UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

New results on asymptotic properties of likelihood estimators with persistent data for small and large T

Juodis, A.; Sarafidis, V.

DOI
10.1007/s13209-023-00286-y

Publication date
2023

Document Version
Final published version

Published in
SERIEs

License
CC BY

Link to publication

Citation for published version (APA):
Juodis, A., & Sarafidis, V. (2023). New results on asymptotic properties of likelihood estimators with persistent data for small and large *T*. *SERIEs*, *14*(3-4), 435-461. https://doi.org/10.1007/s13209-023-00286-y

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:11 10월 2026

SERIEs (2023) 14:435–461
https://doi.org/10.1007/s13209-023-00286-y

ORIGINAL ARTICLE

New results on asymptotic properties of likelihood estimators with persistent data for small and large *T*

Artūras Juodis[1,2] · **Vasilis Sarafidis**[3,4]

Received: 21 December 2022 / Accepted: 14 July 2023 / Published online: 3 August 2023
© The Author(s) 2023

Abstract

This paper revisits the panel autoregressive model, with a primary emphasis on the unit-root case. We study a class of misspecified Random effects Maximum Likelihood (mRML) estimators when T is either fixed or large, and N tends to infinity. We show that in the unit-root case, for any fixed value of T, the log-likelihood function of the mRML estimator has a single mode at unity as $N \to \infty$. Furthermore, the Hessian matrix of the corresponding log-likelihood function is non-singular, unless the scaled variance of the initial condition is exactly zero. As a result, mRML is consistent and asymptotically normally distributed as N tends to infinity. In the large-T setup, it is shown that mRML is asymptotically equivalent to the bias-corrected FE estimator of Hahn and Kuersteiner (Econometrica 70(4):1639–1657, 2002). Moreover, under certain conditions, its Hessian matrix remains non-singular.

Keywords Dynamic panel data · Maximum likelihood · Monte Carlo simulation

JEL Classification C13 · C33

We are grateful to the Guest Editor Stephane Bonhomme and an anonymous referee for constructive comments that have substantially improved this paper. We would also like to thank Manuel Arellano for his lifetime contributions to econometrics, which have served as an inspiration for the two authors to work in the field of panel data econometrics.

✉ Artūras Juodis
a.juodis@uva.nl

Vasilis Sarafidis
vasilis.sarafidis@bi.no

1 Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

2 Tinbergen Institute, Amsterdam, The Netherlands

3 Brunel University London, London, UK

4 BI Norwegian Business School, Oslo, Norway

1 Introduction

Dynamic panel data analysis is highly popular in econometrics for its ability to capture the dynamics of microeconomic agents (such as households and firms) using a limited number of time series observations. The prevalent approach in the literature has been the autoregressive (AR) panel model with individual-specific intercepts.

Early research on the estimation of the panel AR model utilized unconditional maximum likelihood estimators, treating individual-specific intercepts as random variables. See e.g., Balestra and Nerlove (1966) and Maddala (1971).

During the 1980s, however, a growing awareness of the significance of accounting for heterogeneity across entities led to the emergence of the fixed effects approach. This approach treats individual-specific intercepts as parameters, requiring fewer distributional assumptions compared to the random effects approach. Notwithstanding, a major challenge arises in that the number of parameters increases with the total cross-sectional observations (N).

An early popular way to address the "incidental parameters problem" involved transforming the data by subtracting individual-specific means and then running least-squares. The resulting Within Group (WG) estimator is a Maximum Likelihood (ML) estimator conditional on individual fixed effects. Unfortunately, in dynamic panels the within transformation induces a correlation between lagged dependent variables and idiosyncratic errors, which is non-negligible when T is fixed (Nickell 1981). Thus, the WG estimator is inconsistent as $N \to \infty$.

More recently, several alternative ML approaches have been proposed in the literature to deal with the incidental parameters problem. Many of these methods treat individual-specific effects as fixed but either rely on modifications of the profile likelihood, as in Lancaster (2002) and Dhaene and Jochmans (2016), or they start from the likelihood function of the model transformed in first differences, as in Hsiao et al. (2002) and Hayakawa and Pesaran (2015). Alternative likelihood-based estimators treat individual-specific effects as random variables but make use of Chamberlain-type projections to explicitly model the dependence between these effects and initial conditions (Anderson and Hsiao 1982; Alvarez and Arellano 2003; and Moral-Benito (2013)).

The present paper revisits the transformed maximum likelihood approach (TML) as in Hsiao et al. (2002), and the random effects maximum likelihood estimator (RML) as in Alvarez and Arellano (2003). In addition, we study a class of RML-type estimators that arises by misspecifying (i.e., imposing an incorrect value for) the correlation strength between the initial conditions and the individual-specific intercepts. Thus, the class of misspecified RML (mRML) estimators considered in this paper generalizes (Hahn et al. 2004), whose setting corresponds to the misspecified likelihood that imposes (potentially incorrectly) such correlation to be zero.

We mainly focus on the case where the data are highly persistent, that is, the autoregressive parameter equals unity. This case is important from an empirical point of view because many economic variables exhibit time series properties very close to

random walks. Examples arise in the estimation of production functions, household income and consumer spending, to mention a few.[1]

The contributions of the paper are as follows:

(i) Firstly, we show that in the unit root setting for any fixed value of T, the log-likelihood function of the mRML estimator has a single mode at unity as $N \to \infty$. It is also shown that the Hessian matrix of the corresponding log-likelihood function is non-singular, unless the scaled variance of the initial condition is exactly zero. As a result, the misspecified RML estimator for the autoregressive parameter is consistent and asymptotically normally distributed, as $N \to \infty$ for T fixed. This implies that standard inference procedures are valid. To the best of our knowledge, this is the first result in the literature that shows that a class of mRML estimators has desirable asymptotic properties in the unit-root case for fixed-T.

(ii) Secondly, the paper also provides new insights on the properties of mRML and TML in large N, T samples. In specific, in a stable autoregressive setting, we show that the mRML estimator is asymptotically equivalent to the bias-corrected FE estimator of Hahn and Kuersteiner (2002). This result complements that in Alvarez and Arellano (2003) and Hahn et al. (2004), who show asymptotic equivalence between the RML estimator and the bias-corrected FE estimator.

In a Monte Carlo study, we investigate how informative our asymptotic results are for the finite sample properties of all estimators considered. We find that this asymptotic characterization is only informative about the finite sample behavior of the estimators that have non-singular limiting Hessian matrices. This excludes the TML and RML estimators, which have singular Hessian matrices in the limit.

The remainder of this paper is structured as follows. The next section sets out the panel AR(1) model and specifies the underlying assumptions. Section 3 introduces the misspecified RML approach and links it to the TML and RML approaches. Section 4 provides the asymptotic results of the paper. Section 5 reports finite-sample results from a Monte Carlo study, and a final section concludes. Proofs of all propositions are provided in the Appendix.

2 The linear panel AR(1) model

We consider the following simple AR(1) specification without exogenous regressors[2]:

$$y_{i,t} = \alpha y_{i,t-1} + \eta_i + \varepsilon_{i,t}, \quad \mathrm{E}[\varepsilon_{i,t}|y_{i,0}, \eta_i] = 0, \tag{1}$$

for $i = 1, \ldots, N, t = 1, \ldots, T$ where the true parameter is $\alpha = \alpha_0$. For example, simple linear models of this form played a prominent role in the early econometric literature on the propagation of income shocks to consumption, by decomposing the income shocks into a permanent component (a function of η_i), and a transitory component (a function of $y_{i,t-1}$). Some recent contributions to this literature include

[1] See the discussion in Blundell and Bond (2023).

[2] Time-specific effects can be accommodated by taking the variables in deviations from the cross-sectional mean.

the papers of Botosaru and Sasaki (2018) and Botosaru (2022), while the paper of Arellano et al. (2017) provides the most recent advances in the literature on nonlinear models for incomes shocks.

In this paper, we limit our attention to the stylized setting in Eq. (1), where the parameters of interest α_0 and $\sigma_0^2 = \mathrm{E}[\varepsilon_{i,t}^2]$ can be estimated using the Maximum Likelihood principle. Our prime focus is the behavior of Maximum Likelihood estimators when $\alpha_0 \to 1$.

In this paper, we operate under the conditions of the following Random Effects ML (RML) assumption.

Assumption **RML**: For $\eta_i = (1-\alpha_0)\mu_i$, the vector $(y_{i,0}, \mu_i)$ is i.i.d. across i, with finite fourth moments. $\varepsilon_{i,t}$ is i.i.d. $(0, \sigma^2)$ across all i, t and $\mathrm{E}[\varepsilon_{i,t}^4]$ is finite.

Here we implicitly assume that initial conditions $y_{i,0}$ are observed by the econometrician and can be used in the formulation of the likelihood function. In particular, this assumption imposes restrictions on the joint distribution of $y_{i,0}$ and μ_i. Note that we do not impose any form of stationarity restrictions (mean and/or covariance) on the initial condition. This fact is essential in the derivations of the asymptotic results of the two estimators for $\alpha_0 = 1$ in the remainder of this paper.

