University of Groningen

On Maximum Likelihood Estimation of Dynamic Panel Data Models

Bun, Maurice J G; Carree, Martin A.; Juodis, Arturas

Published in:
Oxford Bulletin of Economics and Statistics

DOI:
[10.1111/obes.12156](https://doi.org/10.1111/obes.12156)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bun, M. J. G., Carree, M. A., & Juodis, A. (2017). On Maximum Likelihood Estimation of Dynamic Panel Data Models. *Oxford Bulletin of Economics and Statistics*, 79(4), 463-494.
<https://doi.org/10.1111/obes.12156>

Copyright

Other than for strictly personal use, 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), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

On Maximum Likelihood Estimation of Dynamic Panel Data Models*

MAURICE J.G. BUN,[†] MARTIN A. CARREE,[‡] and ARTŪRAS JUODIS[§]

[†]*Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands
(e-mail: m.j.g.bun@uva.nl)*

[‡]*School of Business and Economics, Maastricht University, Maastricht, The Netherlands
(e-mail: m.carree@maastrichtuniversity.nl)*

[§]*Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands
(e-mail: a.juodis@rug.nl)*

Abstract

We analyse the finite sample properties of maximum likelihood estimators for dynamic panel data models. In particular, we consider transformed maximum likelihood (TML) and random effects maximum likelihood (RML) estimation. We show that TML and RML estimators are solutions to a cubic first-order condition in the autoregressive parameter. Furthermore, in finite samples both likelihood estimators might lead to a negative estimate of the variance of the individual-specific effects. We consider different approaches taking into account the non-negativity restriction for the variance. We show that these approaches may lead to a solution different from the unique global unconstrained maximum. In an extensive Monte Carlo study we find that this issue is non-negligible for small values of T and that different approaches might lead to different finite sample properties. Furthermore, we find that the Likelihood Ratio statistic provides size control in small samples, albeit with low power due to the flatness of the log-likelihood function. We illustrate these issues modelling US state level unemployment dynamics.

I. Introduction

Dynamic panel data models have a prominent place in applied research and at the same time form a challenging field in econometric theory. Many panel data applications have a relatively small number of time periods T , whereas the cross sectional dimension N is

JEL Classification numbers: C13, C23.

*This paper greatly benefited from comments made by Anindya Banerjee (Editor) and by two anonymous referees. The authors thank Peter Boswijk, Geert Dhaene, David Edgerton, Hugo Kruiniger, Peter Schmidt, Joakim Westerlund, Jeffrey Wooldridge and participants of the Netherlands Econometric Study Group Meeting 2015 in Maastricht, the 21st International Panel Data Conference in Budapest, seminars at University of Amsterdam and Lund University for helpful comments and suggestions. Financial support from the NWO MaGW Grants ‘Causal inference with panel data’ and ‘Likelihood-based inference in dynamic panel data models with endogenous covariates’ is gratefully acknowledged by the first and third authors respectively.

sizeable. It is therefore common to consider the asymptotic behaviour of estimators and corresponding test statistics with T fixed and only N tending to infinity.

A central theme in linear dynamic panel data analysis is the fact that the fixed effects (FE) estimator is inconsistent for fixed T and N large. This inconsistency is referred to as the Nickell (1981) bias, and is an example of the incidental parameters problem. It has therefore become common practice to estimate the parameters of dynamic panel data models by the generalized method of moments (GMM), see Arellano and Bond (1991) and Blundell and Bond (1998). A main reason for using GMM is that it provides asymptotically efficient inference exploiting a minimal set of statistical assumptions. GMM inference has not been without its own problems, however. These include small sample biases in both coefficient and variance estimators, sensitivity to important nuisance parameters and choices regarding the type and number of moment conditions. A large literature has been devoted to adapting the GMM approach to limit the impact of these inherent drawbacks, see Bun and Sarafidis (2015) for a recent overview.

This again has led to an interest in likelihood-based methods that correct for the incidental parameters problem. Some of these methods are based on modifications of the profile likelihood, see Lancaster (2002) and Dhaene and Jochmans (2016). Other methods start from the likelihood function of the first differences, see Hsiao, Pesaran and Tahmiscioglu (2002). Essentially these methods treat the incidental parameters as fixed in estimation. The alternative approach is to assume random effects, but in dynamic models it is then necessary to be explicit about the non-zero correlation between individual-specific effects and initial conditions (Anderson and Hsiao, 1982; Alvarez and Arellano, 2003; Moral-Benito, 2012). Random effects type ML estimators therefore typically exploit Chamberlain (1982)-type projections to model the dependence between individual-specific effects, initial observations and additional covariates.

In this study, we consider 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).¹ There is a close connection between TML and RML in the sense that TML can be expressed as a restricted version of RML. Under suitable regularity conditions, ML estimators are consistent and asymptotically normally distributed (Amemiya, 1985; Kruiniger, 2013). Monte Carlo evidence in Hsiao *et al.* (2002), Alvarez and Arellano (2003) and Hayakawa and Pesaran (2015), suggest that these likelihood-based approaches can serve as viable alternatives to the usual GMM estimators. Just like for GMM, however, the application of ML estimators is not without its own problems.

We address two important issues when implementing ML for dynamic panel data models. First, we show that in the simple setup without time-series heteroscedasticity, both the TML and RML estimators give rise to a cubic first-order condition in the autoregressive parameter. We therefore have either one or three real solutions to the first-order conditions. As a result, even asymptotically the log-likelihood function can be bimodal. This result is different from Kruiniger (2008) and Han and Phillips (2013), who find a quartic equation assuming covariance stationarity.

Second, because both TML and RML can be seen as correlated random effect ML estimators, we address the issue of negative variance estimates as mentioned in Maddala

¹The former estimator is also referred to as fixed effects ML in Kruiniger (2013).

(1971) and Alvarez and Arellano (2003). An important consequence of bimodality is that unconstrained maximization of the log-likelihood may lead to ML estimates which do not satisfy the implicit restriction of non-negative variances. Enforcing this non-negativity constraint may furthermore lead to a boundary solution.

We further investigate the impact of multiple roots and boundary conditions in finite samples in a Monte Carlo study. We consider the finite sample bias and RMSE of coefficient estimators as well as size and power of corresponding t - and LR statistics. We find that, despite the robustness of TML and RML to initial conditions (i.e. they remain consistent as long as initial conditions have finite variances), the finite sample properties of both estimators for small values of T depend heavily on the initial condition. A partial explanation is that the behaviour of the initial condition has a direct effect on the bimodality of the log-likelihood function. We find that when there are three solutions to the first-order condition, the left solution always satisfies the non-negativity restriction, while the right solution violates it. Estimators taking into account non-negativity constraints perform much better than unconstrained counterparts. Furthermore, we find that inference based on the LR statistic is size correct, while t -statistics show large size distortions. Using the data set in Bun and Carree (2005) we show how these theoretical results can influence empirical estimates of US state level unemployment dynamics.

Throughout the analysis, we limit ourselves to an asymptotic analysis in which T is fixed and $N \rightarrow \infty$. When T is large the influence of initial conditions becomes negligible, hence our main results become less relevant. For results with T large, see e.g. Bai (2013). Furthermore, we do not analyse the unit root case (or explosive root). Distribution theory becomes rather different in this case, see Ahn and Thomas (2006) and Kruiniger (2013).

The plan of this study is as follows. In section II, we introduce the Maximum Likelihood estimators for the panel AR(1) model including the cubic first-order condition for the autoregressive parameter, with extensions to dynamic models with additional covariates in subsection ‘Extension to exogenous regressors’. Section III deals with the possibility of multiple solutions and proposes bounded estimation as a solution. Section IV reports the results from the Monte Carlo study, while section V shows the empirical results. Section VI concludes. Proofs of all propositions are provided in the supplementary online appendix.

II. ML estimation for the panel AR(1) model

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

$$y_{i,t} = \eta_i + \phi y_{i,t-1} + \varepsilon_{i,t}, \quad E[\varepsilon_{i,t} | y_{i,0}, \eta_i] = 0, \quad (1)$$

for $i = 1, \dots, N$, $t = 1, \dots, T$. We assume that the idiosyncratic errors $\varepsilon_{i,t}$ are i.i.d. $(0, \sigma^2)$ and that initial conditions $y_{i,0}$ are observed.³ Stacking the observations over time, we can write the AR(1) model for each individual as follows:

$$\mathbf{y}_i = \phi \mathbf{y}_{i-} + \mathbf{1}_T \eta_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,T})', \quad (2)$$

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

³The estimators considered in this paper remain consistent even if in the population some variance components are heteroscedastic over i . In this case both estimators presented should be treated as pseudo maximum likelihood estimators, see also Hayakawa and Pesaran (2015).

with y_i and y_{i-} defined accordingly and $\mathbf{1}_T$ a vector of ones. We follow Kruiniger (2013) to derive the log-likelihood function(s), but final results are identical to those in Hsiao *et al.* (2002) or Alvarez and Arellano (2003). As discussed in e.g. Kruiniger (2013), the crucial assumptions for consistency, asymptotic normality and feasible inference of the likelihood estimation can be summarized as follows:

Assumption TML: $\psi_i \equiv y_{i,0} - \eta_i/(1 - \phi)$ are i.i.d. across i , with finite fourth moments.

Assumption RML: The vector $(y_{i,0}, \eta_i)$ is i.i.d. across i , with finite fourth moments.

Remark 1. Note that Assumption **TML** is somewhat less restrictive than **RML**, because one has to impose distributional assumptions only on a particular linear combination of $y_{i,0}$ and η_i . Assumption **RML**, on the other hand, imposes restrictions on the joint distribution of $y_{i,0}$ and η_i . However, neither of the assumptions imposes any stationarity restrictions (mean and/or covariance) on the initial condition.

To derive the corresponding estimating equations for both estimators, it is convenient to use the Chamberlain (1982) type of projection for η_i :⁴

$$\eta_i = \pi y_{i,0} + v_i, \quad E[v_i y_{i,0}] = 0, \quad v_i \sim \text{i.i.d.}(0, \sigma_v^2). \quad (3)$$

Note that this projection is only necessary for the RML estimator, while the i.i.d. assumption is only imposed for simplicity as the discussed estimators remain consistent even if v_i is heteroscedastic. When we set $\pi = 1 - \phi$ the projection corresponds exactly to the TML framework as in this case $v_i = -(1 - \phi)\psi_i$. In the TML approach, the variance of $v_i = \Delta y_{i,1} - \varepsilon_{i,1}$ is a parameter to be estimated. We therefore only exploit the projection in equation (3) to show the algebraic comparison between RML and TML. The model can be represented as:

$$y_i - \phi y_{i-} = \mathbf{1}_T \pi y_{i,0} + \mathbf{1}_T v_i + \varepsilon_i. \quad (4)$$

Define the combined error term as

$$\mathbf{u}_i \equiv \mathbf{1}_T v_i + \varepsilon_i, \quad (5)$$

then it follows under our assumption

$$E[\mathbf{u}_i] = \mathbf{0}_T, \quad \text{var}[\mathbf{u}_i] = \Sigma = \sigma_v^2 \mathbf{1}_T \mathbf{1}_T' + \sigma^2 \mathbf{I}_T. \quad (6)$$

The variance–covariance structure of Σ is of the usual ‘correlated’ random effects form. Using the matrix inversion and determinant lemmas, we obtain:

$$\Sigma^{-1} = \frac{1}{\sigma^2} \mathbf{I}_T - \frac{1}{\sigma^2} \frac{\sigma_v^2}{\sigma^2 + T \sigma_v^2} \mathbf{1}_T \mathbf{1}_T', \quad |\Sigma| = (\sigma^2)^{T-1} (\sigma^2 + T \sigma_v^2). \quad (7)$$

Denote by $\mathbf{W}_T \equiv \mathbf{I}_T - \frac{1}{T} \mathbf{1}_T \mathbf{1}_T'$ the usual fixed effects projection matrix, then we can write:

$$\Sigma^{-1} = \frac{1}{\sigma^2} \mathbf{W}_T + \frac{1}{T \theta^2} \mathbf{1}_T \mathbf{1}_T', \quad |\Sigma| = (\sigma^2)^{T-1} \theta^2, \quad (8)$$

⁴For simplicity we do not include a constant term in the projection as it would serve as a restricted time effect.

$$\theta^2 \equiv \sigma^2 + T\sigma_v^2. \quad (9)$$

Hence, instead of estimating σ^2 and σ_v^2 , we rather estimate σ^2 and θ^2 . By doing so we do not restrict σ_v^2 to be positive. This parametrization, in one form or the other, has been used in Hsiao *et al.* (2002), Alvarez and Arellano (2003), Ahn and Thomas (2006), and Kruiniger (2008) *inter alia*.

