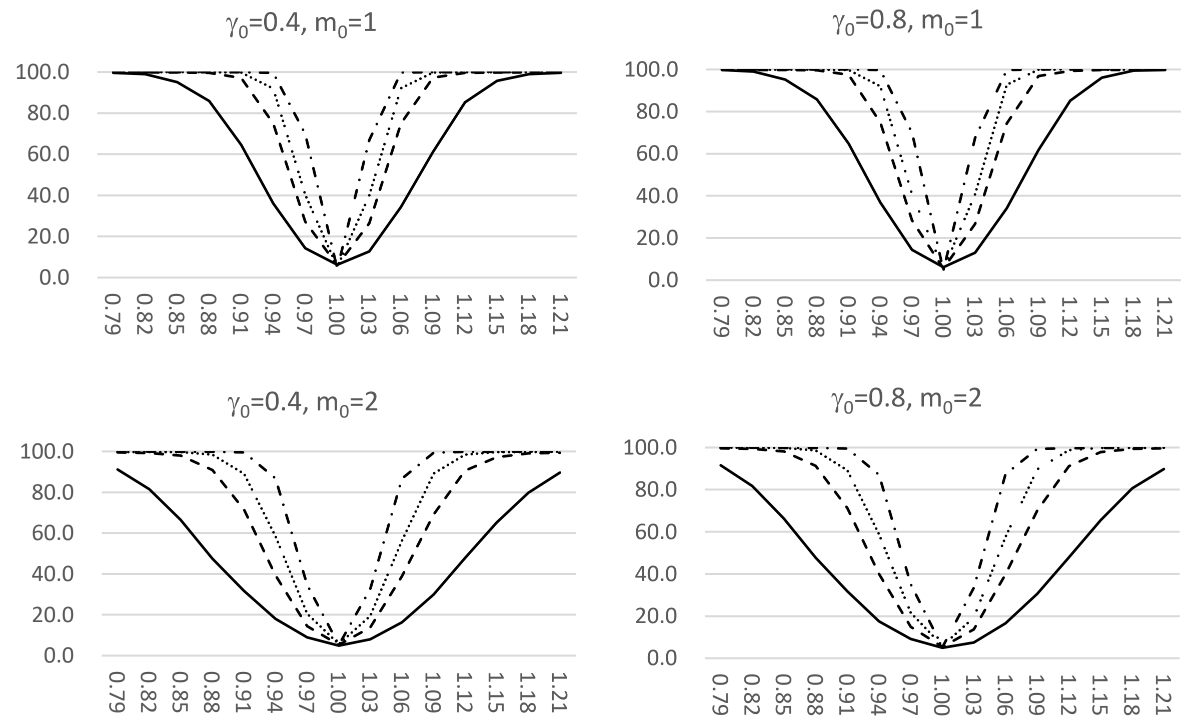
Note: N=100 N=300 N=500 N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with α_N=p/N(T-2) and p=0.05; γ and β are the coefficients of the lagged dependent variable and the x_{it} regressor given in (1). See also the note to Table 1.

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

Panel A: T=5



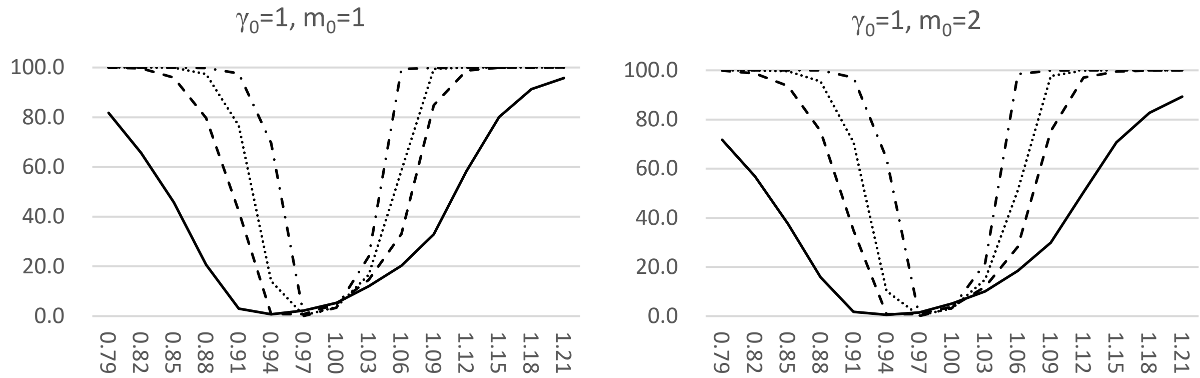
Panel B: T=10



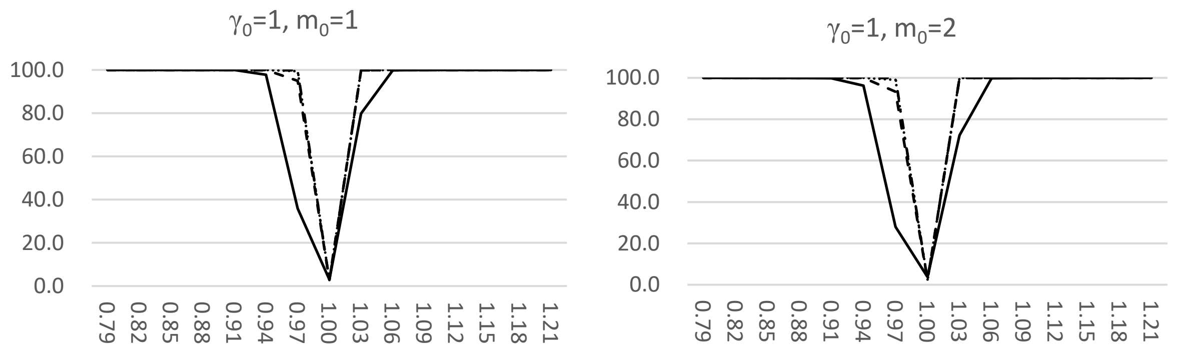
Note: —— N=100 - - - - N=300 ········ N=500 -·-·- N=1000. See also the note to Figure 2a.

Figure 3: Power functions for estimation of γ in the AR(1) model with different values of m and N

Panel A: T=5



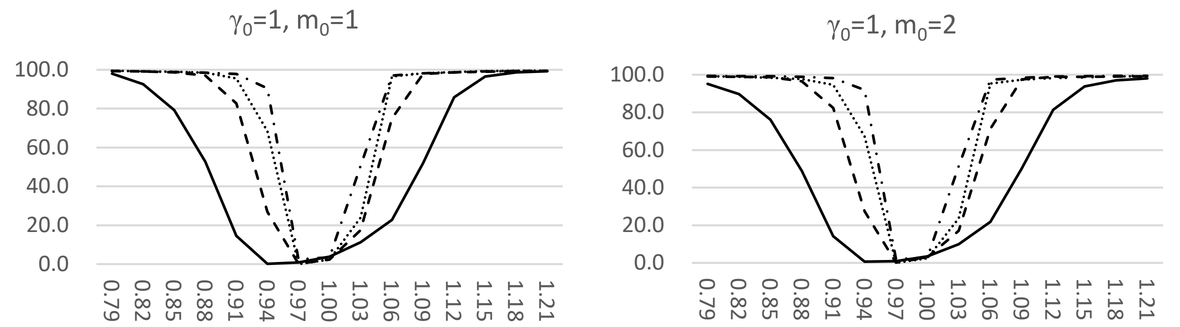
Panel B: T=10



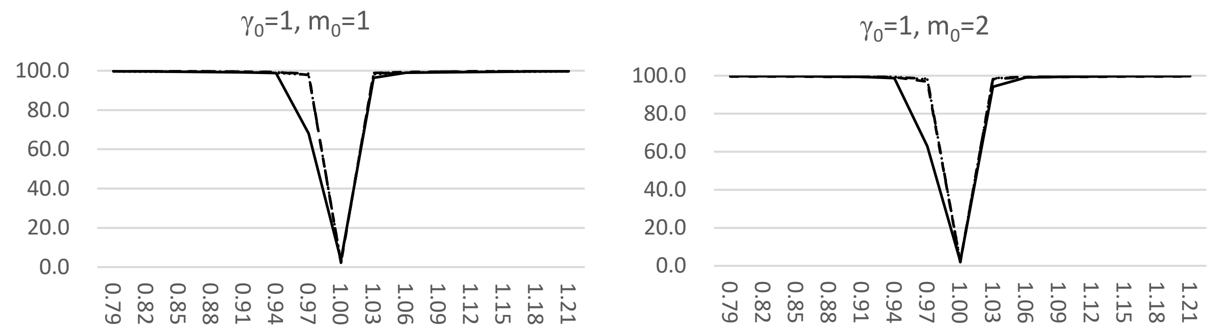
Note: N=100 N=300 N=500 N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with α_N=p/N(T-2) and p=0.05; γ is the coefficient of the lagged dependent variable given in (1) in the absence of the $\mathbf{x}_{it}$ regressors. See also the note to Table 4.

Figure 4a: Power functions for estimation of γ in the ARX(1) model with different values of m and N

Panel A: T=5



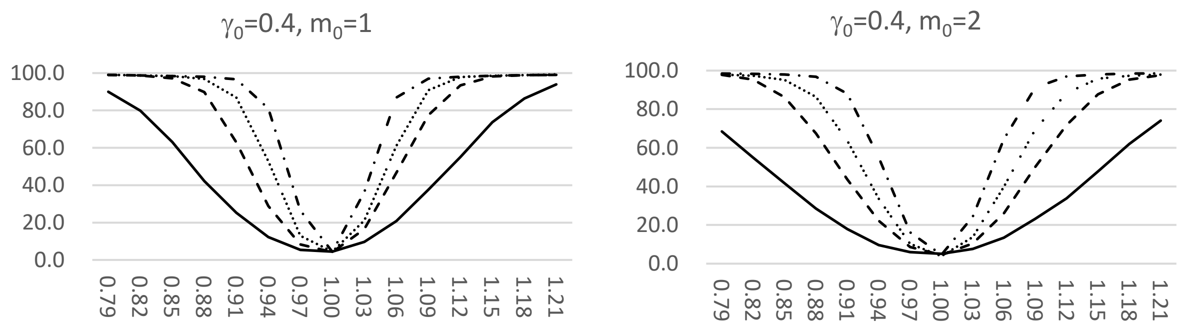
Panel B: T=10



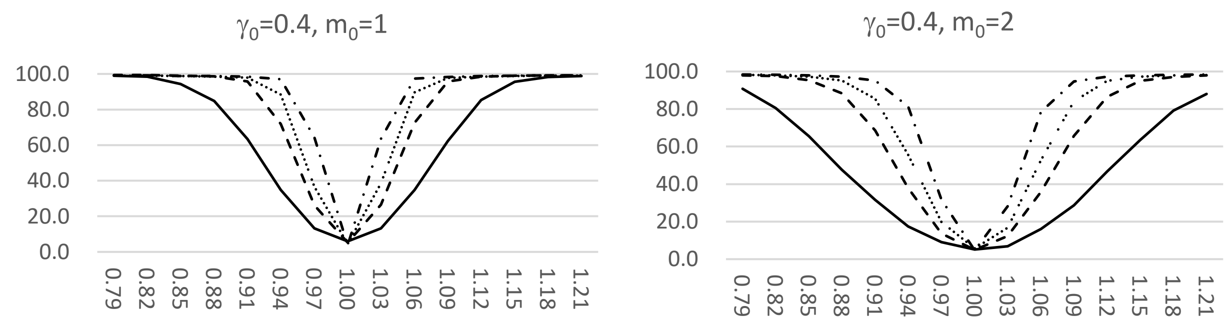
Note: —— N=100 - - - - N=300 ········· N=500 -·-·- N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with α_N=p/N(T-2) and p=0.05; γ and β are the coefficients of the lagged dependent variable and the x_{it} regressor given in (1). See also the note to Table 4.

Figure 4b: 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. See also the note to Figure 4a.

Appendix

A.1 Lemmas and their proofs

Lemma 1 *Consider the composite random variable, ξ_{it}, $i = 1, 2, ..., N$, for $t = 1$ defined by (14), and for $t = 2, 3, ..., T$ defined by (6). Then under Assumptions 1, 2, 3, 5, and 6, the following moment conditions hold:*

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

and

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

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

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

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 $t = 1$, $\xi_{i1} = \widetilde{\mathbf{g}}_1' \boldsymbol{\eta}_i + v_{i1}$, and $\|\widetilde{\mathbf{g}}_1\| < K$ and $\sup_i E|v_{i1}|^{4+\epsilon} < K$ (see (15) and related results). Hence, $\left(E\left|\xi_{it}\right|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}} \leq K$, for $t = 1, 2, ..., T$ and (A.1) follows as required. To establish condition (A.2), using (3) we first note that

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

and by the Minkowski inequality for infinite sums we have

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

for any $p \geq 1$. Set $p = 4 + \epsilon$, and note that under Assumption 5, $\sup_t \|\boldsymbol{\delta}_{x,t}\| < K, \sup_{j,t} |g_{x,jt}| < K$, $\sup_{i,j} E\left\|\boldsymbol{\eta}_{ij,x}\right\|^{4+\epsilon} < K$, $\sup_{i,t} E\|\boldsymbol{\varepsilon}_{it}\|^{4+\epsilon} < K$, 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 $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, ...\xi_{iT})'$, where ξ_{i1} is defined by (14) and ξ_{it}, for $t = 2, 3, ..., T$ are defined by (6). Suppose that the conditions of Lemma 1 hold and T is fixed. Then*

$$\sup_i E\|\boldsymbol{\xi}_i\|^4 < K < \infty, \tag{A.3}$$

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

Proof. To obtain (A.3) note that

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

Then by Minkowski's inequality we have

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

and since $\sup_i E(|\xi_{it}|^{4+\epsilon}) < K$ for $t = 1, 2, ..., T$ from result (A.1) of Lemma 1, result (A.3) follows noting that T is fixed. To establish (A.4), note that $\Delta\mathbf{W}_i = (\mathbf{I}_T, \mathbf{Z}_i, \Delta\widetilde{\mathbf{y}}_{i,-1}) = (\mathbf{I}_T, \mathbf{Z}_i, \mathbf{L}\Delta\mathbf{y}_i)$, where $\Delta\widetilde{\mathbf{y}}_{i,-1} = (0, \Delta y_{i1}, ..., \Delta y_{i,T-1})$, $\mathbf{Z}_i$ and $\Delta\mathbf{y}_i$ are given by (31) and (30), and

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

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

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

Also $\|\mathbf{Z}_i\|^2 = \|\Delta\mathbf{x}_{i1}\|^2 + 2\sum_{t=2}^{T}\|\Delta\mathbf{x}_{it}\|^2$, 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 $\sup_i E\|\mathbf{Z}_i\|^4 < K$. Similarly, using (30), we have

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

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

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

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

Lemma 3 *Consider the model given by (17) and let*

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

Define

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

and suppose that Assumptions 1-6 and parts (i)-(ii) of Assumption 7 hold. Then

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

where

$$\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0) = \operatorname{tr}\left\{\left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}) - \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)\right]\mathbf{C}(\boldsymbol{\psi}, \gamma_0)\right\} \tag{A.8}$$

and

$$\mathbf{C}\left(\boldsymbol{\psi},\gamma_0\right)=\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\begin{pmatrix} 0 & 0 & \cdots & 0 & 0\\ 1 & 0 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \gamma_0^{T-3} & \gamma_0^{T-4} & \cdots & 0 & 0\\ \gamma_0^{T-2} & \gamma_0^{T-3} & \cdots & 1 & 0 \end{pmatrix}. \tag{A.9}$$

Furthermore

$$E_0\left[\mathbf{d}_i(\boldsymbol{\psi}_0,\boldsymbol{\varphi}_0)\right]=\mathbf{0},\ \textit{for } i=1,2,...,N, \tag{A.10}$$

$$\mathbf{b}_N\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)=\frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)\overset{a.s.}{\to}\mathbf{b}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)=\left[\mathbf{0},\mathbf{0},-\kappa\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right)\right]', \tag{A.11}$$

$$\mathbf{b}_N\left(\boldsymbol{\psi}_0,\boldsymbol{\varphi}_0\right)=\frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\overset{a.s.}{\to}\mathbf{0}, \tag{A.12}$$

and

$$\boldsymbol{\Sigma}_{N,\xi}\left(\boldsymbol{\psi}_0\right)=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)'\overset{a.s.}{\to}\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right). \tag{A.13}$$

Proof. Under (17),

$$\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)=\Delta\mathbf{y}_i-\Delta\mathbf{W}_i\boldsymbol{\varphi}_0=\mathbf{G}_0\boldsymbol{\eta}_{0i}+\mathbf{r}_{0i}, \tag{A.14}$$

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 $\boldsymbol{\psi}=\boldsymbol{\psi}_0$. It is now easily seen that $E_0\left[\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\right]=\mathbf{0}$, and $Var\left[\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\right]=E_0\left[\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}_0\right)\right]=\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right)$. Also under Assumptions 1-6, $\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}\right)=\mathbf{G}\boldsymbol{\eta}_i+\mathbf{r}_i$ 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 $\Delta\mathbf{W}_i=(\mathbf{I}_T,\mathbf{Z}_i,\Delta\widetilde{\mathbf{y}}_{i,-1})$, where $\mathbf{I}_T$ is the identity matrix of order T,

$$\mathbf{Z}_i=\begin{pmatrix} \Delta\mathbf{x}_i' & 0\\ \mathbf{0} & \Delta\mathbf{x}_{i2}'\\ \vdots & \vdots\\ \mathbf{0} & \Delta\mathbf{x}_{iT}' \end{pmatrix},\ \Delta\widetilde{\mathbf{y}}_{i,-1}=\begin{pmatrix} 0\\ \Delta y_{i1}\\ \vdots\\ \Delta y_{i,T-1} \end{pmatrix},$$

and note that $\Delta\widetilde{\mathbf{y}}_{i,-1}=\mathbf{L}\Delta\mathbf{y}_i$, where $\mathbf{L}$ is given by (A.5). Also, using (30) and evaluating it at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ we have

$$\Delta\mathbf{y}_i=\mathbf{B}\left(\gamma_0\right)^{-1}\left(\mathbf{Z}_i\boldsymbol{\delta}_0+\mathbf{d}_0\right)+\mathbf{B}\left(\gamma_0\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right), \tag{A.15}$$

where $\boldsymbol{\delta}=\left(\boldsymbol{\pi}',\boldsymbol{\beta}'\right)'$, and $\mathbf{B}\left(\gamma\right)$ is defined by (26). Consider now (A.6), and note that

$$\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)=\boldsymbol{\Delta}\mathbf{W}_i'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)=\begin{pmatrix} \boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\\ \mathbf{Z}_i'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\\ \boldsymbol{\Delta}\mathbf{y}_i'\mathbf{L}'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right) \end{pmatrix}=\begin{pmatrix} \mathbf{d}_{1i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)\\ \mathbf{d}_{2i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)\\ d_{3i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right) \end{pmatrix}. \tag{A.16}$$

Further, using (A.15), write $d_{3i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)$ as

$$\begin{aligned} d_{3i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right) &= \left[\mathbf{B}\left(\gamma_0\right)^{-1}\left(\mathbf{Z}_i\boldsymbol{\delta}_0+\mathbf{d}_0\right)+\mathbf{B}\left(\gamma_0\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\right]'\mathbf{L}'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right) \\ &= \left(\mathbf{Z}_i\boldsymbol{\delta}_0+\mathbf{d}_0\right)'\mathbf{B}\left(\gamma_0\right)'^{-1}\mathbf{L}'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)+\boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}_0\right)\mathbf{B}\left(\gamma_0\right)'^{-1}\mathbf{L}'\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right). \end{aligned} \tag{A.17}$$

Also under Assumptions 1, 3, and 5, $\mathbf{Z}_i$ and $\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)$ are cross-sectionally independently distributed, and $E_0\left[\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\right]=\mathbf{0}$. Hence

$$E_0\left[\mathbf{d}_{1i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)\right]=\mathbf{0},\text{ and }E_0\left[\mathbf{d}_{2i}\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right)\right]=\mathbf{0},\text{ for all }i, \tag{A.18}$$

and

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

Also, using (27) and (A.5), we have

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

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

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

where $\mathbf{C}(\boldsymbol{\psi},\gamma_0) = \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}\,\mathbf{LB}(\gamma_0)^{-1}$. Using (A.19) and (A.18) now yields (A.7), as required. Result (A.10) then follows immediately, noting that $E_0\left[d_{3i}(\boldsymbol{\psi}_0,\boldsymbol{\varphi}_0)\right] = \operatorname{tr}\left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)\,\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1}\,\mathbf{LB}(\gamma_0)^{-1}\right] = \operatorname{tr}\left[\mathbf{LB}(\gamma_0)^{-1}\right] = 0$. To establish (A.11), we first note that $\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)$, for $i = 1, 2, ..., N$ are independently distributed, and therefore conditional on $\mathbf{Z}_i$, $\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)$ 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 $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E_0\left[\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)\right]$, it is sufficient to show that $\sup_i E_0\left\|\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)\right\|^2 < K$. 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 $E\left\|\xi_i\right\|^2 < K < \infty$ and noting that by assumption 7(ii) $\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})^{-1}$ is positive definite for all $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$, then

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

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

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

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

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

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

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

$$\sup_i E_0\left\|\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)\right\|^2 = \sup_i E_0\left\|\Delta\mathbf{W}_i'\boldsymbol{\Sigma}_{\xi}\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\right\|^2 < K,$$

which establishes that $\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)$ is uniformly L_2-bounded, besides being cross-sectionally independent. Hence,

$$\mathbf{b}_N\left(\boldsymbol{\psi},\boldsymbol{\varphi}_0\right) = N^{-1}\sum_{i=1}^{N}\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0) \overset{a.s.}{\to} \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}E_0\left[\mathbf{d}_i(\boldsymbol{\psi},\boldsymbol{\varphi}_0)\right] = \left[\mathbf{0},\mathbf{0},-\kappa\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right)\right]',$$

which establishes (A.11). Result (A.12) follows from the above by setting $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ and noting from (A.10) that $E_0\left[\mathbf{d}_i(\boldsymbol{\psi}_0,\boldsymbol{\varphi}_0)\right] = \mathbf{0}$. Finally, since $\sup_i E_0\left\|\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}_0\right)\right\|^2 < K$, for a finite T (see result (A.3) of Lemma 2), and by assumption $\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}_0\right)$ are distributed independently over i, then

$$\boldsymbol{\Sigma}_{N,\xi}\left(\boldsymbol{\psi}_0\right) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)' \overset{a.s.}{\to} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E_0\left[\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)'\right],$$

and result (A.13) follows, since $E_0\left[\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}_0\right)\right] = \boldsymbol{\Sigma}_{\xi}\left(\boldsymbol{\psi}_0\right)$. ∎