Next, consider the Mundlak (1978)-Chamberlain (1982) type of projection for η_i[3]:

$$\eta_i = \pi y_{i,0} + v_i, \quad \mathrm{E}[v_i y_{i,0}] = 0, \quad v_i \sim i.i.d.(0, \sigma_v^2). \tag{2}$$

Different likelihood-based estimators discussed in this paper will primarily differ in the way they treat the π parameter or an associated function of π. For example, when we set $\pi = 1 - \alpha$ the projection corresponds exactly to the TML (Transformed Maximum Likelihood) framework of Hsiao et al. (2002). On the other hand, for the RML approach as in Alvarez and Arellano (2003), the π is treated as an unrestricted parameter to be estimated.

The conditional AR(1) model in Eq. (1) can be rewritten in the following stacked form:

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i-} + \boldsymbol{\iota}_T \eta_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'. \tag{3}$$

Alternatively, using the projection device, the model can also be described as follows[4]:

$$\boldsymbol{R}\boldsymbol{y}_i = (\boldsymbol{e}_1 \alpha + \boldsymbol{\iota}_T \pi) y_{i,0} + \boldsymbol{u}_i, \tag{4}$$

[3] We do not include a constant term in the projection as it would serve as a restricted time effect. In most applications, data will be used in deviations from cross-sectional means, thus allowing for unrestricted time effects.

[4] Bai (2013b) considers a similar conditional maximum likelihood estimator with a possible factor structure in the error term $\boldsymbol{\varepsilon}_i$. See Juodis and Sarafidis (2018) for a review on the literature on dynamic panels with common factor residuals, and Juodis and Sarafidis (2022) for an alternative treatment of this model based on GMM.

where $\boldsymbol{R} = \boldsymbol{I}_T - \boldsymbol{L}_T\alpha$, $\boldsymbol{e}_1$ is the first column of the $\boldsymbol{I}_T$ matrix, and $\boldsymbol{u}_i \equiv \boldsymbol{\iota}_T v_i + \boldsymbol{\varepsilon}_i$.[5] This correlated random effects decomposition can then be directly used to formulate the (quasi-) log-likelihood in the next section.

3 Maximum likelihood estimation approaches

3.1 The log-likelihood function

The derivations hereby mostly follow those in Bun et al. (2017) and have been adapted appropriately for the purpose of this paper. Note that

$$\mathrm{E}[\boldsymbol{u}_i] = \boldsymbol{0}_T, \quad \mathrm{var}[\boldsymbol{u}_i] = \boldsymbol{\Sigma} = \sigma_v^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \sigma^2 \boldsymbol{I}_T, \tag{5}$$

where the variance–covariance structure of $\boldsymbol{\Sigma}$ is of the usual random effects (or Generalized Least Squares) form. The quasi-log-likelihood function for some individual i is defined as:

$$\ell_i(\boldsymbol{\kappa}) = -\frac{1}{2}\Big((T-1)\log(\sigma^2) + \log(\theta^2) + \big[(\boldsymbol{y}_i - \alpha \boldsymbol{y}_{i-} - \boldsymbol{\iota}_T \pi y_{i,0})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \alpha \boldsymbol{y}_{i-} - \boldsymbol{\iota}_T \pi y_{i,0})\big]\Big), \tag{6}$$

where $\boldsymbol{\kappa} = (\alpha, \pi, \sigma^2, \sigma_v^2)'$. This function is the true likelihood function if $\boldsymbol{u}_i$ is a multivariate Gaussian vector.

Given the specific structure of the $\boldsymbol{\Sigma}$ matrix, the above expression can be substantially simplified. For example, using the notation in Bun et al. (2017) (e.g., $\widetilde{y}_{i,t} = y_{i,t} - \bar{y}_i$ and $\ddot{y}_i = \bar{y}_i - y_{i,0}$, $\ddot{y}_{i-} = \bar{y}_{i-} - y_{i,0}$) and defining $\rho = \pi - (1-\alpha)$, we obtain the following final expression for the log-likelihood function (after summing over all individual log-likelihood functions):

$$\ell(\boldsymbol{\kappa}) = -\frac{N}{2}\left((T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t} - \alpha \widetilde{y}_{i,t-1})^2 + \frac{T}{N\theta^2}\sum_{i=1}^{N}(\ddot{y}_i - \alpha \ddot{y}_{i-} - \rho y_{i,0})^2\right), \tag{7}$$

where $\boldsymbol{\kappa} = (\alpha, \pi, \sigma^2, \theta^2)'$ with $\theta^2 = \sigma^2 + T\sigma_v^2$. As it is extensively discussed in Bun et al. (2017), the parameters $(\sigma^2, \theta^2, \rho)$ can be concentrated out as:

[5] The lag-operator matrix $\boldsymbol{L}_T$ is defined such that for any $[T \times 1]$ vector $\boldsymbol{x} = (x_1, \ldots, x_T)'$, $\boldsymbol{L}_T \boldsymbol{x} = (0, x_1, \ldots, x_{T-1})'$.

$$\ell^c(\alpha) = -\frac{N}{2}\left((T-1)\log\left(\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t}-\alpha\widetilde{y}_{i,t-1})^2\right)\right.$$
$$\left.+\log\left(\frac{T}{N}\sum_{i=1}^{N}(\dot{y}_i-\alpha\dot{y}_{i-})^2\right)\right). \tag{8}$$

Here we define $\dot{y}_i$ and $\dot{y}_{i-}$ as follows:

$$\dot{y}_i = \ddot{y}_i - y_{i,0}\frac{\sum_{i=1}^{N}\ddot{y}_i y_{i,0}}{\sum_{i=1}^{N} y_{i,0}^2}, \quad \dot{y}_{i-} = \ddot{y}_{i-} - y_{i,0}\frac{\sum_{i=1}^{N}\ddot{y}_{i-} y_{i,0}}{\sum_{i=1}^{N} y_{i,0}^2}, \tag{9}$$

The above characterization of the log-likelihood function is highly appealing to empirical researchers, as the numerical/computational burden decreases dramatically. Moreover, a simple grid search-based procedures can be used to investigate the curvature of the likelihood, see Sect. 3.3.

Remark 1 Note that the TML log-likelihood function of Hsiao et al. (2002) and Bun et al. (2017) is obtained by setting $\rho = 0$ in Eq. (7), such that $\dot{y}_i = \ddot{y}_i$ and $\dot{y}_{-i} = \ddot{y}_{i-}$.

3.2 The misspecified RML approach

The TML and the RML approaches can be viewed as two special cases in the way the ρ parameter is being handled in estimation. In particular, TML sets $\rho = 0$, whereas RML estimates ρ (or π for that mater) freely, without imposing any restrictions. An alternative approach, that we label as the misspecified RML (mRML) approach, uses a more general formulation for the π parameter. In particular, we consider $\pi(\phi)$ such that

$$\pi(\phi) = \pi = (1-\alpha)\phi, \tag{10}$$

where $\phi \in \mathbb{R}$ denotes some arbitrary a priori chosen scalar. Note that since the population value of the correlation between $y_{i,0}$ and η_i corresponds to a specific "true" value of ϕ, (say) ϕ_0, setting $\phi \neq \phi_0$ implies a misspecification of the correlation between the initial condition and the individual-specific effect. The term "misspecified ML estimator" was first used by Hahn et al. (2004), who studied the properties of this estimator for the special case where $\phi = 0$.

The only exception for the appropriateness of this terminology is setting $\phi = 1$ (corresponding to the TML approach), as in that case the TML estimator is known to be fixed-T consistent for all $|\alpha_0| \leq 1$. For all other values of ϕ, the mRML estimator is not generally fixed-T consistent for $\alpha_0 < 1$, as we formally show in Sect. 4.

The concentrated log-likelihood function of the mRML estimator (for any ϕ) is given by:

$$\ell^{c}_{mRML}(\alpha) = -\frac{N}{2}\left((T-1)\log\left(\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t}-\alpha\widetilde{y}_{i,t-1})^2\right)\right) - \frac{N}{2}\left(\log\left(\frac{T}{N}\sum_{i=1}^{N}(\dot{y}_i(\phi)-\alpha\dot{y}_{i-}(\phi))^2\right)\right), \tag{11}$$

where we now set $\dot{y}_i(\phi) = \overline{y}_i - \phi y_{i,0}$ and $\dot{y}_{i-}(\phi) = \overline{y}_{i-} - \phi y_{i,0}$. From this formulation, the mRML approach with $\phi = 0$ can be alternatively motivated as a special case of the approach studied by Bai (2013a) if one erroneously assumes that $y_{i,0} = 0\ \forall i$, while in reality the initial conditions are non-zero.

3.3 The problem of multiple solutions

Consider the first derivative of the concentrated log-likelihood function for all estimators considered above. Let

$$\widehat{\sigma}^2(\phi) = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t}-\alpha\widetilde{y}_{i,t-1}\right)^2, \tag{12}$$

$$\widehat{\theta}^2(\phi) = \frac{T}{N}\sum_{i=1}^{N}(\dot{y}_i(\phi)-\alpha\dot{y}_{i-}(\phi))^2, \tag{13}$$

then the first derivative of the concentrated log-likelihood function is given by:

$$\frac{\mathrm{d}\ell^{c}(\alpha)}{\mathrm{d}\alpha} = \frac{1}{\widehat{\sigma}^2(\alpha)}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}(\widetilde{y}_{i,t}-\alpha\widetilde{y}_{i,t-1}) + \frac{T}{\widehat{\theta}^2(\alpha)}\sum_{i=1}^{N}\dot{y}_{i-}(\phi)(\dot{y}_i(\phi)-\alpha\dot{y}_{i-}(\phi)). \tag{14}$$

In particular, any solution of the corresponding first-order conditions (FOC) should satisfy:

$$\widehat{\theta}^2(\alpha)\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}(\widetilde{y}_{i,t}-\alpha\widetilde{y}_{i,t-1}) + \widehat{\sigma}^2(\alpha)T\sum_{i=1}^{N}\dot{y}_{i-}(\dot{y}_i(\phi)-\alpha\dot{y}_{i-}(\phi)) = 0. \tag{15}$$

Given that $\widehat{\sigma}^2(\alpha)$ and $\widehat{\theta}^2(\alpha)$ are quadratic in α, it is not difficult to see that the FOC are cubic in α. Thus, for any value of T and any realization of $\{\boldsymbol{y}_i\}_{i=1}^{N}$ there will be at least one and at most three solutions to Eq. (15).