Now we define the quasi log-likelihood function for some individual i (up to a constant):

$$\ell_i(\boldsymbol{\kappa}) = -\frac{1}{2} \left((T-1)\log(\sigma^2) + \log(\theta^2) + [(y_i - \phi y_{i-} - \mathbf{1}_T \pi y_{i,0})' \boldsymbol{\Sigma}^{-1} (y_i - \phi y_{i-} - \mathbf{1}_T \pi y_{i,0})] \right), \quad (10)$$

where $\boldsymbol{\kappa} = (\phi, \pi, \sigma^2, \theta^2)'$. This function is the true likelihood function if u_i are jointly normal.⁵ Now using the fact that $\mathbf{W}_T \mathbf{1}_T = \mathbf{0}_T$, we can write:

$$\begin{aligned} \ell_i(\boldsymbol{\kappa}) = & -\frac{1}{2} \left((T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{\sigma^2} [(y_i - \phi y_{i-})' \mathbf{W}_T (y_i - \phi y_{i-})] \right. \\ & \left. + \frac{T}{\theta^2} (\bar{y}_i - \phi \bar{y}_{i-} - \pi y_{i,0})^2 \right), \end{aligned} \quad (11)$$

where $\bar{y}_i \equiv (1/T) \sum_{t=1}^T y_{i,t}$ and $\bar{y}_{i-} \equiv (1/T) \sum_{t=1}^T y_{i,t-1}$. Furthermore, we define $\tilde{y}_{i,t} \equiv y_{i,t} - \bar{y}_i$, $\tilde{y}_{i,t-1} \equiv y_{i,t-1} - \bar{y}_{i-}$, $\ddot{y}_i \equiv \bar{y}_i - y_{i,0}$, $\ddot{y}_{i-} \equiv \bar{y}_{i-} - y_{i,0}$ and $\rho \equiv \pi - (1 - \phi)$. This implies the following final expression for the log-likelihood function (after summing over all individual log-likelihood functions):

$$\begin{aligned} \ell(\boldsymbol{\kappa}) = & -\frac{N}{2} \left((T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1})^2 \right. \\ & \left. + \frac{T}{N\theta^2} \sum_{i=1}^N (\ddot{y}_i - \phi \ddot{y}_{i-} - \rho y_{i,0})^2 \right). \end{aligned} \quad (12)$$

The log-likelihood function for observations in first-differences (also known as transformed log-likelihood in Hsiao *et al.*, 2002 and Juodis, 2017) is obtained by setting $\rho = 0$.

The parameters σ^2 , θ^2 (and ρ for RML) can be concentrated out, resulting in (up to a constant):

$$\ell^c(\phi) = -\frac{N}{2} \left((T-1)\log \hat{\sigma}^2(\phi) + \log \hat{\theta}^2(\phi) \right), \quad (13)$$

where:

$$\hat{\sigma}^2(\phi) = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1})^2, \quad \hat{\theta}^2(\phi) = \frac{T}{N} \sum_{i=1}^N (\ddot{y}_i - \phi \ddot{y}_{i-})^2. \quad (14)$$

Here for the random effects log-likelihood function we defined $\dot{y}_i$ and $\dot{y}_{i-}$:

$$\dot{y}_i \equiv \ddot{y}_i - y_{i,0} \frac{\sum_{i=1}^N \ddot{y}_i y_{i,0}}{\sum_{i=1}^N y_{i,0}^2}, \quad \dot{y}_{i-} \equiv \ddot{y}_{i-} - y_{i,0} \frac{\sum_{i=1}^N \ddot{y}_{i-} y_{i,0}}{\sum_{i=1}^N y_{i,0}^2}, \quad (15)$$

⁵ Alternatively, following Hsiao *et al.* (2002) one can estimate $\omega \equiv 1 + \sigma_v^2/\sigma^2 = \text{var}(\Delta y_{i,1})/\sigma^2$. This parametrization ensures that $\theta^2 > 0$ or equivalently $\omega > 1 - \frac{1}{T}$.

while for the log-likelihood in first differences $\dot{y}_i \equiv \ddot{y}_i$ and $\dot{y}_{i-} \equiv \ddot{y}_{i-}$. Thus equation (13) shows that both RML and TML estimators provide an in-built bias-correction term ($\log \hat{\theta}^2(\phi)$) for the usual fixed effects log-likelihood function for small T .

The likelihood function in (13) is defined for all values of $\phi \in \mathbb{R}$, hence from a theoretical and computational point of view there are no reasons to consider a restricted parameter space for estimation. Nevertheless, some studies (Hsiao *et al.*, 2002, p. 135 and Hayakawa and Pesaran, 2015, p. 123, footnote 14) restrict $\phi \in (-1; 1)$. This may have consequences for the finite sample properties of the resulting estimators, as we shall see below. Furthermore, the fact that the likelihood function is defined over the whole real line, is in contrast with the likelihood function in Kruiniger (2008) and Han and Phillips (2013). In these studies, covariance stationarity of the initial conditions is imposed, hence the likelihood function is naturally defined only for $-1 < \phi \leq 1$.

The FOC (first order condition) for the autoregressive parameter ϕ can now be expressed in the following way:

$$\frac{d\ell^c(\phi)}{d\phi} = \frac{1}{\hat{\sigma}^2(\phi)} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1} (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1}) + \frac{T}{\hat{\theta}^2(\phi)} \sum_{i=1}^N \dot{y}_{i-} (\dot{y}_i - \phi \dot{y}_{i-}) = 0, \quad (16)$$

or alternatively:

$$\hat{\sigma}^2(\phi) \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1} (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1}) + \hat{\sigma}^2(\phi) T \sum_{i=1}^N \dot{y}_{i-} (\dot{y}_i - \phi \dot{y}_{i-}) = 0. \quad (17)$$

Given that $\hat{\sigma}^2(\phi)$ and $\hat{\theta}^2(\phi)$ are quadratic in ϕ it is not difficult to see that the FOC is cubic in ϕ . Thus for any value of T and any realization of $\{y_i\}_{i=1}^N$ there will be at least one and at most three real solutions to equation (17). For a general value of T there is no easy formula for the solutions,⁶ but in the next section we will obtain interesting analytical results for three-wave panels. In any case the solutions of the cubic equation can be found without any need of explicit numerical maximization. One can simply use root finder algorithms based on the eigenvalues of the companion matrix.

For some reason, the possibility of multiple solutions is mostly forgotten when discussing both likelihood estimators. An exception is Hayakawa and Pesaran (2015) who observe that the TML log-likelihood function can have more than one solution asymptotically. However, their results suggest the presence of two solutions asymptotically, which is incomplete in the light of our result in equation (17). Here we show multiple solutions can also occur in finite samples, and that both the TML and RML log-likelihood functions have one or two local maxima. More importantly, as we show later in section ‘Extension to exogenous regressors’, this result is unaffected if strictly exogenous regressors are added to the model.

Given the structure of the log-likelihood function, we can easily specify the interval for all solutions $\hat{\phi}$. In particular, we have the following:

⁶ One can use trigonometric formulas to obtain all solutions exactly without any numerical efforts, but analytical results using those solutions are far from trivial.

Figure 1. Concentrated (asymptotic) transformed maximum likelihood (TML) log-likelihood function for $T = 6$. In this figure the dotted line represents the within group part of the log-likelihood function, while the dashed line is the between group part. The solid line, which stands for the TML log-likelihood function, is the sum of dashed and dotted lines. The global maximum of this function is located at the true value $\phi_0 = 0.5$

Lemma 1. For any N and T all solutions of equation (17) lie in the following interval:

$$\hat{\phi} \in (\hat{\phi}_W, \hat{\phi}_B) = \left(\frac{\sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t} \tilde{y}_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1}^2}, \frac{\sum_{i=1}^N \dot{y}_i \dot{y}_{i-}}{\sum_{i=1}^N \dot{y}_{i-}^2} \right). \quad (18)$$

Furthermore, this result continues to hold if $N \rightarrow \infty$.

The proof of this lemma follows directly from the fact that the log-likelihood function is a sum of two quasi-concave functions with different maxima. The lower bound of this interval is the fixed effects ML estimator (also known as within group or LSDV estimator). The upper bound can be interpreted as a ‘quasi-between’ estimator. Figure 1 illustrates this. It is well known that $\text{plim}_{N,T \rightarrow \infty} \hat{\phi}_W = \phi$, but that for fixed T the within estimator has a negative bias, see Nickell (1981). Furthermore, it is straightforward to show that $\text{plim}_{N,T \rightarrow \infty} \hat{\phi}_B = 1$ because $\bar{y}_i$ and $\bar{y}_{i-}$ converge to the same value as T goes to infinity.

Next, we investigate the asymptotic behaviour of the interval in Lemma 1. As the lower bound ($\hat{\phi}_W$) is the same for both the RML and TML estimators we are primarily interested in the upper bound ($\hat{\phi}_B$), which is different for the two estimators. The result is summarized in the following proposition.⁷

Proposition 1. The probability limits of the estimators from Lemma 1 are as follows:

$$\text{plim}_{N \rightarrow \infty} \hat{\phi}_W < \phi_0 < \text{plim}_{N \rightarrow \infty} \hat{\phi}_B^{\text{RML}} \leq \text{plim}_{N \rightarrow \infty} \hat{\phi}_B^{\text{TML}}. \quad (19)$$

Thus the upper bound for RML is asymptotically no larger than for TML. The interval for possible values for $\hat{\phi}_{\text{RML}}$ is narrower than the corresponding interval for TML. This

⁷ From now on, where necessary to avoid confusion, we will use the subscript 0 to denote the true values of the parameters, e.g. ϕ_0, σ_0^2 .

result can be expected given that the RML estimator is found to be more efficient than TML (at least in the model with Gaussian errors), see Alvarez and Arellano (2004) and Kruiniger (2013).