Lemma 4 *Consider the average log-likelihood function*

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

$\bar{\ell}_N\left(\boldsymbol{\theta}\right) = N^{-1}\ell_N\left(\boldsymbol{\theta}\right)$ *and* $\ell_N\left(\boldsymbol{\theta}\right)$ *is defined by (22). Then under Assumptions 1-7 we have*

$$\bar{\ell}_N\left(\boldsymbol{\theta}_0\right) \overset{a.s.}{\to} -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\xi}\left(\boldsymbol{\psi}_0\right)\right| - \frac{T}{2}, \tag{A.23}$$

and

$$\begin{aligned}\bar{\ell}_N\left(\boldsymbol{\theta}\right) \overset{a.s.}{\to}& -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_{\xi}\left(\boldsymbol{\psi}\right)\right| - \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_{\xi}\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\Sigma}_{\xi}\left(\boldsymbol{\psi}_0\right)\right] \\ &-\frac{1}{2}\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right)'\mathbf{A}\left(\boldsymbol{\psi}\right)\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right) - \left(\gamma-\gamma_0\right)\kappa\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right),\end{aligned} \tag{A.24}$$

where $\kappa\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right)$ *is defined by (A.8). Also*

$$\bar{\ell}_N\left(\boldsymbol{\theta}_0\right) - \bar{\ell}_N\left(\boldsymbol{\theta}\right) \overset{a.s.}{\to} \lim_{N\to\infty}E_0\left[\bar{\ell}_N\left(\boldsymbol{\theta}_0\right) - \bar{\ell}_N\left(\boldsymbol{\theta}\right)\right] \geq 0, \tag{A.25}$$

where

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

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

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

where

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

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). Result (A.24) 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), and the Kullback–Leibler type information inequality. ■

Lemma 5 *Consider the average log-likelihood function defined by (32) and (24):*

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

and suppose that Assumptions 1 to 6 and parts (i)-(ii) of Assumption 7 hold. Denote the average score function by $\bar{\mathbf{s}}_N(\boldsymbol{\theta}) = \partial\bar{\ell}_N(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$*. Then*

$$\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \overset{a.s}{\to} \mathbf{0}, \tag{A.29}$$

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

where

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

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

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

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

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

$$\widehat{\mathbf{J}}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\omega}_i\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{\omega}_i'\left(\widehat{\boldsymbol{\theta}}\right), \tag{A.34}$$

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

Proof. Let $\bar{\mathbf{s}}_N(\boldsymbol{\theta}) = \left(\bar{\mathbf{s}}'_{N,\varphi}(\boldsymbol{\theta}), \bar{\mathbf{s}}'_{N,\psi}(\boldsymbol{\theta})\right)'$, $\boldsymbol{\psi} = \left(\psi_1, \psi_2, ..., \psi_{n_\psi}\right)'$, 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 $j = 1, 2, ..., n_\psi$. Using (A.6), and result (A.12) of Lemma 3, it then readily follows that

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

Also

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

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

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

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

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

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

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

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

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

where $C = \mathrm{tr}\left[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)^{-1} \frac{\partial \boldsymbol{\Sigma}_\xi(\boldsymbol{\psi}_0)}{\partial \psi_j}\right]$. 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) $\sup_i E\left\|\boldsymbol{\xi}_i(\boldsymbol{\varphi}_0)\right\|^4 < K$. Therefore,

$\sup_i E \left\|\boldsymbol{\nu}_i\left(\boldsymbol{\theta}_0\right)\right\|^2 < K$. 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 $\bar{\mathbf{s}}_{N,\psi}(\boldsymbol{\theta}_0) \overset{a.s}{\rightarrow} \mathbf{0}$, and together with (A.35) yields $\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \overset{a.s}{\rightarrow} \mathbf{0}$, as required. Consider now the limiting distribution of $\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0)$ and note that

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

where $\boldsymbol{\omega}_i\left(\boldsymbol{\theta}_0\right) = \left(\mathbf{d}_i'(\boldsymbol{\theta}_0), \boldsymbol{\nu}_i'\left(\boldsymbol{\theta}_0\right)\right)'$, and it is already established that $\boldsymbol{\omega}_i\left(\boldsymbol{\theta}_0\right)$ 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 have[11] $\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \rightarrow_d N\left[\mathbf{0}, \mathbf{J}\left(\boldsymbol{\theta}_0\right)\right]$,where $\mathbf{J}\left(\boldsymbol{\theta}_0\right)$ is given by (A.31), as required. Consistency of $\widehat{\mathbf{J}}\left(\widehat{\boldsymbol{\theta}}\right)$ for $\mathbf{J}\left(\boldsymbol{\theta}_0\right)$ follows from consistency of $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, and the independence of $\boldsymbol{\omega}_i\left(\boldsymbol{\theta}_0\right)$ over i. ■

A.2 Proofs of Propositions and Theorems

Proof of Proposition 1.

Recall that $\boldsymbol{\theta} = \left(\boldsymbol{\varphi}', \boldsymbol{\psi}'\right)'$, and $\boldsymbol{\varphi} = \left(\boldsymbol{\lambda}', \gamma\right)'$, and using (32) note that

$$\bar{\ell}_N\left(\boldsymbol{\lambda}, \gamma, \boldsymbol{\psi}\right) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)\right| - \frac{1}{2N}\sum_{i=1}^N \boldsymbol{\xi}_i'(\boldsymbol{\lambda}, \gamma)\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\xi}_i(\boldsymbol{\lambda}, \gamma).$$

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

$$\bar{\ell}_N\left(\boldsymbol{\lambda}_0, \gamma_0, \boldsymbol{\psi}_0\right) - \bar{\ell}_N\left(\boldsymbol{\lambda}, \gamma, \boldsymbol{\psi}\right) \overset{a.s.}{\rightarrow} \lim_{N\rightarrow\infty} E_0\left[\bar{\ell}_N\left(\boldsymbol{\lambda}_0, \gamma_0, \boldsymbol{\psi}_0\right) - \bar{\ell}_N\left(\boldsymbol{\lambda}, \gamma, \boldsymbol{\psi}\right)\right] \geq 0, \tag{A.37}$$

$$2\lim_{N\rightarrow\infty} E_0\left[\bar{\ell}_N\left(\boldsymbol{\lambda}_0, \gamma_0, \boldsymbol{\psi}_0\right) - \bar{\ell}_N\left(\boldsymbol{\lambda}, \gamma, \boldsymbol{\psi}\right)\right] = \chi\left(\boldsymbol{\psi}, \boldsymbol{\psi}_0\right) + \left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\right)'\mathbf{A}\left(\boldsymbol{\psi}\right)\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\right) + 2\left(\gamma - \gamma_0\right)\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}_0), \tag{A.38}$$

where

$$\chi\left(\boldsymbol{\psi}, \boldsymbol{\psi}_0\right) = \text{tr}\left[\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right)\right] - \ln\left(\left|\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right)\right| / \left|\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)\right|\right) - T, \tag{A.39}$$

and

$$\kappa\left(\boldsymbol{\psi}, \boldsymbol{\psi}_0\right) = \text{tr}\left\{\left[\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right) - \boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right)\right]\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)^{-1}\mathbf{L}\mathbf{B}\left(\gamma_0\right)^{-1}\right\}, \tag{A.40}$$

with $\mathbf{B}\left(\gamma_0\right)^{-1}$ given by (27) evaluated at γ_0, $\mathbf{L}$ is a matrix lag operator defined by (A.5) and $\mathbf{A}\left(\boldsymbol{\psi}\right)$ is defined by (33). Denote the eigenvalues of $\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_0\right)$ and $\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)$ by λ_{0t} and λ_t ($t = 1, 2, ..., T$), respectively (note that $\lambda_{0t} > 0$ and $\lambda_t > 0$) and write $\chi\left(\boldsymbol{\psi}, \boldsymbol{\psi}_0\right)$ as

$$\chi\left(\boldsymbol{\psi}, \boldsymbol{\psi}_0\right) = \sum_{t=1}^T \left[\left(\lambda_{0t}/\lambda_t\right) - \ln\left(\lambda_{0t}/\lambda_t\right) - 1\right].$$

Also note that $\left(\lambda_{0t}/\lambda_t\right) - \ln\left(\lambda_{0t}/\lambda_t\right) - 1 \geq 0$ with the equality holding if and only if $\lambda_{0t} = \lambda_t$, for all t, or equivalently if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$.[12] Therefore, $\chi\left(\boldsymbol{\psi}, \boldsymbol{\psi}_0\right) \geq 0$, with equality holding if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Furthermore, since by Assumption 7 (iii) $\mathbf{A}\left(\boldsymbol{\psi}\right)$ is a positive definite matrix, then

$$\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\right)'\mathbf{A}\left(\boldsymbol{\psi}\right)\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\right) \geq \lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}\right)\right]\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\right)'\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\right),$$

[11] See, for example, White (2001, Theorem 5.10).

[12] Note that for any $x > 0$, $\ln(x) \leq x - 1$. Here $x = \lambda_{0t}/\lambda_t > 0$.

where $\lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}\right)\right] > 0$. It is clear that the first two terms of (A.38) can not be negative, but the same is not true of the third term, $\left(\gamma-\gamma_0\right)\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)$, and therefore, global identification of γ_0 can not be guaranteed. Consider now the almost sure probability limit of $\bar{\ell}_N\left(\boldsymbol{\varphi}_0,\boldsymbol{\psi}_0\right)-\bar{\ell}_N\left(\boldsymbol{\varphi},\boldsymbol{\psi}\right)$ on the set $\boldsymbol{\Theta}_\epsilon=\boldsymbol{\Theta}_\varphi\times\mathcal{N}_\epsilon(\gamma_0)\times\boldsymbol{\Theta}_\psi$, for some small positive ϵ, where $\mathcal{N}_\epsilon(\gamma_0)$ is defined by Definition 1. We now establish that there exists $\epsilon>0$ for which this limit can be zero if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. To see this consider the first and the third terms of (A.38) together, and note that $\chi\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right)+2\left(\gamma-\gamma_0\right)\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)=0$ if $\boldsymbol{\psi}=\boldsymbol{\psi}_0$. In such a case

$$2\lim_{N\to\infty}E_0\left[\bar{\ell}_N\left(\boldsymbol{\varphi}_0,\boldsymbol{\psi}_0\right)-\bar{\ell}_N\left(\boldsymbol{\varphi},\boldsymbol{\psi}_0\right)\right]\geq\frac{1}{2}\lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}_0\right)\right]\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right)'\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right),$$

and $\bar{\ell}_N\left(\boldsymbol{\varphi}_0,\boldsymbol{\psi}_0\right)-\bar{\ell}_N\left(\boldsymbol{\varphi},\boldsymbol{\psi}_0\right)\overset{a.s.}{\to}0$, if and only if $\lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}_0\right)\right]\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right)'\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right)=0$, which implies $\boldsymbol{\varphi}=\boldsymbol{\varphi}_0$, as required since by assumption $\lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}_0\right)\right]>0$. Consider now the case where $\boldsymbol{\psi}\neq\boldsymbol{\psi}_0$, and note that $\chi\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right)>0$, and $\left|\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)\right|>0$, and therefore on $\boldsymbol{\Theta}_\epsilon$ we have

$$\left|\left(\gamma-\gamma_0\right)\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)\right|\leq\left|\left(\gamma-\gamma_0\right)\right|\left|\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)\right|<\epsilon\left|\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)\right|.$$

Also note that under Assumptions 1, 2 and 3, $\left\|\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}\right)\right\|<K$ for all $\boldsymbol{\psi}\in\boldsymbol{\Theta}_\psi$, and it is readily seen that $\left|\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)\right|<K$. Hence, on $\boldsymbol{\Theta}_\epsilon$ there must exist $\epsilon>0$, such that $\chi\left(\boldsymbol{\psi},\boldsymbol{\psi}_0\right)+2\left(\gamma-\gamma_0\right)\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0)\geq0$, and hence

$$2\lim_{N\to\infty}E_0\left[\bar{\ell}_N\left(\boldsymbol{\varphi}_0,\boldsymbol{\psi}_0\right)-\bar{\ell}_N\left(\boldsymbol{\varphi},\boldsymbol{\psi}\right)\right]\geq\frac{1}{2}\lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}\right)\right]\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right)'\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right).$$

Once again since by assumption $\lambda_{\min}\left[\mathbf{A}\left(\boldsymbol{\psi}\right)\right]>0$ for all values of $\boldsymbol{\psi}\in\boldsymbol{\Theta}_\psi$, then on $\boldsymbol{\Theta}_\epsilon$ there exists $\epsilon>0$ such that $\boldsymbol{\varphi}=\boldsymbol{\varphi}_0$, and hence $\boldsymbol{\psi}=\boldsymbol{\psi}_0$, almost surely. ■

Proof of Theorem 1. For the proof of consistency it suffices to show here that under the assumptions of the theorem, $\bar{C}_N\left(\boldsymbol{\theta}\right)=-2\bar{\ell}_N\left(\boldsymbol{\theta}\right)\overset{a.s.}{\to}\bar{C}\left(\boldsymbol{\theta}\right)$ uniformly on $\boldsymbol{\Theta}_\epsilon$ (see Section 5). From results in Lemma 4 (see (A.25) and (A.26)) it follows that $\bar{C}_N\left(\boldsymbol{\theta}\right)=-2\bar{\ell}_N\left(\boldsymbol{\theta}\right)\overset{a.s.}{\to}\bar{C}\left(\boldsymbol{\theta}\right)$ for every $\boldsymbol{\theta}\in\boldsymbol{\Theta}_\epsilon$, where

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

and

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

and the term $C\left(\boldsymbol{\psi}_0\right)$ does not depend on $\boldsymbol{\theta}$. Since under our assumptions $\bar{\ell}_N\left(\boldsymbol{\theta}\right)$ is continuous in $\boldsymbol{\theta}$, this pointwise result holds uniformly on $\boldsymbol{\Theta}_\epsilon$ by the uniform law of large numbers, so long as the dominance condition

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

holds; see for example Pötscher and Prucha (2001, Theorem 23). Since T is finite, it is sufficient to show that

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

We have

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

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

$$\begin{aligned}\sup_{\psi\in\boldsymbol{\Theta}_\psi}|\ln|\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})|| &\leq \sup_{\psi\in\boldsymbol{\Theta}_\psi}\mathrm{tr}[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})]-T\\ &\leq \sup_{\psi\in\boldsymbol{\Theta}_\psi}\left(\textstyle\sum_{t=1}^T\lambda_t[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})]\right)-T\\ &\leq T\sup_{\psi\in\boldsymbol{\Theta}_\psi}(\lambda_{\max}[\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})])-T\leq T(c_{\max}-1)<\infty.\end{aligned}$$

For the first term, defining $\boldsymbol{\Theta}_\varphi=\boldsymbol{\Theta}_\lambda\times\mathcal{N}_\epsilon(\gamma_0)$, we have

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

Further

$$\begin{aligned}E\sup_{\varphi\in\boldsymbol{\Theta}_\varphi}\left\|\boldsymbol{\xi}_i(\boldsymbol{\varphi})\right\|^2 &= E\sup_{\varphi\in\boldsymbol{\Theta}_\varphi}\left\|\Delta\mathbf{y}_i-\Delta\mathbf{W}_i\boldsymbol{\varphi}\right\|^2\\ &\leq E\left\|\Delta\mathbf{y}_i\right\|^2+E\left\|\Delta\mathbf{W}_i\right\|^2\sup_{\varphi\in\boldsymbol{\Theta}_\varphi}\left\|\boldsymbol{\varphi}\right\|^2.\end{aligned}$$

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

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

$$\bar{\ell}_N(\boldsymbol{\theta})-\bar{\ell}_N(\boldsymbol{\theta}_0)=(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\,\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0)-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\,\mathbf{H}_N(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}_0),\tag{A.41}$$

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 (A.37) and (A.38) we have

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

Hence, in view of (A.41) we must also have

$$(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\,\mathbf{H}_N(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\overset{a.s.}{\to}\chi(\boldsymbol{\psi},\boldsymbol{\psi}_0)+(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)'\,\mathbf{A}(\boldsymbol{\psi})(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)+2(\gamma-\gamma_0)\,\kappa(\boldsymbol{\psi},\boldsymbol{\psi}_0).\tag{A.42}$$

But it has aleardy been established by Proposition 1 that on $\boldsymbol{\Theta}_\epsilon$ the right hand side of (A.42) can be equal to zero if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, and hence we must also have

$$\mathbf{H}_N(\bar{\boldsymbol{\theta}})\overset{a.s.}{\to}\mathbf{H}(\boldsymbol{\theta}_0),\tag{A.43}$$

where $\mathbf{H}(\boldsymbol{\theta}_0)$ *must* be a positive definite matrix given by $\mathbf{H}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} E_0\left[-\partial^2\bar{\ell}_N(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right]$. Applying the mean value theorem to $\bar{\mathbf{s}}_N(\widehat{\boldsymbol{\theta}})$ around $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ we have

$$\mathbf{0} = \sqrt{N}\bar{\mathbf{s}}_N(\widehat{\boldsymbol{\theta}}) = \sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) - \mathbf{H}_N(\breve{\boldsymbol{\theta}})\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

where $\breve{\boldsymbol{\theta}}$ lies on a line segment joining $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Then,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{H}_N^{-1}(\breve{\boldsymbol{\theta}})\left[\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0)\right].$$

Since $\breve{\boldsymbol{\theta}}$ lies between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$ is almost surely locally consistent for $\boldsymbol{\theta}_0$ on the set $\boldsymbol{\Theta}_\epsilon$ so is $\breve{\boldsymbol{\theta}}$, and as in (A.43) above $\mathbf{H}_N(\breve{\boldsymbol{\theta}}) \overset{a.s.}{\to} \mathbf{H}(\boldsymbol{\theta}_0)$. In addition, using result (A.30) of Lemma 5, we have $\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \to_d N[\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0)]$, where $\mathbf{J}(\boldsymbol{\theta}_0)$ is given by (A.31). Hence

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d N(\mathbf{0}, \mathbf{V}_\theta).$$

where $\mathbf{V}_\theta$ has the familiar sandwich form

$$\mathbf{V}_\theta = \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\theta}_0)\mathbf{H}^{-1}(\boldsymbol{\theta}_0).$$

■

Proof of Theorem 2. Denote the exactly identified estimator of $\boldsymbol{\theta}$ (under H_1) by $\widehat{\boldsymbol{\theta}}_{m_{\max}}$ with its dimension $n_\theta^* = 3 + T(k+1) + k + (T-2)(T+3)/2$, and the constrained estimator of $\boldsymbol{\theta}$ under $H_0 : m = m_0 < T-2$ by $\widehat{\boldsymbol{\theta}}_{m_0}$. The latter estimator is obtained under $\mathbf{q}_0(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{q}_0(\boldsymbol{\theta})$ is the $r_0 \times 1$ vector of restrictions on $\ell_N(\boldsymbol{\theta})$, the log-likelihood function defined by (22), implied by setting $m = m_0$. Since $\widehat{\boldsymbol{\theta}}_{m_0}$ is the constrained estimator of $\boldsymbol{\theta}$ under $H_0 : \mathbf{q}_0(\boldsymbol{\theta}) = \mathbf{0}$, by using the results from constrained optimization (see, for example, Davidson (2000, pp.289-290)), we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_0} - \boldsymbol{\theta}_0\right) \overset{a}{\sim} -\mathbf{F}_0\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \tag{A.44}$$

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

$$\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \overset{d}{\to} N(\mathbf{0}, \mathbf{J}_0) \tag{A.45}$$

and

$$\mathbf{F}_0 = \mathbf{H}_0^{-1} - \mathbf{H}_0^{-1}\mathbf{Q}_0'\left(\mathbf{Q}_0\mathbf{H}_0^{-1}\mathbf{Q}_0'\right)^{-1}\mathbf{Q}_0\mathbf{H}_0^{-1}.$$

Also for the unconstrained estimator $\widehat{\boldsymbol{\theta}}_{m_{\max}}$, using result (34) in Section 5, we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_{\max}} - \boldsymbol{\theta}_0\right) \overset{a}{\sim} \mathbf{H}_0^{-1}\sqrt{N}\bar{\mathbf{s}}_N(\boldsymbol{\theta}_0) \tag{A.46}$$

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

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

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

$$2\left[\ell_N\left(\widehat{\boldsymbol{\theta}}_{m_0}\right) - \ell_N\left(\widehat{\boldsymbol{\theta}}_{m_{\max}}\right)\right] = -\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_0} - \widehat{\boldsymbol{\theta}}_{m_{\max}}\right)'\left(\frac{-1}{N}\frac{\partial^2\ell_N(\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)'\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_0} - \widehat{\boldsymbol{\theta}}_{m_{\max}}\right). \tag{A.47}$$

Since $\widehat{\boldsymbol{\theta}}_{m_{\max}}$ and $\widehat{\boldsymbol{\theta}}_{m_0} \xrightarrow{p} \boldsymbol{\theta}_0$ under $m = m_0$, we have $\bar{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ and

$$2\left[\ell_N\left(\widehat{\boldsymbol{\theta}}_{m_{\max}}\right) - \ell_N\left(\widehat{\boldsymbol{\theta}}_{m_0}\right)\right] \overset{a}{\sim} \sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_0} - \widehat{\boldsymbol{\theta}}_{m_{\max}}\right)' \mathbf{H}_0 \sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_0} - \widehat{\boldsymbol{\theta}}_{m_{\max}}\right).$$

Using (A.44) and (A.46), we have the following result:

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{m_{\max}} - \widehat{\boldsymbol{\theta}}_{m_0}\right) \overset{a}{\sim} \left(\mathbf{H}_0^{-1} - \mathbf{F}_0\right)\sqrt{N}\bar{\mathbf{s}}_N\left(\boldsymbol{\theta}_0\right) = \left(\mathbf{H}_0^{-1} - \mathbf{F}_0\right)\mathbf{J}_0^{1/2}\mathbf{z}_n\left(\boldsymbol{\theta}_0\right) \tag{A.48}$$

where $\mathbf{z}_n\left(\boldsymbol{\theta}_0\right) = \mathbf{J}_0^{-1/2}\sqrt{N}\bar{\mathbf{s}}_N\left(\boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_{n_\theta}\right)$, which follows from (A.45). Then, using (A.48) in (A.47), we have

$$2\left[\ell_N\left(\widehat{\boldsymbol{\theta}}_{m_{\max}}\right) - \ell_N\left(\widehat{\boldsymbol{\theta}}_{m_0}\right)\right] \overset{a}{\sim} \mathbf{z}_n\left(\boldsymbol{\theta}_0\right)' \mathbf{A}_0 \mathbf{z}_n\left(\boldsymbol{\theta}_0\right)$$

where

$$\mathbf{A}_0 = \mathbf{J}_0^{1/2}\left(\mathbf{H}_0^{-1} - \mathbf{F}_0\right)\mathbf{H}_0\left(\mathbf{H}_0^{-1} - \mathbf{F}_0\right)\mathbf{J}_0^{1/2} = \mathbf{J}_0^{1/2}\mathbf{H}_0^{-1}\mathbf{Q}_0'\left(\mathbf{Q}_0\mathbf{H}_0^{-1}\mathbf{Q}_0'\right)^{-1}\mathbf{Q}_0\mathbf{H}_0^{-1}\mathbf{J}_0^{1/2}.$$