As noted by Alvarez and Arellano (2004, 2022), the TML estimator might suffer from issues related to non-identification for $T = 2$. Bun et al. (2017) and Juodis

(2018a) built upon those results and obtained further insights on the properties of the distribution of the TML and RML estimators for stationary data. Among other things, they note that the TML approach is more susceptible to generating bimodal finite sample distributions of the corresponding estimator.[6]

As the mRML estimator shares the same structure of the concentrated log-likelihood function Eq. (15), the finite sample distribution of the estimator might be bimodal. However, the choice of ϕ might play a non-trivial effect in determining the shape of the corresponding log-likelihood function.

4 Asymptotic results

Section 4.1 analyzes asymptotic properties of the mRML estimator when T is fixed. We shall focus on the unit-root case, $\alpha_0 = 1$. Before we embark on the asymptotic analysis for this case, we present the following general (although negative) result for $|\alpha_0| < 1$:

Proposition 1 *Let $\nabla_\alpha \ell$ denote the score of the mRML log-likelihood function with respect to α evaluated at true $\boldsymbol{\kappa}_0$. Then, for any T and $|\alpha_0| < 1$:*

$$\mathrm{E}[\nabla_\alpha \ell] = \mathcal{O}(N). \tag{16}$$

Moreover, $\mathrm{E}[\nabla_\alpha \ell] = 0$ *if and only if* $\mathrm{E}[y_{i,0}(\mu_i - \phi y_{i,0})](\phi - 1) = 0$.

Proposition 1 shows that for $|\alpha_0| < 1$, there exist only two values of ϕ that can guarantee fixed-T consistency of the mRML estimator; either $\phi = 1$, which corresponds to the TML approach, or the value of ϕ corresponding to the infeasible (unknown) correlation coefficient between the initial conditions and the individual-specific effects.

4.1 Fixed-T results for the unit-root case

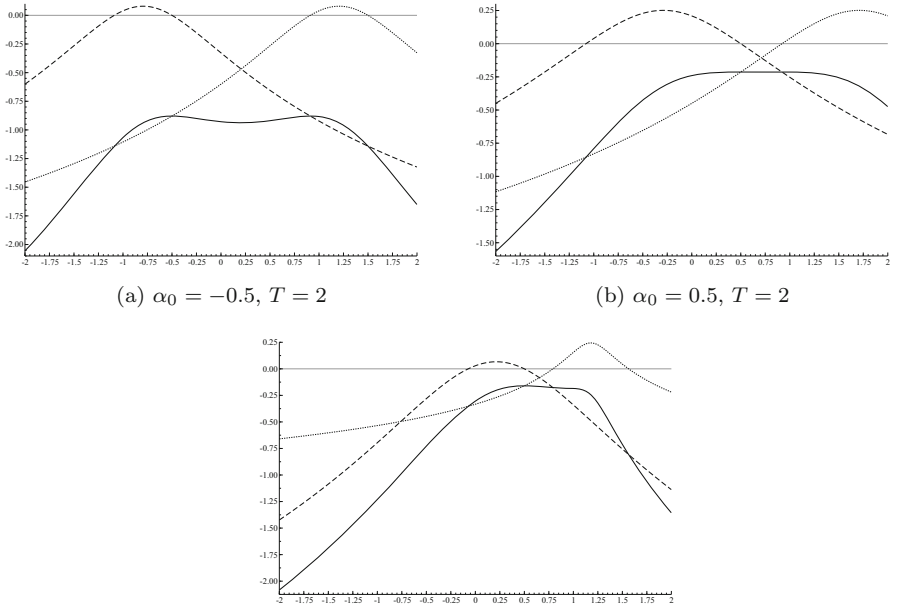
Our main result of this section is formulated in Proposition 2.

Proposition 2 *For any fixed-T as $N \to \infty$, the log-likelihood function corresponding to the mRML estimator is unimodal at the point $\alpha_0 = 1$, for any fixed value of ϕ.*

This proposition extends the analytical and numerical results obtained by Bun et al. (2017), which apply only to $|\alpha_0| < 1$ for the TML approach. Note that since the log-likelihood function corresponding to TML can be deduced from mRML by setting $\phi = 1$, the result of Proposition 2 also applies to TML.

In order to grasp the intuition of the above proposition, Fig. 1 revisits some of the results in Juodis (2018a), which correspond to TML. One can observe from this figure that while for $|\alpha_0| < 1$ the asymptotic concentrated log-likelihood function is bimodal, the second mode is always at $\alpha = 1$, and the first mode naturally approaches

[6] The apparent robustness of the RML approach can be motivated from the fact that it explicitly minimizes the magnitude of the "between group" part of the log-likelihood function for any given value α.



(a) $\alpha_0 = -0.5$, $T = 2$ (b) $\alpha_0 = 0.5$, $T = 2$

(c) $\alpha_0 = 0.5$, $T = 6$

Fig. 1 Concentrated asymptotic log-likelihood function for TML. In all figures, the first mode is at the corresponding true value α_0, while the second mode is located at $\alpha = 1$. The initial observation is from the covariance stationary distribution. The dashed line represents the Within Group part of the log-likelihood function, while the dotted line the Between Group part. The solid line, which stands for the log-likelihood function, is a sum of dashed and dotted lines

the second one as $\alpha_0 \to 1$.[7] Thus, for the true value of $\alpha_0 = 1$ the two modes collapse into one and the log-likelihood function of mRML is unimodal.

When it comes to the actual shape of the log-likelihood function of mRML in the unit-root case, it turns out that this is of standard form, unless $\phi = 1$. To see this, let $\widetilde{\sigma}^2_{y_0}(\phi) \equiv T(1-\phi)^2 \,\mathrm{E}[y^2_{i,0}]$ be the scaled second moment of the initial condition. Moreover, let the true value of θ^2 be defined as $\theta_0^2 = \sigma_0^2 + (1-\alpha_0)(\mu_i - \phi y_{i,0})$, with $\phi = 1$ as the special case of the TML estimator. Obviously, for $\alpha_0 = 1$ the value of θ_0^2 is the same irrespective of ϕ.

Using this notation, the following result is obtained for the Hessian of the mRML estimator.

Proposition 3 (Singularity mRML) *The $\mathcal{H}_\ell(\phi)$ matrix is equal to:*

$$\mathcal{H}_\ell(\phi) = \frac{T-1}{2\sigma_0^2} \begin{pmatrix} T\sigma_0^2 + \frac{2}{T-1}\widetilde{\sigma}^2_{y_0}(\phi) & -1 & 1 \\ -1 & \frac{1}{\sigma_0^2} & 0 \\ 1 & 0 & \frac{1/(T-1)}{\sigma_0^2} \end{pmatrix}. \tag{17}$$

[7] Bun et al. (2017) shows that bimodality and the location of the second mode in this case is primarily determined by the properties of the initial observation $y_{i,0}$.

Moreover, this matrix is singular for fixed-T if and only if $\widetilde{\sigma}^2_{y_0}(\phi) = 0$.

The proof of this proposition largely follows the proof strategy of Theorem 1 in Juodis (2018b), once appropriately modified for the setting at hand.[8] To the best of our knowledge, Proposition 3 is the first result in the literature that proves that the mRML estimator has desirable asymptotic properties for fixed-T in the unit-root case $\alpha_0 = 1$. In particular, as a corollary of the two results presented above, the mRML estimator is consistent and asymptotically normal (subject to the usual other regularity conditions, e.g., compactness of the parameter space).

Since setting $\phi = 1$ implies $\widetilde{\sigma}^2_{y_0}(\phi) = 0$, it is straightforward to see that the Hessian matrix for the TML estimator is singular. This result is well known in the literature, see e.g., Ahn and Thomas (2006); Kruiniger (2013) or Juodis (2018b).[9] Hence, the corresponding asymptotic distribution of the TML estimator is non-standard and non-normal. In particular, following the results in Roznitzky et al. (2000) and Dovonon and Hall (2018), one can show that

$$\sqrt[4]{N}(\widehat{\alpha} - 1) = \mathcal{O}_P(1), \tag{18}$$

where the asymptotic distribution is determined by the higher-order expansion of the likelihood function. As such, the limiting distribution of TML is asymmetric. Due to such non-standard properties of TML, it appears to us that there are no approaches suggested in the literature that can be used to construct uniformly valid confidence intervals for $\alpha_0 \in (-1; 1]$.