Extension to exogenous regressors

For most empirically relevant applications the AR(1) model specification is too restrictive and incomplete. In this subsection, we therefore extend our analysis to an ARX(1) model including additional strictly exogenous regressors. Recently, the model of this type has been used by Moral-Benito (2012), who investigates the determinants of country level growth. Similarly, as a robustness check for GMM results, the RML type estimator was used by Jones and Tarp (2016) in studying the effects of aid on political institutions. For ease of exposition, we consider the following simplified version with one additional regressor:

$$y_{i,t} = \eta_i + \phi y_{i,t-1} + \beta x_{i,t} + \varepsilon_{i,t}, \quad E[\varepsilon_{i,t} | x_{i,0}, \dots, x_{i,T}, y_{i,0}, \eta_i] = 0. \quad (20)$$

Then using stacked notation for individual i we have:

$$\mathbf{y}_i = \phi \mathbf{y}_{i-} + \beta \mathbf{x}_i + \mathbf{1}_T \eta_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{x}_i = (x_{i,1}, \dots, x_{i,T})', \quad (21)$$

with other quantities defined as in the model without exogenous regressors. We continue by using the Chamberlain (1982) type of projection for η_i as in Bai (2013) on not only $y_{i,0}$, but now also all the lags and leads of $x_{i,t}$:

$$\eta_i = \boldsymbol{\pi}' \mathbf{w}_i + v_i, \quad E[v_i \mathbf{w}_i] = \mathbf{0}, \quad \mathbf{w}_i = (y_{i,0}, x_{i,0}, \mathbf{x}_i')'. \quad (22)$$

The main implication is that the combined error term $\mathbf{u}_i$ can be represented as follows:

$$\mathbf{u}_i = \mathbf{y}_i - \phi \mathbf{y}_{i-} - \beta \mathbf{x}_i - \mathbf{1}_T \boldsymbol{\pi}' \mathbf{w}_i. \quad (23)$$

The variance–covariance matrix remains the same as in the pure AR(1) model without exogenous regressors. The quasi log-likelihood function over all individuals is then given by (up to a constant):

$$\begin{aligned} -\frac{2}{N} \ell(\boldsymbol{\kappa}) &= (T-1) \log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1} - \beta \tilde{x}_{i,t})^2 \\ &\quad + \frac{T}{N\theta^2} \sum_{i=1}^N (\bar{y}_i - \phi \bar{y}_{i-} - \beta \bar{x}_i - \boldsymbol{\pi}' \mathbf{w}_i)^2, \end{aligned} \quad (24)$$

where $\boldsymbol{\kappa} = (\phi, \beta, \sigma^2, \theta^2, \boldsymbol{\pi})'$.

Similarly to the model without $x_{i,t}$ the TML estimator can be expressed as a restricted version (in terms of the parameter restrictions) of a more general RML estimator. Without loss of generality, we can rewrite the last component of the log-likelihood function as:

$$\frac{T}{N\theta^2} \sum_{i=1}^N (\bar{y}_i - \phi \bar{y}_{i-} - \beta \bar{x}_i - \boldsymbol{\pi}' \mathbf{w}_i)^2 = \frac{T}{N\theta^2} \sum_{i=1}^N (\ddot{y}_i - \phi \ddot{y}_{i-} - \beta \ddot{x}_i - \boldsymbol{\rho}' \mathbf{z}_i)^2, \quad (25)$$

where $\mathbf{z}_i \equiv (y_{i,0}, x_{i,0}, \Delta x_{i,1}, \dots, \Delta x_{i,T})'$ and $\ddot{x}_i \equiv \bar{x}_i - x_{i,0}$. Furthermore, using e.g. theorem 3.1 in Juodis (2016) the last component of the TML log-likelihood function is given by the following:

$$\frac{T}{N\theta^2} \sum_{i=1}^N (\ddot{y}_i - \phi \ddot{y}_{i-} - \beta \ddot{x}_i - \pi'_\Delta \Delta \mathbf{x}_i)^2, \quad \Delta \mathbf{x}_i = (\Delta x_{i,1}, \dots, \Delta x_{i,T})'. \quad (26)$$

Hence, by setting the first two components of the $\boldsymbol{\rho}$ vector to zero we obtain the TML estimator as the restricted version of the RML estimator.⁸

Irrespective of the estimator considered it is not difficult to see that, because $\ddot{x}_i \in \text{Span}(\Delta \mathbf{x}_i)$, one can concentrate out the $\boldsymbol{\rho}/\pi_\Delta$ parameter such that the last component of the log-likelihood function in (26) does not contain the β parameter. Therefore, the (concentrated) log-likelihood function can be expressed as follows:

$$\begin{aligned} -\frac{2}{N} \ell(\boldsymbol{\kappa}) = & (T-1) \log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1} - \beta \tilde{x}_{i,t})^2 \\ & + \frac{T}{N\theta^2} \sum_{i=1}^N (\dot{y}_i - \phi \dot{y}_{i-})^2, \end{aligned} \quad (27)$$

where we defined:

$$\dot{y}_i \equiv \ddot{y}_i - \left(\sum_{i=1}^N \ddot{y}_i \Delta \mathbf{x}'_i \right) \left(\sum_{i=1}^N \Delta \mathbf{x}_i \Delta \mathbf{x}'_i \right)^{-1} \Delta \mathbf{x}_i, \quad (28)$$

$$\dot{y}_i \equiv \ddot{y}_i - \left(\sum_{i=1}^N \ddot{y}_i \mathbf{z}'_i \right) \left(\sum_{i=1}^N \mathbf{z}_i \mathbf{z}'_i \right)^{-1} \mathbf{z}_i, \quad (29)$$

for the TML and RML estimators respectively (similarly for $\dot{y}_{i-}$). One can also concentrate out the β parameter from the first component of the log-likelihood function. After subsequent concentration of σ^2 and θ^2 , the resulting log-likelihood function is then of the same structure as in equation (13). The cubic FOC in equation (17) follows directly from that.

Summarizing, in this section, we argued that in the model augmented with exogenous regressors the FOC of the TML/RML estimators again is cubic in the autoregressive parameter ϕ . Our derivations above rely upon the fact that the full Chamberlain (1982) projection has been used, rather than the restricted Mundlak (1978) projection. Without going into further discussion, we state that for the TML estimator the results above do not carry over if one uses the Mundlak (1978) projection instead. However, they continue to be valid for the RML estimator if one does not include $x_{i,0}$ when exploiting the Mundlak (1978) projection.⁹

Remark 2. Here, we assume that the regressor $x_{i,t}$ is strictly exogenous. If this assumption cannot be justified, and additional regressors are only weakly exogenous one has to modify the log-likelihood function. Particularly, one can no longer use the Chamberlain (1982) projection on all leads and lags and instead one has to project on initial observations only. From an analytical point of view, this implies that one cannot concentrate out the β parameter and hence the result of this section does not trivially extend. For a more detailed analysis, see the supplementary online appendix.

⁸ Note that the interpretation of TML as restricted version of RML is only valid if one includes $x_{i,0}$ in $\mathbf{w}_i$.

⁹ Or if one does not impose that the coefficient for $x_{i,0}$ is identical to the one for $x_{i,1}, \dots, x_{i,T}$.

III. Multiple solutions and bounded estimation

The possibility of having one or three solutions to the cubic equation (17) has important consequences. We first characterize the solutions for the case in which analytical results can be derived, that of three-wave panels and TML. We then proceed to the case of general T . Finally, we discuss a procedure of bounded estimation.

Finite sample results for $T = 2$

For general values of T , we can only specify in which interval the solutions of the cubic equation lie, as described in Lemma 1. We show that for $T = 2$ and the Transformed log-likelihood function (i.e. $\rho = 0$), this result can be sharpened and a simple analytic expression for the ML estimator can be derived.

Observe that for $T = 2$ we have for $\hat{\sigma}^2(\phi)$ and $\hat{\theta}^2(\phi)$ as defined in equation (14):

$$\hat{\sigma}^2(\phi) = \frac{1}{2N} \sum_{i=1}^N (\Delta y_{i,2} - \phi \Delta y_{i,1})^2, \quad \hat{\theta}^2(\phi) = \frac{1}{2N} \sum_{i=1}^N (\Delta y_{i,2} - (\phi - 2)\Delta y_{i,1})^2. \quad (30)$$

We have then the following expression for the TML log-likelihood function:

Proposition 2. For $T = 2$ the log-likelihood function for the TML estimator is given by:

$$\ell^c(\phi) = -\frac{N}{2} \left[\log \left(\left[\hat{\sigma}^2(\phi) + \left(\frac{1}{N} \sum_{i=1}^N \Delta y_{i,1} \Delta y_{i,2} - \phi \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,1})^2 \right) \right]^2 + d \right) \right], \quad (31)$$

where d does not depend on ϕ but only on data.

The polynomial inside the $\log(\cdot)$ expression in Proposition 2 is symmetric around the point $\tilde{\phi} = \hat{\phi}_W + 1$. The FOC is:

$$\frac{1}{2\hat{\sigma}^2\hat{\theta}^2} \left(\sum_{i=1}^N (\Delta y_{i,2} - \phi \Delta y_{i,1})(\Delta y_{i,2} - (\phi - 2)\Delta y_{i,1}) \right) \left(\sum_{i=1}^N (\Delta y_{i,2} - (\phi - 1)\Delta y_{i,1})\Delta y_{i,1} \right) = 0.$$

The solutions are given by $\tilde{\phi} = \hat{\phi}_W + 1$ and for $D > 0$:

$$\hat{\phi}^{(l)} = \tilde{\phi} - \sqrt{D}, \quad \hat{\phi}^{(r)} = \tilde{\phi} + \sqrt{D},$$

where $D \equiv 1 + \hat{\phi}_W^2 - \frac{\sum_{i=1}^N (\Delta y_{i,2})^2}{\sum_{i=1}^N (\Delta y_{i,1})^2}$ is the discriminant of the quadratic part of the score. The first derivative of the concentrated likelihood consists of a linear and quadratic part. The latter implies either zero or two more solutions for ϕ on top of the intermediate case $\tilde{\phi} = \hat{\phi}_W + 1$. Furthermore, setting this quadratic part equal to zero can be recognized as the FOC of the bias corrected FE estimator as in Bun and Carree (2005), for $T = 2$. The quantities that determine the asymptotic behaviour of the solutions have the following limits:

$$\text{plim}_{N \rightarrow \infty} \hat{\phi}_W = \phi_0 - \frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})}, \quad \text{plim}_{N \rightarrow \infty} D = \left(1 - \frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})} \right)^2. \quad (32)$$

Combining these results, it follows that $\hat{\phi}^{(l)}$ is consistent, i.e.

$$\text{plim}_{N \rightarrow \infty} \hat{\phi}^{(l)} = \phi_0. \quad (33)$$

The solutions $\tilde{\phi}$ and $\hat{\phi}^{(r)}$ are inconsistent unless $\text{var}(\Delta y_{i,1}) = \sigma_0^2$, i.e. $\sigma_v^2 = 0$. The connections between the solutions to the FOC can be further summarized as follows:

Corollary 1. For $T = 2$ and TML the following holds:

$$\ell^c(\hat{\phi}^{(l)}) = \ell^c(\hat{\phi}^{(r)}), \quad (34)$$

$$\hat{\sigma}_v^2(\hat{\phi}^{(l)}) > 0 > \hat{\sigma}_v^2(\hat{\phi}^{(r)}), \quad (35)$$

$$\hat{\theta}^2(\tilde{\phi}) = \hat{\sigma}^2(\tilde{\phi}), \quad (36)$$

$$\hat{\phi}_B = \hat{\phi}_W + 2. \quad (37)$$

However, this result does not hold in general for RML and/or $T > 2$.

