UvA ECONOMETRICS

Discussion Paper: 2014/04

On Maximum Likelihood estimation of dynamic panel data models

Maurice J.G. Bun, Martin A. Carree and Artūras Juodis

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

On Maximum Likelihood estimation of dynamic panel data models[☆]

Maurice J.G. Bun[a,b], Martin A. Carree[c], Artūras Juodis[a,b,*]

[a] *Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE, Amsterdam, The Netherlands*
[b] *Tinbergen Institute, The Netherlands*
[c] *School of Business and Economics, Maastricht University, Maastricht, The Netherlands*

Abstract

We analyze 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 boundary solution different from the unique global unconstrained maximum. In an extensive Monte Carlo study we find that this boundary solution issue is non-negligible for small values of T and that different approaches might lead to substantially 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 modeling U.S. state level unemployment dynamics.

Keywords: Bimodality, Boundary Solution, Dynamic Panel Data, Maximum Likelihood.
JEL: C13, C23.

[☆] 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.

[*] Corresponding author. E.mail: a.juodis@uva.nl.

December 16, 2014

1. 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 sizeable. It is therefore common to consider the semi-asymptotic behavior 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 (2012). Other methods start from the likelihood function of the first differences, see Hsiao et al. (2002) and Binder et al. (2005). 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)). Random effects type ML estimators therefore typically exploit Chamberlain (1982) type of 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. Monte Carlo evidence provided in both studies, 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.

In this study 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 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 random effect ML estimators, we address the issue of negative variance estimates as mentioned in Maddala (1971), Alvarez and Arellano (2003) and Han and Phillips (2013). An important consequence of bimodality is that unconstrained maximization of the log-likelihood may lead to ML estimates which do not satisfy the restriction of non-negative variances. We show that when there are three solutions to the first-order condition, the left solution almost always satisfies the non-negativity restriction, while the right solution violates it. Enforcing the non-negativity constraint may furthermore lead to a boundary solution (Maddala (1971), Alvarez and Arellano (2003)).

We further investigate the impact of multiple roots and boundary conditions in finite samples in a Monte Carlo study. We consider 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 the TML and RML to initial conditions, the finite sample prop-

[1] Because the former is derived conditional on the initial observations and individual specific effects, it is also referred to as fixed effects ML in Kruiniger (2013).

erties of both estimators for small values of T depend heavily on the initial condition. A partial explanation is that the behavior of the initial condition has direct effect on the bimodality of the log-likelihood function. 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 dataset in Bun and Carree (2005) we show how these theoretical results can influence empirical estimates of U.S. state level unemployment dynamics.

Throughout the analysis we limit ourselves to an asymptotic analysis in which T is fixed and $N \to \infty$. 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 analyze the unit root case. 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 2 we introduce the Maximum Likelihood estimators for the panel AR(1) model including the cubic first-order condition for the autoregressive parameter. Section 3 deals with the possibility of multiple solutions and proposes bounded estimation as a solution. Section 4 contains the extension to dynamic models with additional covariates. Section 5 reports the results from the Monte Carlo study, while Section 6 shows the empirical results. Section 7 concludes.

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

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

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

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. Stacking the observations over time, we can write the AR(1) model for each individual as:

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

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

with $\boldsymbol{y}_i$ and $\boldsymbol{y}_{i-}$ defined accordingly and $\boldsymbol{\imath}_T$ 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), Alvarez and Arellano (2003) or Binder et al. (2005).

We continue by using the Chamberlain (1982) type of projection for η_i:[3]

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

Note that this projection is only necessary for the RML estimator. When we set $\pi = 1 - \phi$ the projection corresponds exactly to the TML framework as in this case $\Delta y_{i,1}$ does not depend on $y_{i,0}$. In the TML approach the variance of $v_i = \Delta y_{i,1} - \varepsilon_{i,1}$ is a parameter to be estimated.[4] We therefore only exploit the projection in (3) to show the algebraic comparison between RML and TML. The model can be represented as:[5]

$$\boldsymbol{R}\boldsymbol{y}_i = (\boldsymbol{e}_1\phi + \boldsymbol{\imath}_T\pi)y_{i,0} + (\boldsymbol{\imath}_T v_i + \boldsymbol{\varepsilon}_i), \tag{4}$$

where $\boldsymbol{R} = \boldsymbol{I}_T - \boldsymbol{L}_T\phi$ and $\boldsymbol{e}_1$ is the first column of the $\boldsymbol{I}_T$ matrix. Thus conditionally on $y_{i,0}$:

$$\mathrm{E}[\boldsymbol{R}\boldsymbol{y}_i | y_{i,0}] = (\boldsymbol{e}_1\phi + \boldsymbol{\imath}_T\pi)y_{i,0}, \quad \mathrm{var}[\boldsymbol{R}\boldsymbol{y}_i | y_{i,0}] = \boldsymbol{\Sigma} = \sigma_v^2 \boldsymbol{\imath}_T \boldsymbol{\imath}_T' + \sigma^2 \boldsymbol{I}_T. \tag{5}$$

The variance-covariance structure of $\boldsymbol{\Sigma}$ is of the usual random effects form. Using the matrix inversion and determinant lemmas, we obtain:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\boldsymbol{I}_T - \frac{1}{\sigma^2}\frac{\sigma_v^2}{\sigma^2 + T\sigma_v^2}\boldsymbol{\imath}_T\boldsymbol{\imath}_T', \quad |\boldsymbol{\Sigma}| = (\sigma^2)^{T-1}(\sigma^2 + T\sigma_v^2). \tag{6}$$

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

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\boldsymbol{W}_T + \frac{1}{T\theta^2}\boldsymbol{\imath}_T\boldsymbol{\imath}_T', \quad |\boldsymbol{\Sigma}| = (\sigma^2)^{T-1}\theta^2, \tag{7}$$

$$\theta^2 \equiv \sigma^2 + T\sigma_v^2. \tag{8}$$

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

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

[4]Alternatively, following Hsiao et al. (2002) one can estimate $\omega \equiv 1 + \sigma_v^2/\sigma^2 = \mathrm{var}\,(\Delta y_{i,1})/\sigma^2$.

[5]Bai (2013) considers a similar conditional maximum likelihood estimator with a possible factor structure in the error term $\boldsymbol{\varepsilon}_i$.

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):

$$\ell_i(\boldsymbol{\kappa}) = -\frac{1}{2}\left((T-1)\log(\sigma^2) + \log(\theta^2) + \left[(\boldsymbol{y}_i - \phi\boldsymbol{y}_{i-} - \boldsymbol{\imath}_T\pi y_{i,0})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \phi\boldsymbol{y}_{i-} - \boldsymbol{\imath}_T\pi y_{i,0})\right]\right), \tag{9}$$

where $\boldsymbol{\kappa} = (\phi, \pi, \sigma^2, \theta^2)'$. This function is the true likelihood function if $(\boldsymbol{\imath}_T v_i + \boldsymbol{\varepsilon}_i)$ are jointly normal.[6] Now using the fact that $\boldsymbol{W}_T\boldsymbol{\imath}_T = \mathbf{0}_T$, we can write:

$$\begin{aligned}\ell_i(\boldsymbol{\kappa}) = -\frac{1}{2}\Bigg((T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{\sigma^2}\left[(\boldsymbol{y}_i - \phi\boldsymbol{y}_{i-})'\boldsymbol{W}_T(\boldsymbol{y}_i - \phi\boldsymbol{y}_{i-})\right] \\ + \frac{T}{\theta^2}(\bar{y}_i - \phi\bar{y}_{i-} - \pi y_{i,0})^2\Bigg),\end{aligned} \tag{10}$$

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):

$$\begin{aligned}\ell(\boldsymbol{\kappa}) = -\frac{N}{2}\Bigg((T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{y}_{i,t} - \phi\tilde{y}_{i,t-1})^2 \\ + \frac{T}{N\theta^2}\sum_{i=1}^{N}(\ddot{y}_i - \phi\ddot{y}_{i-} - \rho y_{i,0})^2\Bigg).\end{aligned} \tag{11}$$

The concentrated log-likelihood function for observations in first-differences (also known as Transformed log-likelihood in Hsiao et al. (2002) and Juodis (2014)) is obtained by setting $\rho = 0$. Thus both RML and TML estimators provide an in-built bias-correction term for the usual fixed effects log-likelihood function.

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

$$\ell^c(\phi) = -\frac{N}{2}\left((T-1)\log\hat{\sigma}^2(\phi) + \log\hat{\theta}^2(\phi)\right), \tag{12}$$

[6] This parametrization ensures that $\theta^2 > 0$ or equivalently $\omega > 1 - \frac{1}{T}$ as in Hsiao et al. (2002).

where:

$$\hat{\sigma}^2(\phi) = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1})^2, \quad \hat{\theta}^2(\phi) = \frac{T}{N} \sum_{i=1}^{N} (\dot{y}_i - \phi \dot{y}_{i-})^2. \tag{13}$$

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

$$\dot{y}_i \equiv \ddot{y}_i - y_{i,0} \frac{\sum_{i=1}^{N} \ddot{y}_i y_{i,0}}{\sum_{i=1}^{N} y_{i,0}^2}, \quad \dot{y}_{i-} \equiv \ddot{y}_{i-} - y_{i,0} \frac{\sum_{i=1}^{N} \ddot{y}_{i-} y_{i,0}}{\sum_{i=1}^{N} y_{i,0}^2}, \tag{14}$$

while for the log-likelihood in first differences $\dot{y}_i \equiv \ddot{y}_i$ and $\dot{y}_{i-} \equiv \ddot{y}_{i-}$.