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

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

where $z_j \sim IIDN\left(0,1\right)$, as required. ■

Proof of Theorem 3. Consider the type I error of the test and note that

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

where $z_i \sim IIDN(0,1)$. Now using Lemma A1 of the theory supplement to Chudik et al. (2018) we have

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

Therefore, since $w_i > 0$

$$\alpha_N \le \sum_{i=1}^{h} \Pr\left(z_i^2 > \left(hw_i\right)^{-1} c_N^2(h)\right) \le h \sup_i \Pr\left(z_i^2 > \theta_i^2 c_N^2(h)\right), \tag{A.49}$$

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

$$\begin{aligned} \Pr\left(z_i^2 > \theta_i^2 c_N^2(h)\right) &= 1 - \Pr\left(-\theta_i\left|c_N(h)\right| \le z_i \le \theta_i\left|c_N(h)\right|\right) \\ &= 2\Phi\left(-\theta_i\left|c_N(h)\right|\right). \end{aligned}$$

Using this result in (A.49) we have

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

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

$$\alpha_N \le 2h\left[1 - \Phi\left(\theta_{\min}\left|c_N(h)\right|\right)\right] = 2h\Phi\left(-\theta_{\min}\left|c_N(h)\right|\right).$$

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

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

and it follows that

$$\alpha_N\leq h\exp\left[-\frac{1}{2}\theta_{\min}^2c_N^2(h)\right]=h\exp\left[-\frac{c_N^2(h)}{2hw_1}\right],$$

which ensures that as $N\to\infty$, $\alpha_N\to 0$, so long as $c_N^2(h)\to\infty$. Also due to the monotonicity property of $\Phi(.)$, we have (for α_N sufficiently small) $\theta_{\min}|c_N(h)|\leq\Phi^{-1}\left(1-\frac{\alpha_N}{2h}\right)$, or $c_N^2(h)\leq\theta_{\min}^{-2}\left[\Phi^{-1}\left(1-\frac{\alpha_N}{2h}\right)\right]^2$. But by Lemma 3 of BPS, $\left[\Phi^{-1}\left(1-\frac{\alpha_N}{2h}\right)\right]^2\leq 2\ln\left(\frac{h}{\alpha_N}\right)$, and

$$c_N^2(h)\leq 2\theta_{\min}^{-2}\ln\left(\frac{h}{\alpha_N}\right)=2w_1h\ln\left(\frac{h}{\alpha_N}\right). \tag{A.50}$$

Consider now the type II error of the test and note that

$$\begin{aligned}\beta_N &= \Pr\left(\mathcal{X}_N\leq c_N^2(h)\,|H_1\right)=\Pr\left(\sum_{i=1}^h w_i\chi_i^2(1,\mu_{i,N}^2)\leq c_N^2(h)\right)\\ &= \Pr\left(\sum_{i=1}^h w_i\left(z_i-\mu_{i,N}\right)^2\leq c_N^2(h)\right).\end{aligned}$$

Since $w_1=\max_i(w_i)$, then $\sum_{i=1}^h w_i\left(z_i-\mu_{i,N}\right)^2\leq w_1\sum_{i=1}^h\left(z_i-\mu_{i,N}\right)^2$, and

$$\begin{aligned}\beta_N &= \Pr\left(\sum_{i=1}^h w_i\left(z_i-\mu_{i,N}\right)^2\leq c_N^2(h)\right)\leq\Pr\left(w_1\sum_{i=1}^h\left(z_i-\mu_{i,N}\right)^2\leq c_N^2(h)\right)\\ &= \Pr\left(\sum_{i=1}^h\left(z_i-\mu_{i,N}\right)^2\leq\frac{c_N^2(h)}{w_1}\right)=\Pr\left(\chi^2(h,\mu_N^2)\leq\frac{c_N^2(h)}{w_1}\right),\end{aligned}$$

where $\chi^2(h,\mu_N^2)$ is a non-central chi-squared random variable with h degrees of freedom and the non-centrality parameter, $\mu_N^2=\sum_{i=1}^h\mu_{i,N}^2$. To obtain the rate at which β_N tends to zero with N, we use the normal approximation proposed by Sankaran (1959) for non-central chi-square distributions given by[13]

$$\beta_N\leq\Pr\left(\chi^2(h,\mu_N^2)\leq\frac{c_N^2(h)}{w_1}\right)\approx\Phi\left(\frac{\left(\frac{c_N^2(h)}{w_1\left(h+\mu_N^2\right)}\right)^{s_N}-\left\{1+s_NA_N\left[s_N-1-0.5(2-s_N)A_NB_N\right]\right\}}{s_N\sqrt{2A_N}(1+0.5A_NB_N)}\right),$$

where

$$\begin{aligned}s_N &= 1-\frac{2}{3}\frac{\left(h+\mu_N^2\right)\left(h+3\mu_N^2\right)}{\left(h+2\mu_N^2\right)^2},\\ A_N &= \frac{h+2\mu_N^2}{\left(h+\mu_N^2\right)^2},\qquad B_N=(s_N-1)(1-3s_N).\end{aligned}$$

Since, h and w_1 are fixed in N, then $A_N=\ominus(\mu_N^{-2})$, $s_N=1/2+O(\mu_N^{-2})$, $B_N=1/4+O(\mu_N^{-2})$ and it readily follows that as $N\to\infty$, $\beta_N\to 0$ if $c_N^2(h)/\mu_N^2\to 0$ as $c_N(h)$ and $\mu_N\to\infty$. ■

[13] Also see Patnaik (1949) and Abdel-Aty (1954) for other approximations.

Proof of Theorem 4. Consider the event $\{\widehat{m} > m_0\}$ where m_0 is the true number of factors. For this event to be true it must be the case that for some $t \in \{1, 2, ..., T-2\}$, at a certain stage in the sequential estimation the null hypothesis of the true number of factors is rejected. That is,

$$\begin{aligned}
\Pr(\widehat{m} &> m_0) \leq P(\exists t,\ m_0 \text{ is rejected} \,|H_{T-2,t-1}) \\
&\leq \sum_{t=1}^{T-2} \Pr(\mathcal{LR}_N(T-2, t-1) > c^2_{N,T-2,t-1}(h) \,|H_{T-2,t-1}),
\end{aligned}$$

where $c^2_{N,T-2,t-1}(h)$ denotes the critical value of the test. For any given t, using the result for the type I error of the test in Theorem 3, as $N \to \infty$, we have

$$\alpha_N = \Pr(\mathcal{LR}_N(T-2, t-1) > c^2_{N,T-2,t-1}(h) \,|H_{T-2,t-1}) = \Pr\left(\sum_{i=1}^{h} w_i z_i^2 > c^2_{N,T-2,t-1}(h)\right) \to 0,$$

if $c^2_{N,T-2,t-1}(h) \to \infty$, from which it follows that (recall that $z_i \sim IIDN(0,1)$)

$$\Pr(\widehat{m} > m_0) \leq (T-2) \max_{1\leq t\leq T-2} P\left[\mathcal{LR}_N(T-2, t-1) > c^2_{N,T-2,t-1}(h) \,|H_{T-2,t-1}\right] \to 0. \tag{A.51}$$

Next consider the event $\{\widehat{m} < m_0\}$, and note that

$$\begin{aligned}
\Pr(\widehat{m} &< m_0) = \Pr\left(\max_{1\leq t\leq T-2} \mathcal{LR}_N(T-2, t-1) \leq c^2_{N,T-2,t-1}(h) \,|H_{T-2,t-1} \text{ is false}\right) \\
&\leq \sum_{t=1}^{T-2} \Pr\left(\mathcal{LR}_N\left(T-2, t-1\right) \leq c^2_{N,T-2,t-1}(h) \,|H_{T-2,t-1} \text{ is false}\right).
\end{aligned} \tag{A.52}$$

From Theorem 3 for the type II error of the test we have that as $N \to \infty$

$$\begin{aligned}
\beta_N &= \Pr\left(\mathcal{LR}_N\left(T-2, t-1\right) \leq c^2_{N,T-2,t-1}(h) \,|H_{T-2,t-1} \text{ is false}\right) \\
&= \Pr\left(\sum_{i=1}^{h} w_i \chi_i^2(1, \mu^2_{i,N}) \leq c^2_{N,T-2,t-1}(h)\right) \to 0.
\end{aligned}$$

But from (A.52), it readily follows that since $\beta_N \to 0$ as $N \to \infty$, $\Pr(\widehat{m} < m_0) \to 0$ which together with (A.51) establishes the desired result. ■

An 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

February 2020

S.1 Introduction

This supplement is organised as follows: Section S.2 outlines the eigenvalue approach used for computing the QML estimator. Section S.3 gives the derivations of the initial values used for the Monte Carlo (MC) analysis. Section S.4 provides details for the computation of the GMM estimators. Sections S.5 and S.6 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 $\sigma_{\mathrm{v}}^2 = 1$. The results for other values, $\sigma_{\mathrm{v}}^2 = \{0.5, 1.5\}$, are very similar and are available upon request.

Section S.7 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.

S.2 An eigenvalue approach for computing the QML estimator

To compute the QML estimators consider the log-likelihood function given by (23) and note that since $\mathbf{\Omega}$ is a positive definite matrix and $\mathbf{QQ}'$ is rank deficient (recall that $m < T$), we have $|\mathbf{\Omega} + \mathbf{QQ}'| = |\mathbf{\Omega}| \left|\mathbf{I}_m + \mathbf{Q}'\mathbf{\Omega}^{-1}\mathbf{Q}\right|$, and using the Woodbury matrix identity

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

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

$$\mathbf{A} = \mathbf{I}_m + \mathbf{Q}'\mathbf{\Omega}^{-1}\mathbf{Q}. \tag{S.2}$$

Using the above results in (23), and after some simplification the quasi-log-likelihood function can be written as

$$N^{-1}\ell_N(\boldsymbol{\theta}) \propto -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\mathbf{\Omega}| - \frac{1}{2}\ln|\mathbf{A}| - \frac{1}{2\sigma^2}\left[\operatorname{tr}\left(\mathbf{B}_N\mathbf{\Omega}^{-1}\right) - \operatorname{tr}\left(\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{QA}^{-1}\mathbf{Q}'\mathbf{\Omega}^{-1}\right)\right], \tag{S.3}$$

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

$$\mathbf{B}_N(\boldsymbol{\varphi}) = N^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i(\boldsymbol{\varphi})\boldsymbol{\xi}_i'(\boldsymbol{\varphi}). \tag{S.4}$$

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

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

and using $\mathbf{Q}'\mathbf{\Omega}^{-1}\mathbf{Q} = \mathbf{A} - \mathbf{I}_m$ from (S.2), we have

$$\mathbf{A}^{-1} = \mathbf{I}_m - \mathbf{P}'\mathbf{P}. \tag{S.5}$$

Similarly,

$$\operatorname{tr}\left(\mathbf{B}_N\mathbf{\Omega}^{-1}\mathbf{QA}^{-1}\mathbf{Q}'\mathbf{\Omega}^{-1}\right) = \sigma^2\operatorname{tr}\left[\mathbf{P}'\mathbf{C}_N(\phi)\mathbf{P}\right],$$

where

$$\mathbf{C}_N(\boldsymbol{\phi}) = \sigma^{-2}\boldsymbol{\Omega}^{-1/2}\mathbf{B}_N(\boldsymbol{\varphi})\boldsymbol{\Omega}^{-1/2}, \tag{S.6}$$

and $\boldsymbol{\phi} = (\boldsymbol{\varphi}', \omega, \sigma^2)'$.

Using the above results, the quasi-log-likelihood function given by (S.3) can now be written as

$$N^{-1}\ell_N(\boldsymbol{\phi}, \mathbf{P}) \propto -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln[1 + T(\omega - 1)] + \frac{1}{2}\ln\left|\mathbf{I}_m - \mathbf{P}'\mathbf{P}\right| - \frac{1}{2}\left\{\mathrm{tr}\left[\mathbf{C}_N(\boldsymbol{\phi})\right] - \mathrm{tr}\left[\mathbf{P}'\mathbf{C}_N(\boldsymbol{\phi})\mathbf{P}\right]\right\}. \tag{S.7}$$

While, as mentioned earlier, the transformation from $\mathbf{Q}$ to $\mathbf{P}$ is carried out for analytical convenience, $\mathbf{P}$ is still not identified. It is easily seen that the value of $\ell_N(\boldsymbol{\phi}, \mathbf{P})$ is invariant to the orthonormal transformation of $\mathbf{P}$. To see this consider the transformation $\widetilde{\mathbf{P}} = \mathbf{P}\boldsymbol{\Xi}$, where $\boldsymbol{\Xi}$ is an $m \times m$ orthonormal matrix such that $\boldsymbol{\Xi}'\boldsymbol{\Xi} = \mathbf{I}_m$. Then it is readily verified that $N^{-1}\ell_N(\boldsymbol{\phi}, \mathbf{P}) = N^{-1}\ell_N\left(\boldsymbol{\phi}, \widetilde{\mathbf{P}}\right)$. Hence, $\mathbf{P}$ (or $\widetilde{\mathbf{P}}$) is identified only up to an $m \times m$ orthonormal rotation matrix. Let $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m)$, 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 $rank(\mathbf{P}) = m$, then $\mathbf{P}'\mathbf{P}$ can be diagonalised by an orthonormal transformation, and without loss of generality we can impose the following $m(m-1)/2$ orthogonality conditions

$$\mathbf{p}_t'\mathbf{p}_s = 0, \text{ for all } s \neq t = 1, 2, ..., m. \tag{S.8}$$

Under these restrictions the quasi-log-likelihood function, (S.7), simplifies to

$$N^{-1}\ell_N(\boldsymbol{\phi}, \mathbf{P}) \propto -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln[1 + T(\omega - 1)] + \frac{1}{2}\sum_{t=1}^{m}\ln\left(1 - \mathbf{p}_t'\mathbf{p}_t\right) + \frac{1}{2}\sum_{t=1}^{m}\mathbf{p}_t'\mathbf{C}_N(\boldsymbol{\phi})\mathbf{p}_t - \frac{1}{2}\mathrm{tr}\left[\mathbf{C}_N(\boldsymbol{\phi})\right]. \tag{S.9}$$

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

$$\mathbf{C}_N(\boldsymbol{\phi})\widehat{\mathbf{p}}_t - \left(\frac{1}{1 - \widehat{\mathbf{p}}_t'\widehat{\mathbf{p}}_t}\right)\widehat{\mathbf{p}}_t = \mathbf{0}, \quad \text{for } t = 1, 2, ..., m, \tag{S.10}$$

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

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

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

$$N^{-1}\ell_N(\boldsymbol{\phi}; m) \propto -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln[1 + T(\omega - 1)] - \frac{1}{2}\sum_{t=1}^{m}\ln\left[\lambda_t(\boldsymbol{\phi})\right] + \frac{1}{2}\sum_{t=1}^{m}\left[\lambda_t(\boldsymbol{\phi}) - 1\right] - \frac{1}{2}\sum_{t=1}^{T}\lambda_t(\boldsymbol{\phi}), \tag{S.11}$$

where $\lambda_t(\boldsymbol{\phi})$ is the t^{th} eigenvalue of $\mathbf{C}_N(\boldsymbol{\phi})$, given by (S.6). This concentrated quasi log-likelihood function can now be maximised with respect to $\boldsymbol{\phi} = (\boldsymbol{\varphi}', \omega, \sigma^2)'$. The QML estimators, $\widehat{\lambda}_t(\boldsymbol{\phi})$, can then be computed using the QML estimator of $\boldsymbol{\phi}$ and their corresponding variance covariance matrix can be computed using the delta method.

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

S.3 Steady state distribution of y_{it} in the stationary case

Consider the panel data model

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

where

$$\zeta_{it} = \sum_{\ell=1}^{m} \eta_{\ell i} f_{\ell t} + u_{it}, \tag{S.12}$$

$$x_{it} = \alpha_{xi} + \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} + \mathrm{v}_{it}, \tag{S.13}$$

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

$$\mathrm{v}_{it} = \rho_x \mathrm{v}_{i,t-1} + \left(1 - \rho_x^2\right)^{1/2} \varepsilon_{it}, \;\; |\rho_x| < 1, \text{ for } t = 1, ..., T, \tag{S.14}$$

$$\varepsilon_{it} \sim IIDN(0, \sigma_{i\mathrm{v}}^2), \text{ and } \mathrm{v}_{i0} \sim IIDN(0, \sigma_{i\mathrm{v}}^2),$$

which ensures that $Var(\mathrm{v}_{it}) = \sigma_{i\mathrm{v}}^2$. Further,

$$f_{\ell t} = \rho_{\ell f} f_{\ell,t-1} + \left(1 - \rho_{f\ell}^2\right)^{1/2} \varepsilon_{f\ell t}, \; \varepsilon_{f\ell t} \sim IID\mathcal{N}(0, 1),$$

with $f_{\ell,0} = 0$, for $\ell = 1, 2, ..., m$, and $t = 1, .., T$. 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

$$E\left(\mathrm{v}_{it}\right) = 0, \;\; E\left(\mathbf{f}_t\right) = 0, \; Cov\left(\mathrm{v}_{i,t-j}\mathrm{v}_{i,t-j'}\right) = \sigma_{i\mathrm{v}}^2 \rho_x^{|j-j'|} \text{ and } Cov\left(\mathbf{f}_{t-j}\mathbf{f}_{t-j'}\right) = \rho_f^{|j-j'|}\mathbf{I}_m. \tag{S.15}$$

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

$$y_{it} = \widetilde{\alpha}_i + \delta_t + \gamma y_{i,t-1} + \beta \mathrm{v}_{it} + \mathbf{c}_i'\mathbf{f}_t + u_{it}, \tag{S.16}$$

where

$$\widetilde{\alpha}_i = \alpha_i + \beta \alpha_{xi}, \tag{S.17}$$

$$\mathbf{c}_i'\mathbf{f}_t = \sum_{\ell=1}^{m} \eta_{\ell i} f_{\ell t} + \beta \sum_{\ell=1}^{m_x} \vartheta_{\ell i} f_{\ell t} = \sum_{\ell=1}^{\max(m, m_x)} c_{\ell i} f_{\ell t}, \tag{S.18}$$

where $c_{\ell i}$ for all i and $\ell = 1, 2, ...,\max(m, m_x)$, are defined implicitly. Using the results in (S.15), and noting that $\mathbf{f}_t, u_{it'}$ and v_{is} are mutually uncorrelated for all values of t, t' and s, it then follows, conditional on $\widetilde{\alpha}_i$ and $\mathbf{c}_i$, that (without loss of generality we set $\delta_t = 0$)

$$E\left(y_{it}|\widetilde{\alpha}_i, \mathbf{c}_i\right) = \gamma E\left(y_{i,t-1}|\widetilde{\alpha}_i, \mathbf{c}_i\right) + \widetilde{\alpha}_i \tag{S.19}$$

$$\begin{aligned} Var\left(y_{it}|\widetilde{\alpha}_i, \mathbf{c}_i\right) &= \gamma^2 Var\left(y_{i,t-1}|\widetilde{\alpha}_i, \mathbf{c}_i\right) + \beta^2 Var\left(\mathrm{v}_{it}|\widetilde{\alpha}_i, \mathbf{c}_i\right) + \mathbf{c}_i' Cov\left(\mathbf{f}_t\mathbf{f}_t'\right)\mathbf{c}_i + \sigma^2 \\ &\quad + 2\gamma Cov\left(y_{i,t-1}, \mathbf{c}_i'\mathbf{f}_t|\widetilde{\alpha}_i, \mathbf{c}_i\right) + 2\gamma\beta Cov\left(y_{i,t-1}, \mathrm{v}_{it}|\widetilde{\alpha}_i, \mathbf{c}_i\right). \end{aligned} \tag{S.20}$$

Also, the steady state values of the covariances in the above expression are given by (upon using (S.15))

$$\begin{aligned} Cov\left(y_{i,t-1}, \mathbf{c}_i'\mathbf{f}_t|\widetilde{\alpha}_i, \mathbf{c}_i\right) &= \sum_{j=0}^{\infty} \gamma^j \mathbf{c}_i' E\left(\mathbf{f}_{t-j-1}\mathbf{f}_t'\right)\mathbf{c}_i = \left(\mathbf{c}_i'\mathbf{c}_i\right) \sum_{j=0}^{\infty} \rho_f^{j+1}\gamma^j = \frac{\left(\mathbf{c}_i'\mathbf{c}_i\right)\rho_f}{1 - \gamma\rho_f}, \\ Cov\left(y_{i,t-1}, \mathrm{v}_{it}|\widetilde{\alpha}_i, \mathbf{c}_i\right) &= \beta\sigma_{i\mathrm{v}}^2 \sum_{j=0}^{\infty} \gamma^j E\left(\mathrm{v}_{i,t-j-1}\mathrm{v}_{it}\right) = \beta\sigma_{i\mathrm{v}}^2 \sum_{j=0}^{\infty} \rho_x^{j+1}\gamma^j = \frac{\beta\rho_x\sigma_{i\mathrm{v}}^2}{1 - \gamma\rho_x}. \end{aligned}$$

Using the above results in (S.20) and noting that in steady state $E(y_{it}|\widetilde{\alpha}_i, \mathbf{c}_i) = E(y_{i0}|\widetilde{\alpha}_i, \mathbf{c}_i)$ and $Var(y_{it}|\widetilde{\alpha}_i, \mathbf{c}_i) = Var(y_{i0}|\widetilde{\alpha}_i, \mathbf{c}_i)$ we have

$$E\left(y_{it}|\widetilde{\alpha}_i, \mathbf{c}_i\right) = \mu_{i0} = \frac{\alpha_i + \beta\alpha_{xi}}{1-\gamma}, \tag{S.21}$$

$$Var(y_{it}|\widetilde{\alpha}_i, \mathbf{c}_i) = \sigma_{i0}^2 = \frac{\sigma^2 + \mathrm{a}_x\beta^2\sigma_{\mathrm{iv}}^2 + \mathrm{a}_f\mathrm{a}_i}{1-\gamma^2}, \tag{S.22}$$

where

$$\mathrm{a}_i = \mathbf{c}_i'\mathbf{c}_i = \sum_{\ell=1}^{m}\eta_{\ell i}^2 + \beta^2\sum_{\ell=1}^{m_x}\vartheta_{\ell i}^2 + 2\beta\textstyle\sum_{\ell=1}^{\min(m,m_x)}\eta_{\ell i}\vartheta_{\ell i}, \tag{S.23}$$

and

$$\mathrm{a}_x = \left(\frac{1+\gamma\rho_x}{1-\gamma\rho_x}\right), \text{ and } \mathrm{a}_f = \left(\frac{1+\gamma\rho_f}{1-\gamma\rho_f}\right). \tag{S.24}$$