Corollary 1 states that, if the cubic equation has only one solution, it is a corner solution as the estimate for $\sigma_v^2(\tilde{\phi}) = 0$ in this case. The equality of the likelihood for $\hat{\phi}^{(l)}$ and $\hat{\phi}^{(r)}$ would imply that both can be considered ‘maximum likelihood’. However, the second is inconsistent and leads to a negative estimate of σ_v^2 . This result suggests that by considering always the ‘left’ solution (in the case of the two competing local maxima), the problem regarding the global non-identification can be avoided.

To illustrate the relevance of the occurrence of three solutions we derived the probability of a positive discriminant by assuming normality.

Corollary 2. For $T = 2$ and under joint normality of the data, the probability of $D > 0$ (two maxima) is given by:

$$\Pr(D > 0) = F \left(\frac{N}{N-1} \left(2 \frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})} - \left(\frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})} \right)^2 \right)^{-1} \right)_{(N-1, N)}, \quad (38)$$

where $F(\cdot)_{(N-1, N)}$ is the CDF of an F distributed random variable with $(N-1, N)$ degrees of freedom.

The term inside $F(\cdot)_{(N-1, N)}$ is always larger than 1 and consequently $\Pr(D > 0) \geq 0.5$ as $\text{var}(\Delta y_{i,1}) \geq \sigma_0^2$. If the initial observation is generated from a stationary process then $\frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})} = (1 + \phi_0)/2$. To allow for unrestricted initial condition we define the following relative variance ratio α_0 :

$$\alpha_0 \equiv \frac{1 - \phi_0^2}{\sigma_0^2} \text{var} \left(y_{i,0} - \frac{\eta_i}{(1 - \phi_0)} \right), \Rightarrow \text{var}(\Delta y_{i,1}) = \sigma_0^2 \left(\alpha_0 \frac{1 - \phi_0}{1 + \phi_0} + 1 \right), \quad (39)$$

Figure 2. Probability of $D > 0$ with $N = 50$ on the left and $N = 250$ on the right. $\phi_0 \in [0; 0.95]$ and $\alpha_0 \in [0.0; 1.05]$ such that $\alpha_0 = 1$ if the initial observation is covariance stationary. $\Pr(D > 0)$ then depends on N , ϕ_0 and α_0 . It can be easily seen that $\Pr(D > 0)$ is a decreasing function of ϕ_0 and a increasing function of α_0 . In Figure 2, we provide two graphs to illustrate how this probability depends on the population parameters.

Asymptotic results for $T > 2$ and TML

In this subsection, we extend the analysis of one or three solutions of the TML FOC in equation (17) to $T > 2$. Before proceeding we define the following quantities:

$$\tilde{a}_E \equiv E \left[\sum_{t=1}^T \hat{y}_{i,t-1}^2 \right], \quad \dot{a}_E \equiv E [T \dot{y}_{i-}^2], \quad \xi \equiv \sum_{t=0}^{T-2} (T-t-1) \phi_0^t, \quad x \equiv \phi_0 - \phi. \quad (40)$$

Using this notation we can express the asymptotic solutions of the FOC for general T as:

Proposition 3. In addition to $x = 0$, the two non-zero ($x \neq 0$) asymptotic solutions of equation (17) (if they exist) are implicitly defined by:

$$\begin{aligned} x^2 \frac{T}{T-1} (\dot{a}_E \tilde{a}_E) + x \frac{\xi}{T} \left(\left(\frac{2T-1}{T-1} \theta_0^2 \tilde{a}_E \right) - \left(\frac{T+1}{T-1} \sigma_0^2 \dot{a}_E \right) \right) \\ + (\theta_0^2 \tilde{a}_E + \sigma_0^2 \dot{a}_E) - \frac{2\xi^2}{T(T-1)} \theta_0^2 \sigma_0^2 = 0. \end{aligned} \quad (41)$$

The existence of non-zero solutions to this equation in the simple AR(1) model depends on two parameters: the autoregressive coefficient ϕ_0 and the relative variance parameter α_0 . The reason is that the solutions in Proposition 3 are invariant to σ_0^2 , because all quantities of interest are multiplicative in σ_0^2 :

$$\tilde{a}_E = \left(\sum_{t=0}^{T-1} \phi_0^{2t} - \frac{1}{T} \left(\sum_{t=0}^{T-1} \phi_0^t \right)^2 \right) \frac{\alpha_0 \sigma_0^2}{1 - \phi_0^2} + \sigma_0^2 \sum_{t=0}^{T-2} \left(\sum_{j=0}^t \phi_0^j \left(\phi_0^j - \frac{1}{T} \left(\sum_{j=0}^t \phi_0^j \right) \right) \right), \quad (42)$$

$$\dot{a}_E = \frac{\alpha_0 \sigma_0^2 \xi^2}{T} \frac{1 - \phi_0}{1 + \phi_0} + \frac{\sigma_0^2}{T} \sum_{t=0}^{T-2} \left(\sum_{j=0}^t \phi_0^j \right)^2, \quad (43)$$

Figure 3. The sign of the discriminant for $T=3$ on the left and $T=5$ in the right graph. 3 is for positive discriminant and thus three solutions to FOC, while 1 is for negative discriminant and one solution. $\phi_0 \in [0; 0.99]$ and $\alpha_0 \in [0.8; 1.05]$

$$\theta_0^2 = T\sigma_0^2 \left(\alpha_0 \frac{1-\phi_0}{1+\phi_0} + \frac{1}{T} \right). \quad (44)$$

Given that $x=0$ ($\phi=\phi_0$) is always a solution to FOC one can invoke e.g. theorem 4.1.2 in Amemiya (1985) to conclude that there exists a (local) maximizer of the log-likelihood function that is consistent and asymptotically normal for any fixed T (including $T=2$).¹⁰

To gain further insight into the quadratic equation in Proposition 3 for general $T>2$ we investigate the sign of the discriminant numerically for different values of T . In Figure 3 we present two plots of the sign of the discriminant. Here ‘3’ indicates that the discriminant is positive (thus bimodality), while ‘1’ implies that the discriminant is negative and the log-likelihood function is asymptotically unimodal. We present results for $T=3$ and $T=5$. For higher values of T the border between three and one solution to the FOC approaches the $\alpha_0=1$ line from below.¹¹ There is a major change from $T=2$ to $T>2$ in the set of values (ϕ_0, α_0) for which in the limit there is a positive discriminant value. For $T=2$ all values of (ϕ_0, α_0) for which $\alpha_0 > 0$ and $\phi_0 < 1$ have a positive discriminant when $N \rightarrow \infty$. This set is obviously smaller already for $T=3$.

Figure 3 shows that as T increases the interval of $\alpha_0 < 1$, that results in a positive discriminant, shrinks. It can be shown numerically for the relevant range of values for α_0 and ϕ_0 that for $\alpha_0 \geq 1$ the discriminant is always positive. The behaviour for $\alpha_0 \geq 1$ can be explained, using ‘large α_0 ’ asymptotics of equation (41), summarized in the next corollary.

Corollary 3. For any ϕ_0 and T , as $\alpha_0 \rightarrow \infty$ the quadratic equation in (41) always has two solutions. Furthermore, the corresponding local maxima of the log-likelihood function are such that:

¹⁰ A similar argument for a consistent local maximum is used by Anderson and Hsiao (1982).

¹¹ The fact that the properties of the estimator depend on the properties of the initial condition are not surprising, as it is well known that all dynamic panel data estimators are sensitive to this issue (see e.g. Sevestre and Trognon 1985).

$$\begin{aligned}\text{plim}_{N \rightarrow \infty} \hat{\phi}^{(l)} &= \phi_0 \\ \text{plim}_{N \rightarrow \infty} \hat{\phi}^{(r)} &= \phi_0 + \frac{T}{\xi} + \mathcal{O}(\alpha_0^{-1}) = \text{plim}_{N \rightarrow \infty} \hat{\phi}_B + \mathcal{O}(\alpha_0^{-1}).\end{aligned}$$

Here $\text{plim}_{N \rightarrow \infty} \hat{\phi}^{(l)}$ is the ‘left’ local maximum, while $\text{plim}_{N \rightarrow \infty} \hat{\phi}^{(r)}$ is the ‘right’ one.

This result suggests, that for a sufficiently large α_0 , the log-likelihood function is bi-modal. More importantly, the two local maxima converge to the boundaries of the interval in Lemma 1. Also note that because $\text{plim}_{N \rightarrow \infty} \hat{\phi}_W = \phi_0 + \mathcal{O}(\alpha_0^{-1})$, the difference between $\text{plim}_{N \rightarrow \infty} \hat{\phi}^{(l)}$ and $\text{plim}_{N \rightarrow \infty} \hat{\phi}_W$ is of order $\mathcal{O}(\alpha_0^{-1})$.

Our results indicate that multiple solutions are possible for $T > 2$ even if N becomes large. However, they are silent to which of the maximum is the global one. In particular, the results of Proposition 3 and Corollary 3 do not say that the local maximum that corresponds to $x = 0$ (i.e. $\phi = \phi_0$) is a unique global one. Despite our best efforts, we could not prove this claim analytically. However, extensive numerical simulations for a fine grid of α_0 and ϕ_0 show that the global maximum is always attained at the left maximum $\text{plim}_{N \rightarrow \infty} \hat{\phi}^{(l)}$, and only that solution always satisfies the non-negativity restriction for a positive variance estimate.

The existence of a unique global maximizer for $T > 2$ follows from the observation that, as T increases, the TML log-likelihood function is dominated by the within group part (which is unimodal for any T). Figure 4 illustrates how the shape of the concentrated log-likelihood function for $N \rightarrow \infty$ depends on T . Note that it is the ‘left’ solution that starts to dominate, as the lower bound of the interval in Lemma 1 approaches ϕ_0 from below. The explanation behind this behaviour lies in the fact that

$$\begin{aligned}\mathbb{E}[\hat{\theta}^2(\phi)] &= \mathcal{O}(T), \\ \mathbb{E}[\hat{\sigma}^2(\phi)] &= \mathcal{O}(1),\end{aligned}$$

for any finite ϕ . Hence the between group part of the log-likelihood function is:

$$-\frac{N}{2} \log(\hat{\theta}^2(\phi)) = \mathcal{O}_P(N \log(T)). \quad (45)$$

The within group part, on the other hand, is of a higher order of magnitude, as:

$$-\frac{N(T-1)}{2} \log(\hat{\sigma}^2(\phi)) = \mathcal{O}_P(NT). \quad (46)$$

Asymptotic dominance of the within group part follows directly as $\mathcal{O}_P((NT)^{-1}N \log(T)) = \mathcal{O}_P(T^{-1} \log(T)) = o_p(1)$. Despite the fact that the between group component is dominated (as $T \rightarrow \infty$), the rate at which it converges to zero is sufficiently slow. That ensures that the between part is able to provide an implicit bias-correction for the FOC of the FE estimator. Similar conclusions also apply to the RML estimator. For further results on the large T distribution of the TML estimator, please refer to Bai (2013), where it is shown that the TML/RML estimator is asymptotically equivalent to the bias-corrected FE estimator of Hahn and Kuersteiner (2002).