The likelihood function in (12) is defined for all values of $\phi \in \mathbb{R}$, hence from 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); Hayakawa and Pesaran (2014)) 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 stationarity has been assumed, 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:

$$\frac{\mathrm{d}\ell^c(\phi)}{\mathrm{d}\phi} = \frac{1}{\hat{\sigma}^2(\phi)} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t-1}(\tilde{y}_{i,t} - \phi \tilde{y}_{i,t-1}) + \frac{T}{\hat{\theta}^2(\phi)} \sum_{i=1}^{N} \dot{y}_{i-}(\dot{y}_i - \phi \dot{y}_{i-}) = 0, \tag{15}$$

or alternatively:

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

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 $\{\boldsymbol{y}_i\}_{i=1}^N$ there will be at least one and at most three solutions to (16). For general value of T there is no easy formula for the solutions, 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 fact of possible multiple solutions is mostly forgotten when discussing both maximum likelihood estimators. An exemption is Hayakawa and Pesaran (2014) who observe that the TML log-likelihood function can have more than one solution asymptotically. Here we show that also in finite samples this is possible, and that both the TML and RML log-likelihood functions have one or two local maxima. More importantly, this result is unaffected if strictly exogenous regressors are added to the model as in Section 4.

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

Corollary 1. *For any N and T all solutions of* (16) *lie in the following interval:*

$$\hat{\phi} \in (\hat{\phi}_W, \hat{\phi}_B) = \left(\frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t} \tilde{y}_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t-1}^2}, \frac{\sum_{i=1}^{N} \dot{y}_i \dot{y}_{i-}}{\sum_{i=1}^{N} \dot{y}_{i-}^2} \right). \tag{17}$$

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

The proof of this corollary 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. It is well known that $\text{plim}_{N,T\to\infty} \hat{\phi}_W = \phi$, but that for fixed T the within estimator has a negative bias, see Nickell (1981). Furthermore, it is straightforward to show that $\text{plim}_{N,T\to\infty} \hat{\phi}_B = 1$ because $\bar{y}_i$ and $\bar{y}_{i-}$ converge to the same value as T goes to infinity.

Next, we investigate the asymptotic behavior of the interval in Corollary 1. As the lower bound ($\hat{\phi}_W$) is the same for both estimators we are primarily interested in the upper bound ($\hat{\phi}_B$), which is different between estimators. The result is summarized in the following Proposition.

Proposition 1. *The probability limits of the quasi-between estimators from Corollary 1 are:*

$$\operatorname*{plim}_{N\to\infty} \hat{\phi}_B^{RML} \leq \operatorname*{plim}_{N\to\infty} \hat{\phi}_B^{TML}. \tag{18}$$

Thus the upper bound for RML is no larger than for TML. The interval for possible values for $\hat{\phi}_{RML}$ is narrower than the corresponding interval for TML. This result can be expected given that the RML estimator is found to be more efficient than TML, see Kruiniger (2013).

3. Multiple solutions and bounded estimation

The possibility of having one or three solutions to the cubic equation (16) 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 suggest a procedure of bounded estimation as in Maddala (1971).

3.1. Three-wave panel and the Transformed ML estimator

For general values of T we can only specify in which interval the solutions of the cubic equation lie, as described in Corollary 1. 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 (13):

$$\hat{\sigma}^2(\phi) = \frac{1}{2N}\sum_{i=1}^{N}\left(\Delta y_{i,2} - \phi\Delta y_{i,1}\right)^2, \quad \hat{\theta}^2(\phi) = \frac{1}{2N}\sum_{i=1}^{N}\left(\Delta y_{i,2} - (\phi-2)\Delta y_{i,1}\right)^2. \tag{19}$$

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:*

$$\ell^c(\phi) = -\frac{N}{2}\left(\log\left(\left(\hat{\sigma}^2(\phi) + \left(\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i,1}\Delta y_{i,2} - \phi\frac{1}{N}\sum_{i=1}^{N}(\Delta y_{i,1})^2\right)\right)^2 + d\right)\right), \tag{20}$$

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). Consistency[7] of $\hat{\phi}^{(l)}$ follows directly from $\text{plim}_{N\to\infty} \hat{\phi}_W = \phi_0 - \frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})}$ and $\text{plim}_{N\to\infty} D = \left(1 - \frac{\sigma_0^2}{\text{var}(\Delta y_{i,1})}\right)^2$. 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 relationships between the solutions to FOC can be further summarized as follows:

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

$$\ell^c(\hat{\phi}^{(l)}) = \ell^c(\hat{\phi}^{(r)}), \tag{21}$$

$$\hat{\sigma}_v^2(\hat{\phi}^{(l)}) > 0 > \hat{\sigma}_v^2(\hat{\phi}^{(r)}), \tag{22}$$

$$\hat{\theta}^2(\tilde{\phi}) = \hat{\sigma}^2(\tilde{\phi}), \tag{23}$$

$$\hat{\phi}_B = \hat{\phi}_W + 2. \tag{24}$$

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

Corollary 2 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.[8] 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. To illustrate the relevance

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

[8] While discussing the properties of the panel VAR estimator, Juodis (2014) observed that for the AR(1) with $T = 2$ case the results of the previous Corollary hold asymptotically. In this paper, we show that this result is exact if the quadratic equation has a positive discriminant. Furthermore, Juodis (2014) investigates the location of the second mode asymptotically and shows that the location of it depends on initialization of $y_{i,0}$.

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

Corollary 3. *For $T = 2$ and under joint normality of the data, the probability of $D > 0$ (two maxima) is given by:*

$$\Pr(D > 0) = F\left(\frac{N}{N-1}\left(2\frac{\sigma_0^2}{\operatorname{var}(\Delta y_{i,1})} - \left(\frac{\sigma_0^2}{\operatorname{var}(\Delta y_{i,1})}\right)^2\right)^{-1}\right)_{(N-1,N)}, \tag{25}$$

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

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

$$\alpha_0 \equiv \frac{1-\phi_0^2}{\sigma_0^2}\operatorname{var}\left(y_{i,0} - \frac{\eta_i}{(1-\phi_0)}\right), \Rightarrow \quad \operatorname{var}(\Delta y_{i,1}) = \sigma_0^2\left(\alpha_0\frac{1-\phi_0}{1+\phi_0} + 1\right), \tag{26}$$

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. Below we provide two graphs to illustrate how this probability depends on the population parameters.

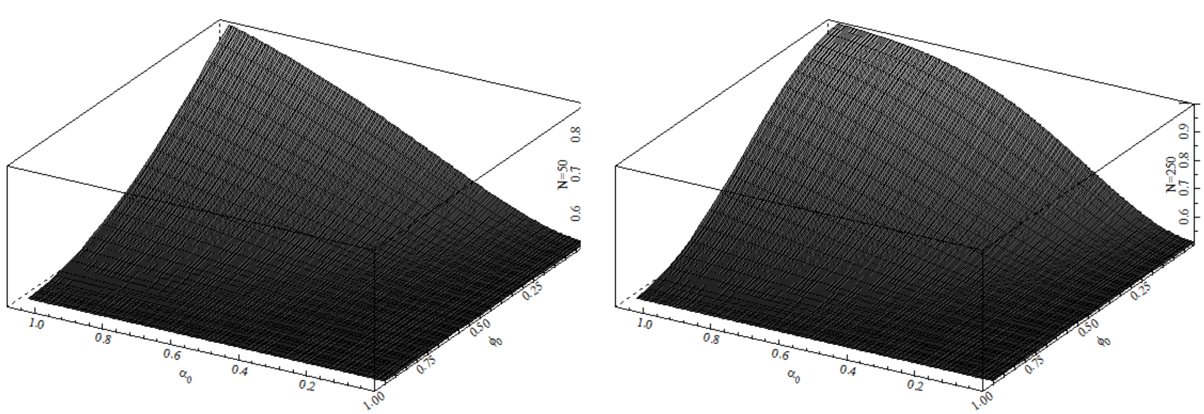


Figure 1: Probability of $D > 0$ with $N = 50$ on the left and $N = 250$ on the right. $\phi_0 \in [0; 0.95]$ and $\alpha_0 \in [0.0; 1.05]$

3.2. Further Asymptotic Results for $T > 2$ and TML

In this subsection we extend the question of one or three solutions to the FOC to $T > 2$. We consider the extent to which asymptotically the discriminant of (16) of TML is positive or negative. Before proceeding we define the following quantities:

$$\tilde{a}_E \equiv \mathrm{E}\left[\sum_{t=1}^{T} \tilde{y}_{i,t-1}^2\right], \quad \dot{a}_E \equiv \mathrm{E}\left[T\dot{y}_{i-}^2\right], \quad \xi \equiv \sum_{t=0}^{T-2}(T-t-1)\phi_0^t, \quad x \equiv \phi_0 - \phi. \tag{27}$$

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

Proposition 3. *The two non-trivial ($x \neq 0$) asymptotic solutions of* (16) *(if they exist) are implicitly defined by:*

$$x^2 \frac{T}{T-1}\left(\dot{a}_E \tilde{a}_E\right) + x\left(\frac{\xi}{T}\left(\theta_0^2 \frac{\tilde{a}_E}{T-1} - \sigma_0^2 \dot{a}_E\right) + \frac{2\xi}{T}\left(\theta_0^2 \tilde{a}_E - \frac{\sigma_0^2 \dot{a}_E}{T-1}\right)\right)$$
$$+ \left(\theta_0^2 \tilde{a}_E + \sigma_0^2 \dot{a}_E\right) - \frac{2\xi^2}{T(T-1)}\theta_0^2\sigma_0^2 = 0. \tag{28}$$