S.4 The GMM approach

Let us consider a GMM approach to estimate the dynamic panel data model with interactive effects:

$$y_{it} = \alpha_i + \mathbf{w}_{it}'\boldsymbol{\delta} + \boldsymbol{\lambda}_i'\mathbf{f}_t + \varepsilon_{it}, \qquad (i = 1, 2, ..., N; t = 1, 2, ..., T) \tag{S.25}$$

where $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}_{it}')'$, $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$, $\boldsymbol{\lambda}_i = (\lambda_{1i}, ..., \lambda_{mi})'$ and $\mathbf{f}_t = (f_{1t}, ..., f_{mt})'$ 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) considers 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 $\boldsymbol{\lambda}_i'\mathbf{f}_t$ in (S.25), we have the following alternative expression

$$y_{it} = \mathbf{w}_{it}'\boldsymbol{\delta} + \widetilde{\boldsymbol{\lambda}}_i'\widetilde{\mathbf{f}}_t + \varepsilon_{it},$$

where $\widetilde{\boldsymbol{\lambda}}_i = (\alpha_i, \lambda_{1i}, ..., \lambda_{mi})'$ and $\widetilde{\mathbf{f}}_t = (1, f_{1t}, ..., f_{mt})'$. The model in matrix notation can be written as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \widetilde{\mathbf{F}}\widetilde{\boldsymbol{\lambda}}_i + \boldsymbol{\varepsilon}_i, \tag{S.26}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ and $\widetilde{\mathbf{F}} = (\widetilde{\mathbf{f}}_1, ..., \widetilde{\mathbf{f}}_T)'$ is a $T \times \widetilde{m}$ matrix. Define $\widetilde{\boldsymbol{\Psi}} = \widetilde{\mathbf{F}}\bar{\mathbf{F}}^{-1}$ where $\bar{\mathbf{F}} = (\widetilde{\mathbf{f}}_{T-\widetilde{m}+1}, ..., \widetilde{\mathbf{f}}_T)'$. To separately identify $\widetilde{\mathbf{F}}$ from $\widetilde{\boldsymbol{\lambda}}_i$, $\widetilde{m}^2$ restrictions are imposed on the factors such that $\widetilde{\mathbf{F}} = (\boldsymbol{\Psi}', \mathbf{I}_{\widetilde{m}})'$ where $\boldsymbol{\Psi}$ is a $(T-\widetilde{m})\times\widetilde{m}$ matrix of unrestricted parameters obtained as the first $T - \widetilde{m}$ rows of $\widetilde{\boldsymbol{\Psi}}$. Let $\mathbf{H}_Q = (\mathbf{I}_{T-\widetilde{m}}, -\boldsymbol{\Psi})'$, so that $\mathbf{H}_Q'\widetilde{\mathbf{F}} = (\mathbf{I}_{T-\widetilde{m}}, -\boldsymbol{\Psi})(\boldsymbol{\Psi}', \mathbf{I}_{\widetilde{m}})' = \mathbf{0}_{(T-\widetilde{m})\times\widetilde{m}}$. Then, pre-multiplying equation (S.26) 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

$$\begin{aligned}\dot{\mathbf{y}}_i &= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \boldsymbol{\Psi}\ddot{\mathbf{y}}_i - \boldsymbol{\Psi}\ddot{\mathbf{W}}_i\boldsymbol{\delta} + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{\varepsilon}}_i \\ &= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \left(\mathbf{I}_{T-\widetilde{m}} \otimes \ddot{\mathbf{y}}_i'\right) vec(\boldsymbol{\Psi}) - \left(vec(\ddot{\mathbf{W}}_i)' \otimes \mathbf{I}_{T-\widetilde{m}}\right) vec(\boldsymbol{\delta}' \otimes \boldsymbol{\Psi}) + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{\varepsilon}}_i,\end{aligned} \tag{S.27}$$

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

The t^{th} equation is given by

$$y_{it} = \boldsymbol{\delta}'\mathbf{w}_{it} + \boldsymbol{\psi}_t'\ddot{\mathbf{y}}_i - \boldsymbol{\psi}_t'\ddot{\mathbf{W}}_i\boldsymbol{\delta} + v_{it}, \qquad (i = 1, ..., N; t = 1, ..., T - \widetilde{m}), \tag{S.28}$$

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 $(k+1)(T-\widetilde{m})(T-\widetilde{m}+1)/2 + k(T-\widetilde{m})\widetilde{m}$ moment conditions given by $E[\mathbf{z}_{it}v_{it}] = \mathbf{0}$, for $t = 1, ..., T-\widetilde{m}$, where $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-1}, \mathbf{x}_{i1}', ..., \mathbf{x}_{it}', \mathbf{x}_{i,T-\widetilde{m}+1}', ..., \mathbf{x}_{iT}')'$. In addition to the commonly used instruments $(y_{i0}, ..., y_{i,t-1}, \mathbf{x}_{i1}', ..., \mathbf{x}_{it}')$, we also use $\mathbf{x}_{i,T-\widetilde{m}+1}', ..., \mathbf{x}_{iT}'$ as instruments since they are included in the regressor $\ddot{\mathbf{W}}$. In matrix notation the moment conditions can be written as $E\left[\mathbf{Z}_i'\mathbf{v}_i(\boldsymbol{\theta})\right] = \mathbf{0}$, where $\mathbf{Z}_i = diag(\mathbf{z}_{i1}', ..., \mathbf{z}_{i,T-\widetilde{m}}')$, $\mathbf{v}_i(\boldsymbol{\theta}) = (v_{i1}, ..., v_{i,T-\widetilde{m}})'$ and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\psi}')'$ with $\boldsymbol{\psi} = vec(\boldsymbol{\Psi})$.

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

$$\widehat{\boldsymbol{\theta}}_{QD1} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i(\boldsymbol{\theta})'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{v}_i(\boldsymbol{\theta})\right), \tag{S.29}$$

and

$$\widehat{\boldsymbol{\theta}}_{QD2} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i(\boldsymbol{\theta})'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{v}_i(\widehat{\boldsymbol{\theta}}_{QD1})\mathbf{v}_i(\widehat{\boldsymbol{\theta}}_{QD1})'\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{v}_i(\boldsymbol{\theta})\right). \tag{S.30}$$

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

$$Var(\widehat{\boldsymbol{\theta}}_{QD1}) = N^{-1}\left(\widehat{\mathbf{G}}_{QD1}'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{G}}_{QD1}\right)^{-1}\widehat{\mathbf{G}}_{QD1}'\widehat{\mathbf{W}}^{-1}\widehat{\boldsymbol{\Omega}}_{QD1}\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{G}}_{QD1}\left(\widehat{\mathbf{G}}_{QD1}'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{G}}_{QD1}\right)^{-1} \tag{S.31}$$

$$Var(\widehat{\boldsymbol{\theta}}_{QD2}) = N^{-1}\left(\widehat{\mathbf{G}}_{QD2}'\widehat{\boldsymbol{\Omega}}_{QD2}^{-1}\widehat{\mathbf{G}}_{QD2}\right)^{-1}, \tag{S.32}$$

where $\widehat{\mathbf{G}}_j = \partial\bar{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j)/\partial\boldsymbol{\theta}'$ for $j = QD1, QD2$, with $\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j) = \mathbf{Z}_i'\mathbf{v}_i(\widehat{\boldsymbol{\theta}}_j)$ and $\bar{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j) = N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j)$, $\widehat{\mathbf{W}} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i$, and $\widehat{\boldsymbol{\Omega}}_j = N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j)\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j)'$. The derivatives involved in $\widehat{\mathbf{G}}_j$ are computed numerically.

Approach 2: Quasi-differencing after first-differencing

Taking the first-differences of model (S.25) to remove α_i we have

$$\Delta y_{it} = \Delta\mathbf{w}_{it}'\boldsymbol{\delta} + \boldsymbol{\lambda}_i'\Delta\mathbf{f}_t + \Delta\varepsilon_{it}, \qquad (i = 1, 2, ..., N; t = 2, 3, ..., T)$$

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

$$\Delta\mathbf{y}_i = \Delta\mathbf{W}_i\boldsymbol{\delta} + \Delta\mathbf{F}\boldsymbol{\lambda}_i + \Delta\boldsymbol{\varepsilon}_i, \tag{S.33}$$

where $\Delta\mathbf{y}_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$, $\Delta\mathbf{W}_i = (\Delta\mathbf{w}_{i2}, ..., \Delta\mathbf{w}_{iT})'$, $\Delta\boldsymbol{\varepsilon}_i = (\Delta\varepsilon_{i2}, ..., \Delta\varepsilon_{iT})'$ and $\Delta\mathbf{F} = (\Delta\mathbf{f}_2, ..., \Delta\mathbf{f}_T)'$ is a $(T-1) \times m$ matrix. Define $\widetilde{\boldsymbol{\Phi}} = \Delta\mathbf{F}\left(\overline{\Delta\mathbf{F}}\right)^{-1}$ where $\overline{\Delta\mathbf{F}} = (\Delta\mathbf{f}_{T-m+1}, ..., \Delta\mathbf{f}_T)'$. 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 $\widetilde{\mathbf{\Phi}}$. Let $\mathbf{H}_D = (\mathbf{I}_{T-1-m}, -\mathbf{\Phi})'$, so that $\mathbf{H}_D'\Delta\mathbf{F} = (\mathbf{I}_{T-1-m}, -\mathbf{\Phi})(\mathbf{\Phi}', \mathbf{I}_m)' = \mathbf{0}_{(T-1-m)\times m}$. Then, pre-multiplying equation (S.33) 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

$$\begin{aligned}\Delta\dot{\mathbf{y}}_i &= \Delta\dot{\mathbf{W}}_i\boldsymbol{\delta} + \mathbf{\Phi}\Delta\ddot{\mathbf{y}}_i - \mathbf{\Phi}\Delta\ddot{\mathbf{W}}_i\boldsymbol{\delta} + \dot{\boldsymbol{\varepsilon}}_i - \mathbf{\Phi}\Delta\ddot{\boldsymbol{\varepsilon}}_i \\ &= \Delta\dot{\mathbf{W}}_i\boldsymbol{\delta} + \left(\mathbf{I}_{T-1-m}\otimes\Delta\ddot{\mathbf{y}}_i'\right)vec(\mathbf{\Phi}) - \left(vec(\Delta\ddot{\mathbf{W}}_i)'\otimes\mathbf{I}_{T-1-m}\right)vec(\boldsymbol{\delta}'\otimes\mathbf{\Phi}) + \Delta\dot{\boldsymbol{\varepsilon}}_i - \mathbf{\Phi}\Delta\ddot{\boldsymbol{\varepsilon}}_i,\end{aligned}$$

where $\Delta\dot{\mathbf{y}}_i = (\Delta y_{i2}, ..., \Delta y_{i,T-m})'$, $\Delta\ddot{\mathbf{y}}_i = (\Delta y_{i,T-m+1}, ..., \Delta y_{iT})'$, $\Delta\dot{\mathbf{W}}_i = (\Delta\mathbf{w}_{i2}, ..., \Delta\mathbf{w}_{i,T-m})'$, $\Delta\ddot{\mathbf{W}}_i = (\Delta\mathbf{w}_{i,T-m+1}, ..., \Delta\mathbf{w}_{iT})'$, $\mathbf{\Phi}' = (\boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_{T-m})$, $\Delta\dot{\boldsymbol{\varepsilon}}_i = (\Delta\varepsilon_{i2}, ..., \Delta\varepsilon_{i,T-m})'$, and
$\Delta\ddot{\boldsymbol{\varepsilon}}_i = (\Delta\varepsilon_{i,T-m+1}, ..., \Delta\varepsilon_{iT})'$.

The t^{th} equation is given by

$$\Delta y_{it} = \boldsymbol{\delta}'\Delta\mathbf{w}_{it} + \boldsymbol{\phi}_t'\Delta\ddot{\mathbf{y}}_i - \boldsymbol{\phi}_t'\Delta\ddot{\mathbf{W}}_i\boldsymbol{\delta} + \Delta v_{it}, \qquad (i = 1, ..., N; t = 2, ..., T-m), \tag{S.34}$$

where $\Delta v_{it} = (\Delta\varepsilon_{it} - \boldsymbol{\phi}_t'\Delta\ddot{\boldsymbol{\varepsilon}}_i)$. 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 $t = 2, ..., T-m$, where $\mathbf{z}_{it} = (y_{i0}, ..., y_{i,t-1}, \mathbf{x}_{i0}', \mathbf{x}_{i1}'..., \mathbf{x}_{it}', \mathbf{x}_{i,T-m+1}', ..., \mathbf{x}_{iT}')'$. In addition to the commonly used instruments $(y_{i0}, ..., y_{i,t-1}, \mathbf{x}_{i0}', ..., \mathbf{x}_{it}')$, we also use $\mathbf{x}_{i,T-m+1}', ..., \mathbf{x}_{iT}'$ 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 $E\left[\mathbf{Z}_i'\Delta\mathbf{v}_i(\boldsymbol{\theta})\right] = \mathbf{0}$, where $\mathbf{Z}_i = diag(\mathbf{z}_{i2}', ..., \mathbf{z}_{i,T-m}')$, $\Delta\mathbf{v}_i(\boldsymbol{\theta}) = (\Delta v_{i2}, ..., \Delta v_{i,T-m})'$ and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\phi}')'$ with $\boldsymbol{\phi} = vec(\mathbf{\Phi})$.

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

$$\widehat{\boldsymbol{\theta}}_{FD1} = \arg\min_{\boldsymbol{\theta}}\left(\frac{1}{N}\sum_{i=1}^N\Delta\mathbf{v}_i(\boldsymbol{\theta})'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^N\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N\mathbf{Z}_i'\Delta\mathbf{v}_i(\boldsymbol{\theta})\right), \tag{S.35}$$

and

$$\widehat{\boldsymbol{\theta}}_{FD2} = \arg\min_{\boldsymbol{\theta}}\left(\frac{1}{N}\sum_{i=1}^N\Delta\mathbf{v}_i(\boldsymbol{\theta})'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^N\mathbf{Z}_i'\Delta\mathbf{v}_i(\widehat{\boldsymbol{\theta}}_{FD1})\Delta\mathbf{v}_i(\widehat{\boldsymbol{\theta}}_{FD1})'\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N\mathbf{Z}_i'\Delta\mathbf{v}_i(\boldsymbol{\theta})\right). \tag{S.36}$$

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

$$Var(\widehat{\boldsymbol{\theta}}_{FD1}) = N^{-1}\left(\widehat{\mathbf{G}}_{FD1}'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{G}}_{FD1}\right)^{-1}\widehat{\mathbf{G}}_{FD1}'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{\Omega}}_{FD1}\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{G}}_{FD1}\left(\widehat{\mathbf{G}}_{FD1}'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{G}}_{FD1}\right)^{-1} \tag{S.37}$$

$$Var(\widehat{\boldsymbol{\theta}}_{FD2}) = N^{-1}\left(\widehat{\mathbf{G}}_{FD2}'\widehat{\mathbf{\Omega}}_{FD2}^{-1}\widehat{\mathbf{G}}_{FD2}\right)^{-1}, \tag{S.38}$$

where $\widehat{\mathbf{G}}_j = \partial\bar{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j)/\partial\boldsymbol{\theta}'$ for $j = FD1, FD2$, with $\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j) = \mathbf{Z}_i'\Delta\mathbf{v}_i(\widehat{\boldsymbol{\theta}}_j)$ and $\bar{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j) = N^{-1}\sum_{i=1}^N\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j)$, $\widehat{\mathbf{W}} = N^{-1}\sum_{i=1}^N\mathbf{Z}_i'\mathbf{Z}_i$, and $\widehat{\mathbf{\Omega}}_j = N^{-1}\sum_{i=1}^N\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j)\mathbf{g}_i(\widehat{\boldsymbol{\theta}}_j)'$. The derivatives involved in $\widehat{\mathbf{G}}_j$ 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 $\widetilde{m} = 1$, and $\mathbf{L}_1 = (\mathbf{I}_{\widetilde{m}}, \mathbf{0}_{\widetilde{m}})$ and $\mathbf{L}_2 = (\mathbf{0}_{\widetilde{m}}, \mathbf{I}_{\widetilde{m}})$ for $\widetilde{m} > 1$. Also, define $\check{\mathbf{y}}_i = \left(y_{i,T-\widetilde{m}}, y_{i,T-\widetilde{m}+1}, .., y_{iT}\right)' = \left(y_{i,T-\widetilde{m}}, \ddot{\mathbf{y}}_i'\right)'$. Then, noting that $\ddot{\mathbf{W}}_i = \left(\ddot{\mathbf{y}}_{i,-1}, \ddot{\mathbf{X}}_{it}\right)$ where $\ddot{\mathbf{y}}_{i,-1} = \left(y_{i,T-\widetilde{m}}, y_{i,T-\widetilde{m}+1}, .., y_{iT-1}\right)'$, $\ddot{\mathbf{y}}_i = \mathbf{L}_2\check{\mathbf{y}}_i$ and $\ddot{\mathbf{y}}_{i,-1} = \mathbf{L}_1\check{\mathbf{y}}_i$, we have

$$
\begin{aligned}
\dot{\mathbf{y}}_i &= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \boldsymbol{\Psi}\ddot{\mathbf{y}}_i - \boldsymbol{\Psi}\ddot{\mathbf{W}}_i\boldsymbol{\delta} + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{\varepsilon}}_i \\
&= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \boldsymbol{\Psi}\mathbf{L}_2\check{\mathbf{y}}_i - \boldsymbol{\Psi}\left(\gamma\mathbf{L}_1\check{\mathbf{y}}_i + \ddot{\mathbf{X}}_i\boldsymbol{\beta}\right) + \dot{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{\varepsilon}}_i \\
&= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \boldsymbol{\Psi}\left(\mathbf{L}_2 - \gamma\mathbf{L}_1\right)\check{\mathbf{y}}_i - \boldsymbol{\Psi}\ddot{\mathbf{X}}_i\boldsymbol{\beta} + \mathbf{v}_i \\
&= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \boldsymbol{\Upsilon}\check{\mathbf{y}}_i - \boldsymbol{\Psi}\ddot{\mathbf{X}}_i\boldsymbol{\beta} + \mathbf{v}_i \\
&= \dot{\mathbf{W}}_i\boldsymbol{\delta} + \left(\mathbf{I}_{T-\widetilde{m}} \otimes \check{\mathbf{y}}_i'\right) vec(\boldsymbol{\Upsilon}') - \left(vec(\ddot{\mathbf{X}}_i)' \otimes \mathbf{I}_{T-\widetilde{m}}\right) vec(\boldsymbol{\beta}' \otimes \boldsymbol{\Psi}) + \mathbf{v}_i \\
&= \widetilde{\mathbf{X}}_i\boldsymbol{\pi} + \mathbf{v}_i
\end{aligned}
$$

where $\boldsymbol{\Upsilon} = \boldsymbol{\Psi}\left(\mathbf{L}_2 - \gamma\mathbf{L}_1\right)$, $\mathbf{X}_i = \left(\dot{\mathbf{W}}_i, \left(\mathbf{I}_{T-\widetilde{m}} \otimes \check{\mathbf{y}}_i'\right), -\left(vec(\ddot{\mathbf{X}}_i)' \otimes \mathbf{I}_{T-\widetilde{m}}\right)\right)$ and $\boldsymbol{\pi} = \left(\boldsymbol{\delta}', vec(\boldsymbol{\Upsilon}')', vec(\boldsymbol{\beta}' \otimes \boldsymbol{\Psi})'\right)' = (\boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \boldsymbol{\pi}_3')'$ with $\boldsymbol{\pi}_1 = \boldsymbol{\delta}, \boldsymbol{\pi}_2 = vec(\boldsymbol{\Upsilon}')$, $\boldsymbol{\pi}_3 = vec(\boldsymbol{\beta}' \otimes \boldsymbol{\Psi})$. We consider this particular model rather than the original model (S.27) because perfect multicollinearity between $\ddot{\mathbf{y}}_i$ and $\ddot{\mathbf{W}}_i$ occurs in (S.27) when $\widetilde{m} > 1$. Since this is a linear model in $\boldsymbol{\pi}$ with moment conditions $E\left[\mathbf{Z}_i'\mathbf{v}_i(\boldsymbol{\pi})\right] = \mathbf{0}$, a closed form solution is obtained as

$$
\begin{aligned}
\widehat{\boldsymbol{\pi}} &= \left[\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\widetilde{\mathbf{X}}_i\right)\right]^{-1} \\
&\quad \times \left[\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{Z}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\dot{\mathbf{y}}_i\right)\right].
\end{aligned}
$$

Hence, $\widehat{\boldsymbol{\pi}}_1$ and $\widehat{\boldsymbol{\pi}}_2$ are consistent estimates of $\boldsymbol{\delta}$ and $vec(\boldsymbol{\Upsilon}')$, respectively. To recover $\boldsymbol{\Psi}$ from the estimate of $\boldsymbol{\Upsilon}$, since

$$
vec\left(\boldsymbol{\Upsilon}'\right) = vec\left(\left(\mathbf{L}_2 - \gamma\mathbf{L}_1\right)'\boldsymbol{\Psi}'\right) = \left(\mathbf{I}_{T-\widetilde{m}} \otimes \left(\mathbf{L}_2 - \gamma\mathbf{L}_1\right)'\right) vec\left(\boldsymbol{\Psi}'\right) = \mathbf{A}vec\left(\boldsymbol{\Psi}'\right),
$$

$vec\left(\boldsymbol{\Psi}'\right)$ is obtained as $vec\left(\boldsymbol{\Psi}'\right) = \left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'vec\left(\boldsymbol{\Upsilon}'\right)$. In the computation of the nonlinear GMM estimators, estimates of $\boldsymbol{\delta}$ and $vec\left(\boldsymbol{\Psi}'\right)$ are obtained from $\widehat{\boldsymbol{\pi}}_1$ and $\widehat{\boldsymbol{\pi}}_2$ 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, jth element of $vec\left(\boldsymbol{\Psi}'\right)$.

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) 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

$$\gamma_{ini} \sim U(-0.999, 0.999), \ \psi_{j,ini} \sim \psi_{j,0} \times U(-0.5, 0.5), \quad (j = 1, ..., (T - \widetilde{m})\widetilde{m})$$

for approach 1 and

$$\gamma_{ini} \sim U(-0.999, 0.999), \ \psi_{j,ini} \sim \psi_{j,0} \times U(-0.5, 0.5), \quad (j = 1, ..., (T - 1 - m)m)$$

for approach 2 where $\psi_{j,0}$ is the true value of ψ_j.

S.5 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)

κ^2	0.25			0.5			1			2		
m_0	0	1	2	0	1	2	0	1	2	0	1	2
						$T=5$						
N	$\gamma_0 = 0.4$											
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
						$T=10$						
	$\gamma_0 = 0.4$											
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: $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 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)

$T = 5$

	$\kappa^2 = 0.25$									$\kappa^2 = 0.5$								
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

A2(i): Bias(×100) and RMSE(×100) of γ for the AR(1) model, using the estimated number of factors, $\widehat{m}$

	Bias(×100)				RMSE(×100)				Bias(×100)				RMSE(×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
	$\gamma_0 = 0.4$								$\gamma_0 = 0.8$							
	$T = 5$															
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
	$T = 10$															
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.