Figure 4. Within group log-likelihood (dotted line) and the transformed maximum likelihood log-likelihood function (solid line). The global maximum of the latter function is located at the true value $\phi_0 = 0.5$. (a) $T = 3$, (b) $T = 5$, (c) $T = 9$, (d) $T = 20$

Remark 3. All results of this section carry through even if $\sigma_{0,i}^2$ and $\text{var}(y_{i,0})$ are individual specific. In particular, for the heteroscedastic case we redefine quantities in Proposition 3 in the following way:

$$\sigma_0^2 \equiv \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_{0,i}^2, \quad \alpha_0 \equiv \frac{1 - \phi_0^2}{\sigma_0^2} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{var} \left(y_{i,0} - \frac{\eta_i}{(1 - \phi_0)} \right).$$

As a result, the two key parameters α_0 and ϕ_0 that determine the shape of the log-likelihood function are the same, but slightly redefined. In this way this paper also completes the analysis of Hayakawa and Pesaran (2015), who do not fully investigate the shape of the TML likelihood function in such a setting.

Constrained estimation

Aside from the suggested ‘take left’ procedure in case of bimodality to avoid a negative estimate of σ_v^2 , we may also use restricted ML estimation. Consider the following reparametrization $\delta = \frac{\sigma^2}{\theta^2}$, so that in the population $\delta \in (0; 1]$ because by definition $\theta^2 = \sigma^2 + T\sigma_v^2$. In order to take this population restriction into account, we consider the following reformulated log-likelihood function

$$\ell(\boldsymbol{\kappa}) = -\frac{N}{2} \left(T \log(\sigma^2) - \log(\delta) + \frac{1}{N\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1})^2 + \delta \frac{T}{N\sigma^2} \sum_{i=1}^N (\ddot{y}_i - \phi \ddot{y}_{i-} - \rho y_{i,0})^2 \right), \quad (47)$$

for $\boldsymbol{\kappa} = (\phi, \sigma^2, \delta, \rho)'$. The corresponding concentrated log-likelihood function in terms of the δ parameter is:

$$-\frac{2}{N} \ell^c(\delta) = T \log \left[\left(\tilde{c} - 2\phi(\delta)\tilde{b} + \phi^2(\delta)\tilde{a} \right) + \delta \left(\dot{c} - 2\phi(\delta)\dot{b} + \phi^2(\delta)\dot{a} \right) \right] - \log \delta, \quad (48)$$

where $\phi(\delta) = \frac{\tilde{b} + \delta \dot{b}}{\tilde{a} + \delta \dot{a}}$ and

$$\begin{aligned} \tilde{a} &= \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1}^2, & \tilde{b} &= \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t} \tilde{y}_{i,t-1}, & \tilde{c} &= \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t}^2, \\ \dot{a} &= \frac{T}{N} \sum_{i=1}^N \dot{y}_{i-}^2, & \dot{b} &= \frac{T}{N} \sum_{i=1}^N \dot{y}_i \dot{y}_{i-}, & \dot{c} &= \frac{T}{N} \sum_{i=1}^N \dot{y}_i^2. \end{aligned}$$

The expression for the concentrated log-likelihood can be further simplified as:

$$-\frac{2}{N} \ell^c(\delta) = T \log \left[(\tilde{c} + \delta \dot{c}) - \frac{(\tilde{b} + \delta \dot{b})^2}{\tilde{a} + \delta \dot{a}} \right] - \log \delta. \quad (49)$$

Similarly to equation (17), the FOC for the log-likelihood function in (49) is cubic in δ .

One can relate the likelihood function in (49) to equation (2.3) in Maddala (1971), who investigates the occurrence of the boundary solution $\delta = 1$ for this type of likelihood function.

Remark 4. Note that Balestra and Nerlove (1966) and Maddala (1971), assume that for all i one has $y_{i,0} = 0$. If this restriction is indeed true, the two log-likelihood functions are identical and $\hat{\phi}^{\text{RML}} = \hat{\phi}^{\text{TML}}$. On the other hand, if this restriction is not satisfied, the resulting estimator is inconsistent for any fixed value of T , with cubic FOC and has an asymptotic bias of order $\mathcal{O}_p(T^{-2})$. Hence, that estimator also efficiently removes the leading $\mathcal{O}_p(T^{-1})$ term of the bias of the FE estimator. This estimator is labelled a ‘misspecified random effects estimator’ in Hahn, Kuersteiner and Cho (2004), where the authors provide asymptotic results for this estimator as $N, T \rightarrow \infty$ (jointly).¹²

One can see that the necessary and sufficient condition for $\delta = 1$ to be a local (constrained) maximum is:

$$\left. \frac{d\ell^c(\delta)}{d\delta} \right|_{\delta=1} > 0. \quad (50)$$

¹² As mentioned in Arellano and Bonhomme (2009) the same conclusion does not generally hold for the model with strictly exogenous regressors.

In this case one sets $\delta = 1$ and the corresponding estimate of ϕ is given by:

$$\phi(1) = \frac{\tilde{b} + \dot{b}}{\tilde{a} + \dot{a}}. \quad (51)$$

We know that for $T = 2$ and in the case of the TML estimator this $\phi(1)$ is exactly the middle solution $\tilde{\phi} = \hat{\phi}_w + 1$. In all other cases, this solution will differ from the unique global unconstrained maximum (in the one solution case). For example, if $y_{i,0} = 0$ for all i , one can recognize $\phi(1)$ as the pooled OLS estimator of ϕ , which is known to be positively biased. In general, $\phi(\delta)$ is a weighted sum of ‘within’ and ‘quasi-between’ estimators and thus belongs to the interval of Lemma 1. Furthermore, the weight of the ‘within’ estimator is monotonically decreasing in δ , because:

$$\phi(\delta) = \hat{\phi}_w q(\delta) + \hat{\phi}_B (1 - q(\delta)), \quad q(\delta) = \frac{\tilde{a}}{\tilde{a} + \delta \dot{a}}, \quad q'(\delta) < 0. \quad (52)$$

Hence, if the global maximum $\hat{\phi}$ does not satisfy the non-negativity constraint it is always non-smaller than $\phi(1)$ (assuming $\hat{\phi}_w < \hat{\phi}_B$). In the Monte Carlo section of this paper, we will investigate the finite sample properties of the TML and RML estimators that use $\phi(1)$ as estimate at the boundary of the parameter space.

Remark 5. Although not addressed in this paper, the use of the boundary solution $\phi(1)$ might lead to a non-standard inference problem in finite samples. Confidence intervals based on inverting the Likelihood Ratio statistic have to be handled with care as for some values of the null hypothesis $H_0 : \phi = \phi_0$, the likelihood ratio statistic can be non-positive.

Connections with the literature

Early contributions to the literature on ML estimation of the dynamic panel data models include the papers of Balestra and Nerlove (1966), Maddala (1971) and Anderson and Hsiao (1982). Although not thoroughly analysed in this paper, a restricted version of the TML/RML estimators can be considered if one is willing to impose further restrictions on the distribution of the initial condition. For example, Kruiniger (2008) and Han and Phillips (2013) analyse the TML likelihood function in more detail under the assumption of covariance stationarity. Rather than treating θ^2 as a free parameter in estimation they impose:

$$\theta^2 = \sigma^2 \left(1 + T \frac{1 - \phi}{1 + \phi} \right). \quad (53)$$

They show that, after concentrating out the other model parameters, the FOC for the autoregressive coefficient ϕ is a quartic equation. Our analysis can therefore be seen as a generalization of these studies to a more general form of the initial conditions resulting in a cubic FOC. However, the fact that FOC is found to be quartic under stationarity, also implies that in finite samples one can get boundary estimates at $\phi = 1$, as no interior optima can exist.

The TML and RML estimators in this study are also related to the adjusted profile likelihood of Dhaene and Jochmans (2016)¹³ in the sense that the estimators provide a bias-correction to the standard FE OLS estimator, which is inconsistent for small T (Nickell, 1981). The estimating equation of this bias-corrected FE estimator is of the following form:

$$\frac{1}{N\hat{\sigma}^2(\phi)} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1}(\tilde{y}_{i,t} - \phi\tilde{y}_{i,t-1}) + \frac{1}{T} \xi(\phi, T) = 0, \quad (54)$$

where, as in section ‘Asymptotic results for $T > 2$ and TML’, $\xi(\phi, T) = \sum_{t=0}^{T-2} (T-t-1)\phi^t$. For $T=2$ equation (54) is quadratic in ϕ and coincides with the quadratic polynomial in section III, and therefore is equal to the bias-corrected FE of Bun and Carree (2005). For $T > 2$, $\xi(\phi, T)$ is a polynomial of order $T-2$, and consequently the order of the polynomial in the FOC is T . Using the terminology of Arellano and Bonhomme (2009), one can conclude that the estimator of Dhaene and Jochmans (2016) uses a non-data dependent prior to bias-correct the FE estimator. The RML estimator discussed in this paper, on the other hand, can be seen as the bias-corrected Random Effects estimator, that corrects for the non-zero correlation between η_i and y_{i0} .

IV. Monte Carlo simulations

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:

$$y_{i,t} = \phi y_{i,t-1} + (1 - \phi)\mu_i + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, 1), \quad t = 1, \dots, T. \quad (55)$$

$$y_{i,0} = \gamma\mu_i + \varepsilon_{i,0}, \quad \varepsilon_{i,0} \sim \mathcal{N}\left(0, \frac{\zeta}{1 - \phi^2}\right), \quad \mu_i \sim \mathcal{N}(0, \sigma_\mu^2). \quad (56)$$

Mean (effect) 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 both $\gamma = \zeta = 1$. 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. For the TML estimator the only important parameter is α as defined in equation (39) as it measures the deviation from covariance stationarity.

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:

$$N = \{50, 250\}, \quad T = \{3, 7\}, \quad \gamma = \{0.5, 1.0\}, \quad \sigma_\mu = \{1, 3\}, \quad \phi = \{0.5, 0.8\},$$

while $\zeta = 1$. We report mean, median, IQR (interquartile range) and RMSE for the following coefficient estimators (TML/RML):¹⁴

¹³ Dhaene and Jochmans (2016) extend the setup of Lancaster (2002) and show that the adjusted profile likelihood can also be viewed as a penalized log-likelihood proposed by Bester and Hansen (2009), or as a bias-corrected estimator of Bun and Carree (2005).

¹⁴ We do not provide results for the T(R)ML estimator where one imposes restrictions directly on the parameter space of ϕ . The common restriction of this type is $\phi \in (-1; 1)$ as considered e.g. in Hsiao *et al.* (2002, p. 135) and Hayakawa and Pesaran (2015). Furthermore, unlike Hsiao *et al.* (2002, p. 135) we do not neglect Monte Carlo replications with negative variance estimates, as it will bias the results.

- (i) T(R)ML based on the global maximum (T(R)MLg), where always the global maximum is selected.
- (ii) T(R)ML based on the ‘left’ maximum (T(R)MLl), that takes into account the non-negativity restriction only if there are two competing local maxima.
- (iii) T(R)ML with the imposed boundary condition $\phi(1)$ (T(R)MLb) from equation (51), as in section ‘Constrained estimation’.