The existence of non-trivial 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. Under this reparametrization the solutions in Proposition 3 are invariant to σ_0^2, because all quantities of interest are multiplicative in σ_0^2:

$$\tilde{a}_E = \left(\sum_{t=0}^{T-1}\phi_0^{2t} - \frac{1}{T}\left(\sum_{t=0}^{T-1}\phi_0^t\right)^2\right)\frac{\alpha_0\sigma_0^2}{1-\phi_0^2} + \sigma_0^2\sum_{t=0}^{T-2}\left(\sum_{j=0}^{t}\phi_0^j\left(\phi_0^j - \frac{1}{T}\left(\sum_{j=0}^{t}\phi_0^j\right)\right)\right), \tag{29}$$

$$\dot{a}_E = \frac{\alpha_0\sigma_0^2\xi^2}{T}\frac{1-\phi_0}{1+\phi_0} + \frac{\sigma_0^2}{T}\sum_{t=0}^{T-2}\left(\sum_{j=0}^{t}\phi_0^j\right)^2, \tag{30}$$

$$\theta_0^2 = T\sigma_0^2\left(\alpha_0\frac{1-\phi_0}{1+\phi_0} + \frac{1}{T}\right). \tag{31}$$

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 2 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 $T = \{3, 5\}$. For higher values of T the border between three and one solution to the FOC approaches the $\alpha_0 = 1$ line from below. 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 $N \to \infty$. This set is obviously smaller already for $T = 3$.

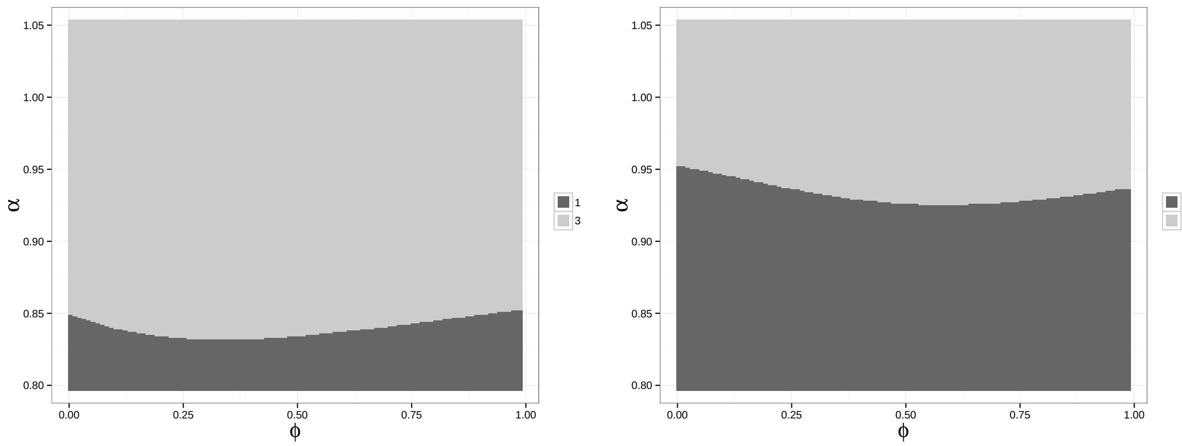


Figure 2: The sign of the discriminant for $T = 3$ on the left and $T = 5$ on the right graph. 3 is for positive discriminant and thus three solutions to FOC, while 1 is for negative discriminant and one solution. $\phi_0 \in [0; 0.99]$ and $\alpha_0 \in [0.8; 1.05]$

Figure 2 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. These results show that multiple solutions are possible for $T > 2$ even if N becomes large.

3.3. Connection with Maddala (1971)

Aside from the suggested "take left" procedure in case of bimodality to avoid negative 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 concentrated log-likelihood function in terms of the δ parameter:

$$-\frac{2}{N}\ell^c(\delta) = T \log\left[\left(\tilde{c} - 2\phi(\delta)\tilde{b} + \phi^2(\delta)\tilde{a}\right) + \delta\left(\dot{c} - 2\phi(\delta)\dot{b} + \phi^2(\delta)\dot{a}\right)\right] - \log\delta, \qquad (32)$$

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

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

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

$$-\frac{2}{N}\ell^c(\delta) = T\log\left[(\tilde{c}+\delta\dot{c}) - \frac{(\tilde{b}+\delta\dot{b})^2}{\tilde{a}+\delta\dot{a}}\right] - \log\delta. \tag{33}$$

One can recognize this expression as equation (2.3) in Maddala (1971). Similarly to (16), the FOC for this log-likelihood function is cubic in δ. Maddala (1971) investigates the occurrence of the boundary solution $\delta = 1$ for this likelihood function. Particularly one can see that the necessary and sufficient condition for $\delta = 1$ to be a local maximum is:

$$\left.\frac{\mathrm{d}\ell^c(\delta)}{\mathrm{d}\delta}\right|_{\delta=1} > 0. \tag{34}$$

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

$$\phi(1) = \frac{\tilde{b}+\dot{b}}{\tilde{a}+\dot{a}}. \tag{35}$$

We know that for $T = 2$ and in 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 Corollary 1. Furthermore, the weight of "within" estimator is monotonically decreasing in δ, because:

$$\phi(\delta) = \hat{\phi}_W q(\delta) + \hat{\phi}_B(1-q(\delta)), \quad q(\delta) = \frac{\tilde{a}}{\tilde{a}+\delta\dot{a}}, \quad q'(\delta) < 0. \tag{36}$$

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

4. 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. For ease of exposition, we consider the following simplified version with one additional regressor:

$$y_{i,t} = \eta_i + \phi y_{i,t-1} + \beta x_{i,t} + \varepsilon_{i,t}, \quad \mathrm{E}[\varepsilon_{i,t}|x_{i,0}, \ldots, x_{i,T}, y_{i,0}, \eta_i] = 0. \tag{37}$$

Then using stacked notation for individual i we have:

$$\boldsymbol{y}_i = \phi \boldsymbol{y}_{i-} + \beta \boldsymbol{x}_i + \boldsymbol{\imath}_T \eta_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'. \tag{38}$$

We continue by using the Chamberlain (1982) type of projection for η_i as in Kruiniger (2006) and Bai (2013) on not only $y_{i,0}$, but now also all the lags and leads of $x_{i,t}$:

$$\eta_i = \boldsymbol{\pi}' \boldsymbol{w}_i + v_i, \quad \mathrm{E}[v_i \boldsymbol{w}_i] = \boldsymbol{0}, \quad \boldsymbol{w}_i = (y_{i,0}, x_{i,0}, \boldsymbol{x}_i')'. \tag{39}$$

The main implication is that the conditional model can be represented as:

$$\boldsymbol{R}\boldsymbol{y}_i = \boldsymbol{e}_1 \phi y_{i,0} + \beta \boldsymbol{x}_i + \boldsymbol{\imath}_T \boldsymbol{\pi}' \boldsymbol{w}_i + (\boldsymbol{\imath}_T v_i + \boldsymbol{\varepsilon}_i). \tag{40}$$

Thus conditionally on $\boldsymbol{w}_i$:

$$\mathrm{E}[\boldsymbol{R}\boldsymbol{y}_i|\boldsymbol{w}_i] = \boldsymbol{e}_1 \phi y_{i,0} + \beta \boldsymbol{x}_i + \boldsymbol{\imath}_T \boldsymbol{\pi}' \boldsymbol{w}_i, \tag{41}$$

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

$$\begin{aligned} -\frac{2}{N}\ell(\boldsymbol{\kappa}) = (T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{y}_{i,t} - \phi\tilde{y}_{i,t-1} - \beta\tilde{x}_{i,t})^2 \\ + \frac{T}{N\theta^2}\sum_{i=1}^{N}(\bar{y}_i - \phi\bar{y}_{i-} - \beta\bar{x}_i - \boldsymbol{\pi}'\boldsymbol{w}_i)^2, \end{aligned} \tag{42}$$

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 second component of the log-likelihood function as:

$$\frac{T}{N\theta^2}\sum_{i=1}^{N}(\bar{y}_i - \phi\bar{y}_{i-} - \beta\bar{x}_i - \boldsymbol{\pi}'\boldsymbol{w}_i)^2 = \frac{T}{N\theta^2}\sum_{i=1}^{N}(\ddot{y}_i - \phi\ddot{y}_{i-} - \beta\ddot{x}_i - \boldsymbol{\rho}'\boldsymbol{z}_i)^2, \tag{43}$$

where $\boldsymbol{z}_i \equiv (y_{i,0}, x_{i,0}, \Delta x_{i,1}, \dots, \Delta x_{i,T})'$ and $\ddot{x}_i \equiv \bar{x}_i - x_{i,0}$. Furthermore, using e.g. Theorem 3.1 in Juodis (2014) the second component of the TML log-likelihood function is given by:

$$\frac{T}{N\theta^2}\sum_{i=1}^{N}(\ddot{y}_i - \phi\ddot{y}_{i-} - \beta\ddot{x}_i - \boldsymbol{\pi}_{\Delta}'\Delta\boldsymbol{x}_i)^2, \quad \Delta\boldsymbol{x}_i = (\Delta x_{i,1}, \dots, \Delta x_{i,T})'. \tag{44}$$

Hence, by setting first two components of the $\boldsymbol{\rho}$ vector to zero we obtain the TML estimator as the restricted version of the RML estimator.[9]