A2(ii): Size(×100) of γ for the AR(1) model, using the estimated number of factors, $\widehat{m}$

	$T = 5$								$T = 10$							
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
	$\gamma_0 = 0.4$				$\gamma_0 = 0.8$				$\gamma_0 = 0.4$				$\gamma_0 = 0.8$			
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(×100) and RMSE(×100) of γ and β for the ARX(1) model, using the estimated number of factors, $\widehat{m}$ ($\sigma_{\mathrm{v}}^2 = 1$)

	Bias(×100)				RMSE(×100)				Bias(×100)				RMSE(×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
					$\gamma_0 = 0.4$								$\gamma_0 = 0.8$			
									$T = 5$							
	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	1.85
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
									$T = 10$							
	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) model, using the estimated number of factors, $\widehat{m}$ ($\sigma_v^2 = 1$)

	$T=5$								$T=10$							
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
	$\gamma_0 = 0.4$				$\gamma_0 = 0.8$				$\gamma_0 = 0.4$				$\gamma_0 = 0.8$			
	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).

QML and GMM estimates: Bias, RMSE and Size for the AR(1) Case

In what follows results 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.

Table A2(v): Bias($\times 100$) and RMSE($\times 100$) of γ for the QML and GMM estimators in the case of the AR(1) model, using the true number of factors, m_0 ($T = 10, \kappa^2 = 1$)

	Bias ($\times 100$)					RMSE ($\times 100$)				
	QML	GMM				QML	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 A2(ii).

Table A2(vi): Size($\times 100$) of γ for the QML and GMM estimators in the case of the AR(1) model, using the true number of factors, m_0 ($T = 10$, $\kappa^2 = 1$)

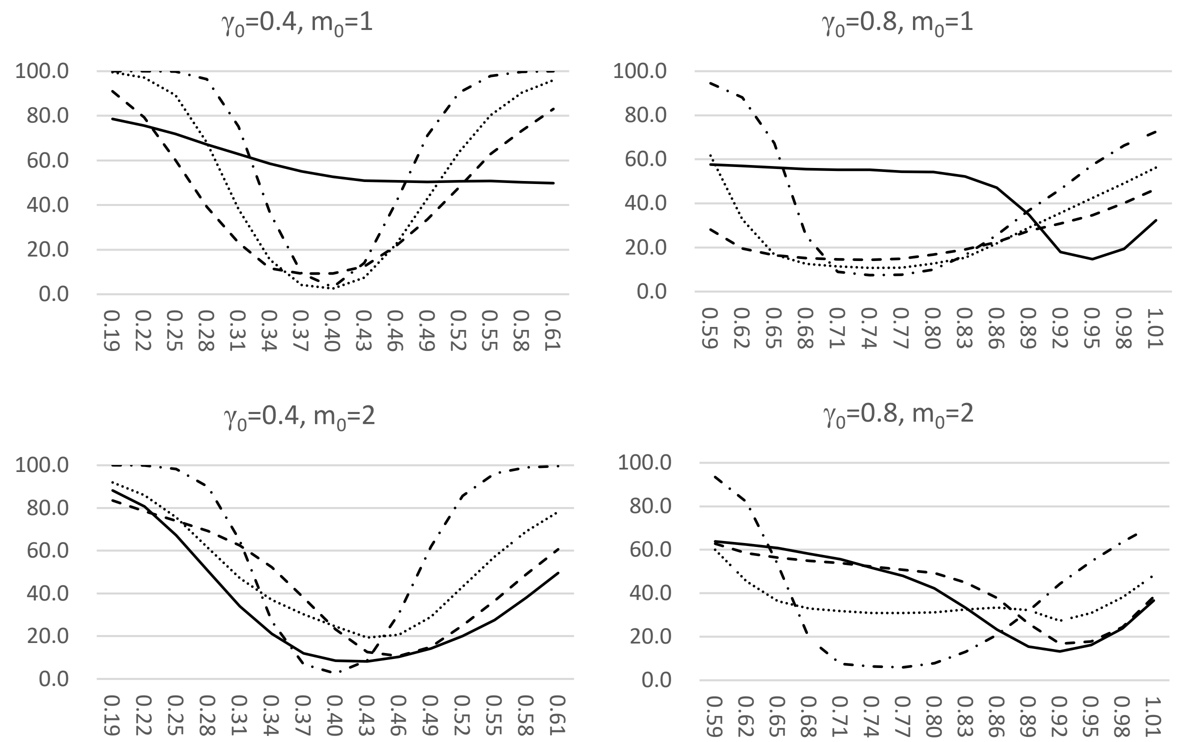
	QML	GMM				QML	GMM			
		QD1	QD2	FD1	FD2		QD1	QD2	FD1	FD2
m_0	1					2				
N	$\gamma_0 = 0.4$									
100	6.5	95.5	98.4	97.7	100.0	4.8	73.9	81.8	51.5	71.0
300	5.8	95.3	98.7	97.9	100.0	4.9	64.2	70.2	34.1	50.2
500	5.3	95.1	99.6	97.8	100.0	3.7	54.5	61.8	22.5	38.0
1000	5.3	92.2	99.5	97.8	100.0	5.3	41.1	48.4	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 A2(v).

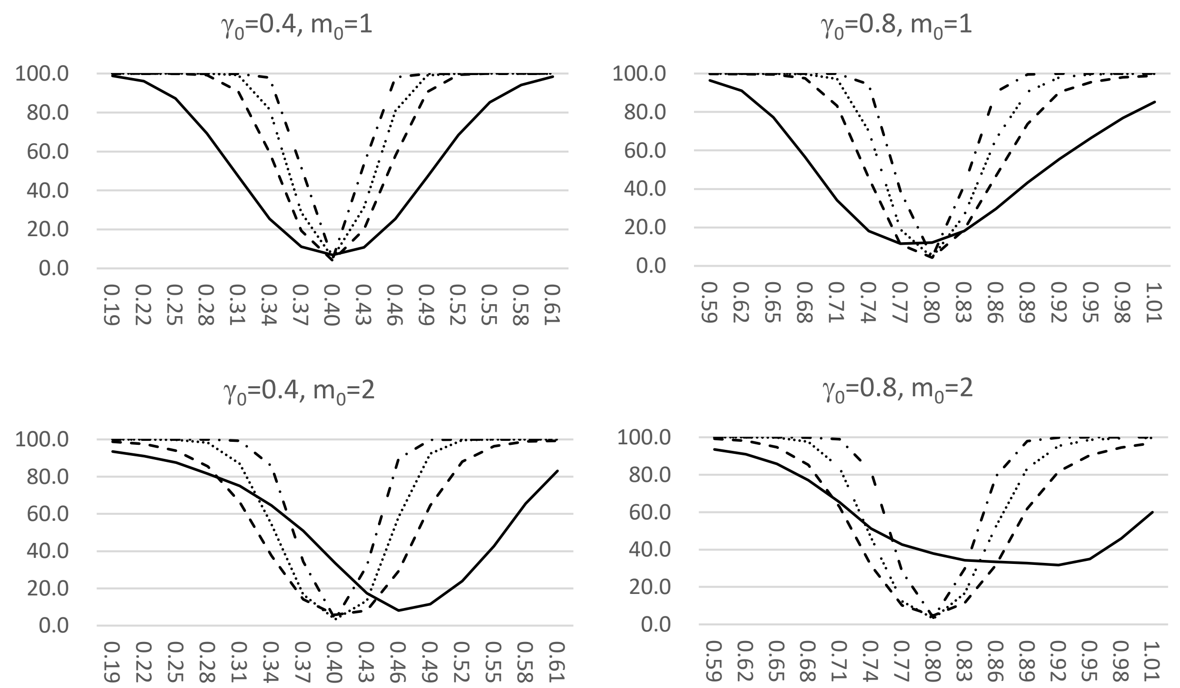
A3: Power Functions

Figure A3(i): Power functions for estimation of γ in the AR(1) model with different values of m and N (κ^2=0.25)

Panel A: T=5



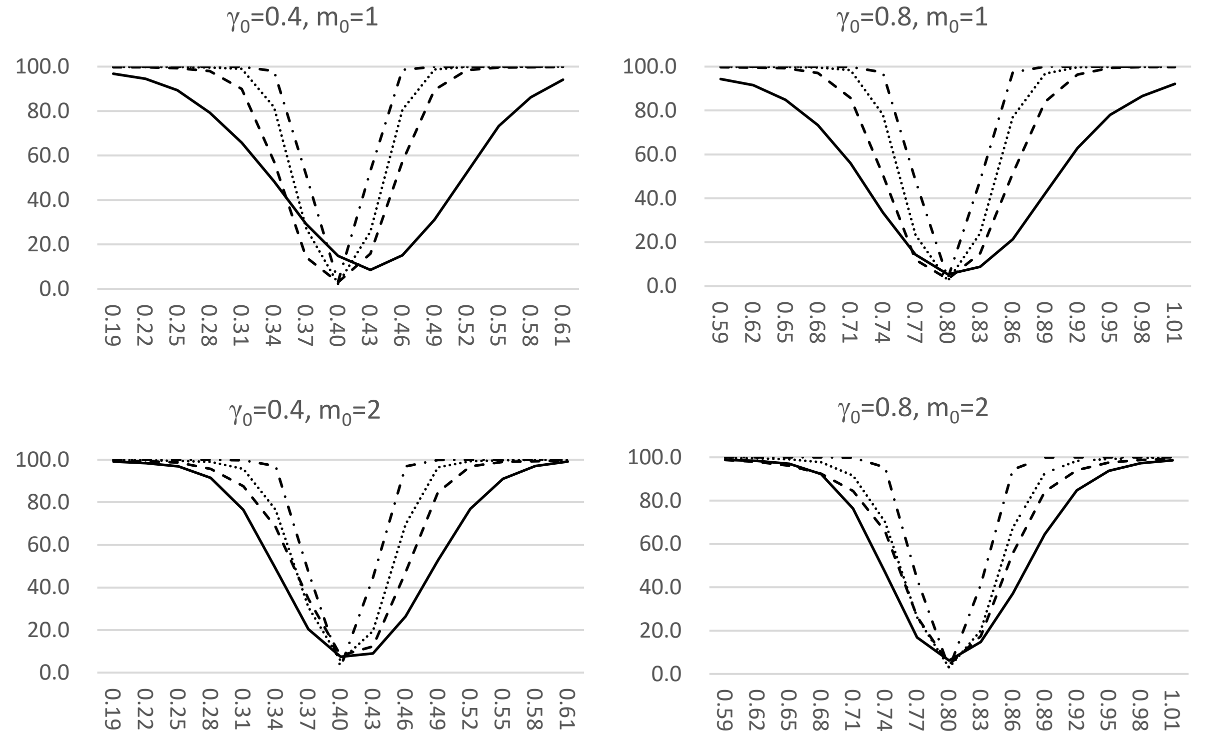
Panel B: T=10



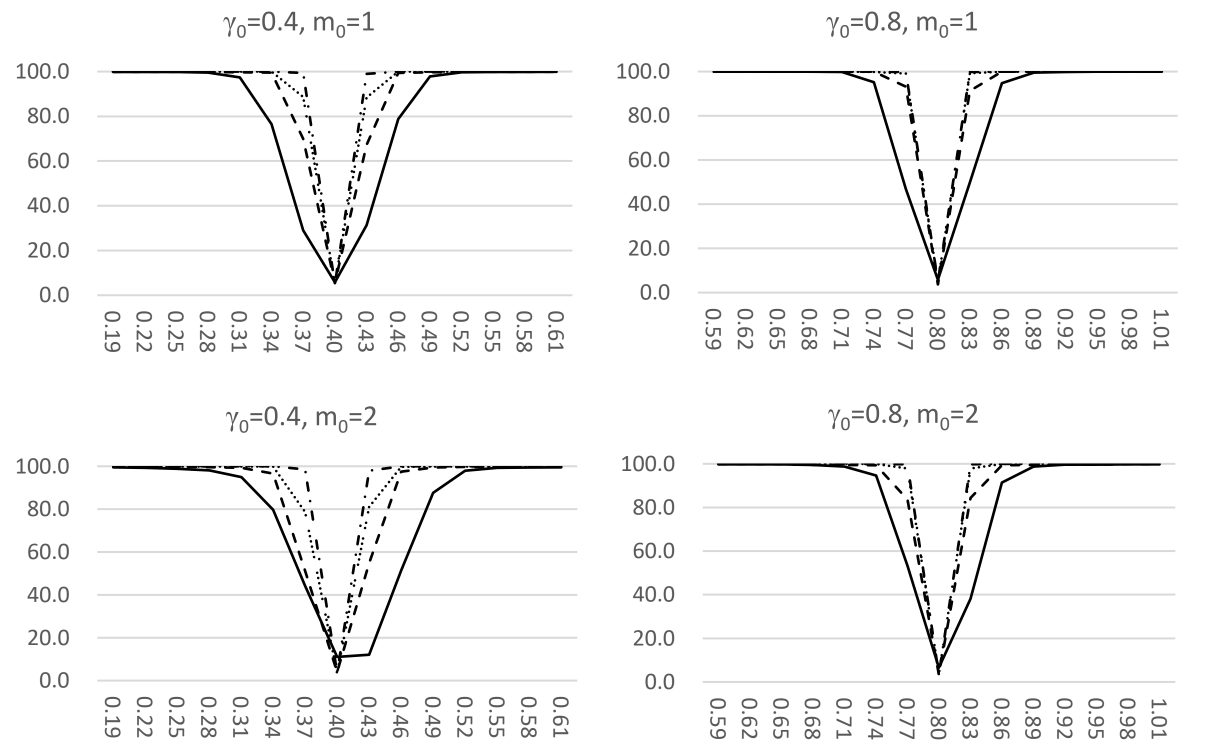
Note: N=100 N=300 N=500 N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with α_N=p/N(T-2) and p=0.05; γ is the coefficient of the lagged dependent variable given in (1) in the absence of the $\mathbf{x}_{it}$ regressors. See also the note to Table 1.

Figure A3(ii): Power functions for estimation of γ in the ARX(1) model with different values of m and N (κ^2=0.25)

Panel A: T=5



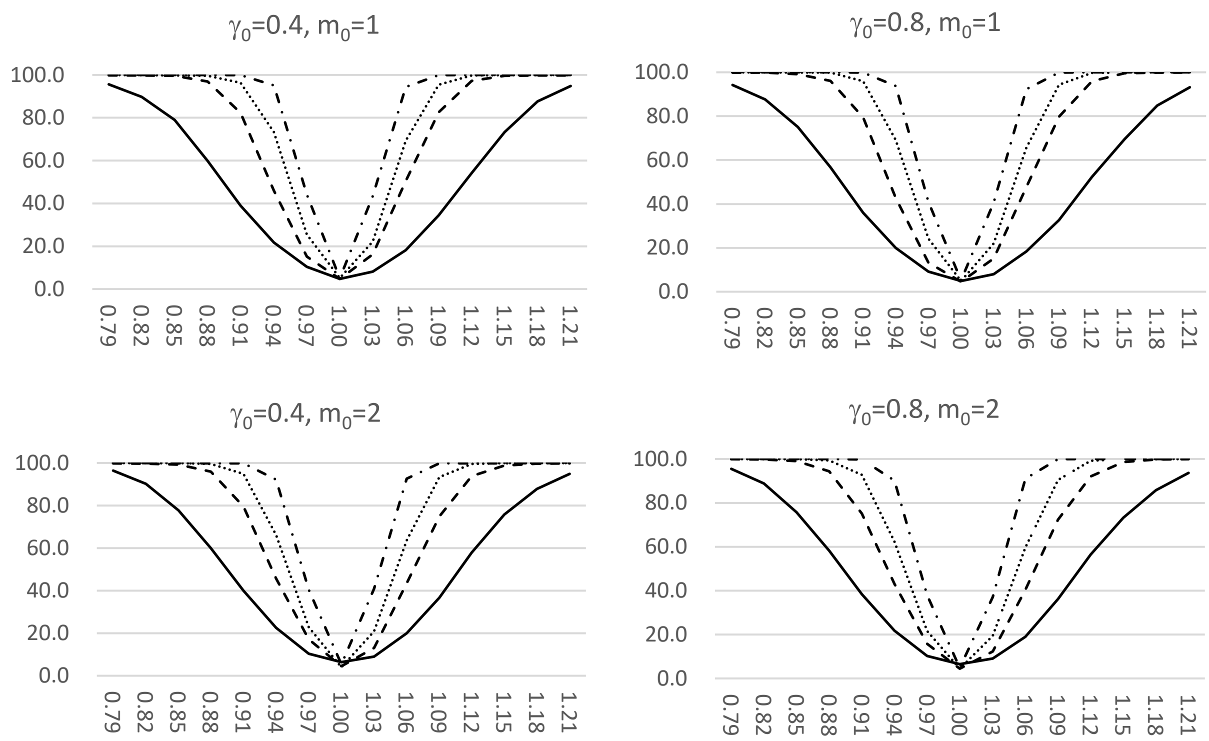
Panel B: T=10



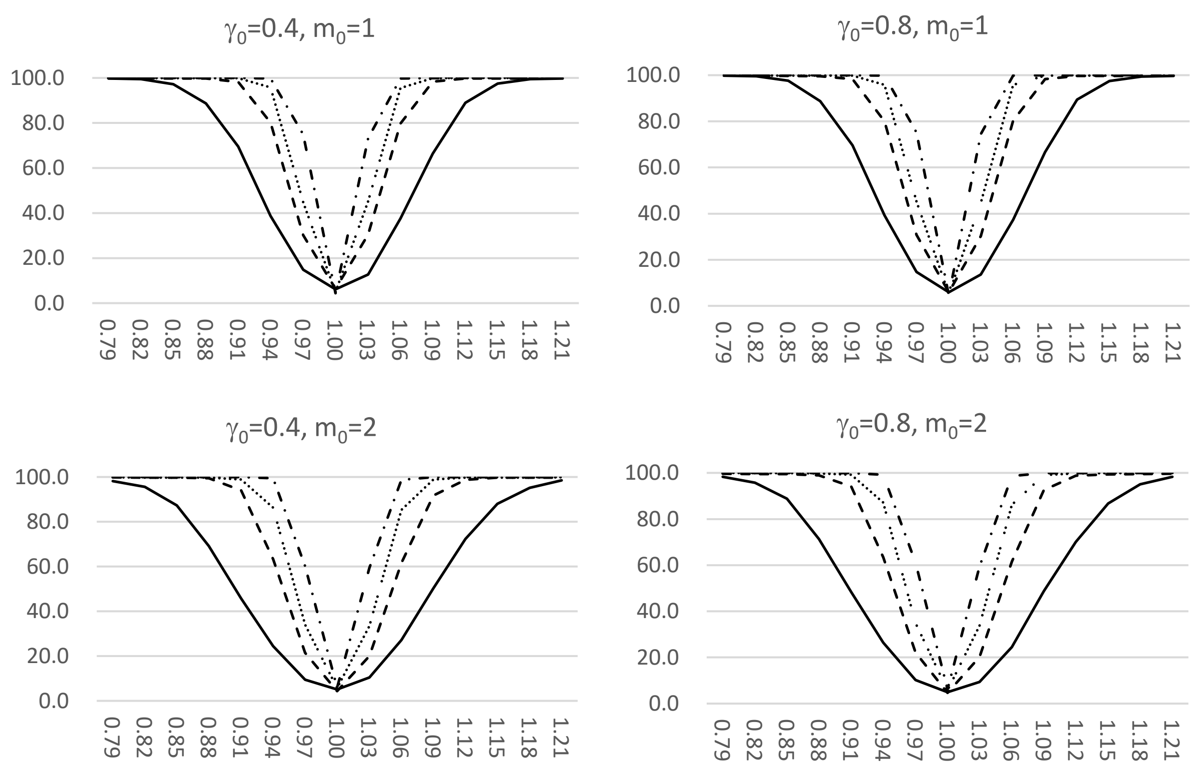
Note: —— N=100 - - - - N=300 .... N=500 -·-·- N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with α_N=p/N(T-2) and p=0.05; γ and β are the coefficients of the lagged dependent variable and the x_{it} regressor given in (1). See also the note to Table 1.

Figure A3(iii): Power functions for estimation of β in the ARX(1) model with different values of m and N (κ^2=0.25)

Panel A: T=5



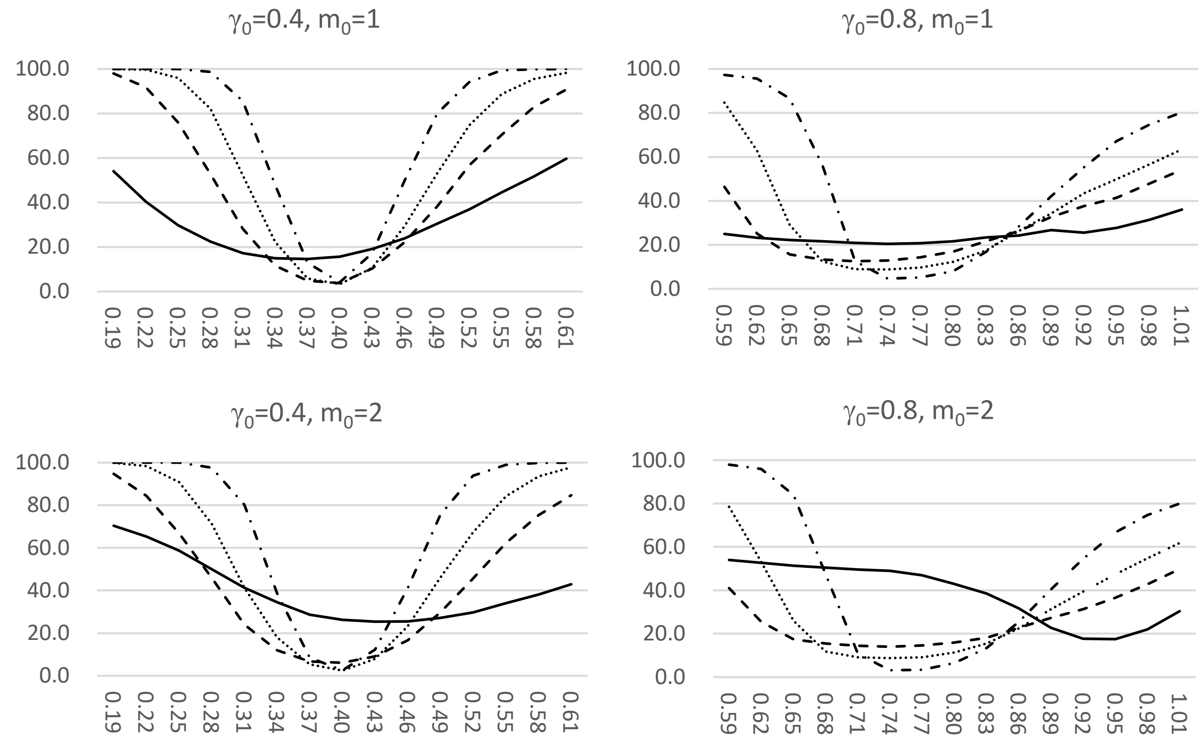
Panel B: T=10



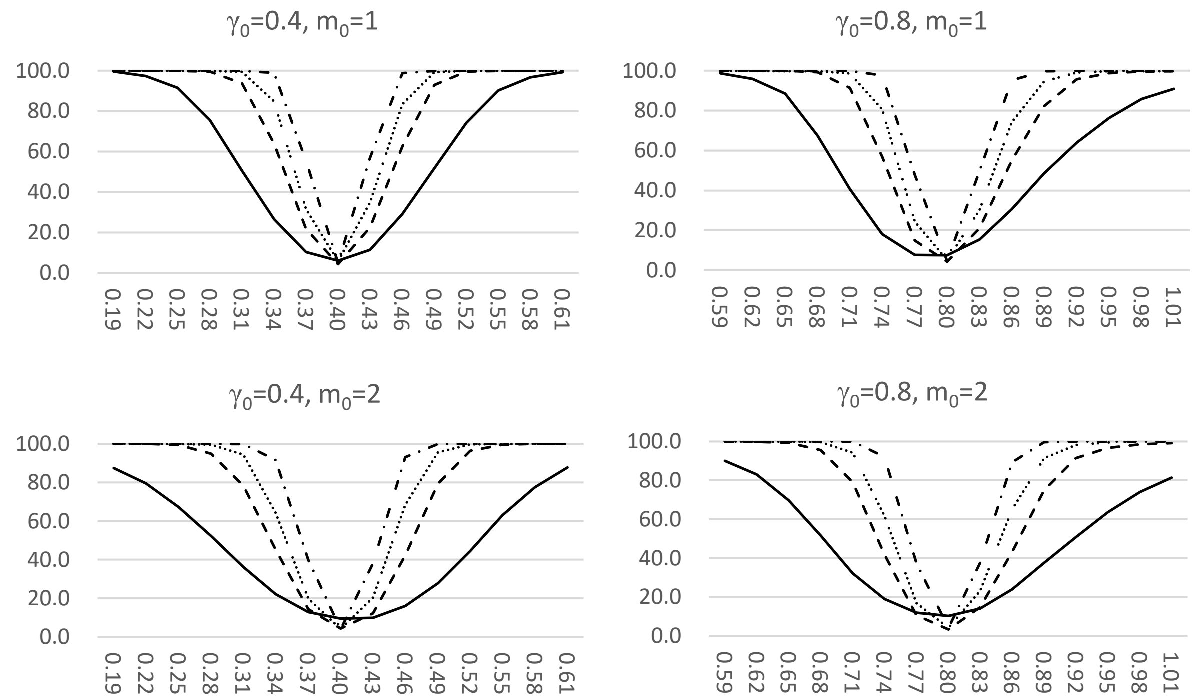
Note: —— N=100 - - - - N=300 .... N=500 -·-·- N=1000. See also the note to Figure A3(ii).