Remark 2 We note that the TML approach is not the only fixed-T consistent "bias-corrected" FE-type approach that suffers from the singularity of the limiting Hessian matrix for $\alpha_0 = 1$. It is well known in the literature that the standard bias-corrected FE estimator, as studied by Lancaster (2002); Bun and Carree (2005); Dhaene and Jochmans (2016), and Kruiniger (2018), shares this property at $\alpha_0 = 1$. For more details, we refer to Kruiniger (2018). Thus, we are not aware of any estimator that would satisfy all three requirements below: (i) it is consistent for all $\alpha_0 \in (-1; 1]$; (ii) consistency does not depend on the stationarity of the initial condition; and (iii) it has asymptotic normal distribution.

Remark 3 The result in Proposition 3 might seem at odds with the unit-root results in Norkutė and Westerlund (2021), who use the factor analytic approach of Bai (2013a) to construct their estimator. The main difference between their approach and the approach in this paper is that their explicit model is of the error-components structure:

$$y_{i,t} = v_i + u_{i,t}, \quad u_{i,t} = \alpha_0 u_{i,t-1} + \varepsilon_{i,t}. \tag{19}$$

While the two coincide asymptotically when $|\alpha_0| < 1$, this is not the case when $\alpha_0 = 1$. In particular, their results build upon the assumption that $\mathrm{E}[v_i^2] > 0$. As such, their

[8] Hence, the corresponding result in Juodis (2018b) can be seen as a special case of Proposition 3.

[9] An analogous result for the RML approach is also available in the literature and was derived by Kruiniger (2013).

results are not uniformly valid when the true individual heterogeneity is degenerate. Hence, the desirable finite sample properties of the proposed procedure (as compared to the standard FE-type methods, e.g., Moon et al. (2007) and Juodis and Westerlund (2019)) are achieved at the expense of non-uniformity with respect to this nuisance parameter.

4.2 Large-T results

4.2.1 Asymptotic equivalence in the stationary case

As shown by Alvarez and Arellano (2003), when $N, T \to \infty$ and $|\alpha_0| < 1$, the RML estimator is asymptotically equivalent to the bias-corrected FE estimator of Hahn and Kuersteiner (2002) (provided that $N^3/T \to 0$ is satisfied for the latter).

In what follows, we show that the same conclusion can be reached for the mRML estimator for any ϕ (thus also the TML estimator). The intuition for this result is fairly simple. Consider the likelihood function in Eq. (8). It is fairly easy to see that as $N, T \to \infty$ (irrespective of the relative magnitude):

$$\ell^c(\alpha) = -\frac{(T-1)N}{2} \log\left(\frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=1}^{T} (\widetilde{y}_{i,t} - \alpha \widetilde{y}_{i,t-1})^2\right) + \mathcal{O}_P(N \log(T)), \tag{20}$$

uniformly for all α and ϕ. Hence, the large-T consistency of all estimators follows from large-T consistency of the corresponding FE estimator. That is, both the RML and the mRML estimators provide an in-built bias-correction term for the standard fixed effects log-likelihood function. As both approaches handle bias-correction rather differently for T fixed, the underlying asymptotic properties depend on the way ρ (or π) is handled, i.e., whether it is estimated or it is fixed. On the other hand, for large-T this choice is mostly inconsequential, as it can be expected from the expansion in Eq. (20). Our next result formalizes this conjecture.

Proposition 4 *Under assumption RML, as $N, T \to \infty$ such that $N/T^3 \to 0$:*

$$\sqrt{NT}(\widehat{\alpha}(\phi) - \alpha_0) \to N(0, 1 - \alpha_0^2), \tag{21}$$

for all $|\alpha_0| < 1$ and any constant value of ϕ such that $\pi = (1-\alpha)\phi$.

Hence, the class of RML estimators indexed by ϕ is asymptotically equivalent to the bias-corrected FE estimator of Hahn and Kuersteiner (2002). This result is not unexpected, as the mRML specification can be seen as a "bias reducing prior" for η_i using the terminology of Arellano and Bonhomme (2009).

Proposition 4 extends the analogous result in Hahn et al. (2004), which was proven for the special case $\phi = 0$. In particular, their setting corresponds to the misspecified likelihood where one incorrectly assumes that $\mathrm{E}[\eta_i y_{i,0}] = 0$,[10] when in fact this is

[10] As e.g., in Maddala (1971).

not the case. Here we have shown that the specific choice of ϕ is inconsequential for the asymptotic distribution of the estimator as long as $N/T^3 \to 0$.

4.2.2 Singularity of the Hessian matrix of mRML in the unit-root case when *T* is large

The results in Proposition 3 have been derived for any fixed value of T. It is then natural to wonder what happens when $T \to \infty$. For the Hessian matrix to be non-singular also as $T \to \infty$, it is necessary that $\widetilde{\sigma}_0^2 = \mathcal{O}(T^{1+\beta})$ or alternatively $\mathrm{E}[y_{i,0}^2] = \mathcal{O}(T^{\beta})$, where $\beta \geq 1$. Observe that elements of $\boldsymbol{\mathcal{H}}_\ell$ are not of the same order of magnitude. Thus, heuristically, it can be expected that the corresponding coefficients in $\boldsymbol{\kappa}$ will exhibit different rates of convergence if we let $T \to \infty$. This heuristic is formalized in our next proposition based on sequential limit theory where $N \to \infty$ first followed by $T \to \infty$.

Proposition 5 (Singularity mRML large-T) *Let* $(N, T)_{seq} \to \infty$*, then the scaled mRML Hessian matrix is non-singular if and only if* $\beta \geq 1$.

We conjecture that equivalent result can be also proven rigorously using the joint limit theory with $N, T \to \infty$, but we leave this question for future research. Furthermore, we will not attempt to properly characterize the asymptotic distribution of the misspecified RML estimator in this case, as it would involve a complete characterization of all components in the Taylor's expansion of the log-likelihood function.

We expect that our previous result in Proposition 3 is useful to characterize the asymptotic distribution of mRML, provided that $\widetilde{\sigma}_0^2$ is not too small.

Remark 4 The non-vanishing effect of the initial condition $y_{i,0}$ in this setting can be compared with the similar result obtained by Juodis and Poldermans (2021) in the unit-root non-stationary setting for the Backwards Orthogonal Deviations (BOD) estimator of Everaert (2013). In that setting, the initial condition is not asymptotically dominant, but it has a variance reduction effect. Note that the rate $\mathrm{E}[y_{i,0}^2] = \mathcal{O}(T^{\beta})$ can be achieved if the process $y_{i,t}$ has a distant or infinite past, see e.g., Westerlund (2016).

5 Monte Carlo study

5.1 The setup

In this section, we investigate the finite sample performance of the various estimators and corresponding test statistics using simulated data. In particular, we consider the following panel AR(1) model:

$$y_{i,t} = \alpha y_{i,t-1} + (1-\alpha)\mu_i + \varepsilon_{i,t}; \quad \varepsilon_{i,t} \sim \mathcal{N}(0,1); \quad t = 1, \ldots, T. \tag{22}$$

$$y_{i,0} = \gamma \mu_i + \varepsilon_{i,0}; \quad \varepsilon_{i,0} \sim \mathcal{N}\left(0, \zeta^2\right); \quad \mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right). \tag{23}$$

Mean-stationarity of $y_{i,t}$ is achieved for designs with $\gamma = 1$, while the process $y_{i,t}$ is covariance stationary if and only if $\gamma = 1$ and $\zeta^2 = 1 - \alpha^2$. The actual value of σ_μ^2

is irrelevant for the TML estimator as long as $\gamma = 1$, but for the RML estimator this parameter is always important.

As we are interested in setups with $\alpha_0 \approx 1$, we will set $\zeta^2 = 1$, so that the process is never covariance stationary. Moreover, we restrict our attention to mean stationary settings with $\gamma = 1$. Even for the simple AR(1) model, the parameter space is already very large. We have tried to cover its most relevant part by considering the following parameter settings:

$$N = \{50, 200, 500\}, \quad T = \{5, 10, 20\}, \quad \alpha_0 = \{0.5, 0.9, 1.0\}.$$

We consider three estimation approaches, namely TML, RML and mRML with $\phi = 0$. For all three approaches, we report two types of estimators: (i) based on global maximum of the objective function; (ii) based on local-maximum of the WG mode. The second option is the suggest "left" rule-of-thumb by Bun et al. (2017). We report the mean bias, the median bias and the RMSE for all estimators. Moreover, for all estimators we report the fraction of replications the objective function is found to be unimodal. For all estimators, we use root finder algorithms based on the eigenvalues of the companion matrix to obtain the maximum likelihood estimates in all three cases.[11] The number of Monte Carlo replications is set to 4000.

5.2 Results

The estimation results are summarized in Tables 1, 2, 3, while in Table 4 we summarize the unimodality properties of the three approaches considered.

To begin with, consider the results in Table 1. Initially we focus on $\alpha_0 = 0.5$. One observes that the mRML estimator is the one with the largest bias. This is not surprising given that the correlation between the initial condition and the individual-specific effect is misspecified.[12] However, this bias quickly disappears as T increases to at least $T = 10$. This observation is consistent with the results of Sect. 4.2. The two fixed-T consistent estimators for $|\alpha_0| < 1$, RML and TML, exhibit much smaller bias than mRML, with most of the bias being present due to the bimodality of their finite sample distributions. Such bias can be effectively mitigated using the "left" option.