Irrespective of the estimator considered it is not difficult to see that, because $\ddot{x}_i \in Span(\Delta\boldsymbol{x}_i)$, one can concentrate out the $\boldsymbol{\rho}/\boldsymbol{\pi}_{\Delta}$ parameter such that the second component of the log-likelihood function in (44) does not contain the β parameter. Therefore, the (concentrated) log-likelihood function can be expressed as:

$$\begin{aligned}-\frac{2}{N}\ell(\boldsymbol{\kappa}) = (T-1)\log(\sigma^2) + \log(\theta^2) + \frac{1}{N\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{y}_{i,t} - \phi\tilde{y}_{i,t-1} - \beta\tilde{x}_{i,t})^2 \\ + \frac{T}{N\theta^2}\sum_{i=1}^{N}(\dot{y}_i - \phi\dot{y}_{i-})^2,\end{aligned} \tag{45}$$

where we defined:

$$\dot{y}_i \equiv \ddot{y}_i - \left(\sum_{i=1}^{N}\ddot{y}_i\Delta\boldsymbol{x}_i'\right)\left(\sum_{i=1}^{N}\Delta\boldsymbol{x}_i\Delta\boldsymbol{x}_i'\right)^{-1}\Delta\boldsymbol{x}_i, \tag{46}$$

$$\dot{y}_i \equiv \ddot{y}_i - \left(\sum_{i=1}^{N}\ddot{y}_i\boldsymbol{z}_i'\right)\left(\sum_{i=1}^{N}\boldsymbol{z}_i\boldsymbol{z}_i'\right)^{-1}\boldsymbol{z}_i, \tag{47}$$

for 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 (12). The cubic FOC in (16) follows directly from that.

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

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.[10]

5. 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:

$$y_{i,t} = \phi y_{i,t-1} + (1-\phi)\mu_i + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0,1), \quad t = 1, \ldots, T. \tag{48}$$

$$y_{i,0} = \gamma\mu_i + \varepsilon_{i,0}, \quad \varepsilon_{i,0} \sim \mathcal{N}\left(0, \frac{\zeta}{1-\phi^2}\right), \quad \mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right). \tag{49}$$

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 (26) as it measures the deviation from the 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):

- T(R)ML based on the global maximum (T(R)MLg), where always the global maximum is selected.

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

- T(R)ML based on the "left" maximum (T(R)MLl), that takes into account the non-negative restriction only if there two competing local maxima.
- T(R)ML with the boundary condition $\phi(1)$ as in (35), imposed (T(R)MLb) if the discriminant is negative and the estimator based on "left" solution does not satisfy the non-negativity constraint.

Note that in calculating coefficient estimators we refrained from using numerical optimization techniques. As mentioned earlier, exploiting root finder algorithms one can find solutions to the cubic first-order condition.[11]

Regarding inference we consider empirical rejection frequencies based on two sided t- and LR statistics.[12] 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 probably 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 the Appendix.

5.1. 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 and 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

[11] As implemented by e.g. the $roots(\cdot)$ function in Matlab.

[12] For the t-test we exploit the usual "sandwich" covariance matrix estimator.

the aforementioned study. Furthermore, in terms of RMSE exploiting the boundary solution $\phi(1)$ is almost always better than the "left" estimator. However, in some cases (for small N) this choice has a negative effect on mean and median bias. 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 found that in all cases where the cubic FOC had three solutions, the "left" solution always satisfied the non-negativity constraint, while the "right" solution never satisfies it. Hence for replications with three solutions only the "left" solution is natural. This is an important observation, because it suggests that there is always at most one interior maximum.

Figure 3 further illustrates how different ways of dealing with the boundary condition shapes finite sample distributions of coefficient estimators. The fact that most of the

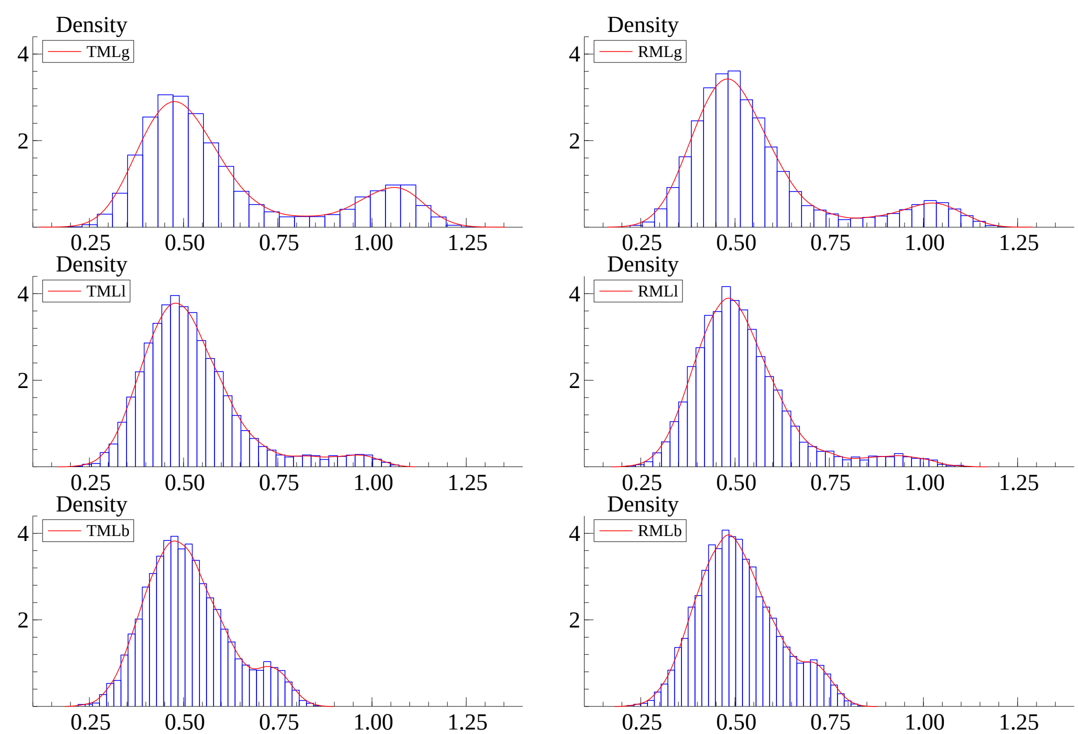


Figure 3: Finite sample distribution of the TML/RML estimators for $N = 250, T = 3, \phi = 0.5$ with covariance stationary initialization of $y_{i,0}$.

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

(2014)) either do not address the non-negativity variance issue at all or only mention it without further exploring its consequences is somewhat puzzling. As we can see for most designs substantial gains in terms of RMSE can be achieved when using RMLb/TMLb rather than RMLl/TMLl (or especially RMLg/TMLg).

5.2. 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 nominal ones. The results for the t-statistic deteriorate for $\phi = 0.8$, which is related to the non-standard behavior 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.

6. Empirical illustration

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

$$u_{i,t} = \phi u_{i,t-1} + \beta g_{i,t-1} + \eta_i + \tau_t + \varepsilon_{i,t}. \tag{50}$$

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 $t-1$.[13] The annual panel data cover the years 1991-2000 for all U.S.

[13]Some evidence on strict exogeneity of $g_{i,t-1}$ is provided in Bun and Carree (2005), hence the model may serve as an ARX(1) example.

states (including Washington D.C, 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 behavior 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.

6.1. 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.[14] 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 $RMLg = RMLl = RMLb$, 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 $\hat{\phi}^{(l)} = 0.506$.

6.2. 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.[15]

---

[14] The bias corrected FE estimates are $\hat{\phi} = 0.615$ and $\hat{\beta} = -0.057$. Furthermore, our results are in line with the results of Lokshin (2008) obtained for the TML.

[15] When $T = 2$ we report the "left" solution for TMLg as both its solutions are of the same log-likelihood value.

Table 1: TML and RML estimates for the ARX(1) model.

	TMLg		TMLl		RMLg		RMLl	
T	$\hat{\phi}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\beta}$
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

The results for the RML estimator, however, are considerably less stable. 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 A.4 (see the 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.

Table 2: TML and 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

7. 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 of a positive 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. 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 U.S. 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.

Bibliography

Ahn, S. C. and G. M. Thomas (2006): "Likelihood Based Inference for Dynamic Panel Data Models," Unpublished Manuscript.

Alvarez, J. and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71(4), 1121–1159.

Anderson, T. W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Bai, J. (2013): "Fixed-Effects Dynamic Panel Models, a Factor Analytical Method," *Econometrica*, 81, 285–314.

Binder, M., C. Hsiao, and M. H. Pesaran (2005): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Root and Cointegration," *Econometric Theory*, 21, 795–837.

Blundell, R. W. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Bun, M. J. G. and M. A. Carree (2005): "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business & Economic Statistics*, 23(2), 200–210.

Bun, M. J. G. and V. Sarafidis (2015): "Dynamic Panel Data Models," in *The Oxford Handbook of Panel Data*, ed. by B. H. Baltagi, Oxford University Press, chap. 3.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Dhaene, G. and K. Jochmans (2012): "An Adjusted Profile Likelihood for Non-Stationary Panel Data Models with Fixed Effects," Working paper.

Han, C. and P. C. B. Phillips (2013): "First Difference Maximum Likelihood and Dynamic Panel Estimation," *Journal of Econometrics*, 175, 35–45.

Hayakawa, K. and M. H. Pesaran (2014): "Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models," Working Paper.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150.

Juodis, A. (2014): "First Difference Transformation in Panel VAR models: Robustness, Estimation and Inference," UvA-Econometrics working paper 2013/06.