Note that in calculating coefficient estimators we refrained from using numerical optimization techniques. As mentioned earlier, exploiting root finding algorithms one can find solutions to the cubic first-order condition.¹⁵

Regarding inference we consider empirical rejection frequencies based on two sided *t*- and LR statistics.¹⁶ We address both size and power. Due to the possible flatness of the profile log-likelihood functions induced by the bimodality, inference quality based on the two classical tests might differ substantially. The *t*- or Wald test critically depends on a quadratic approximation of the likelihood, which may cause problems when the likelihood is flat. The LR test is likely better behaved under the null hypothesis, but the flatness of the likelihood will influence its power. Below we summarize some general patterns that arise from the various tables with simulation results in Appendix A and the supplementary online appendix.

Estimation

Regarding coefficient estimation we find that both T(R)MLl and T(R)MLb perform substantially better than always choosing the global maximum of the likelihood function (T(R)MLg). This point is especially relevant for TML and results from not taking the non-negativity constraint into consideration. The ‘right’ instead of ‘left’ solution to the FOC may sometimes provide the global maximum of the likelihood, but choosing it causes serious bias. Therefore, in terms of bias and of RMSE both TMLg and RMLg are dominated by ‘left’ estimators or by exploiting the boundary condition.

When we do not consider T(R)MLg, we find little difference between RML and TML. There is some tendency of RML to dominate TML, confirming results in Kruiniger (2013). However, we do not observe any substantial problems for TML when σ_μ increases, unlike the aforementioned study. Furthermore, in terms of RMSE exploiting the boundary solution is almost always better than the ‘left’ estimator. However, in some cases (for small *N* and/or large ϕ) this choice can have a negative effect on mean and median bias. This observation is in line with the theoretical discussion at the end of section ‘Constrained estimation’. As *N* and *T* increase the discrepancy becomes negligible. Also the distributions of all estimators tend to be asymmetric as illustrated by the discrepancy between mean and the median.

Finally, we find that in all cases where the cubic FOC had three solutions, the ‘left’ solution always satisfies the non-negativity constraint, while the ‘right’ solution never satisfies it. Furthermore, in those cases the global constrained maximum is also achieved at the ‘left’ solution, and not at the boundary. Hence for replications with three solutions

¹⁵ As implemented by e.g. the *roots(.)* function in Matlab.

¹⁶ For the *t*-test we exploit the usual ‘sandwich’ covariance matrix estimator, e.g. as in Hayakawa and Pesaran (2015) or Juodis (2016).

Figure 5. Finite sample distribution of the transformed maximum likelihood (TML)/random effects maximum likelihood (RML) estimators for $N = 250$, $T = 3$, $\phi = 0.5$ with covariance stationary initialization of y_{i0}

only the ‘left’ solution is natural. This is an important observation, because it suggests that there is always at most one interior maximum in the constrained optimization problem.

Figure 5 further illustrates how different ways of dealing with the boundary condition shapes finite sample distributions of coefficient estimators. The fact that most of the studies that consider RML and/or TML estimation (e.g. Hsiao *et al.*, 2002; Alvarez and Arellano, 2003; Ahn and Thomas, 2006; Kruiniger, 2008; Hayakawa and Pesaran, 2015) either do not address the negative variance issue at all or only mention it without further exploring its consequences is somewhat puzzling. As we can see for many designs substantial gains in terms of RMSE can be achieved when using RMLl/TMLl rather than RMLg/TMLg. Finally, in most cases IQR for TML is larger than the corresponding value for RML, thus confirming the asymptotic results presented in Proposition 1.

Inference

Regarding inference with the t - and LR statistics, we observe that the LR test provides reasonable size control, but the power properties are poor for small N and T . Given the asymmetry of the likelihood function, especially the power for alternatives larger than the null hypothesis is negligible and comparable to size. The power of the LR test improves significantly with a large sample size, and especially for larger T . TMLb/RMLb tends to be undersized in comparison to TMLl/RMLl.

Furthermore, inference based on the t -test in samples with small N and T is unreliable as the actual rejection frequencies are substantially higher than the nominal ones. The results for the t -statistic deteriorate for $\phi = 0.8$, which is related to the non-standard behaviour of the TML/RML estimator when ϕ is local-to-unity, see Kruiniger (2013) for related results. The LR test is much less affected by the value of ϕ .

Finally, TML and RML statistics are similar in terms of empirical size with RML statistics having higher power. In terms of empirical size we can rank test statistics in the following order: $g > l > b$. These differences slowly disappear as N and/or T get larger.

V. Empirical illustration

In this section, we study the behaviour of the TML and RML estimators exploiting data from Bun and Carree (2005), who considered the following model for unemployment at the US state level:

$$u_{i,t} = \phi u_{i,t-1} + \beta g_{i,t-1} + \eta_i + \tau_t + \varepsilon_{i,t}. \quad (57)$$

Here $u_{i,t}$ is the unemployment rate in state i at time t and $g_{i,t-1}$ is the real economic growth rate at time $t - 1$.¹⁷ The annual panel data cover the years 1991–2000 for all US states (including Washington DC, hence $N = 51$). We present estimation results for the model including the growth regressor (Table 1) and the pure AR(1) specification (Table 2). In order to investigate how the behaviour of the log-likelihood function for both estimators changes as T increases we consider estimates over an increasing window. Thus results for $T = 2$ are obtained based on years 1998–2000, $T = 3$ exploits the period 1997–2000, etc. All models are estimated based on data in deviations from cross-sectional means to filter out the time effects. To illustrate the finite sample properties of T(R)MLg, T(R)MLl and T(R)MLb, we present coefficient estimates for varying T . To save space, we only report T(R)MLl estimates when T(R)MLl = T(R)MLb.

TABLE 1
Transformed maximum likelihood (TML) and random effects maximum likelihood (RML) estimates for the ARX(1) model

T	TMLg		TMLl		RMLg		RMLl	
	$\hat{\phi}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\beta}$
2	1.422	0.020	0.506	0.003	0.493	0.003	0.493	0.003
3	0.429	−0.006	0.429	−0.006	0.532	−0.008	0.532	−0.008
4	0.492	−0.026	0.492	−0.026	0.562	−0.025	0.562	−0.025
5	0.451	−0.031	0.451	−0.031	0.489	−0.030	0.489	−0.030
6	0.511	−0.036	0.511	−0.036	0.531	−0.035	0.531	−0.035
7	0.511	−0.038	0.511	−0.038	0.517	−0.038	0.517	−0.038
8	0.577	−0.041	0.577	−0.041	0.587	−0.040	0.587	−0.040
9	0.617	−0.057	0.617	−0.057	0.641	−0.055	0.641	−0.055

¹⁷ Empirical evidence on strict exogeneity of $g_{i,t-1}$ is provided in Bun and Carree (2005).

ARX(1) model

Using all time periods ($T = 9$) the estimation results based on TML and RML are very similar to the estimates in Bun and Carree (2005) obtained using the bias-corrected FE estimator.¹⁸ The similarity between the bias corrected FE results as found by Bun and Carree (2005) and TML/RML estimators is not surprising, given that all three estimators correct for the bias in the FE estimator using some bias adjustment procedure.

The results in Table 1 show that for RML estimation $\text{RMLg} = \text{RMLl} = \text{RMLb}$, irrespective of T . Hence, the global maximum is always achieved at the ‘left’ solution that satisfies the non-negativity restriction, which amounts to $\theta^2 \geq \sigma^2$. The same holds for TML estimation, with the clear exception of the $T = 2$ case. There the global maximum is attained at $\hat{\phi}^{(r)} = 1.422$, which is substantially larger than $\hat{\phi}^{(l)} = 0.506$.

In the supplementary online appendix, we present estimation results, where $g_{i,t-1}$ is assumed to be weakly exogenous. The results are very close to the estimates imposing strict exogeneity.

AR(1) model

The empirical results for the pure AR(1) model without $g_{i,t-1}$ are reported in Table 2. As can be seen, the estimation results obtained from TML are quite stable irrespective of the time horizon under consideration. Furthermore, in all cases we find that the global maximum of the log-likelihood function is attained at the left maximum.¹⁹

The results for the RML estimator, however, are considerably less stable.²⁰ For $T = \{2, 5, 7, 8\}$ all three RML estimators are identical and are very close to the TML estimator. In two other cases, i.e. $T = \{6, 9\}$, the global maximum is obtained at the ‘right’ solution and not at the ‘left’ one. The result is a large difference between the RMLg and RMLl/RMLb. Because the ‘left’ solution satisfies the non-negativity constraint, the RMLl and RMLb estimators are identical. Finally, for $T = \{3, 4\}$ there exists one solution only, which does not satisfy the non-negativity constraint. RMLl and RMLb produce therefore markedly different estimates, with the latter actually being quite close to the corresponding TMLl estimates. In Figure S1 of the supplementary online appendix, we provide detailed plots of the concentrated log-likelihood function of this model for different values of T . Based on these plots, we can see how small increments in the length of the time series change the shape of the concentrated log-likelihood function. Furthermore, the log-likelihood functions for both estimators are relatively flat where likelihoods are both unimodal and bimodal. Finally, we see in all cases that the second mode of RML is smaller in absolute value as compared to that of TML.

¹⁸ The bias corrected FE estimates are $\hat{\phi} = 0.615$ and $\hat{\beta} = -0.057$.

¹⁹ When $T = 2$ we report the ‘left’ solution for TMLg as both its solutions are of the same log-likelihood value.

²⁰ The fact that the properties of the estimator change with the inclusion of the additional exogenous regressors is in-line with some earlier results in the literature, e.g. Trognon (1978) mention that additional regressors help to reduce to lower the frequency at which the boundary solution can occur.

TABLE 2
Transformed maximum likelihood (TML) and random effects maximum likelihood (RML) estimates for the AR(1) model

T	$TMLg$	$TMLl$	$RMLg$	$RMLl$	$RMLb$
2	0.502	0.502	0.516	0.516	0.516
3	0.514	0.514	0.750	0.750	0.613
4	0.522	0.522	0.950	0.950	0.667
5	0.461	0.461	0.514	0.514	0.514
6	0.553	0.553	1.062	0.596	0.596
7	0.545	0.545	0.565	0.565	0.565
8	0.613	0.613	0.632	0.632	0.632
9	0.671	0.671	1.054	0.695	0.695

Practical considerations

Based on the theoretical analysis, simulation results and the empirical estimates a number of remarks can be made regarding the implementation of TML and RML estimators in practice. First, as the FOC in equation (17) is cubic in the autoregressive parameter, its solutions can be found without any need of explicit numerical maximization. One can simply use root finder algorithms based on the eigenvalues of the companion matrix. Iterative procedures such as the Newton-Raphson method, which is typically used (e.g. Hsiao *et al.*, 2002, p. 111), can therefore be avoided. Second, if one nevertheless relies on numerical optimization techniques, it is important to choose different starting values to minimize the possibility that some solutions to the FOC are not found. Our Lemma 1 suggests lower and upper bounds for the grid to be used. Third, in case of three solutions one should simply take the ‘left’ solution. Although we were not able to provide a formal proof of this selection rule, see section ‘Asymptotic results for $T > 2$ and TML’, we find in our simulations and empirical illustration that this solution always satisfies the non-negativity constraint for the variance, and it is therefore numerically identical to the constrained ML estimator. Fourth, in case of one solution only one should check the non-negativity constraint. If it is violated, the constrained estimator might be preferred.