Regarding $\alpha_0 = 0.9$ and $\alpha_0 = 1.0$, we note that the bias of the mRML estimator becomes comparable to that of the RML/TML approaches and becomes nearly negligible in the unit-root setting. Moreover, for $\alpha_0 = 1.0$ the mRML estimator has smaller RMSE, as predicted by the potentially faster convergence rate of the estimator in this case (provided that $\mathrm{E}[y_{i,0}^2]$ is sizeable).

Next, we consider the bimodality properties of the three estimators. From Table 4, it is clear that the behavior of the RML and TML estimators differs dramatically between the stable setting of $|\alpha_0| < 1$ and the unit-root setup $\alpha_0 = 1$. In the latter case, even for large N, T, in almost 40% of the replications the likelihood functions are bimodal. This is in sharp contrast with the theoretical predictions from Proposition 2.

[11] This ensures that all our results can be obtained within 5 min.

[12] Of course, the magnitude of the bias depends on the extent of misspecification, i.e., the deviation between the implied correlation with $\phi = 0$ and the true correlation.

Table 1 Monte Carlo results for $\alpha_0 = 0.5$

N	T	Mean bias						RMSE					
		TML	RML	mRML	TML(L)	RMl(L)	mRML(L)	TML	RML	mRML	TML(L)	RMl(L)	mRML(L)
50	5	0.06	0.03	0.32	0.02	0.01	0.31	0.21	0.15	0.33	0.14	0.14	0.33
50	10	−0.00	−0.00	0.12	−0.00	−0.00	0.08	0.06	0.06	0.20	0.06	0.06	0.15
50	20	−0.00	−0.00	0.01	−0.00	−0.00	0.01	0.03	0.03	0.03	0.03	0.03	0.03
200	5	0.01	0.00	0.33	0.00	0.00	0.33	0.08	0.06	0.33	0.06	0.06	0.33
200	10	0.00	0.00	0.08	0.00	0.00	0.06	0.03	0.03	0.12	0.03	0.03	0.08
200	20	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.02	0.02	0.02	0.02	0.02
500	5	0.00	0.00	0.33	0.00	0.00	0.33	0.04	0.04	0.33	0.04	0.04	0.33
500	10	−0.00	−0.00	0.06	−0.00	−0.00	0.06	0.02	0.02	0.07	0.02	0.02	0.06
500	20	−0.00	−0.00	0.01	−0.00	−0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01

TML, RML, and mRML refer to the Transformed ML, Random Effects ML, and the misspecified Random Effects ML estimators, respectively. TML(L), RML(L), and mRML(L), are the versions of the same estimator with "left" rule of Bun et al. (2017) implemented. RMSE is the Root Mean Squared Error of the corresponding estimators over 4000 Monte Carlo replications

Table 2 Monte Carlo results for $\alpha_0 = 0.9$

N	T	Mean bias						RMSE					
		TML	RML	mRML	TML(L)	RMl(L)	mRML(L)	TML	RML	mRML	TML(L)	RMl(L)	mRML(L)
50	5	0.02	0.01	0.05	−0.05	−0.05	0.05	0.15	0.16	0.07	0.14	0.15	0.07
50	10	0.01	0.01	0.05	−0.01	−0.01	0.05	0.09	0.08	0.06	0.08	0.08	0.06
50	20	0.01	0.01	0.05	0.00	0.00	0.05	0.05	0.04	0.06	0.04	0.04	0.06
200	5	0.02	0.01	0.05	−0.02	−0.02	0.05	0.11	0.11	0.06	0.10	0.10	0.06
200	10	0.01	0.01	0.06	0.00	0.00	0.06	0.06	0.06	0.06	0.05	0.05	0.06
200	20	0.00	0.00	0.05	0.00	0.00	0.05	0.02	0.02	0.06	0.02	0.02	0.06
500	5	0.02	0.01	0.05	−0.01	−0.01	0.05	0.08	0.08	0.06	0.08	0.08	0.06
500	10	0.01	0.00	0.06	0.00	0.00	0.06	0.04	0.04	0.06	0.04	0.04	0.06
500	20	0.00	0.00	0.06	0.00	0.00	0.06	0.01	0.01	0.06	0.01	0.01	0.06

See Table 1

Table 3 Monte Carlo results for $\alpha_0 = 1.0$

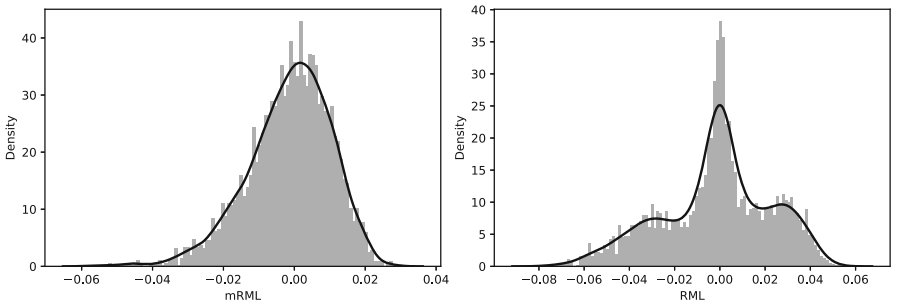
N	T	Mean bias						RMSE					
		TML	RML	mRML	TML(L)	RMl(L)	mRML(L)	TML	RML	mRML	TML(L)	RMl(L)	mRML(L)
50	5	−0.00	−0.01	−0.01	−0.08	−0.08	−0.01	0.14	0.15	0.05	0.15	0.15	0.05
50	10	−0.01	−0.00	−0.01	−0.05	−0.04	−0.01	0.07	0.07	0.04	0.08	0.08	0.04
50	20	−0.00	−0.00	−0.01	−0.02	−0.02	−0.01	0.04	0.04	0.03	0.04	0.04	0.03
200	5	−0.00	−0.00	−0.00	−0.06	−0.06	−0.00	0.10	0.10	0.02	0.10	0.10	0.02
200	10	−0.00	−0.00	−0.00	−0.03	−0.03	−0.00	0.05	0.05	0.02	0.05	0.05	0.02
200	20	−0.00	−0.00	−0.00	−0.01	−0.01	−0.00	0.02	0.02	0.01	0.02	0.03	0.01
500	5	−0.00	−0.00	−0.00	−0.05	−0.04	−0.00	0.07	0.08	0.01	0.08	0.08	0.01
500	10	−0.00	−0.00	−0.00	−0.02	−0.02	−0.00	0.04	0.04	0.01	0.04	0.04	0.01
500	20	−0.00	−0.00	−0.00	−0.01	−0.01	−0.00	0.02	0.02	0.01	0.02	0.02	0.01

See Table 1

Table 4 Unimodality analysis

N	T	Fraction of replications where the objective function is unimodal								
		$\alpha_0 = 0.5$			$\alpha_0 = 0.9$			$\alpha_0 = 1.0$		
		TML	RML	mRML	TML	RML	mRML	TML	RML	mRML
50	5	0.60	0.89	0.99	0.65	0.72	1.00	0.58	0.64	1.00
50	10	0.69	0.98	0.70	0.76	0.80	0.99	0.56	0.62	1.00
50	20	0.75	0.99	0.93	0.86	0.90	0.98	0.58	0.64	0.98
200	5	0.69	0.99	1.00	0.71	0.77	1.00	0.55	0.60	1.00
200	10	0.88	1.00	0.76	0.84	0.90	1.00	0.56	0.61	1.00
200	20	0.94	1.00	1.00	0.97	0.98	1.00	0.57	0.61	1.00
500	5	0.83	1.00	1.00	0.76	0.83	1.00	0.55	0.59	1.00
500	10	0.97	1.00	0.86	0.91	0.95	1.00	0.55	0.59	1.00
500	20	0.99	1.00	1.00	1.00	1.00	1.00	0.56	0.59	1.00

See Table 1



(a) The mRML estimator. (b) The RML estimator.

Fig. 2 The finite sample distributions of the mRML and RML estimators for $N = 200$, $T = 20$, $\alpha_0 = 1$

The misspecified likelihood function, on the other hand, is mostly unaffected by the exact value of the α_0 parameter.

Finally, one may wonder whether our results support the theoretical prediction in Proposition 3 or not. In Fig. 2, we summarize the finite sample distributions of RML and mRML estimators for a given choice of design parameters. The results are fairly revealing on the differences between the finite sample distributions of the two estimators. In particular, while the mRML estimator has a distinct unimodal distribution (even if asymmetric), the finite sample distribution of the RML estimator is distinctively non-standard and asymmetric. While not presented here, the results in Table 4 also indicate that in this setting the results of the mRML estimator are unchanged when one considers the "left" option of the estimator. The same is not true, however, for the RML estimator that is bimodal in 40% of the Monte Carlo replications.

6 Conclusions

The present paper studied a class of misspecified Random effect Maximum Likelihood estimators. The misspecification arises by imposing the wrong value for the correlation strength between the initial condition and the individual-specific intercepts. As a special case, we have analyzed the asymptotic behavior of the transformed maximum likelihood approach as in Hsiao et al. (2002).

We have shown that for any fixed value of T, the log-likelihood function of the mRML estimator has a single mode at the true value as $N \to \infty$. In addition, the Hessian matrix of the corresponding log-likelihood function is non-singular, unless the scaled variance of the initial condition is exactly zero. As a result, mRML is consistent and asymptotically normally distributed, as $N \to \infty$ for T fixed. Thus, standard inference procedures are valid. To the best of our knowledge, this is the first result in the literature that shows that a class of mRML estimators has desirable asymptotic properties in the unit-root case for fixed-T.