Kruiniger, H. (2006): "Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Condition," Working paper 582, Queen Marry, University of London.

——— (2008): "Maximum Likelihood Estimation and Inference Methods for the Covariance Stationary Panel AR(1)/Unit Root Model," *Journal of Econometrics*, 144, 447–464.

——— (2013): "Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Conditions," *Journal of Econometrics*, 173, 175–188.

Lancaster, T. (2002): "Orthogonal Parameters and Panel Data," *Review of Economic Studies*, 69, 647–666.

Lokshin, B. (2008): "A Monte Carlo Comparison of Alternative Estimators for Dynamic Panel Data Models," *Applied Economics Letters*, 15, 15–18.

Maddala, G. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39, 341–35.

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.

Appendix A. Proofs, Monte Carlo results and Figures

Proof Proposition 1. To prove that the quasi-between estimator for RML is asymptotically smaller than for TML, it is sufficient to show that:

$$\theta_T^2 \dot{a}_\infty^R - \theta_R^2 \dot{a}_\infty^T \geq 0,$$

as $\text{plim}_{N\to\infty} \hat{\phi}_B^{(j)} = \phi_0 + \frac{1}{T}\frac{\theta_j^2}{\dot{a}_\infty^j}\xi$, for $j = \{T, R\}$ that follows from the fact that $\phi = \phi_0$ is always a solution to FOC asymptotically (for more details refer to the proof of Proposition 3) so that the maximum likelihood estimator is consistent. Here $\dot{a}_\infty^{(j)} = \text{plim}_{N\to\infty} \frac{T}{N}\sum_{i=1}^N \dot{y}_{i-}^2$ for $j = \{T, R\}$. Observe that for any T:

$$\xi \equiv \sum_{t=0}^{T-2}(T-t-1)\phi_0^t \geq 0,$$

$$\theta_T^2 = \sigma_0 + T(\text{var}(\Delta y_{i,1}) - \sigma_0^2) = \sigma_0^2 + T\left(\text{E}\left((\phi_0 - 1 + \pi_0)y_{i,0} + v_i + \varepsilon_{i,1}\right)^2 - \sigma_0^2\right),$$

$$\theta_R^2 = \sigma_0^2 + T\left(\text{E}\left(v_i\right)^2\right),$$

where as before $\eta_i = \pi_0 y_{i,0} + v_i$ and $\pi_0 = \frac{\text{E}[y_{i,0}\eta_i]}{\text{E}[y_{i,0}^2]}$. The difference is thus $\theta_T^2 - \theta_R^2 = T\left((\phi_0 - 1 + \pi_0)^2\,\text{E}[y_{i,0}^2]\right) \geq 0$. Regarding $\dot{a}_\infty^R$ we have:

$$\dot{a}_\infty^R = \dot{a}_\infty^T - T\frac{(\text{E}[\ddot{y}_{i-}y_{i,0}])^2}{\text{E}[y_{i,0}^2]},$$

$$\text{E}[\ddot{y}_{i-}y_{i,0}] = \frac{1}{T}\xi(\phi_0 - 1 + \pi_0)\,\text{E}[y_{i,0}^2].$$

While for $\dot{a}_\infty^T$:[16]

$$\begin{aligned}
\dot{a}_\infty^T &= \left(\frac{\xi^2}{T}(\phi_0-1)^2\right)\text{E}\left(y_{i,0} - \frac{\eta_i}{1-\phi_0}\right)^2 + \frac{\sigma_0^2}{T}\sum_{t=0}^{T-2}\left(\sum_{j=0}^{t}\phi_0^j\right)^2, \\
&= \left(\frac{\xi^2}{T}(\phi_0-1)^2\right)\text{E}\left(y_{i,0}\left(1 - \frac{\pi_0}{1-\phi_0}\right) - \frac{v_i}{1-\phi_0}\right)^2 + \frac{\sigma_0^2}{T}\sum_{t=0}^{T-2}\left(\sum_{j=0}^{t}\phi_0^j\right)^2, \\
&= \left(\frac{1}{T}(\phi_0 - 1 + \pi_0)^2\xi^2\right)\text{E}[y_{i,0}^2] + \frac{\xi^2}{T}\,\text{E}[v_i^2] + \frac{\sigma_0^2}{T}\sum_{t=0}^{T-2}\left(\sum_{j=0}^{t}\phi_0^j\right)^2,
\end{aligned}$$

[16] For derivations of this term please refer to Lemma 2 in the Appendix of Juodis (2014).

as $\mathrm{E}[v_i y_{i,0}] = 0$, which implies that:

$$\dot{a}_\infty^R = \frac{\xi^2}{T}\,\mathrm{E}[v_i^2] + \frac{\sigma_0^2}{T}\sum_{t=0}^{T-2}\left(\sum_{j=0}^{t}\phi_0^j\right)^2 = \frac{1}{T}\left(\xi^2\,\mathrm{E}[v_i^2] + \sigma_0^2 q\right),$$

where q is implicitly defined. Denote $w \equiv (\phi_0 - 1 + \pi_0)^2\,\mathrm{E}[y_{i,0}^2] \geq 0$, then:

$$\begin{aligned}
\theta_T^2 \dot{a}_\infty^R - \theta_R^2 \dot{a}_\infty^T &= Tw\dot{a}_\infty^R - \theta_R^2 \frac{\xi^2}{T} w \\
&= w\left(\xi^2\,\mathrm{E}[v_i^2] + \sigma_0^2 q\right) - w\xi^2\left(\mathrm{E}[v_i^2] + \frac{1}{T}\sigma_0^2\right) \\
&= w\sigma_0^2\left(q - \frac{\xi^2}{T}\right) > 0,
\end{aligned}$$

where the last result follows as an implication of the Jensen's inequality. □

Proof Proposition 2. Using the variables defined in Section 3.3, we note that for $T = 2$ and TML:

$$\tilde{a} = \dot{a}, \quad \dot{b} = \tilde{b} + 2\tilde{a}, \quad \dot{c} = \tilde{c} + 4(\tilde{a} + \tilde{b}).$$

and thus $\hat{\theta}^2(\phi) = \hat{\sigma}^2(\phi) + 4(\tilde{a}(1-\phi) + \tilde{b})$. Furthermore, for $T = 2$ we have from (12) that $\ell^c(\phi) \propto \log(\hat{\theta}^2(\phi)\hat{\sigma}^2(\phi))$ with:

$$\begin{aligned}
\hat{\theta}^2(\phi)\hat{\sigma}^2(\phi) &= \hat{\sigma}^4(\phi) + 4\hat{\sigma}^2(\tilde{b} - \phi\tilde{a}) + 4\hat{\sigma}^2(\phi)\tilde{a} \\
&= \left(2(\tilde{b} - \phi\tilde{a}) + \hat{\sigma}^2(\phi)\right)^2 - 4(\tilde{b} - \phi\tilde{a})^2 + 4\hat{\sigma}^2(\phi)\tilde{a} \\
&= \left(2(\tilde{b} - \phi\tilde{a}) + \hat{\sigma}^2(\phi)\right)^2 - 4(\tilde{b}^2 - 2\phi\tilde{a}\tilde{b} + \phi^2\tilde{a}^2) + 4(\tilde{a}\tilde{c} - 2\phi\tilde{a}\tilde{b} + \phi^2\tilde{a}^2) \\
&= \left(2(\tilde{b} - \phi\tilde{a}) + \hat{\sigma}^2(\phi)\right)^2 + 4(\tilde{a}\tilde{c} - \tilde{b}^2) \\
&= \left(2(\tilde{b} - \phi\tilde{a}) + \hat{\sigma}^2(\phi)\right)^2 + d,
\end{aligned}$$

where

$$d = \left(\frac{1}{N}\sum_{i=1}^{N}(\Delta y_{i,1})^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}(\Delta y_{i,2})^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i,1}\Delta y_{i,2}\right)^2. \tag{A.1}$$

□

Proof Corollary 2. From Proposition 2 we have that: $\hat{\theta}^2(\phi) - \hat{\sigma}^2(\phi) = 4(\tilde{a}(1-\phi) + \tilde{b})$. Given that $\tilde{\phi} = 1 + \tilde{b}/\tilde{a}$ one can easily see that $\hat{\theta}^2(\tilde{\phi}) - \hat{\sigma}^2(\tilde{\phi}) = 4(\tilde{a}(1-\tilde{\phi}) + \tilde{b}) = 0$. The first and the third parts follow from the symmetry established in Proposition 2, while the last part follows directly from definitions.