VI. Conclusions

We have investigated some finite sample and asymptotic properties of the TML and RML estimators for dynamic panel data models. Both estimators are consistent for fixed T and N large, but in finite samples their actual numerical implementation matters for inference. We showed that in a simple AR(1) model with homoscedastic errors the TML and RML estimators can be obtained as solutions of cubic first-order conditions. We furthermore argued that in some cases the value that maximizes the log-likelihood function is not the best possible solution as it can violate the non-negativity constraint for the variance. Finally, we showed that these results extend to models with additional exogenous regressors.

In a Monte Carlo study, we found that the issue of non-negativity constraints cannot be ignored as it is commonly done in the literature. However, we also found, that for some parameter values the use of a constrained likelihood can have negative effects on

finite sample bias of the corresponding ML estimator. Additionally, the inference based on likelihood-based estimators can be highly misleading, as for small values of N and T we found that t -statistics tend to be substantially oversized. Although inference based on the LR test provides reasonable size control for small N and T , it can result in low power due to possible flatness of the likelihood function.

Finally, we investigated the issues of local maxima and boundary solutions in an empirical analysis of US state level unemployment rates. We have found that in some cases the different treatment of these issues leads to markedly different estimates of the autoregressive parameter.

Appendix A. Monte Carlo results

Tables A1–A6

TABLE A1
Estimation results for $N=50$, $T=3$

	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE
	$\phi=0.5$	$\gamma=0.5$	$\sigma_\mu=1$	$\sigma_\mu=3$	$\phi=0.5$	$\gamma=0.5$	$\sigma_\mu=3$	$\sigma_\mu=1$	$\phi=0.5$	$\gamma=1.0$	$\sigma_\mu=1$	$\sigma_\mu=3$	$\phi=0.5$	$\gamma=1.0$	$\sigma_\mu=1$	$\sigma_\mu=3$
TMlG	0.72	0.68	0.62	0.40	0.82	0.66	0.83	0.53	0.69	0.66	0.57	0.38	0.69	0.66	0.57	0.38
RMlG	0.57	0.52	0.33	0.26	0.75	0.58	0.77	0.47	0.55	0.51	0.32	0.25	0.64	0.58	0.51	0.34
TMlI	0.54	0.50	0.32	0.23	0.52	0.49	0.20	0.18	0.54	0.50	0.34	0.24	0.54	0.50	0.34	0.24
RMlI	0.53	0.50	0.28	0.22	0.53	0.50	0.20	0.19	0.53	0.49	0.29	0.22	0.54	0.49	0.33	0.25
TMlB	0.51	0.50	0.30	0.19	0.52	0.49	0.20	0.16	0.50	0.50	0.30	0.19	0.50	0.50	0.30	0.19
RMlB	0.49	0.49	0.24	0.16	0.52	0.50	0.20	0.16	0.47	0.48	0.22	0.15	0.50	0.49	0.28	0.18
	$\phi=0.8$	$\gamma=0.5$	$\sigma_\mu=1$	$\sigma_\mu=3$	$\phi=0.8$	$\gamma=0.5$	$\sigma_\mu=3$	$\sigma_\mu=1$	$\phi=0.8$	$\gamma=1.0$	$\sigma_\mu=1$	$\sigma_\mu=3$	$\phi=0.8$	$\gamma=1.0$	$\sigma_\mu=1$	$\sigma_\mu=3$
TMlG	0.90	0.93	0.36	0.28	0.95	0.97	0.41	0.31	0.90	0.92	0.36	0.28	0.90	0.92	0.36	0.28
RMlG	0.83	0.82	0.36	0.24	0.95	0.96	0.44	0.31	0.82	0.82	0.36	0.24	0.86	0.87	0.40	0.27
TMlI	0.76	0.77	0.34	0.21	0.78	0.78	0.34	0.20	0.75	0.77	0.34	0.21	0.75	0.77	0.34	0.21
RMlI	0.78	0.77	0.33	0.22	0.79	0.78	0.35	0.22	0.77	0.77	0.33	0.22	0.77	0.77	0.34	0.23
TMlB	0.73	0.76	0.29	0.19	0.77	0.78	0.31	0.19	0.73	0.75	0.29	0.19	0.73	0.75	0.29	0.19
RMlB	0.71	0.73	0.22	0.17	0.76	0.78	0.30	0.18	0.70	0.73	0.22	0.18	0.72	0.74	0.26	0.19

TABLE A2
Estimation results for $N = 250$, $T = 3$

	Mean			Median			IQR			RMSE			Mean			Median			IQR			RMSE			Mean			Median			IQR			RMSE		
	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$			
	$\phi = 0.5$			$\gamma = 0.5$			$\sigma_\mu = 1$			$\phi = 0.5$			$\gamma = 0.5$			$\sigma_\mu = 1$			$\phi = 0.5$			$\gamma = 0.5$			$\sigma_\mu = 1$			$\phi = 0.5$			$\gamma = 0.5$			$\sigma_\mu = 1$		
TMlG	0.63	0.53	0.28	0.29	0.71	0.52	0.75	0.41	0.61	0.52	0.25	0.26	0.61	0.52	0.25	0.26	0.61	0.52	0.25	0.26	0.61	0.52	0.25	0.26	0.61	0.52	0.25	0.26	0.61	0.52	0.25	0.26				
RmLG	0.51	0.49	0.12	0.11	0.58	0.50	0.10	0.26	0.51	0.49	0.13	0.11	0.56	0.50	0.13	0.11	0.56	0.50	0.13	0.11	0.56	0.50	0.13	0.11	0.56	0.50	0.13	0.11	0.56	0.50	0.13	0.11				
TMlI	0.51	0.50	0.13	0.12	0.50	0.49	0.08	0.07	0.52	0.49	0.14	0.14	0.52	0.49	0.14	0.14	0.52	0.49	0.14	0.14	0.52	0.49	0.14	0.14	0.52	0.49	0.14	0.14	0.52	0.49	0.14	0.14				
RmLI	0.51	0.49	0.12	0.10	0.50	0.49	0.08	0.07	0.51	0.49	0.12	0.10	0.52	0.49	0.12	0.10	0.52	0.49	0.12	0.10	0.52	0.49	0.12	0.10	0.52	0.49	0.12	0.10	0.52	0.49	0.12	0.10				
TMlB	0.51	0.50	0.13	0.10	0.50	0.49	0.08	0.06	0.51	0.49	0.14	0.11	0.51	0.49	0.14	0.11	0.51	0.49	0.14	0.11	0.51	0.49	0.14	0.11	0.51	0.49	0.14	0.11	0.51	0.49	0.14	0.11				
RmLB	0.50	0.49	0.12	0.09	0.50	0.49	0.08	0.06	0.50	0.49	0.12	0.08	0.50	0.49	0.12	0.08	0.50	0.49	0.12	0.08	0.50	0.49	0.12	0.08	0.50	0.49	0.12	0.08	0.50	0.49	0.12	0.08				
	$\phi = 0.8$			$\gamma = 0.5$			$\sigma_\mu = 1$			$\phi = 0.8$			$\gamma = 0.5$			$\sigma_\mu = 3$			$\phi = 0.8$			$\gamma = 1.0$			$\sigma_\mu = 1$			$\phi = 0.8$			$\gamma = 1.0$			$\sigma_\mu = 3$		
TMlG	0.89	0.90	0.27	0.20	0.95	0.96	0.34	0.25	0.88	0.90	0.27	0.20	0.88	0.90	0.27	0.20	0.88	0.90	0.27	0.20	0.88	0.90	0.27	0.20	0.88	0.90	0.27	0.20	0.88	0.90	0.27	0.20				
RmLG	0.82	0.80	0.20	0.14	0.94	0.94	0.35	0.24	0.81	0.80	0.20	0.14	0.85	0.80	0.20	0.14	0.85	0.83	0.20	0.14	0.85	0.83	0.20	0.14	0.85	0.83	0.20	0.14	0.85	0.83	0.20	0.14				
TMlI	0.80	0.79	0.23	0.14	0.81	0.79	0.23	0.13	0.79	0.79	0.23	0.14	0.79	0.79	0.23	0.14	0.79	0.79	0.23	0.14	0.79	0.79	0.23	0.14	0.79	0.79	0.23	0.14	0.79	0.79	0.23	0.14				
RmLI	0.80	0.79	0.19	0.13	0.82	0.79	0.23	0.14	0.80	0.79	0.20	0.13	0.80	0.79	0.20	0.13	0.80	0.79	0.20	0.13	0.80	0.79	0.20	0.13	0.80	0.79	0.20	0.13	0.80	0.79	0.20	0.13				
TMlB	0.78	0.79	0.20	0.12	0.80	0.79	0.21	0.12	0.78	0.79	0.20	0.12	0.78	0.79	0.20	0.12	0.78	0.79	0.20	0.12	0.78	0.79	0.20	0.12	0.78	0.79	0.20	0.12	0.78	0.79	0.20	0.12				
RmLB	0.76	0.78	0.12	0.09	0.80	0.79	0.21	0.12	0.76	0.78	0.12	0.09	0.77	0.79	0.12	0.09	0.77	0.79	0.12	0.09	0.77	0.79	0.12	0.09	0.77	0.79	0.12	0.09	0.77	0.79	0.12	0.09				

TABLE A3
t-test results for $N = 50$, $T = 3$

$\phi - \phi_0$	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2
	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 3$	$\phi = 0.5$	$\gamma = 1.0$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 1.0$	$\sigma_\mu = 3$	$\phi = 0.5$	$\gamma = 1.0$	$\sigma_\mu = 3$
TMLg	0.47	0.44	0.44	0.46	0.44	0.46	0.46	0.43	0.42	0.42	0.43	0.46	0.43	0.42	0.42
RMLg	0.32	0.25	0.23	0.28	0.36	0.45	0.36	0.38	0.46	0.58	0.31	0.24	0.22	0.26	0.34
TMLl	0.23	0.20	0.22	0.26	0.32	0.21	0.08	0.10	0.22	0.38	0.25	0.22	0.23	0.26	0.32
RMLl	0.26	0.19	0.18	0.23	0.33	0.23	0.10	0.11	0.22	0.39	0.27	0.20	0.19	0.23	0.32
TMLb	0.17	0.13	0.14	0.20	0.28	0.20	0.07	0.09	0.20	0.37	0.17	0.14	0.14	0.19	0.27
RMLb	0.14	0.07	0.07	0.16	0.28	0.21	0.07	0.08	0.19	0.36	0.13	0.06	0.07	0.16	0.27
	$\phi = 0.8$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.8$	$\gamma = 0.5$	$\sigma_\mu = 3$	$\phi = 0.8$	$\gamma = 1.0$	$\sigma_\mu = 1$	$\phi = 0.8$	$\gamma = 1.0$	$\sigma_\mu = 3$	$\phi = 0.8$	$\gamma = 1.0$	$\sigma_\mu = 3$
TMLg	0.56	0.47	0.36	0.28	0.27	0.58	0.52	0.44	0.33	0.30	0.56	0.47	0.35	0.27	0.27
RMLg	0.40	0.31	0.26	0.27	0.33	0.55	0.48	0.40	0.33	0.32	0.39	0.31	0.26	0.27	0.33
TMLl	0.34	0.28	0.22	0.21	0.29	0.33	0.29	0.25	0.22	0.28	0.34	0.28	0.21	0.21	0.29
RMLl	0.32	0.25	0.21	0.25	0.33	0.32	0.27	0.22	0.22	0.29	0.32	0.24	0.21	0.24	0.33
TMLb	0.29	0.24	0.20	0.21	0.32	0.30	0.27	0.24	0.21	0.29	0.29	0.23	0.20	0.21	0.33
RMLb	0.10	0.07	0.10	0.19	0.33	0.23	0.18	0.15	0.19	0.28	0.10	0.07	0.11	0.20	0.34