Secondly, the paper also provided new insights on the properties of TML and mRML in large-T samples in a stable autoregressive setting. When N, T are both large, the TML estimator is asymptotically equivalent to the bias-corrected FE estimator of Hahn and Kuersteiner (2002). Moreover, for $\alpha_0 = 1$, the Hessian matrix corresponding to the likelihood function of mRML remains non-singular, so long as the scaled variance of the initial conditions is of order $\mathcal{O}(T^{1+\beta})$, $\beta \geq 1$.

In a Monte Carlo study, we have explored how informative our asymptotic results are for the finite sample properties of all estimators considered. We found that this asymptotic characterization is informative about finite sample behavior only for those estimators that have non-singular limiting Hessian matrices. This excludes the TML and RML estimators, which have singular Hessian matrices in the limit.

In this paper, we have limited our attention to the stylized panel AR(1) model. This may be too restrictive for many real-life applications. In our future research, we are planning to extend the present analysis to panel vector autoregressive models, similarly to Binder et al. (2005); Arellano (2016), and Juodis (2018a, b), in order to account for feedback effects from other variables, as it is commonly the case in micro- and macro-economic panels.

Data Availability Statement Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest We have no financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: Proofs

Notation

In what follows we introduce a few terms used in the next derivations.

$$\theta_0^2 = \sigma_0^2 + T(1-\alpha_0)^2\,\mathrm{E}[(\mu_i - \phi y_{i,0})^2], \tag{A.1}$$

$$\xi = \sum_{l=0}^{T-2}(T-1-l)\alpha_0^l, \tag{A.2}$$

$$\widetilde{\sigma}_{y_0}^2(\phi) = T(1-\phi)^2\,\mathrm{E}[y_{i,0}^2], \tag{A.3}$$

$$\Delta_\alpha = \alpha - \alpha_0, \tag{A.4}$$

$$\theta_1 = \frac{(T-1)(2T-1)}{6}\sigma_0^2, \tag{A.5}$$

$$\sigma_1 = \frac{T+1}{6}\sigma_0^2, \tag{A.6}$$

$$\theta_1(\phi) = \widetilde{\sigma}_{y_0}^2(\phi) + \theta_1. \tag{A.7}$$

Here (unless specified otherwise) all terms are defined for all values of $\sigma_0^2, \alpha, \alpha_0, \phi$.

Appendix A.1: Fixed T results

Proof Proposition 1 Let $\nabla_\alpha \ell$ denote the partial derivative of $\ell(\boldsymbol{\kappa})$ with respect to α once evaluated at $\boldsymbol{\kappa}_0$. Consider the element of the score that corresponds to α:

$$\nabla_\alpha \ell = \frac{1}{\sigma_0^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}(\widetilde{\varepsilon}_{i,t}) + \frac{T}{\theta_0^2}\sum_{i=1}^{N}\dot{y}_{i-}(\dot{y}_i(\phi) - \alpha_0\dot{y}_{i-}(\phi)), \tag{A.8}$$

where θ_0^2 is defined as above.

Consider the "between group" part in isolation. Using analogous steps to those in Appendix Lemma A.2 of Juodis (2018a), we expand:

$$\dot{y}_i(\phi) - \alpha_0\dot{y}_{i-}(\phi) = \overline{\varepsilon}_i + (1-\alpha_0)(\mu_i - \phi y_{i,0}). \tag{A.9}$$

Note that from $\mathrm{E}[T\overline{y}_{i-}\overline{\varepsilon}_i] = \xi/T$, we have

$$\begin{aligned} T\dot{y}_{i-}(\phi) &= \xi(1-\alpha_0)(\mu_i - y_{i,0}) - T(\phi-1)y_{i,0} + \sum_{t=1}^{T-1}\sum_{s=0}^{t-1}\alpha_0^s\varepsilon_{i,t-s} \\ &= \xi(1-\alpha_0)(\mu_i - \phi y_{i,0}) + (\phi-1)y_{i,0}(\xi(1-\alpha_0) - T) \\ &\quad + \sum_{t=1}^{T-1}\left(\sum_{s=0}^{T-1-t}\alpha_0^s\right)\varepsilon_{i,t} \end{aligned}$$

$$= \xi(1-\alpha_0)(\mu_i - \phi y_{i,0}) - (\phi-1) y_{i,0} \left(\sum_{s=0}^{T-1} \alpha_0^s \right) + \sum_{t=1}^{T-1} \left(\sum_{s=0}^{T-1-t} \alpha_0^s \right) \varepsilon_{i,t}. \tag{A.10}$$

Thus,

$$\begin{aligned} \mathrm{E}[T \dot{y}_{i-}(\phi)(\dot{y}(\phi)_i - \alpha_0 \dot{y}_{i-}(\phi))] &= \frac{\theta_0^2}{T} \xi - \mathrm{E}[y_{i,0}(\mu_i - \phi y_{i,0})] \left(\sum_{s=0}^{T-1} \alpha_0^s \right) (1-\alpha_0)(\phi-1) \\ &= \frac{\theta_0^2}{T} \xi - \mathrm{E}[y_{i,0}(\mu_i - \phi y_{i,0})] \left(1 - \alpha_0^T\right)(\phi-1), \end{aligned} \tag{A.11}$$

where, as before, $\theta_0^2 = \sigma_0^2 + T(1-\alpha_0)^2 \mathrm{E}[(\mu_i - \phi y_{i,0})^2]$. Using this insight, we can expand the score as follows:

$$\begin{aligned} \nabla_\alpha \ell = \frac{1}{\sigma_0^2} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{\varepsilon}_{i,t} - \mathrm{E}[\widetilde{y}_{i,t-1} \widetilde{\varepsilon}_{i,t}] \right) \\ + \frac{T}{\theta_0^2} \sum_{i=1}^{N} (\dot{y}_{i-}(\phi)(\dot{y}_i(\phi) - \alpha_0 \dot{y}_{i-}(\phi)) - \mathrm{E}[\dot{y}_{i-}(\dot{y}_i(\phi) - \alpha_0 \dot{y}_{i-}(\phi))]) \\ - \frac{N}{\theta_0^2} \mathrm{E}[y_{i,0}(\mu_i - \phi y_{i,0})] \left(1 - \alpha_0^T\right)(\phi-1). \end{aligned} \tag{A.12}$$

From here, it follows immediately that:

$$\mathrm{E}[\nabla_\alpha \ell] = -\frac{N}{\theta_0^2} \mathrm{E}[y_{i,0}(\mu_i - \phi y_{i,0})] \left(1 - \alpha_0^T\right)(\phi-1). \tag{A.13}$$

The conclusion of this proposition follows immediately. □

Proof Proposition 2 In what follows, we consider the probability limit of Eq. (15). Without loss of generality, we set $\sigma^2 = 1$.

The most critical component of this expression (that varies over ϕ) is $\widehat{\theta^2}(\alpha)$, and the corresponding derivative with respect to α. In particular,

$$\widehat{\theta}^2(\alpha) = \frac{T}{N}\sum_{i=1}^{N}(\dot{y}_i(\phi) - \alpha\dot{y}_{i-}(\phi))^2$$
$$= (1-\alpha)^2\frac{T}{N}\sum_{i=1}^{N}(\dot{y}_{i-}(\phi))^2 + 2(1-\alpha)\frac{T}{N}\sum_{i=1}^{N}(\dot{y}_{i-}(\phi))\overline{\varepsilon}_i + \frac{T}{N}\sum_{i=1}^{N}(\overline{\varepsilon}_i)^2. \tag{A.14}$$

Upon inspection of the proof of Theorem 1 in Juodis (2018b), it is clear that the only term affected by ϕ is

$$\mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\overline{y}_{i-} - \phi y_{i,0})^2\right] = \mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\ddot{y}_{i-} - (\phi-1)y_{i,0})^2\right]$$
$$= T(1-\phi)^2\,\mathrm{E}[y_{i,0}^2] + \frac{(T-1)(2T-1)}{6}\sigma_0^2$$
$$= \widetilde{\sigma}_{y_0}^2 + \theta_1. \tag{A.15}$$

The final line follows from the fact that, under our assumptions, in the unit-root setting $\dot{y}_{i-}$ and $y_{i,0}$ are actually independent. This result can then be used to show that $\mathrm{E}[\widehat{\theta}^2(\alpha)]$ equals

$$\mathrm{E}[\widehat{\theta}^2(\alpha)] = (1-\alpha)^2\left(\frac{(T-1)(2T-1)}{6}\right) + (1-\alpha)(T-1) + 1$$
$$= \Delta_\alpha^2\theta_1(\phi) + \Delta_\alpha(T-1) + 1, \tag{A.16}$$

where $\theta_1(\phi)$ is defined as above. Next, we consider the Within-Group transformed components of the log-likelihood function. In particular, note that

$$\widehat{\sigma}^2(\alpha) = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t} - \alpha\widetilde{y}_{i,t-1})^2$$
$$= (1-\alpha)^2\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t-1})^2$$
$$+ (1-\alpha)\frac{2}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t-1})\widetilde{\varepsilon}_{i,t}$$
$$+ \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{\varepsilon}_{i,t})^2. \tag{A.17}$$