These results do not hold for $T > 2$ and/or the RML estimator. For example, observe that for general T we can decompose (for simplicity denote $\tilde{T}_1 = 1/(T-1)$):

$$\hat{\theta}^2(\phi) = \dot{c} - 2\phi\dot{b} + \phi^2\dot{a}, \quad \hat{\sigma}^2(\phi) = \tilde{T}_1\left(\tilde{c} - 2\phi\tilde{b} + \phi^2\tilde{a}\right),$$
$$\hat{\theta}^2(\phi) - \hat{\sigma}^2(\phi) = \phi^2\left(\dot{a} - \tilde{T}_1\tilde{a}\right) - 2\phi\left(\dot{b} - \tilde{T}_1\tilde{b}\right) + \left(\dot{c} - \tilde{T}_1\tilde{c}\right).$$

For $T = 2$ and TML we have $\dot{a} = \tilde{a}$ and the right hand side of the last equation becomes linear in ϕ. Hence, setting the left hand side of the last equation equal to zero and solving for ϕ, there is only one solution given by $\hat{\phi}_W + 1$. For $T > 2$ and/or the RML estimator, the equation $\theta^2(\phi) - \sigma^2(\phi) = 0$ has two solutions of the form:

$$\check{\phi} = \frac{\dot{b} - \tilde{T}_1\tilde{b}}{\dot{a} - \tilde{T}_1\tilde{a}} \pm \sqrt{\frac{(\dot{b} - \tilde{T}_1\tilde{b})^2 - (\dot{a} - \tilde{T}_1\tilde{a})(\dot{c} - \tilde{T}_1\tilde{c})}{(\dot{a} - \tilde{T}_1\tilde{a})^2}}. \tag{A.2}$$

The first order condition in (16), evaluated at the value $\check{\phi}$, becomes proportional to $(\tilde{b} + \dot{b}) - \check{\phi}(\tilde{a} + \dot{a}) \neq 0$. In the point $\check{\phi}$ the first order condition is not zero in this case, hence the corner solution $\hat{\theta}^2(\check{\phi}) = \hat{\sigma}^2(\check{\phi})$ cannot hold. □

Proof Corollary 3. Note that the discriminant $D = 1 - S_{22.1}/S_{11}$, where $S_{22.1} \equiv S_{22} - S_{12}^2/S_{11}$, where S_{ij} is the i,j element of the $\boldsymbol{S} = \sum_{i=1}^{N}((\Delta y_{i,1}, \Delta y_{i,2})'(\Delta y_{i,1}, \Delta y_{i,2}))$ matrix. Under joint normality the elements $S_{22.1}$ and S_{11} are independent $\chi^2(\cdot)$ random variables with respectively $N-1$ and N degrees of freedom. The main result follows after observing that $\mathrm{E}[\mathrm{vech}\,\boldsymbol{S}] = (\mathrm{var}\,(\Delta y_{i,1}), \phi_0\,\mathrm{var}\,(\Delta y_{i,1}) - \sigma_0^2, \phi_0^2\,\mathrm{var}\,(\Delta y_{i,1}) + 2\sigma_0^2(1-\phi_0))'$. □

Proof Proposition 3. Observe that for any value of ϕ (see e.g. Juodis (2014)):

$$\sigma_E^2(x) \equiv \mathrm{E}[\hat{\sigma}^2(\phi)] = \sigma_0^2 + \frac{1}{T-1}\left(x^2\tilde{a}_E - x\frac{2\xi}{T}\sigma_0^2\right),$$
$$\theta_E^2(x) \equiv \mathrm{E}[\hat{\theta}^2(\phi)] = \theta_0^2 + x^2\dot{a}_E + x\frac{2\xi}{T}\theta_0^2,$$

with $\tilde{a}_E, \dot{a}_E, \xi, x$ defined in (27). It is not difficult to see that the asymptotic polynomial is given by:

$$\theta_E^2(x)\left(\tilde{a}_E x - \frac{\sigma_0^2\xi}{T}\right) + \sigma_E^2(x)\left(\dot{a}_E x + \frac{\theta_0^2\xi}{T}\right) = 0.$$

Plugging in the expressions for $\sigma_E^2(x)$ and $\theta_E^2(x)$ into the previous formula:

$$x\left([\theta_E^2(x)\tilde{a}_E + \sigma_E^2(x)\dot{a}_E] + \frac{x}{T}\xi\left[\theta_0^2\frac{\tilde{a}_E}{T-1} - \sigma_0^2\dot{a}_E\right] - \frac{2\xi^2}{T(T-1)}\theta_0^2\sigma_0^2\right) = 0.$$

Note that:

$$\theta_E^2(x)\tilde{a}_E + \sigma_E^2(x)\dot{a}_E = x^2\left(1+\frac{1}{T-1}\right)(\dot{a}_E\tilde{a}_E) + x\frac{2\xi}{T}\left(\theta_0^2\tilde{a}_E - \frac{1}{T-1}\sigma_0^2\dot{a}_E\right) + \left(\theta_0^2\tilde{a}_E + \sigma_0^2\dot{a}_E\right).$$

Combining both expressions and removing the trivial solution $x = 0$ we get:

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

□

Table A.3: Estimation Results for $N = 50$, $T = 3$.

	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$			
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
TMLl	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
RMLl	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
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$			
TMLg	0.90	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
TMLl	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
RMLl	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 A.4: Estimation Results for $N = 50$, $T = 7$.

	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$			
TMLg	0.51	0.49	0.10	0.12	0.52	0.49	0.08	0.15	0.51	0.49	0.10	0.11	0.51	0.49	0.10	0.11
RMLg	0.50	0.49	0.09	0.08	0.52	0.50	0.08	0.13	0.50	0.49	0.10	0.08	0.50	0.49	0.10	0.10
TMLl	0.50	0.49	0.10	0.07	0.49	0.49	0.08	0.06	0.50	0.49	0.10	0.07	0.50	0.49	0.10	0.07
RMLl	0.50	0.49	0.09	0.07	0.49	0.49	0.08	0.06	0.50	0.49	0.10	0.07	0.50	0.49	0.10	0.07
TMLb	0.50	0.49	0.10	0.07	0.49	0.49	0.08	0.06	0.50	0.49	0.10	0.07	0.50	0.49	0.10	0.07
RMLb	0.49	0.49	0.09	0.07	0.49	0.49	0.08	0.06	0.49	0.49	0.10	0.07	0.50	0.49	0.10	0.07
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$			
TMLg	0.87	0.86	0.25	0.16	0.91	0.90	0.30	0.19	0.87	0.85	0.24	0.16	0.87	0.85	0.24	0.16
RMLg	0.81	0.80	0.14	0.11	0.91	0.89	0.30	0.19	0.81	0.80	0.14	0.11	0.84	0.82	0.20	0.14
TMLl	0.81	0.79	0.15	0.11	0.81	0.79	0.13	0.10	0.81	0.79	0.16	0.11	0.81	0.79	0.16	0.11
RMLl	0.80	0.79	0.13	0.10	0.81	0.79	0.13	0.10	0.80	0.79	0.14	0.10	0.81	0.79	0.15	0.11
TMLb	0.79	0.79	0.14	0.08	0.80	0.79	0.13	0.09	0.79	0.79	0.13	0.08	0.79	0.79	0.13	0.08
RMLb	0.77	0.78	0.09	0.07	0.80	0.79	0.13	0.09	0.77	0.78	0.09	0.07	0.78	0.79	0.12	0.08

Table A.5: Estimation Results for $N = 250$, $T = 3$.

	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$			
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
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	0.16	0.20
TMLl	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
RMLl	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	0.14	0.13
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
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	0.51	0.49	0.14	0.10
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$			
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
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	0.83	0.26	0.18
TMLl	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
RMLl	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	0.23	0.15
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
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	0.17	0.11

Table A.6: Estimation Results for $N = 250$, $T = 7$.

	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE	Mean	Median	IQR	RMSE
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$			
TMLg	0.49	0.49	0.04	0.03	0.49	0.49	0.03	0.02	0.49	0.49	0.04	0.03	0.49	0.49	0.04	0.03
RMLg	0.49	0.49	0.04	0.03	0.49	0.49	0.03	0.02	0.49	0.49	0.04	0.03	0.49	0.50	0.04	0.03
TMLl	0.49	0.49	0.04	0.03	0.49	0.49	0.03	0.02	0.49	0.49	0.04	0.03	0.49	0.49	0.04	0.03
RMLl	0.49	0.49	0.04	0.03	0.49	0.49	0.03	0.02	0.49	0.49	0.04	0.03	0.49	0.50	0.04	0.03
TMLb	0.49	0.49	0.04	0.03	0.49	0.49	0.03	0.02	0.49	0.49	0.04	0.03	0.49	0.49	0.04	0.03
RMLb	0.49	0.49	0.04	0.03	0.49	0.49	0.03	0.02	0.49	0.49	0.04	0.03	0.49	0.50	0.04	0.03
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$				$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$			
TMLg	0.84	0.80	0.10	0.11	0.87	0.81	0.26	0.15	0.83	0.80	0.10	0.10	0.83	0.80	0.10	0.10
RMLg	0.80	0.79	0.06	0.05	0.87	0.81	0.24	0.14	0.80	0.79	0.06	0.05	0.81	0.80	0.07	0.07
TMLl	0.80	0.79	0.06	0.06	0.80	0.79	0.05	0.04	0.80	0.79	0.06	0.06	0.80	0.79	0.06	0.06
RMLl	0.80	0.79	0.06	0.04	0.80	0.79	0.05	0.05	0.80	0.79	0.06	0.04	0.80	0.79	0.06	0.06
TMLb	0.80	0.79	0.06	0.05	0.80	0.79	0.05	0.04	0.80	0.79	0.06	0.05	0.80	0.79	0.06	0.05
RMLb	0.79	0.79	0.05	0.03	0.80	0.79	0.05	0.04	0.79	0.79	0.05	0.03	0.80	0.79	0.06	0.04

Table A.7: t- test results for $N = 50$, $T = 3$.

$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$				
TMLg	0.47	0.44	0.44	0.46	0.46	0.49	0.44	0.46	0.54	0.63	0.46	0.43	0.42	0.42	0.43	0.46	0.43	0.42	0.42	0.43
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	0.40	0.36	0.35	0.36	0.40
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	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	0.26	0.22	0.22	0.26	0.33
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	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	0.29	0.15	0.10	0.11	0.16	0.27
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$				
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	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	0.46	0.38	0.31	0.28	0.31
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	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	0.33	0.25	0.21	0.23	0.32
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	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	0.17	0.11	0.13	0.19	0.32

Table A.8: t- test results for $N = 50$, $T = 7$.