Figure A3(iv): Power functions for estimation of γ in the AR(1) model with different values of m and N (κ^2=0.5)

Panel A: T=5



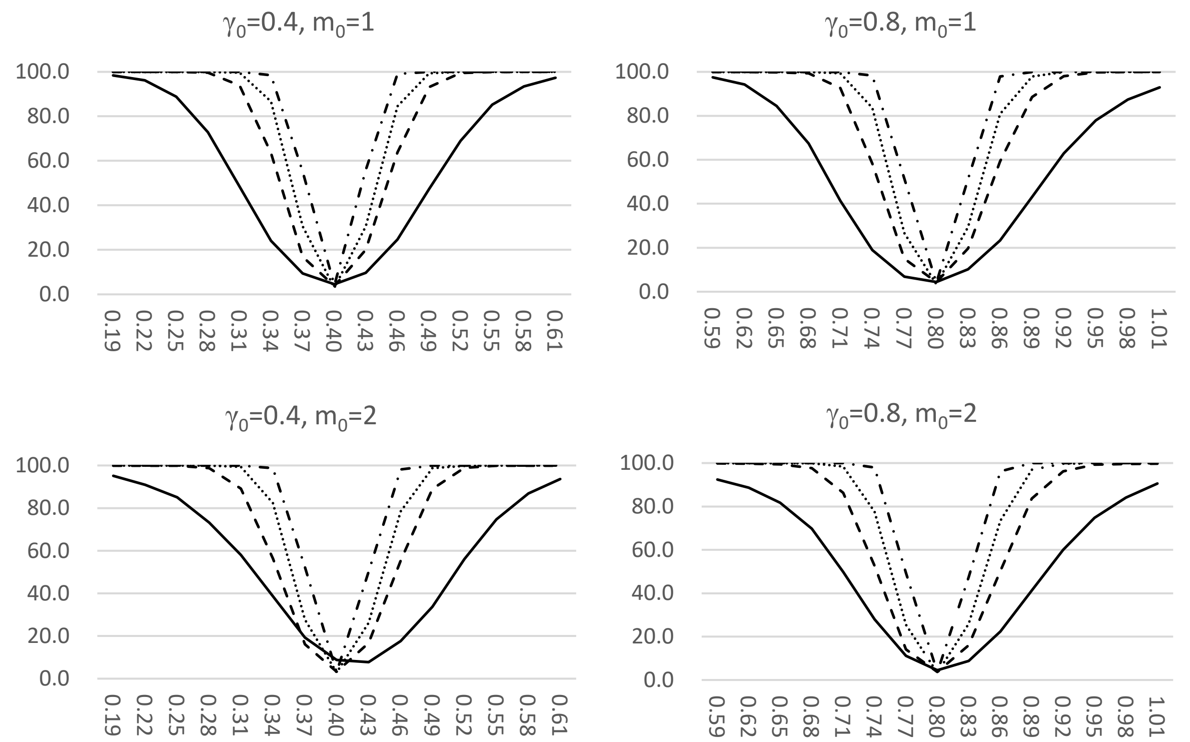
Panel B: T=10



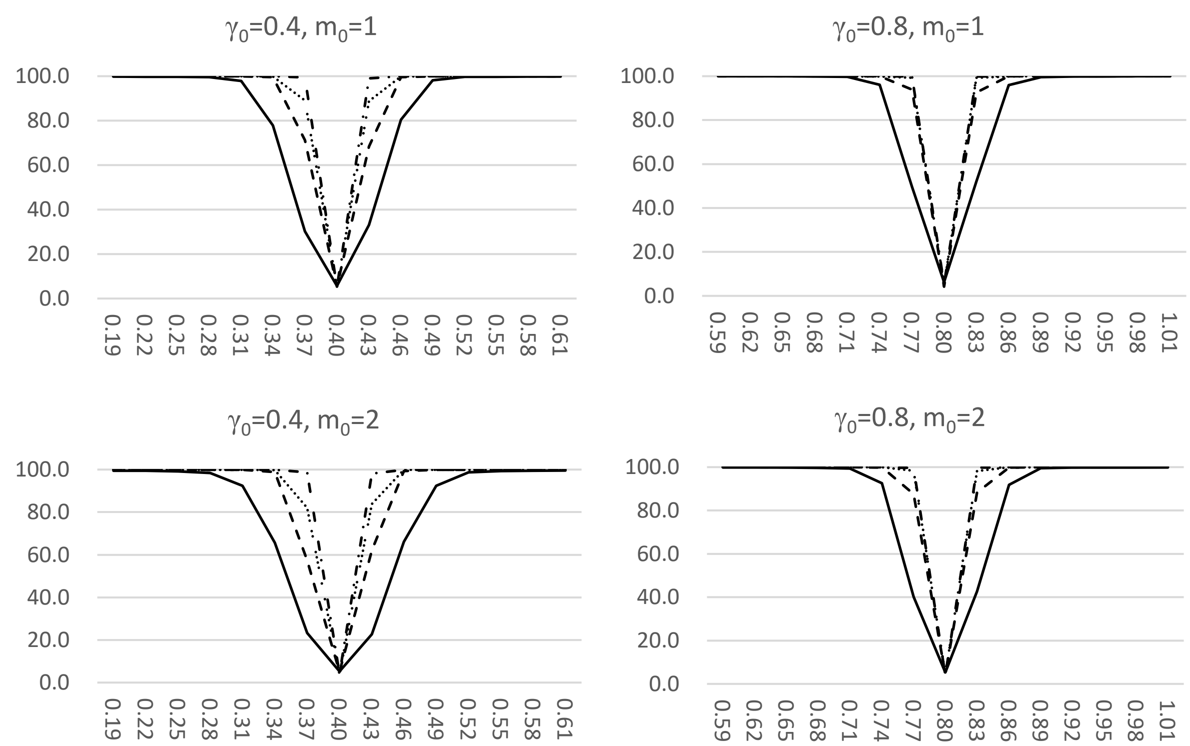
Note: N=100 N=300 N=500 N=1000. See also the note to Table A3(i).

Figure A3(v): Power functions for estimation of γ in the ARX(1) model with different values of m and N (κ^2=0.5)

Panel A: T=5



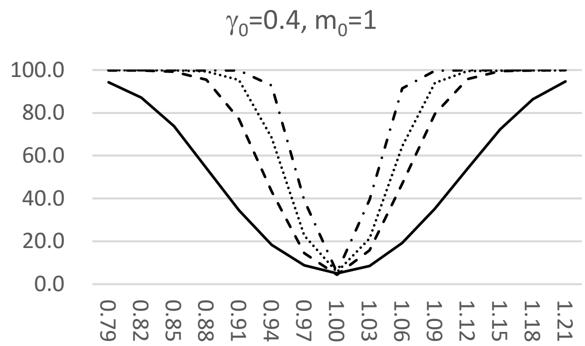
Panel B: T=10

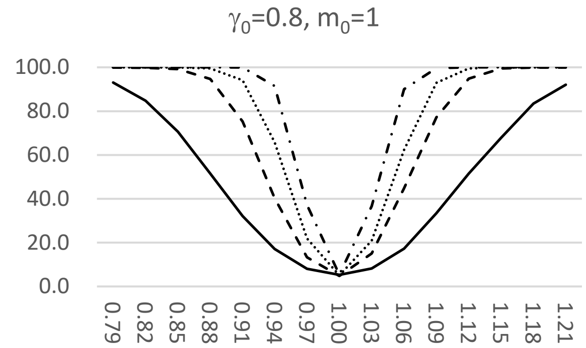


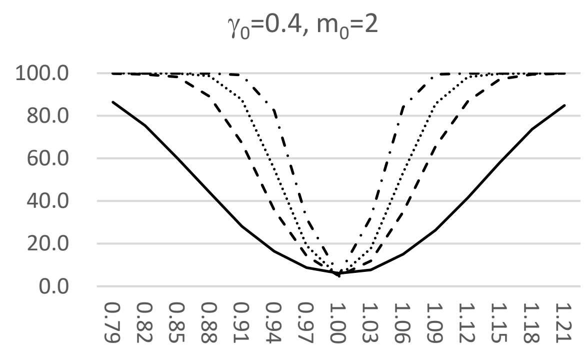
Note: N=100 N=300 N=500 N=1000. See also the note to Table A3(ii).

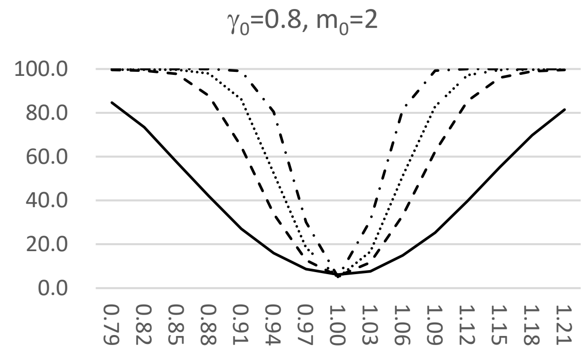
Figure A3(vi): Power functions for estimation of β in the ARX(1) model with different values of m and N (κ^2=0.5)

Panel A: T=5

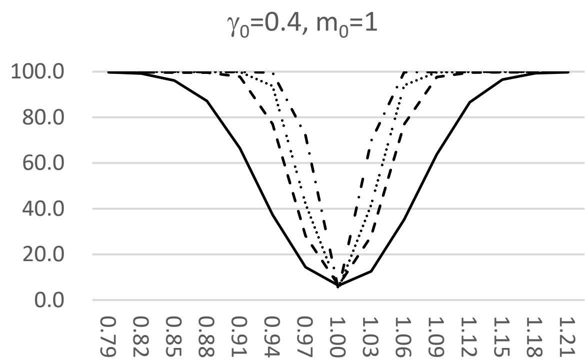


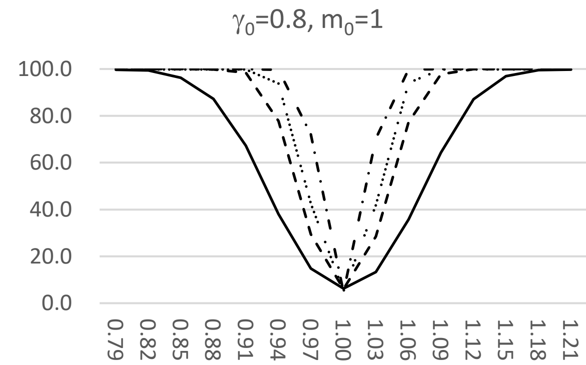


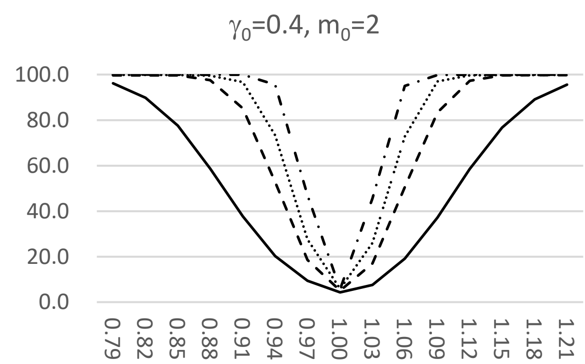


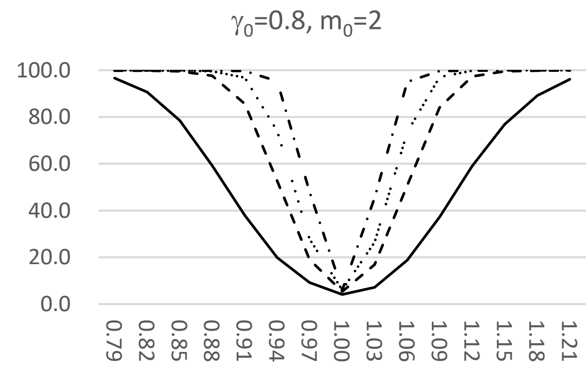


Panel B: T=10





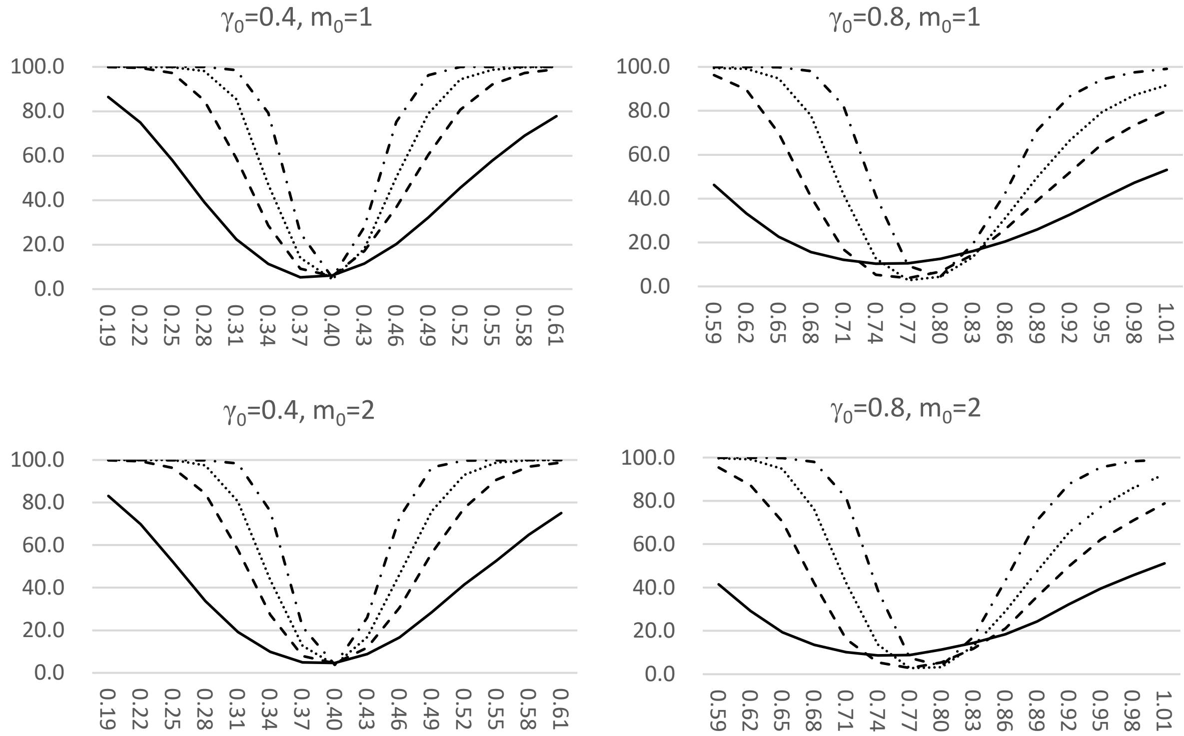




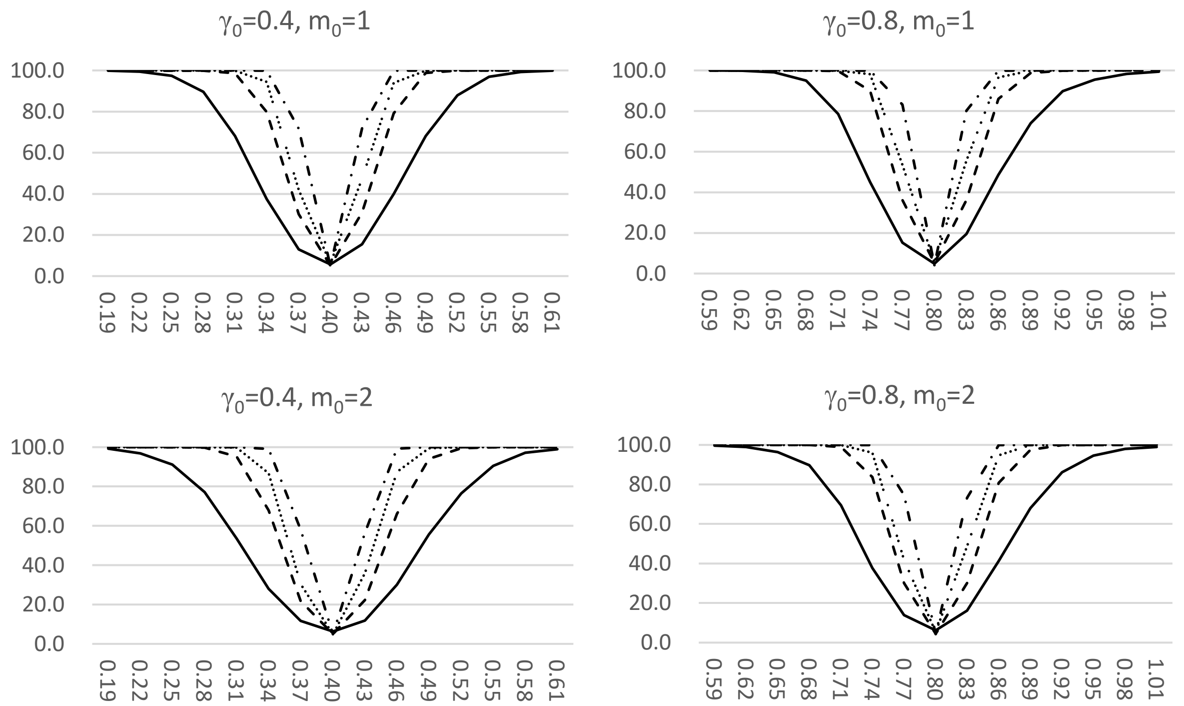
Note: N=100 N=300 N=500 N=1000. See also note to Figure A3(v).

Figure A3(vii): Power functions for estimation of γ in the AR(1) model with different values of m and N (κ^2=2)

Panel A: T=5



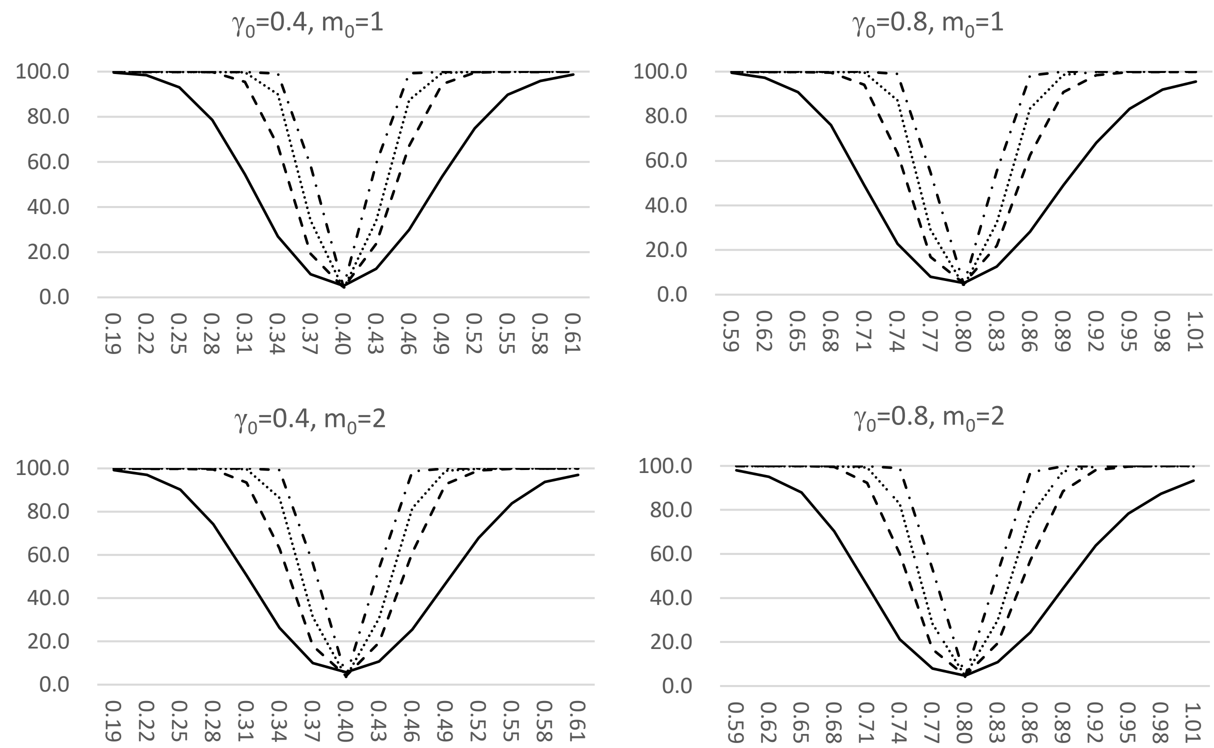
Panel B: T=10



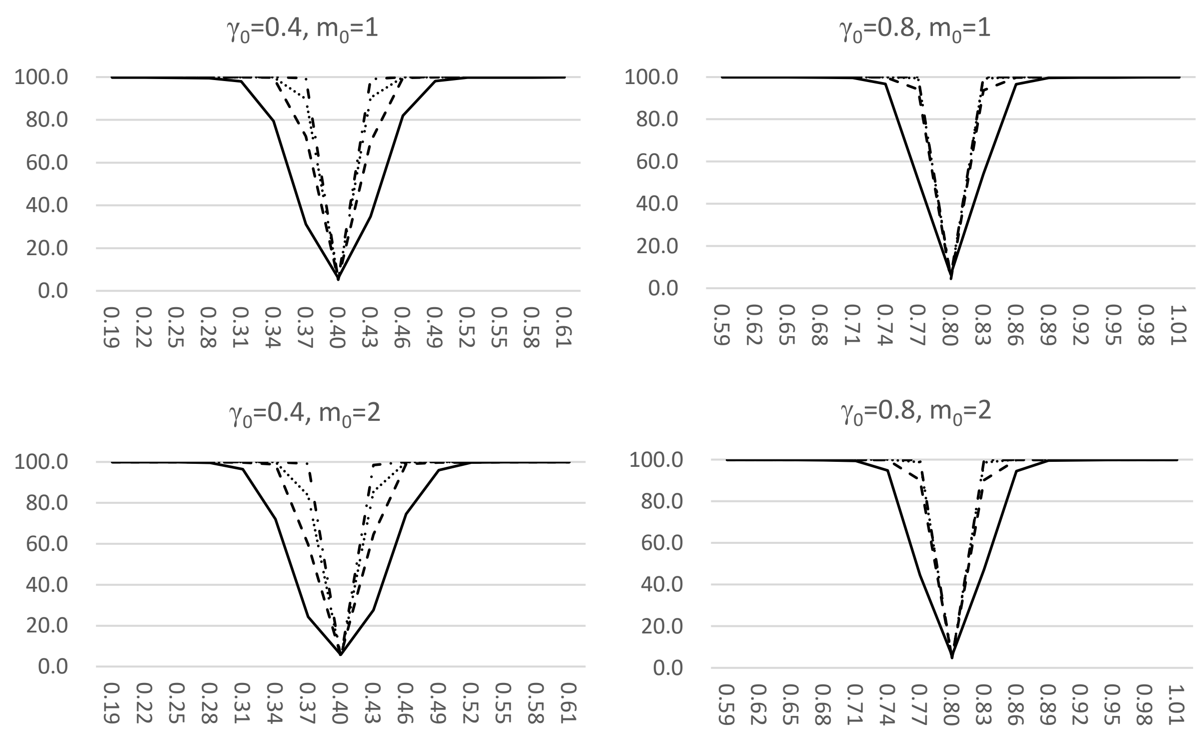
Note: N=100 N=300 N=500 N=1000. See also the note to Table A3(i).