By Theorem 1 in Juodis (2018b), it follows immediately that

$$\begin{aligned}\mathrm{E}[\widehat{\sigma}^2(\alpha)] &= (1-\alpha)^2\left(\frac{T+1}{6}\right) - (1-\alpha) + 1 \\ &= \Delta_\alpha^2\sigma_1 - \Delta_\alpha + 1. \end{aligned} \tag{A.18}$$

Next, we plug in these expressions into the corresponding limiting FOC. From here, it is clear that all solutions Δ_α satisfy

$$\begin{aligned}(T-1)(2\Delta_\alpha\sigma_1 - 1)(\Delta_\alpha^2\theta_1(\phi) + \Delta_\alpha(T-1)+1) + (2\Delta_\alpha\theta_1(\phi) \\ + (T-1))(\Delta_\alpha^2\sigma_1 - \Delta_\alpha + 1) = 0. \end{aligned} \tag{A.19}$$

After a bit of re-arrangement, we obtain

$$\begin{aligned}\Delta_\alpha^3(2T\sigma_1\theta_1(\phi)) + \Delta_\alpha^2((T-1)(2\sigma_1(T-1) - \theta_1(\phi) + \sigma_1) - 2\theta_1) \\ + 2\Delta_\alpha((T-1)(\sigma_1 - 1) + \theta_1(\phi)) = 0. \end{aligned} \tag{A.20}$$

After removing the root $\Delta_\alpha = 0$, we are left with the following polynomial:

$$\begin{aligned}f(\Delta_\alpha) = \Delta_\alpha^2(2T\sigma_1\theta_1(\phi)) + \Delta_\alpha((T-1)(2\sigma_1(T-1) - \theta_1(\phi) + \sigma_1) - 2\theta_1) \\ + 2((T-1)(\sigma_1 - 1) + \theta_1(\phi)). \end{aligned} \tag{A.21}$$

As the coefficient in front of Δ_α^2 is clearly positive, we shall first consider the coefficient in front of Δ_α. Noting that

$$\begin{aligned}(T-1)(2\sigma_1(T-1) - \theta_1 + \sigma_1) - 2\theta_1 &= \sigma_1(2T-1)(T-1) - \theta_1(T+1) \\ &= \frac{(2T-1)(T+1)}{6}(T - 1 - (T-1)) = 0, \end{aligned} \tag{A.22}$$

the quadratic equation simplifies to

$$\begin{aligned}f(\Delta_\alpha) &= \Delta_\alpha^2(2T\sigma_1\theta_1(\phi)) + 2((T-1)(\sigma_1 - 1) + \theta_1(\phi)) - \Delta_\alpha(T+1)(\theta_1(\phi) - \theta_1) \\ &= \Delta_\alpha^2(2T\sigma_1\theta_1) + 2((T-1)(\sigma_1 - 1) + \theta_1) \\ &\quad + \widetilde{\sigma}_{y_0}^2(\Delta_\alpha^2(2T\sigma_1) - \Delta_\alpha(T+1) + 2). \end{aligned} \tag{A.23}$$

It is not difficult to see that

$$(T-1)(\sigma_1 - 1) + \theta_1 = (T-1)\left(\frac{T+1}{6} - 1 + \frac{2T-1}{6}\right) = \frac{T-1}{6}(3T-6) \geq 0, \tag{A.24}$$

for all $T \geq 2$. From here, we conclude that

$$\Delta_\alpha^2(2T\sigma_1\theta_1) + 2((T-1)(\sigma_1 - 1) + \theta_1) > 0, \tag{A.25}$$

for all Δ_α. Similarly, it is easy to see that for all Δ_α:

$$\Delta_\alpha^2 2T\sigma_1 - \Delta_\alpha(T+1) + 2 > 0. \tag{A.26}$$

Hence, we conclude that

$$f(\Delta_\alpha) > 0, \tag{A.27}$$

for all Δ_α. As a result, the only real root of the initial cubic equation is $\Delta_\alpha = 0$, or $\alpha = 1$. □

Proof Proposition 3 The proof of this proposition (with minor changes) is largely based on Theorem 1 in Juodis (2018b). The only element of the Hessian matrix that depends on the actual value of ϕ is the one in the top left corner, i.e., the (1, 1) element, all other elements are the same as $\theta_0^2 = \sigma_0^2$ is invariant to ϕ in the unit-root case.

In particular, this element can be expressed as:

$$\mathcal{H}_\ell^{(1,1)}(\phi) = -\left(\frac{\sigma_1(T-1)}{\sigma_0^2} + \frac{\theta_1(\phi)}{\theta_0^2}\right). \tag{A.28}$$

Note that $\sigma_1(T-1) = \sigma_0^2(T+1)(T-1)/6$ where we use the σ_1 notation as in the proof of Proposition 2. Similarly,

$$\theta_1(\phi) = T(1-\phi)^2\,\mathrm{E}[y_{i,0}^2] + \frac{(T-1)(2T-1)}{6}\sigma_0^2. \tag{A.29}$$

Combining the two expressions, the first result of this proposition follows immediately.

As for the second part of this proposition, this follows directly from the fact that it is only the (1, 1) element that is affected by $\widetilde{\sigma}_{y_0}^2(\phi)$. Thus, using direct calculations, we conclude that

$$|\mathcal{H}_\ell(\phi)| = |\mathcal{H}_\ell(1)| + \frac{\widetilde{\sigma}_{y_0}^2(\phi)}{(T-1)\sigma_0^4}. \tag{A.30}$$

The desired result follows by noticing that $|\mathcal{H}_\ell(1)| = |\mathcal{H}_\ell^{TML}| = 0$. □

Appendix A.2: Large T results

Proof Proposition 4 In what follows, we sketch the main arguments behind the claim of this proposition. First of all, we note that consistency of $\widehat{\alpha}$ can be directly established from Eq. (20). A similar result holds for $\widehat{\sigma}^2$, as a direct implication. Next, in order to

avoid any triangular array arguments, it is easier to analyze the log-likelihood function after concentrating out θ^2 and keeping the parameter vector at $\boldsymbol{\psi} = (\alpha, \sigma^2)'$. From here:

$$\ell(\boldsymbol{\psi}) = -\frac{N}{2}\left((T-1)\log(\sigma^2) + \log\left(\frac{T}{N\theta^2}\sum_{i=1}^{N}(\dot{y}_i - \alpha\dot{y}_{i-})^2\right) + \frac{1}{N\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{y}_{i,t} - \alpha\widetilde{y}_{i,t-1})^2\right), \tag{A.31}$$

where, for simplicity, we drop the dependence on ϕ in the definition of $\dot{y}_i$ and $\dot{y}_{i-}$, as this dependence is inconsequential for the final result of this proposition. Using analogous steps to those used in the proof of Proposition 1, we can show that

$$\begin{aligned}\nabla_\alpha \ell = {} & \frac{1}{\sigma_0^2}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{\varepsilon}_{i,t} - \mathrm{E}[\widetilde{y}_{i,t-1}\widetilde{\varepsilon}_{i,t}]\right) \\ & + \frac{T}{\widehat{\theta^2}(\alpha_0)}\sum_{i=1}^{N}(\dot{y}_{i-}(\dot{y}_i - \alpha_0\dot{y}_{i-}) - \mathrm{E}[\dot{y}_{i-}(\dot{y}_i - \alpha_0\dot{y}_{i-})]) \\ & - \xi\frac{N}{T}\left(\frac{\widehat{\theta}^2(\alpha_0) - \mathrm{E}[\widehat{\theta}^2(\alpha_0)]}{\widehat{\theta}^2(\alpha_0)}\right) \\ & - \frac{N}{\widehat{\theta^2}(\alpha_0)}\mathrm{E}[y_{i,0}(\mu_i - \phi y_{i,0})]\left(1 - \alpha_0^T\right)(\phi - 1).\end{aligned} \tag{A.32}$$

Next, we show that

$$\begin{aligned}\frac{1}{\sqrt{NT}}\nabla_\alpha \ell = {} & \frac{1}{\sigma_0^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{\varepsilon}_{i,t} - \mathrm{E}[\widetilde{y}_{i,t-1}\widetilde{\varepsilon}_{i,t}]\right) \\ & + \mathcal{O}_P(1/\sqrt{T}) + \mathcal{O}_P(1/\sqrt{T}) + \mathcal{O}_P(\sqrt{N/T^3}).\end{aligned} \tag{A.33}$$

As $\mathrm{E}[\widehat{\theta}^2(\alpha_0)] = \sigma_0^2 + T(1-\alpha_0)^2\,\mathrm{E}[(\mu_i - \phi y_{i,0})^2] = \mathcal{O}(T)$, the above result holds provided that

$$\widehat{\theta}^2(\alpha_0) - \mathrm{E}[\widehat{\theta}^2(\alpha_0)] = \mathcal{O}_P(T/\sqrt{N}), \tag{A.34}$$

$$\frac{1}{N}\sum_{i=1}^{N}(\dot{y}_{i-}(\dot{y}_i - \alpha_0\dot{y}_{i-}) - \mathrm{E}[\dot{y}_{i-}(\dot{y}_i - \alpha_0\dot{y}_{i-})]) = \mathcal{O}_P(1/\sqrt{N}). \tag{A.35}$$