$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$				
TMLg	0.79	0.27	0.08	0.35	0.77	0.90	0.38	0.11	0.44	0.88	0.77	0.26	0.08	0.34	0.76	0.77	0.26	0.08	0.34	0.76
RMLg	0.79	0.27	0.06	0.33	0.76	0.90	0.38	0.10	0.44	0.88	0.77	0.26	0.06	0.32	0.75	0.77	0.26	0.07	0.33	0.75
TMLl	0.79	0.26	0.05	0.32	0.76	0.90	0.37	0.06	0.40	0.87	0.77	0.25	0.05	0.31	0.74	0.77	0.25	0.05	0.31	0.74
RMLl	0.79	0.26	0.06	0.32	0.76	0.90	0.38	0.06	0.40	0.87	0.77	0.25	0.06	0.32	0.74	0.77	0.25	0.05	0.31	0.74
TMLb	0.79	0.26	0.05	0.32	0.76	0.90	0.37	0.06	0.40	0.87	0.77	0.25	0.05	0.31	0.74	0.77	0.25	0.05	0.31	0.74
RMLb	0.79	0.26	0.06	0.32	0.76	0.90	0.38	0.06	0.40	0.87	0.77	0.25	0.05	0.32	0.75	0.77	0.24	0.05	0.31	0.74
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$				
TMLg	0.64	0.47	0.42	0.43	0.51	0.72	0.50	0.49	0.54	0.57	0.64	0.46	0.41	0.42	0.51	0.64	0.46	0.41	0.42	0.51
RMLg	0.61	0.32	0.21	0.34	0.63	0.72	0.50	0.48	0.54	0.59	0.60	0.32	0.21	0.34	0.62	0.62	0.39	0.32	0.38	0.55
TMLl	0.51	0.27	0.24	0.32	0.55	0.59	0.21	0.19	0.35	0.58	0.50	0.27	0.24	0.31	0.56	0.50	0.27	0.24	0.31	0.56
RMLl	0.58	0.29	0.18	0.33	0.63	0.59	0.21	0.19	0.35	0.58	0.57	0.29	0.18	0.33	0.63	0.52	0.28	0.22	0.32	0.58
TMLb	0.44	0.18	0.13	0.27	0.56	0.56	0.17	0.14	0.31	0.58	0.43	0.17	0.12	0.27	0.57	0.43	0.17	0.12	0.27	0.57
RMLb	0.46	0.12	0.07	0.30	0.68	0.54	0.15	0.12	0.30	0.58	0.45	0.11	0.06	0.31	0.69	0.42	0.14	0.09	0.27	0.59

Table A.9: t- test results for $N = 250$, $T = 3$.

$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$				
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
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$				
TMLg	0.59	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	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	0.50	0.39	0.29	0.29	0.43
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	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	0.40	0.30	0.22	0.26	0.44
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	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	0.27	0.14	0.09	0.21	0.45

Table A.10: t- test results for $N = 250$, $T = 7$.

$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\mu = 3$				
TMLg	0.99	0.86	0.05	0.84	0.99	1.00	0.94	0.05	0.94	1.00	0.99	0.85	0.05	0.82	0.99	0.99	0.85	0.05	0.82	0.99
RMLg	0.99	0.87	0.05	0.85	0.99	1.00	0.94	0.05	0.94	1.00	0.99	0.85	0.05	0.83	0.99	0.99	0.85	0.05	0.82	0.99
TMLl	0.99	0.86	0.05	0.84	0.99	1.00	0.94	0.05	0.94	1.00	0.99	0.85	0.05	0.82	0.99	0.99	0.85	0.05	0.82	0.99
RMLl	0.99	0.87	0.05	0.85	0.99	1.00	0.94	0.05	0.94	1.00	0.99	0.85	0.05	0.83	0.99	0.99	0.85	0.05	0.82	0.99
TMLb	0.99	0.86	0.05	0.84	0.99	1.00	0.94	0.05	0.94	1.00	0.99	0.85	0.05	0.82	0.99	0.99	0.85	0.05	0.82	0.99
RMLb	0.99	0.87	0.05	0.85	0.99	1.00	0.94	0.05	0.94	1.00	0.99	0.85	0.05	0.83	0.99	0.99	0.85	0.05	0.82	0.99
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\mu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\mu = 3$				
TMLg	0.99	0.57	0.24	0.62	0.80	0.99	0.75	0.32	0.75	0.90	0.99	0.54	0.24	0.60	0.80	0.99	0.54	0.24	0.60	0.80
RMLg	0.99	0.58	0.08	0.61	0.92	0.99	0.74	0.30	0.75	0.91	0.98	0.56	0.08	0.60	0.92	0.99	0.52	0.14	0.58	0.85
TMLl	0.98	0.46	0.11	0.55	0.87	0.99	0.68	0.06	0.64	0.93	0.98	0.44	0.11	0.53	0.86	0.98	0.44	0.11	0.53	0.86
RMLl	0.98	0.57	0.07	0.61	0.92	0.99	0.68	0.06	0.64	0.93	0.98	0.56	0.08	0.60	0.92	0.98	0.48	0.10	0.55	0.87
TMLb	0.96	0.43	0.06	0.52	0.87	0.99	0.68	0.05	0.63	0.93	0.95	0.40	0.07	0.51	0.87	0.95	0.40	0.07	0.51	0.87
RMLb	0.98	0.56	0.04	0.62	0.96	0.99	0.67	0.05	0.63	0.93	0.98	0.54	0.04	0.63	0.96	0.95	0.44	0.06	0.53	0.88

Table A.11: LR test results for $N = 50$, $T = 3$.

$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 3$				
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
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 3$				
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

Table A.12: LR test results for $N = 50$, $T = 7$.

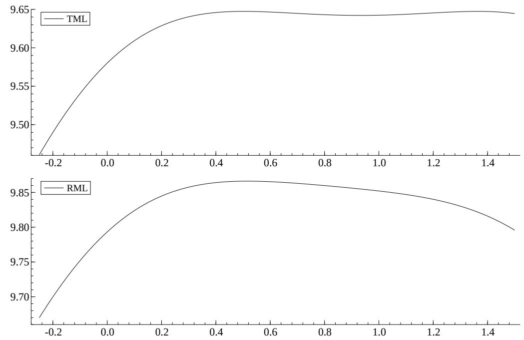
$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 3$				
TMLg	0.82	0.30	0.06	0.25	0.65	0.91	0.39	0.06	0.34	0.83	0.80	0.28	0.05	0.24	0.63	0.80	0.28	0.05	0.24	0.63
RMLg	0.82	0.30	0.04	0.25	0.67	0.91	0.39	0.06	0.34	0.82	0.80	0.28	0.04	0.24	0.66	0.80	0.28	0.05	0.24	0.63
TMLl	0.82	0.29	0.04	0.24	0.64	0.91	0.38	0.05	0.32	0.81	0.80	0.28	0.04	0.23	0.62	0.80	0.28	0.04	0.23	0.62
RMLl	0.82	0.30	0.04	0.24	0.66	0.91	0.38	0.05	0.32	0.81	0.80	0.28	0.04	0.24	0.66	0.80	0.28	0.04	0.23	0.63
TMLb	0.82	0.29	0.04	0.24	0.64	0.91	0.38	0.05	0.32	0.81	0.80	0.28	0.04	0.23	0.62	0.80	0.28	0.04	0.23	0.62
RMLb	0.82	0.30	0.04	0.24	0.66	0.91	0.38	0.05	0.32	0.81	0.80	0.28	0.04	0.24	0.66	0.80	0.28	0.04	0.23	0.63
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 3$				
TMLg	0.70	0.24	0.06	0.12	0.12	0.79	0.30	0.06	0.16	0.19	0.69	0.22	0.06	0.12	0.12	0.69	0.22	0.06	0.12	0.12
RMLg	0.70	0.24	0.07	0.15	0.42	0.79	0.32	0.08	0.18	0.21	0.69	0.23	0.06	0.14	0.43	0.69	0.23	0.07	0.13	0.20
TMLl	0.66	0.21	0.04	0.09	0.10	0.76	0.25	0.04	0.12	0.15	0.65	0.20	0.04	0.09	0.10	0.65	0.20	0.04	0.09	0.10
RMLl	0.69	0.23	0.06	0.14	0.42	0.76	0.25	0.05	0.12	0.16	0.68	0.22	0.06	0.14	0.43	0.66	0.21	0.05	0.11	0.19
TMLb	0.66	0.18	0.03	0.09	0.10	0.76	0.24	0.04	0.12	0.15	0.65	0.17	0.03	0.09	0.10	0.65	0.17	0.03	0.09	0.10
RMLb	0.69	0.18	0.03	0.14	0.43	0.76	0.24	0.04	0.12	0.16	0.68	0.17	0.02	0.13	0.43	0.66	0.18	0.03	0.10	0.19

Table A.13: LR test results for $N = 250$, $T = 3$.