Figure A3(viii): Power functions for estimation of γ in the ARX(1) model with different values of m and N (κ^2=2)

Panel A: T=5



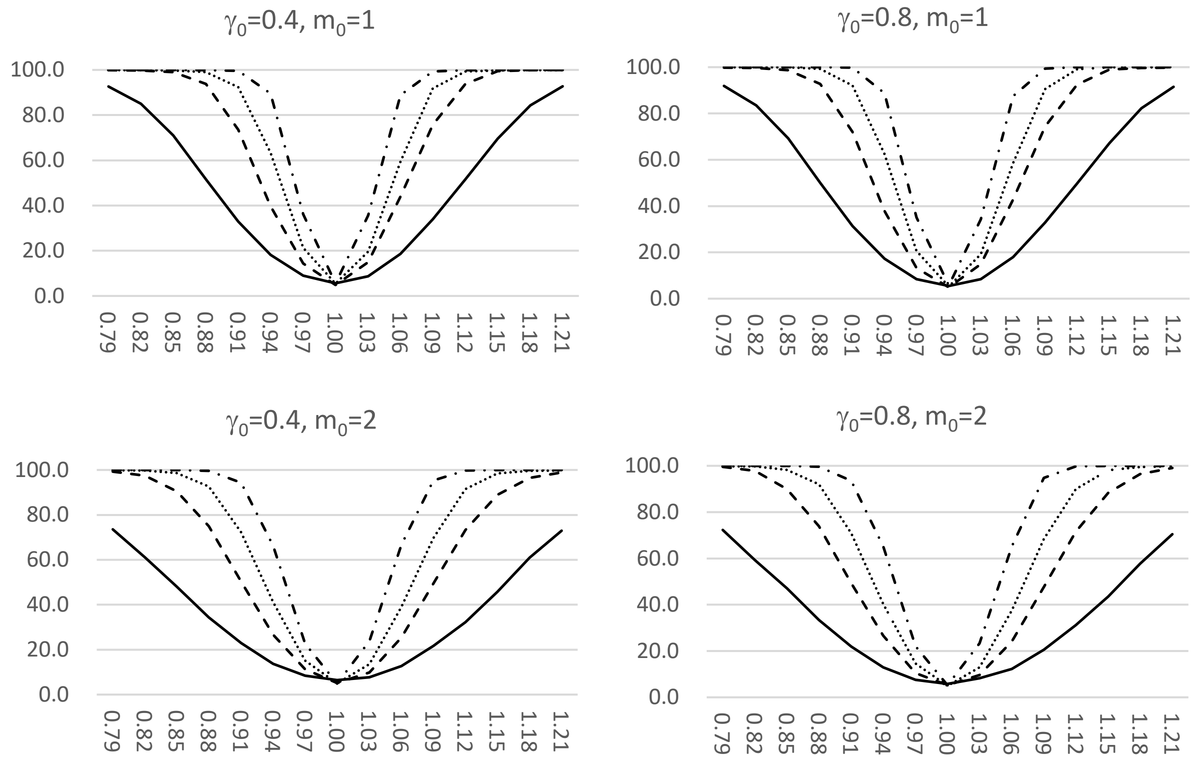
Panel B: T=10



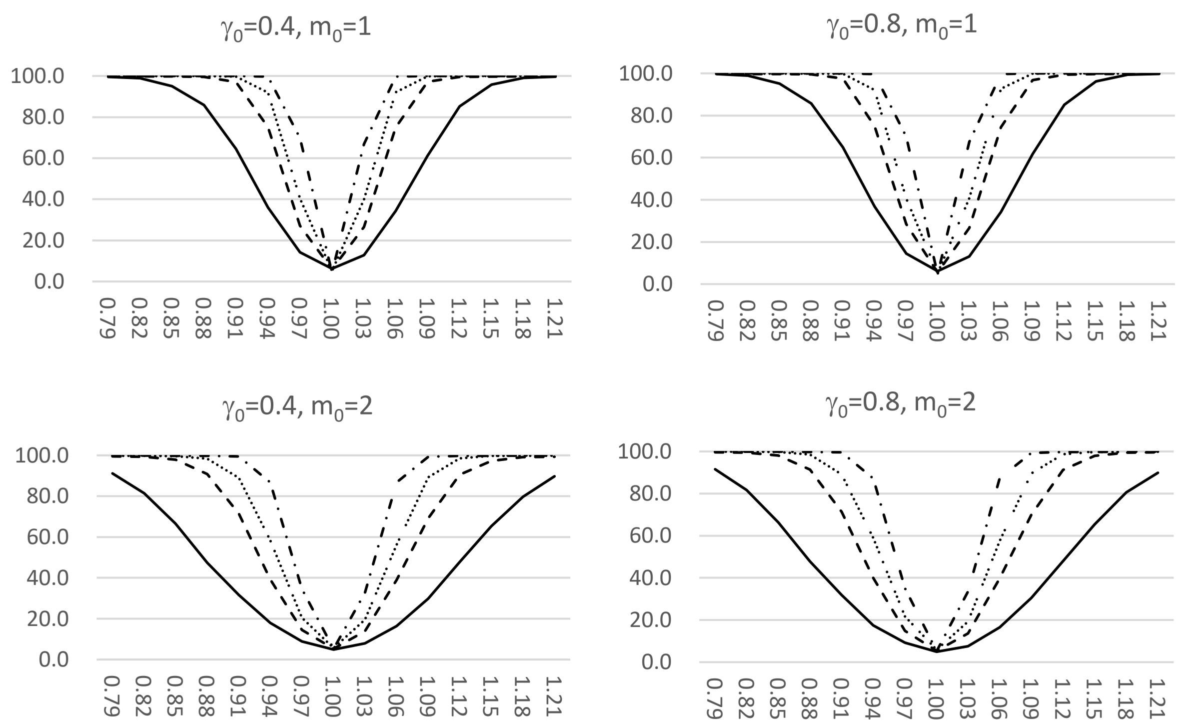
Note: —— N=100 - - - - N=300 ········ N=500 -·-·- N=1000. See also the note to Table A3(ii).

Figure A3(ix): Power functions for estimation of β in the ARX(1) model with different values of m and N (κ^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.6 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)

κ^2	0.25			0.5			1			2		
m_0	0	1	2	0	1	2	0	1	2	0	1	2
N	$T=5$											
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
	$T=10$											
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: $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 with $\alpha_N = \frac{p}{N(T-2)}$ and $p = 0.05$. See also the note to Table 6.

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)

	$T=5$																	
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
	$\kappa^2=1$									$\kappa^2=2$								
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
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
	$T=10$																	
	$\kappa^2=0.25$									$\kappa^2=0.5$								
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
	$\kappa^2=1$									$\kappa^2=2$								
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
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

B2(i): Bias(×100), RMSE(×100) and Size(×100) of γ for the AR(1) model, using the estimated number of factors, $\widehat{m}$

	Bias(×100)				RMSE(×100)				Size(×100)			
κ^2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
					$T = 5$							
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
					$T = 10$							
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 (1) in the absence of the $\mathbf{x}_{it}$ regressors. See also the note to Table B1(i).

Table B2(ii): Bias(×100), RMSE(×100) and Size(×100) of γ and β for the ARX(1) model, using the estimated number of factors, $\widehat{m}$ ($\sigma_v^2 = 1$)

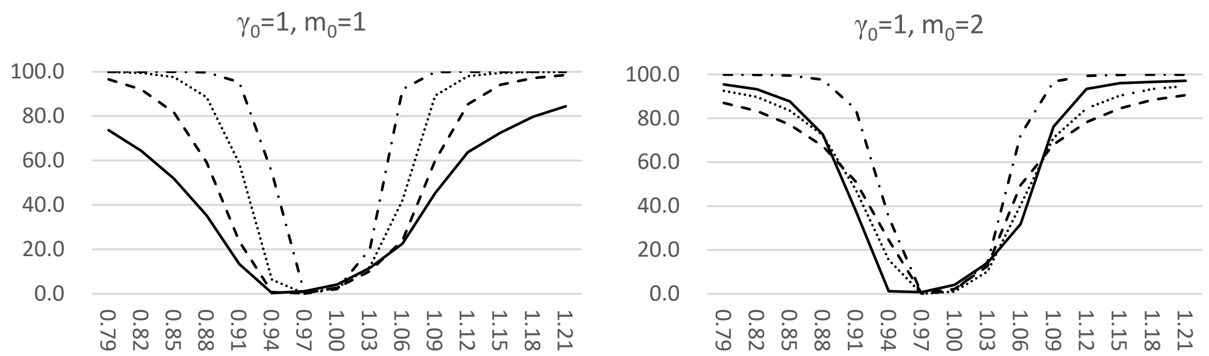
	Bias(×100)				RMSE(×100)				Size(×100)			
κ^2	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
					$T = 5$							
	γ											
N	$m_0 = 0$											
100	-1.28	-1.28	-1.28	-1.28	2.17	2.17	2.17	2.17	3.7	3.7	3.7	3.7
300	-0.77	-0.77	-0.77	-0.77	1.27	1.27	1.27	1.27	3.4	3.4	3.4	3.4
500	-0.58	-0.58	-0.58	-0.58	0.94	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
	$m_0 = 1$											
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
	$m_0 = 2$											
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
	$m_0 = 1$											
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
	$m_0 = 2$											
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
					$T = 10$							
	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
	$m_0 = 1$											
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
	$m_0 = 2$											
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
	$m_0 = 1$											
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
	$m_0 = 2$											
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 (1). See also the note to Table B1(i).

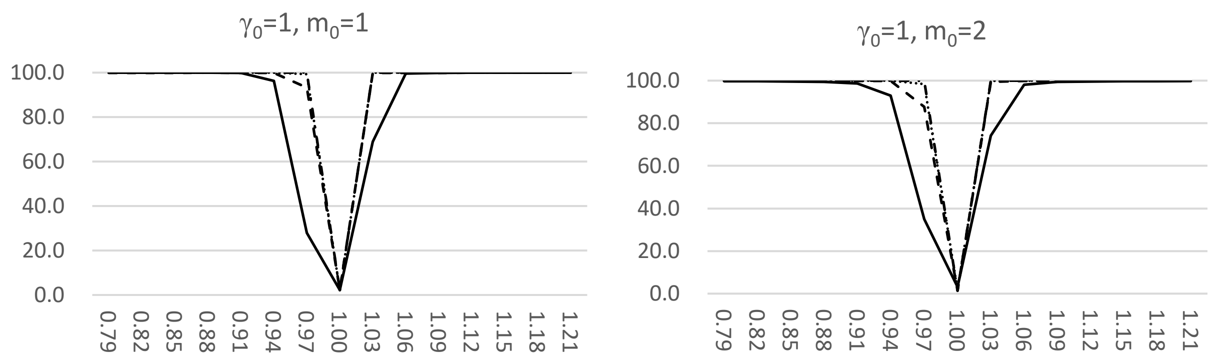
B3: Power Functions

Figure B3(i): Power functions for estimation of γ in the AR(1) model with different values of m and N (κ^2=0.25)

Panel A: T=5



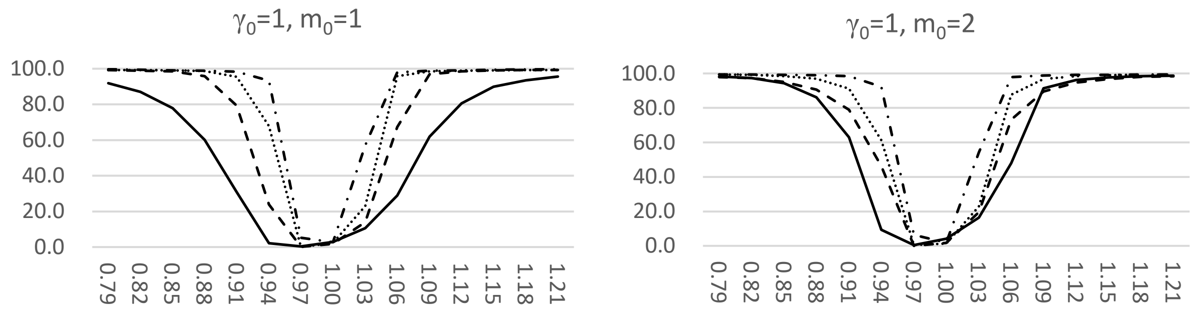
Panel B: T=10



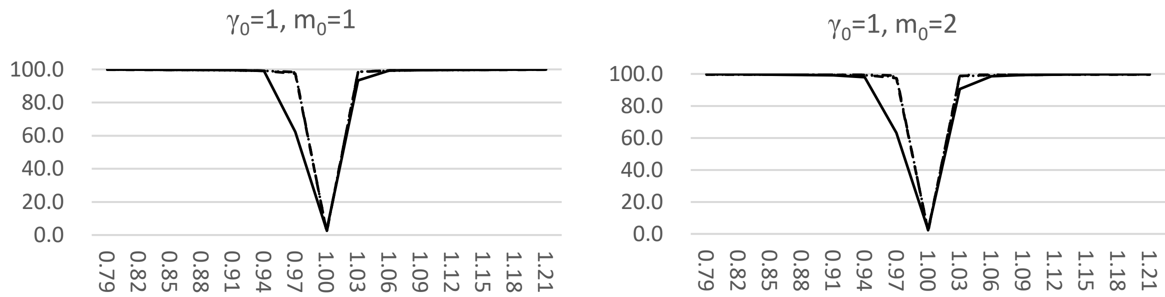
Note: —— N=100 - - - - N=300 ········ N=500 -·-·- N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with α_N=p/N(T-2) and p=0.05; γ is the coefficient of the lagged dependent variable given in (1) in the absence of the $\mathbf{x}_{it}$ regressors. See also the note to Table 4.

Figure B3(ii): Power functions for estimation of γ in the ARX(1) model with different values of m and N (κ^2=0.25)

Panel A: T=5



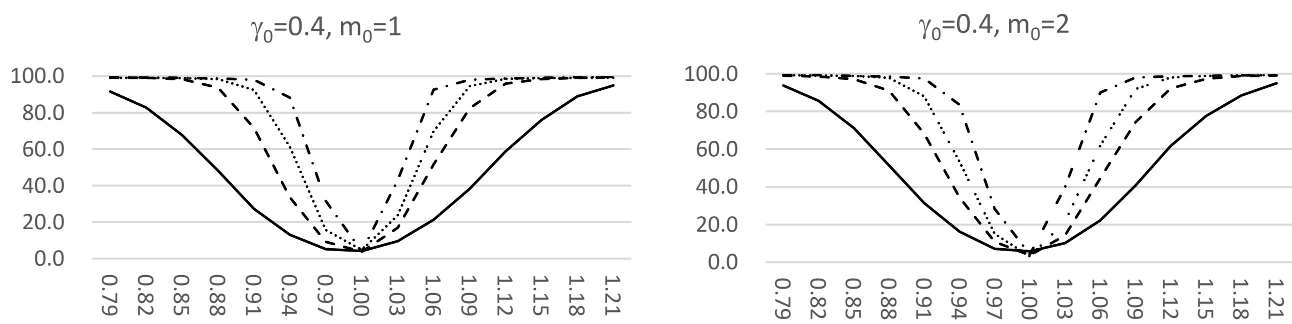
Panel B: T=10



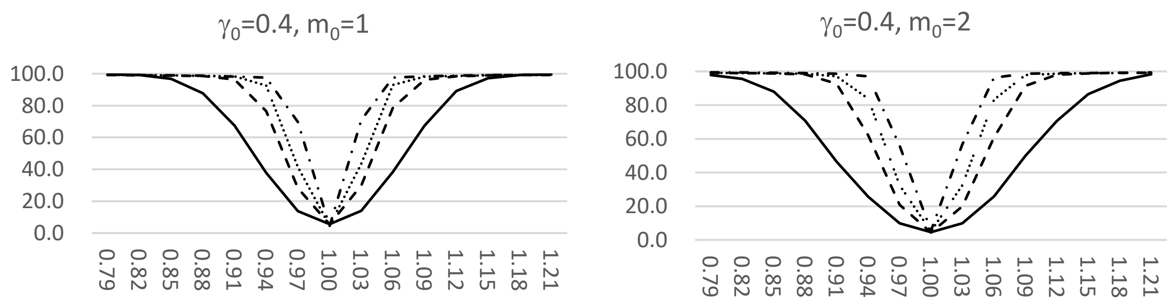
Note: —— N=100 - - - - N=300 ········ N=500 -·-·- N=1000. is estimated using the sequential MTLR procedure described in Section 6.1 with $\alpha_N=p/N(T-2)$ and p=0.05; γ and β are the coefficients of the lagged dependent variable and the x_{it} regressor given in (1). See also the note to Table 4.

Figure B3(iii): Power functions for estimation of β in the ARX(1) model with different values of m and N (κ^2=0.25)

Panel A: T=5



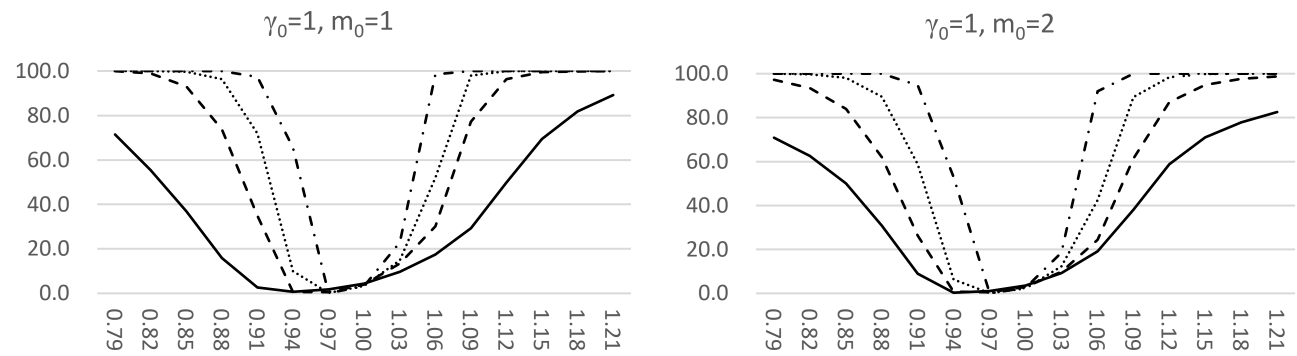
Panel B: T=10



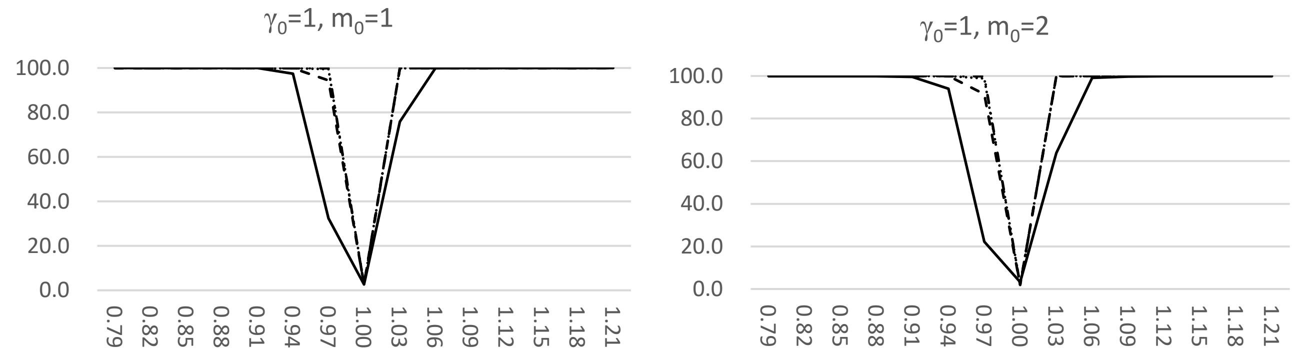
Note: —— N=100 - - - - N=300 ········ N=500 -·-·- N=1000. See also the note to Figure B3(ii).