In particular, let

$$\begin{aligned}\widehat{\theta}^2(\alpha_0) - \mathrm{E}[\widehat{\theta}^2(\alpha_0)] &= \frac{1}{N}\sum_i (T(\overline{\varepsilon}_i)^2 - \sigma_0^2) + \frac{T}{N}\sum_i (1-\alpha_0)^2((\mu_i - \phi y_{i,0})^2 \\ &\quad - \mathrm{E}[(\mu_i - \phi y_{i,0})^2]) \\ &\quad + 2\frac{(1-\alpha_0)^2}{N}\sum_i\sum_t (\mu_i - \phi y_{i,0})\varepsilon_{i,t} \\ &= \frac{1}{NT}\sum_i\sum_t (\varepsilon_{i,t}^2 - \sigma_0^2) + \frac{1}{NT}\sum_i\sum_t\sum_{s\neq t}\varepsilon_{i,t}\varepsilon_{i,s} \\ &\quad + \mathcal{O}_P(T/\sqrt{N}) + \mathcal{O}_P(\sqrt{T/N}) \\ &= \mathcal{O}_P(1/\sqrt{NT}) + \mathcal{O}_P(1/\sqrt{N}) + \mathcal{O}_P(T/\sqrt{N}) + \mathcal{O}_P(\sqrt{T/N}).\end{aligned} \tag{A.36}$$

Here the corresponding orders can be obtained by continuous application of Chebyshev's inequality and the fact that all stochastic elements have a finite fourth moment, while $\{\varepsilon_{i,t}\}$ are independent over all (i,t). Finally, the following conclusion:

$$\frac{1}{N}\sum_{i=1}^{N}(\dot{y}_{i-}(\dot{y}_i - \alpha_0\dot{y}_{i-}) - \mathrm{E}[\dot{y}_{i-}(\dot{y}_i - \alpha_0\dot{y}_{i-})]) = \mathcal{O}_P(1/\sqrt{N}),$$

can be established analogously upon noticing that $\xi/T = O(1)$, and using the fact that

$$\begin{aligned}&\frac{1}{T^2N}\sum_i\sum_{l=1}^{T-1}\sum_{t=1}^{T-1}\left(\sum_{s=0}^{T-1-t}\alpha_0^s\right)\varepsilon_{i,t}\varepsilon_{i,l} - \frac{\xi\sigma_0^2}{T} \\ &= \frac{1}{T^2N}\sum_i\sum_{l=1}^{T-1}\sum_{t=1}^{T-1}\left(\sum_{s=0}^{T-1-t}\alpha_0^s\right)(\varepsilon_{i,t}\varepsilon_{i,l} - \mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,l}]) \\ &= \mathcal{O}_P(1/\sqrt{NT^3}) + \mathcal{O}_P(1/\sqrt{NT^2}),\end{aligned} \tag{A.37}$$

since this rate is the same as with $\frac{1}{NT}\sum_i (T(\overline{\varepsilon}_i)^2 - \sigma_0^2)$. For this result, it is critical that $\left(\sum_{s=0}^{T-1-t}\alpha_0^s\right) = O(1)$ for $|\alpha_0| < 1$. Furthermore, using similar derivations we can further simplify

$$\frac{1}{\sqrt{NT}}\nabla_\alpha \ell = \frac{1}{\sigma_0^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t} + \mathcal{O}_P(1/\sqrt{T}) + \mathcal{O}_P(\sqrt{N/T^3}). \tag{A.38}$$

From here, the main result follows directly using the results in Hahn and Kuersteiner (2002). □

Proof Proposition 5 Define the following normalization matrix:

$$\boldsymbol{S} = \begin{pmatrix} T^{(1\vee\beta)} & 0 & 0 \\ 0 & \sqrt{T} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{A.39}$$

Then, it follows that

$$\boldsymbol{S}^{-1}\boldsymbol{\mathcal{H}}_\ell \boldsymbol{S}^{-1} \to \boldsymbol{\mathcal{H}}_\ell^\infty = \begin{pmatrix} h_{11} & 0 & h_{13} \\ 0 & \frac{1}{2\sigma^4} & 0 \\ h_{13} & 0 & \frac{1}{2\sigma^4} \end{pmatrix}, \tag{A.40}$$

where several cases are available:

$$h_{11} = \begin{cases} 1/2, & \beta < 1; \\ 1/2 + \lim_{T\to\infty} \widetilde{\sigma}_0^2/T^2, & \beta = 1; \\ \lim_{T\to\infty} \widetilde{\sigma}_0^2/T^{1+\beta}, & \beta > 1. \end{cases} \tag{A.41}$$

$$h_{13} = \begin{cases} 1/(2\sigma_0^2), & \beta \le 1; \\ 0, & \beta > 1. \end{cases} \tag{A.42}$$

□

References

Ahn SC, Thomas GM (2006) Likelihood based inference for dynamic panel data models. Mimeo

Alvarez J, Arellano M (2003) The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71(4):1121–1159

Alvarez J, Arellano M (2004) Robust likelihood estimation of dynamic panel data models. Mimeo

Alvarez J, Arellano M (2022) Robust likelihood estimation of dynamic panel data models. J Econom 226:21–61 (annals Issue in Honor of Gary Chamberlain)

Anderson T, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. J Econom 18:47–82

Arellano M (2016) Modeling optimal instrumental variables for dynamic panel data models. Res Econ 70:238–261

Arellano M, Blundell R, Bonhomme S (2017) Earnings and consumption dynamics: a nonlinear panel data framework. Econometrica 85:693–734

Arellano M, Bonhomme S (2009) Robust priors in nonlinear panel data models. Econometrica 77(2):489–536

Bai J (2013) Fixed-effects dynamic panel models, a factor analytical method. Econometrica 81:285–314

Bai J (2013b) Likelihood approach to dynamic panel models with interactive effects. Mimeo

Balestra P, Nerlove M (1966) Pooling cross section and time series data in the estimation of a dynamic model: the demand for Natural Gas. Econometrica 34:585–612

Binder M, Hsiao C, Pesaran MH (2005) Estimation and inference in short panel vector autoregressions with unit root and cointegration. Econom Theor 21:795–837

Blundell R, Bond S (2023) Initial conditions and blundell-bond estimators. J Econom 234:101–110 (jubilee Issue - Celebrating our Fiftieth Anniversary: 1973–2023)

Botosaru I (2022) Time-varying unobserved heterogeneity in earnings shocks. J Econom (forthcoming)

Botosaru I, Sasaki Y (2018) Nonparametric heteroskedasticity in persistent panel processes: an application to earnings dynamics. J Econom 203:283–296

Bun MJG, Carree MA (2005) Bias-corrected estimation in dynamic panel data models. J Bus Econ Stat 23(2):200–210
Bun MJG, Carree MA, Juodis A (2017) On maximum likelihood estimation of dynamic panel data models. Oxford Bull Econ Stat 79:463–494
Chamberlain G (1982) Multivariate regression models for panel data. J Econom 18:5–46
Dhaene G, Jochmans K (2016) Likelihood inference in an autoregression with fixed effects. Econom Theor 32:1178–1215
Dovonon P, Hall A (2018) The asymptotic properties of GMM and indirect inference under second-order identification. J Econom 205:76–111
Everaert G (2013) Orthogonal to backward mean transformation for dynamic panel data models. Econom J 16:179–221
Hahn J, Kuersteiner G (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. Econometrica 70(4):1639–1657
Hahn J, Kuersteiner G, Cho MH (2004) Asymptotic distribution of misspecified random effects estimator for a dynamic panel model with fixed effects when both n and T are large. Econ Lett 84:117–125
Hayakawa K, Pesaran MH (2015) Robust standard errors in transformed likelihood estimation of dynamic panel data models. J Econom 188:111–134
Hsiao C, Pesaran MH, Tahmiscioglu AK (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J Econom 109:107–150
Juodis A (2018) First difference transformation in panel VAR models: robustness, estimation and inference. Econom Rev 37:650–693
Juodis A (2018) Rank based cointegration testing for dynamic panels with fixed T. Empir Econ 55:349–389
Juodis A, Poldermans RW (2021) Backward mean transformation in unit root panel data models. Econ Lett 201:109780
Juodis A, Sarafidis V (2018) Fixed T dynamic panel data estimators with multi-factor errors. Econom Rev 37:893–929
Juodis A, Sarafidis V (2022) An incidental parameters free inference approach for panels with common shocks. J Econom 229:19–54
Juodis A, Westerlund J (2019) Optimal panel unit root testing with covariates. Econom J 22:55–72
Kruiniger H (2013) Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. J Econom 173:175–188
Kruiniger H (2018) A further look at modified ML estimation of the panel AR(1) model with fixed effects and arbitrary initial conditions. Working paper
Lancaster T (2002) Orthogonal parameters and panel data. Rev Econ Stud 69:647–666
Maddala GS (1971) The use of variance components models in pooling cross section and time series data. Econometrica 39:341–358
Moon HR, Perron B, Phillips PCB (2007) Incidental trends and the power of panel unit root tests. J Econom 141:416–459
Moral-Benito E (2013) Likelihood-based estimation of dynamic panels with predetermined regressors. J Bus Econ Stat 31:451–472
Mundlak Y (1978) On the pooling of time series and cross section data. Econometrica 46:69–85
Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 49:1417–1426
Norkutė M, Westerlund J (2021) The factor analytical approach in near unit root interactive effects panels. J Econom 221:569–590
Roznitzky A, Cox DR, Bottai M, Robins J (2000) Likelihoood-based inference with singular information matrix. Bernoulli 6:243–284
Westerlund J (2016) Pooled panel unit root tests and the effect of past initialization. Econom Rev 35:396–427

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.