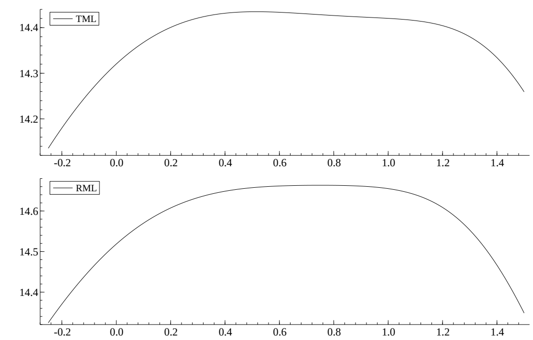
$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 3$				
TMLg	0.73	0.26	0.09	0.18	0.36	0.93	0.43	0.09	0.33	0.73	0.68	0.23	0.08	0.16	0.31	0.68	0.23	0.08	0.16	0.31
RMLg	0.73	0.23	0.05	0.17	0.42	0.93	0.40	0.08	0.31	0.72	0.70	0.21	0.05	0.16	0.41	0.67	0.22	0.07	0.15	0.31
TMLl	0.70	0.21	0.06	0.14	0.32	0.93	0.37	0.05	0.28	0.69	0.65	0.19	0.06	0.13	0.27	0.65	0.19	0.06	0.13	0.27
RMLl	0.73	0.23	0.05	0.16	0.42	0.93	0.38	0.05	0.28	0.70	0.69	0.21	0.05	0.16	0.41	0.66	0.20	0.06	0.14	0.29
TMLb	0.70	0.21	0.05	0.14	0.32	0.93	0.37	0.05	0.28	0.69	0.65	0.19	0.05	0.13	0.27	0.65	0.19	0.05	0.13	0.27
RMLb	0.73	0.23	0.05	0.16	0.42	0.93	0.38	0.05	0.28	0.69	0.69	0.20	0.04	0.15	0.41	0.66	0.19	0.05	0.13	0.29
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 3$				
TMLg	0.33	0.04	0.03	0.06	0.05	0.47	0.10	0.04	0.08	0.09	0.31	0.03	0.04	0.06	0.04	0.31	0.03	0.04	0.06	0.04
RMLg	0.43	0.12	0.05	0.09	0.23	0.49	0.11	0.04	0.09	0.09	0.41	0.12	0.05	0.09	0.23	0.37	0.09	0.05	0.07	0.09
TMLl	0.32	0.04	0.02	0.04	0.04	0.45	0.09	0.02	0.06	0.07	0.30	0.03	0.02	0.04	0.03	0.30	0.03	0.02	0.04	0.03
RMLl	0.42	0.12	0.05	0.09	0.23	0.46	0.10	0.03	0.06	0.08	0.41	0.11	0.05	0.09	0.23	0.35	0.07	0.04	0.06	0.08
TMLb	0.31	0.04	0.02	0.04	0.04	0.45	0.09	0.02	0.06	0.07	0.29	0.03	0.02	0.04	0.03	0.29	0.03	0.02	0.04	0.03
RMLb	0.39	0.06	0.02	0.08	0.23	0.45	0.09	0.03	0.06	0.08	0.37	0.05	0.02	0.08	0.24	0.32	0.04	0.02	0.05	0.08

Table A.14: LR test results for $N = 250$, $T = 7$.

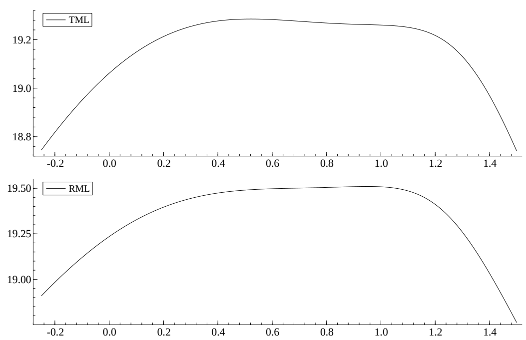
$\phi - \phi_0$	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2	-.2	-.1	**.0**	.1	.2
	$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.5$ $\gamma = 1.0$ $\sigma_\nu = 3$				
TMLg	0.99	0.87	0.04	0.82	0.99	0.99	0.95	0.05	0.93	1.00	0.99	0.86	0.05	0.80	0.99	0.99	0.86	0.05	0.80	0.99
RMLg	0.99	0.88	0.04	0.83	0.99	0.99	0.95	0.04	0.93	1.00	0.99	0.86	0.04	0.81	0.99	0.99	0.86	0.05	0.80	0.99
TMLl	0.99	0.87	0.04	0.82	0.99	0.99	0.95	0.05	0.93	1.00	0.99	0.86	0.05	0.80	0.99	0.99	0.86	0.05	0.80	0.99
RMLl	0.99	0.88	0.04	0.83	0.99	0.99	0.95	0.04	0.93	1.00	0.99	0.86	0.04	0.81	0.99	0.99	0.86	0.05	0.80	0.99
TMLb	0.99	0.87	0.04	0.82	0.99	0.99	0.95	0.05	0.93	1.00	0.99	0.86	0.05	0.80	0.99	0.99	0.86	0.05	0.80	0.99
RMLb	0.99	0.88	0.04	0.83	0.99	0.99	0.95	0.04	0.93	1.00	0.99	0.86	0.04	0.81	0.99	0.99	0.86	0.05	0.80	0.99
	$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 0.5$ $\sigma_\nu = 3$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 1$					$\phi = 0.8$ $\gamma = 1.0$ $\sigma_\nu = 3$				
TMLg	0.99	0.69	0.08	0.33	0.42	0.99	0.80	0.09	0.49	0.60	0.99	0.67	0.08	0.32	0.43	0.99	0.67	0.08	0.32	0.43
RMLg	0.99	0.71	0.05	0.49	0.94	0.99	0.80	0.10	0.50	0.62	0.99	0.70	0.05	0.49	0.95	0.99	0.67	0.06	0.35	0.69
TMLl	0.99	0.67	0.06	0.30	0.41	0.99	0.78	0.05	0.44	0.56	0.99	0.65	0.06	0.29	0.42	0.99	0.65	0.06	0.29	0.42
RMLl	0.99	0.71	0.05	0.49	0.94	0.99	0.78	0.05	0.44	0.57	0.99	0.70	0.05	0.49	0.95	0.99	0.66	0.06	0.34	0.69
TMLb	0.99	0.67	0.04	0.30	0.41	0.99	0.78	0.05	0.44	0.56	0.99	0.65	0.04	0.29	0.42	0.99	0.65	0.04	0.29	0.42
RMLb	0.99	0.71	0.03	0.49	0.93	0.99	0.78	0.05	0.44	0.57	0.99	0.70	0.03	0.49	0.94	0.99	0.66	0.04	0.34	0.69



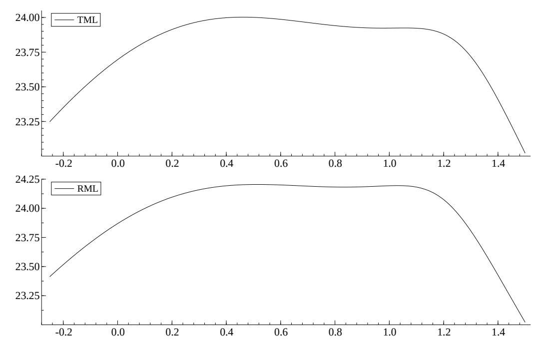
(a) $T = 2$



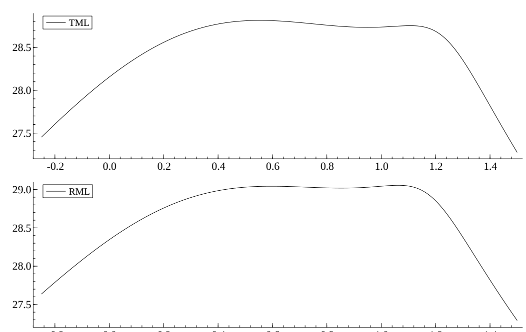
(b) $T = 3$



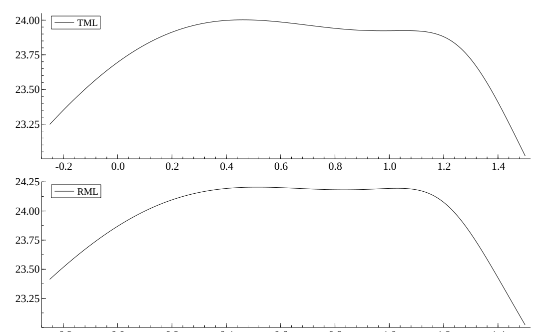
(c) $T = 4$



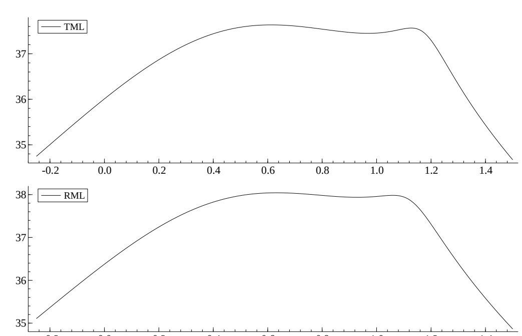
(d) $T = 5$



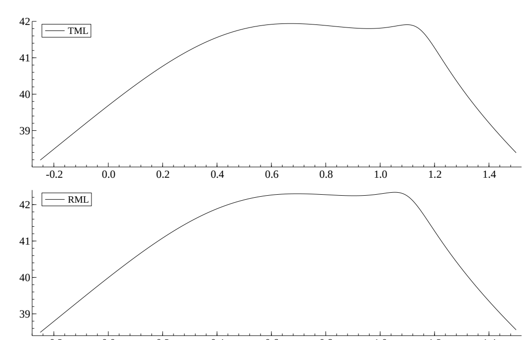
(e) $T = 6$



(f) $T = 7$



(g) $T = 8$



(h) $T = 9$

Figure A.4: Average concentrated log-likelihood function for ϕ in AR(1) model.