Figure B3(iv): Power functions for estimation of γ in the AR(1) model with different values of m and N (κ^2=0.5)

Panel A: T=5



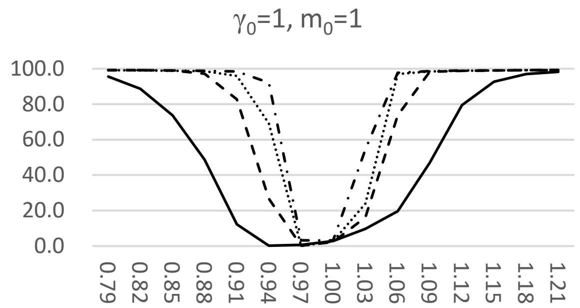
Panel B: T=10

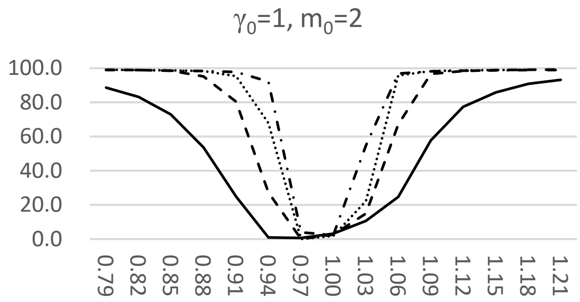


Note: N=100 N=300 N=500 N=1000. See also the note to Figure B3(i).

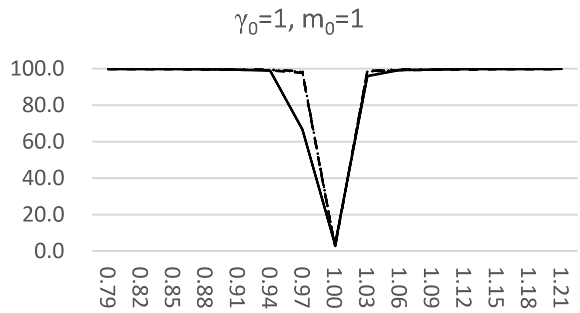
Figure B3(v): Power functions for estimation of γ in the ARX(1) model with different values of m and N (κ^2=0.5)

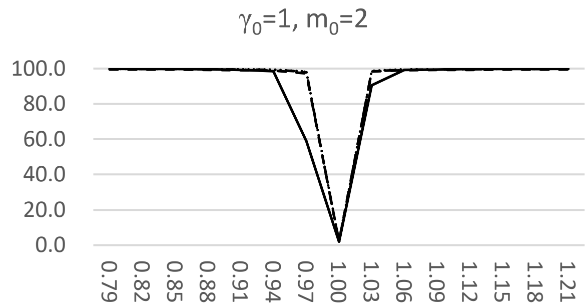
Panel A: T=5





Panel B: T=10

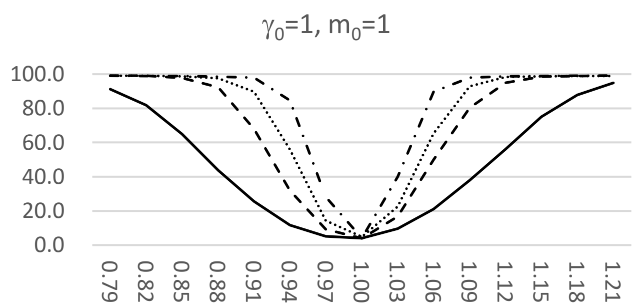


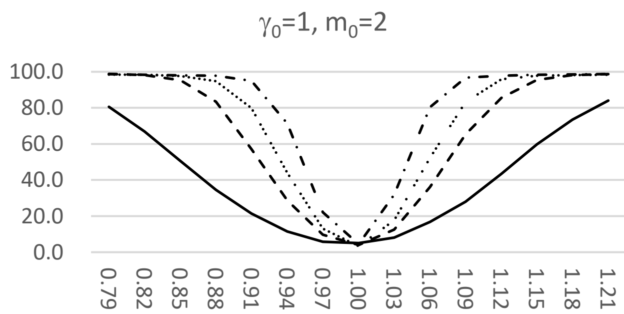


Note: —— N=100 - - - - N=300 ········· N=500 -·-·- N=1000. See also note to Figure B3(ii).

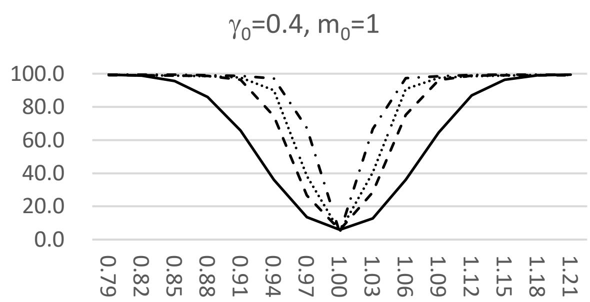
Figure B3(vi): Power functions for estimation of β in the ARX(1) model with different values of m and N (κ^2=0.5)

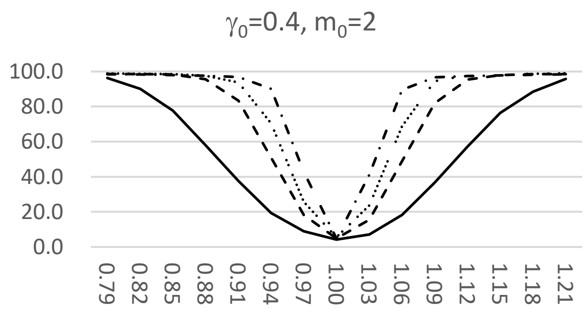
Panel A: T=5





Panel B: T=10

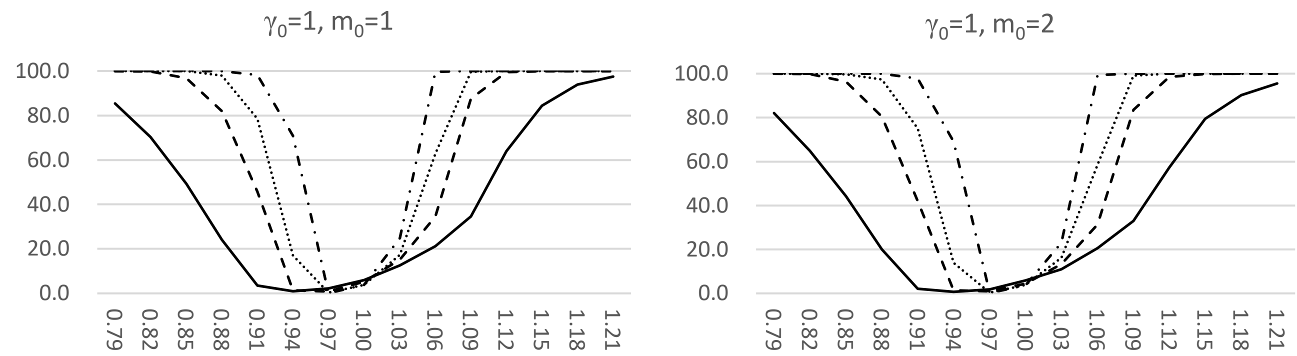




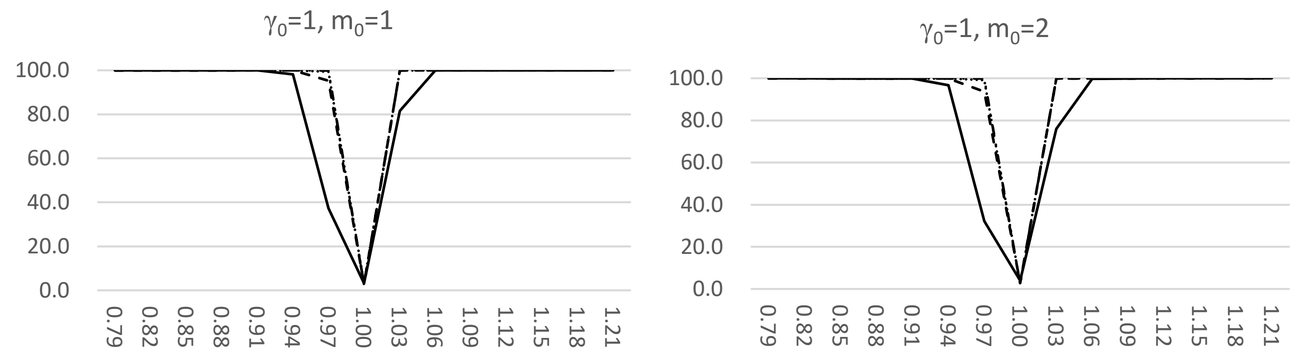
Note: —— N=100 - - - - N=300 ········· N=500 -·-·- N=1000. See also the note to Figure B3(v).

Figure B3(vii): Power functions for estimation of γ in the AR(1) model with different values of m and N (κ^2=2)

Panel A: T=5



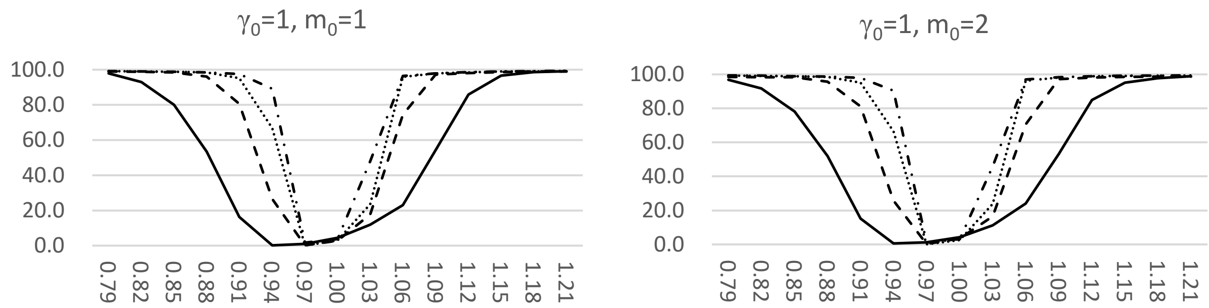
Panel B: T=10



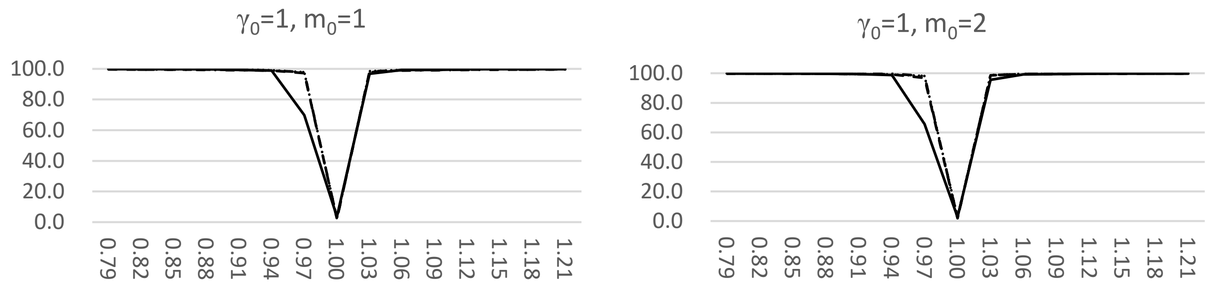
Note: N=100 N=300 N=500 N=1000. See also the note to Figure B3(i).

Figure B3(viii): Power functions for estimation of γ in the ARX(1) model with different values of m and N (κ^2=2)

Panel A: T=5



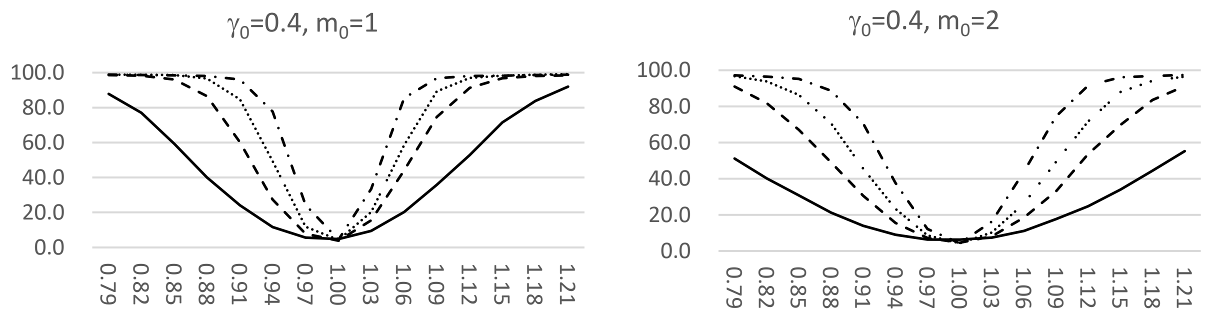
Panel B: T=10



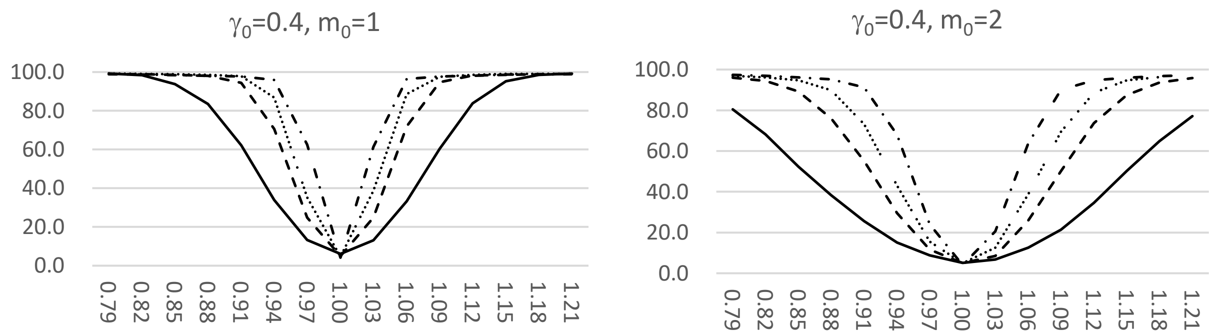
Note: ——N=100 - - - - N=300 .... N=500 -·-·- N=1000. See also the note to Figure B3(ii).

Figure B3(ix): Power functions for estimation of β in the ARX(1) model with different values of m and N (κ^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.7 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 ($\sigma_{\mathrm{v}}^2 = 1$, $\kappa^2 = 1$)

	$T=5$						$T=10$					
m_0	0	1	2	0	1	2	0	1	2	0	1	2
	$\gamma_0=0.4$			$\gamma_0=0.8$			$\gamma_0=0.4$			$\gamma_0=0.8$		
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 $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}$, $\zeta_{it} = \boldsymbol{\eta}_i' \mathbf{f}_t + u_{it}$ for $i = 1, 2, ..., N; t = 1, ..., T$ 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) model, $\beta = 0$. $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 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) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0=0.4$			$\gamma_0=0.8$			$\gamma_0=0.4$			$\gamma_0=0.8$		
N	$m_0=0$											
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 (1) 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) model, using the estimated number of factors, $\widehat{m}$ ($\sigma_{\mathrm{v}}^2 = 1$, $\kappa^2 = 1$)

	$T = 5$						$T = 10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
	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 (1). 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 ($\sigma_{\mathrm{v}}^2 = 1$, $\kappa^2 = 1$)

	$T = 5$						$T = 10$					
m_0	0	1	2	0	1	2	0	1	2	0	1	2
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
N	AR(1)											
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: $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 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) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
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 (1) 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 γ and β for the ARX(1) model, using the estimated number of factors, $\widehat{m}$ ($\sigma_{\mathrm{v}}^2 = 1$, $\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
	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 (1). 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 ($\sigma_v^2 = 1$, $\kappa^2 = 1$)

	$T=5$						$T=10$					
m_0	0	1	2	0	1	2	0	1	2	0	1	2
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
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
	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: $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 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) model, using the estimated number of factors, $\widehat{m}$ ($\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)	Bias ($\times 100$)	RMSE ($\times 100$)	Size ($\times 100$)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
N	$m_0 = 0$											
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 (1) in the absence of the $\mathbf{x}_{it}$ regressors. See also the note to Table C2(iv).

Table C2(vi): Bias(×100), RMSE(×100) and Size (×100) of γ and β for the ARX(1) model, using the estimated number of factors, $\widehat{m}$ ($\sigma_v^2 = 1$, $\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias (×100)	RMSE (×100)	Size (×100)	Bias (×100)	RMSE (×100)	Size (×100)	Bias (×100)	RMSE (×100)	Size (×100)	Bias (×100)	RMSE (×100)	Size (×100)
	$\gamma_0=0.4$			$\gamma_0=0.8$			$\gamma_0=0.4$			$\gamma_0=0.8$		
γ												
N $m_0=0$												
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 (1). See also the note to Table C2(iv).

C3: Correlation of factor loadings and regressors

In this experiment we allow the factor loadings $\boldsymbol{\eta}_i$ in the Monte Carlo design outlined in Section 7.1 to be correlated with the regressors x_{it} according to

$$\eta_{i\ell} = \kappa\sqrt{\frac{2}{m_0}}\left[\left(\sqrt{T}\bar{\mathrm{v}}_i/\sigma_{\mathrm{v}}\right) + \upsilon_{i\ell}\right], \text{ for } \ell = 1, 2, ..., m_0 \tag{S.39}$$

where $\bar{\mathrm{v}}_i = T^{-1}\sum_{t=1}^{T}\mathrm{v}_{it}$, with v_{it} representing the idiosyncratic component of x_{it}, defined by (41), and $\upsilon_{i\ell} \sim IID\mathcal{N}(0,1)$, for $\ell = 1, 2, ..., m_0$. 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 7.1.

Table C3(i): Empirical frequency of correctly selecting the true number of factors, m_0, using the sequential MTLR procedure ($\sigma_{\text{v}}^2 = 1$, $\kappa^2 = 1$)

	$T=5$				$T=10$			
m_0	1	2	1	2	1	2	1	2
	$\gamma_0 = 0.4$		$\gamma_0 = 0.8$		$\gamma_0 = 0.4$		$\gamma_0 = 0.8$	
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 $i = 1, 2, ..., N; t = 1, ..., T$ with $y_{i0} = \mu_{i0} + \sigma_{i0}(u_{i0}/\sigma)$. The factor loadings are generated as $\eta_{i\ell} = \kappa\sqrt{\frac{2}{m_0}}\left[\left(\sqrt{T}\bar{\text{v}}_i/\sigma_{\text{v}}\right) + \upsilon_{i\ell}\right]$, for $\ell = 1, 2, ..., m_0$ where $\bar{\text{v}}_i = T^{-1}\sum_{t=1}^{T}\text{v}_{it}$, and $\upsilon_{i\ell} \sim IID\mathcal{N}(0,1)$, for $\ell = 1, 2, ..., m_0$. In the case of the AR(1) model, $\beta = 0$. $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 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$100) of γ for the AR(1) model, using the estimated number of factors, m, and the true number, m_0 ($\kappa^2 = 1$)

	$T=5$						$T=10$					
	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
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(×100), RMSE(×100) and Size(×100) of γ and β for the ARX(1) model, using the estimated number of factors, m, and the true number, m_0 ($\sigma_v^2 = 1$, $\kappa^2 = 1$)

		$T=5$						$T=10$					
	Bias (×100)	RMSE (×100)	Size (×100)	Bias (×100)	RMSE (×100)	Size (×100)	Bias (×100)	RMSE (×100)	Size (×100)	Bias (×100)	RMSE (×100)	Size (×100)
	$\gamma_0=0.4$			$\gamma_0=0.8$			$\gamma_0=0.4$			$\gamma_0=0.8$		
	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 (1). 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 7.1 to follow a first-order spatial autoregressive process defined by

$$\boldsymbol{\eta}_\ell = a\mathbf{W}\boldsymbol{\eta}_\ell + \mathbf{e}_\ell, \quad \ell = 1, 2, ..., m_0, \tag{S.40}$$

where $\boldsymbol{\eta}_\ell = (\eta_{1\ell}, \eta_{2\ell}, ..., \eta_{N\ell})'$,

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 1/2 & 0 & 1/2 & 0 & & 0 \\ 0 & 1/2 & 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \ddots & 1/2 & 0 \\ \vdots & & & 1/2 & 0 & 1/2 \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}, \tag{S.41}$$

and $\mathbf{e}_\ell = (e_{1\ell}, e_{2\ell}, ..., e_{N\ell})'$. For each i and ℓ, $e_{i\ell}$ are drawn as $IIDN(0, \sigma^2_{e\ell})$. To ensure $N^{-1}\sum_{i=1}^N Var(\eta_{i\ell}) = \frac{\kappa^2}{m_0}$, for $\ell = 1, 2, ..., m_0$ (which corresponds to the case of cross-sectionally independent factor loadings) we set

$$\sigma^2_{e\ell} = \left(\frac{\kappa^2}{m_0}\right)\left\{\frac{N}{\mathrm{tr}\left[(\mathbf{I}_N - a\mathbf{W})^{-1}(\mathbf{I}_N - a\mathbf{W}')^{-1}\right]}\right\}. \tag{S.42}$$

The rest of the parameters are as described in Section 7.1.

Table C4(i): Empirical frequency of correctly selecting the true number of factors, m_0, using the sequential MTLR procedure ($\sigma_{\text{v}}^2 = 1$, $\kappa^2 = 1$)

	$T=5$				$T=10$			
m_0	1	2	1	2	1	2	1	2
	$\gamma_0 = 0.4$		$\gamma_0 = 0.8$		$\gamma_0 = 0.4$		$\gamma_0 = 0.8$	
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 $y_{it} = \alpha_i + \delta_t + \gamma y_{i,t-1} + \beta x_{it} + \zeta_{it}$, $\zeta_{it} = \boldsymbol{\eta}_i'\mathbf{f}_t + u_{it}$, for $i = 1,2,...,N; t = 1,...,T$ with $y_{i0} = \mu_{i0} + \sigma_{i0}(u_{i0}/\sigma)$. The factor loadings $\boldsymbol{\eta}_\ell = (\eta_{1\ell}, \eta_{2\ell}, ..., \eta_{N\ell})'$ are generated as $\boldsymbol{\eta}_\ell = a\mathbf{W}\boldsymbol{\eta}_\ell + \mathbf{e}_\ell$, for $\ell = 1,2,...,m_0$, where $e_\ell = (e_{1\ell}, e_{2\ell}, ..., e_{N\ell})'$, with $a = 0.4$ and $\mathbf{W}$ is specified as in equation (S.41). For each i and ℓ, $e_{i\ell}$ are drawn as $IIDN(0, \sigma_{e\ell}^2)$ with $\sigma_{e\ell}^2 = \left(\frac{\kappa^2}{m_0}\right)\left\{N/\operatorname{tr}\left[(\mathbf{I}_N - a\mathbf{W})^{-1}(\mathbf{I}_N - a\mathbf{W}')^{-1}\right]\right\}$. In the case of the AR(1) model, $\beta = 0$. $\widehat{m}$ is estimated using the sequential MTLR procedure described in Section 6.1 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$)

	$T=5$						$T=10$					
	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)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
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
	$m_0 = 2$											
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 (1) 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$100) of γ and β for the ARX(1) model, using the estimated number of factors, m, and the true number, m_0 ($\sigma_v^2 = 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$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)	Bias ($\times$100)	RMSE ($\times$100)	Size ($\times$100)
	$\gamma_0 = 0.4$			$\gamma_0 = 0.8$			$\gamma_0 = 0.4$			$\gamma_0 = 0.8$		
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 (1). See also the note to Table C4(i).