TABLE A4
t-test results for $N = 250, T = 3$

$\phi - \phi_0$	$\phi = 0.5$		$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 3$	$\phi = 0.5$	$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 1.0$	$\sigma_\mu = 1$	$\phi = 0.5$	$\gamma = 1.0$	$\sigma_\mu = 3$				
	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2
TMLg	0.60	0.28	0.28	0.43	0.64	0.91	0.44	0.31	0.56	0.86	0.53	0.27	0.27	0.40	0.59	0.53	0.27	0.27	0.40	0.59
RMLg	0.57	0.15	0.08	0.26	0.59	0.91	0.36	0.15	0.44	0.83	0.53	0.15	0.08	0.25	0.56	0.50	0.20	0.18	0.33	0.56
TMLl	0.49	0.09	0.09	0.28	0.54	0.90	0.29	0.05	0.36	0.79	0.40	0.11	0.11	0.27	0.51	0.40	0.11	0.11	0.27	0.51
RMLl	0.56	0.14	0.08	0.26	0.58	0.90	0.29	0.05	0.36	0.79	0.52	0.15	0.08	0.25	0.56	0.43	0.12	0.11	0.27	0.52
TMLb	0.47	0.07	0.07	0.25	0.52	0.90	0.29	0.05	0.36	0.79	0.38	0.08	0.08	0.23	0.49	0.38	0.08	0.08	0.23	0.49
RMLb	0.55	0.12	0.06	0.25	0.58	0.90	0.29	0.05	0.36	0.79	0.50	0.12	0.05	0.24	0.56	0.40	0.08	0.06	0.23	0.50
<hr/>																				
$\phi - \phi_0$	$\phi = 0.8$		$\gamma = 0.5$	$\sigma_\mu = 1$	$\phi = 0.8$	$\gamma = 0.8$	$\sigma_\mu = 3$	$\phi = 0.8$	$\gamma = 1.0$	$\sigma_\mu = 1$	$\phi = 0.8$	$\gamma = 1.0$	$\sigma_\mu = 3$							
	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2					
TMLg	0.59	0.53	0.43	0.29	0.35	0.59	0.55	0.51	0.43	0.36	0.58	0.51	0.41	0.28	0.36					
RMLg	0.45	0.28	0.20	0.26	0.46	0.57	0.52	0.47	0.40	0.38	0.44	0.28	0.19	0.26	0.46					
TMLl	0.38	0.33	0.27	0.23	0.40	0.29	0.26	0.27	0.29	0.37	0.38	0.33	0.26	0.23	0.41					
RMLl	0.42	0.25	0.18	0.25	0.46	0.31	0.27	0.25	0.26	0.38	0.42	0.26	0.17	0.25	0.46					
TMLb	0.33	0.28	0.23	0.22	0.45	0.27	0.25	0.26	0.29	0.39	0.33	0.27	0.21	0.22	0.46					
RMLb	0.25	0.06	0.07	0.22	0.50	0.26	0.21	0.20	0.24	0.38	0.24	0.05	0.07	0.23	0.51					

TABLE A5
LR test results for $N = 50$, $T = 3$

$\phi - \phi_0$	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2
	$\phi = 0.5$		$\gamma = 0.5$		$\sigma_v = 1$		$\phi = 0.5$		$\gamma = 0.5$		$\phi = 0.5$		$\gamma = 1.0$		$\sigma_v = 1$		$\phi = 0.5$		$\gamma = 1.0$	
TMLg	0.25	0.09	0.05	0.09	0.13	0.42	0.17	0.08	0.13	0.24	0.22	0.08	0.05	0.08	0.12	0.22	0.08	0.05	0.08	0.12
RMLg	0.25	0.10	0.07	0.09	0.14	0.42	0.18	0.10	0.14	0.25	0.23	0.10	0.06	0.08	0.13	0.24	0.10	0.06	0.09	0.12
TMLl	0.22	0.07	0.03	0.06	0.09	0.36	0.12	0.04	0.08	0.19	0.20	0.06	0.03	0.05	0.09	0.20	0.06	0.03	0.05	0.09
RMLl	0.24	0.10	0.06	0.08	0.12	0.36	0.13	0.05	0.09	0.19	0.22	0.09	0.06	0.07	0.13	0.21	0.08	0.04	0.06	0.10
TMLb	0.20	0.06	0.03	0.06	0.09	0.36	0.12	0.04	0.08	0.19	0.18	0.05	0.03	0.05	0.09	0.18	0.05	0.03	0.05	0.09
RMLb	0.21	0.06	0.03	0.06	0.12	0.36	0.12	0.04	0.09	0.19	0.18	0.05	0.03	0.06	0.12	0.18	0.05	0.03	0.05	0.09
	$\phi = 0.8$		$\gamma = 0.5$		$\sigma_v = 1$		$\phi = 0.8$		$\gamma = 0.5$		$\phi = 0.8$		$\gamma = 1.0$		$\sigma_v = 1$		$\phi = 0.8$		$\gamma = 1.0$	
TMLg	0.05	0.02	0.03	0.04	0.04	0.09	0.03	0.03	0.05	0.05	0.04	0.02	0.04	0.04	0.04	0.04	0.02	0.04	0.04	0.04
RMLg	0.12	0.06	0.06	0.06	0.08	0.11	0.04	0.04	0.06	0.06	0.11	0.06	0.06	0.06	0.08	0.08	0.04	0.05	0.06	0.05
TMLl	0.04	0.01	0.02	0.03	0.03	0.09	0.02	0.02	0.03	0.04	0.04	0.01	0.02	0.03	0.02	0.04	0.01	0.02	0.03	0.02
RMLl	0.11	0.05	0.04	0.05	0.07	0.10	0.02	0.02	0.04	0.04	0.10	0.05	0.04	0.05	0.07	0.06	0.02	0.03	0.04	0.04
TMLb	0.04	0.01	0.02	0.03	0.03	0.08	0.02	0.02	0.03	0.04	0.04	0.01	0.02	0.03	0.02	0.04	0.01	0.02	0.03	0.02
RMLb	0.05	0.01	0.02	0.04	0.07	0.09	0.02	0.02	0.04	0.04	0.05	0.01	0.02	0.04	0.07	0.04	0.01	0.02	0.03	0.04

References

- Ahn, S. C. and Thomas, G. M. (2006). *Likelihood Based Inference for Dynamic Panel Data Models*, Working paper, Arizona State University, Tempe, AZ.
- Alvarez, J. and Arellano, M. (2003). 'The time series and cross-section asymptotics of dynamic panel data estimators', *Econometrica*, Vol. 71, pp. 1121–1159.
- Alvarez, J. and Arellano, M. (2004). *Robust Likelihood Estimation of Dynamic Panel Data Models*, Working paper, CEMFI, Madrid.
- Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Cambridge, MA.
- Anderson, T. W. and Hsiao, C. (1982). 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics*, Vol. 18, pp. 47–82.
- Arellano, M. and Bond, S. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies*, Vol. 58, pp. 277–297.
- Arellano, M. and Bonhomme, S. (2009). 'Robust priors in nonlinear panel data models', *Econometrica*, Vol. 77, pp. 489–536.
- Bai, J. (2013). 'Fixed-effects dynamic panel models, a factor analytical method', *Econometrica*, Vol. 81, pp. 285–314.
- Balestra, P. and Nerlove, M. (1966). 'Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas', *Econometrica*, Vol. 34, pp. 585–612.
- Bester, C. and Hansen, C. B. (2009). 'A penalty function approach to bias reduction in nonlinear panel models with fixed effects', *Journal of Business & Economic Statistics*, Vol. 27, pp. 131–148.
- Blundell, R. W. and Bond, S. (1998). 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics*, Vol. 87, pp. 115–143.
- Bun, M. J. G. and Carree, M. A. (2005). 'Bias-corrected estimation in dynamic panel data models', *Journal of Business & Economic Statistics*, Vol. 23, pp. 200–210.
- Bun, M. J. G. and Sarafidis, V. (2015). 'Dynamic panel data models', in Baltagi B. H. (ed.), *The Oxford Handbook of Panel Data*, Chap. 3, Oxford: Oxford University Press, pp. 76–110.
- Chamberlain, G. (1982). 'Multivariate regression models for panel data', *Journal of Econometrics*, Vol. 18, pp. 5–46.
- Dhaene, G. and Jochmans, K. (2016). 'Likelihood inference in an autoregression with fixed effects', *Econometric Theory*, Vol. 32, pp. 1178–1215.
- Hahn, J. and Kuersteiner, G. (2002). 'Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large', *Econometrica*, Vol. 70, pp. 1639–1657.
- Hahn, J., Kuersteiner, G. and Cho, M. H. (2004). 'Asymptotic distribution of misspecified random effects estimator for a dynamic panel model with fixed effects when both N and T are large', *Economics Letters*, 84, 117–125.
- Han, C. and Phillips, P. C. B. (2013). 'First difference maximum likelihood and dynamic panel estimation', *Journal of Econometrics*, Vol. 175, pp. 35–45.
- Hayakawa, K. and Pesaran, M. H. (2015). 'Robust standard errors in transformed likelihood estimation of dynamic panel data models', *Journal of Econometrics*, Vol. 188, pp. 111–134.
- Hsiao, C., Pesaran, M. H. and Tahmiscioglu, A. K. (2002). 'Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods', *Journal of Econometrics*, Vol. 109, pp. 107–150.
- Jones, S. and Tarp, F. (2016). 'Does foreign aid harm political institutions?', *Journal of Development Economics*, Vol. 118, pp. 266–281.
- Juodis, A. (2017). 'First difference transformation in panel VAR models: robustness, estimation and inference', *Econometric Reviews*, (forthcoming).
- Kruiniger, H. (2008). 'Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model', *Journal of Econometrics*, Vol. 144, pp. 447–464.
- Kruiniger, H. (2013). 'Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions', *Journal of Econometrics*, Vol. 173, pp. 175–188.
- Lancaster, T. (2002). 'Orthogonal parameters and panel data', *Review of Economic Studies*, Vol. 69, pp. 647–666.
- Maddala, G. S. (1971). 'The use of variance components models in pooling cross section and time series data', *Econometrica*, Vol. 39, pp. 341–358.

- Moral-Benito, E. (2012). 'Determinants of economic growth: a bayesian panel data approach', *The Review of Economics and Statistics*, Vol. 94, pp. 566–579.
- Mundlak, Y. (1978). 'On the pooling of time series and cross section data', *Econometrica*, Vol. 46, pp. 69–85.
- Nickell, S. (1981). 'Biases in dynamic models with fixed effects', *Econometrica*, Vol. 49, pp. 1417–1426.
- Sevestre, P. and Trognon, A. (1985). 'A note on autoregressive error components models', *Journal of Econometrics*, Vol. 28, pp. 231–245.
- Trognon, A. (1978). 'Miscellaneous asymptotic properties of ordinary least squares and maximum likelihood estimators in dynamic error components models', *Annales de l'ins*, Vol. 30/31, pp. 631–657.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Further connections with the literature.

Appendix S2. Additional material for weakly exogenous regressors.

Appendix S3. Proofs.

Appendix S4. Monte Carlo results and Figures.

Appendix S5. Bibliography.