HAL
open science

# An Adjusted profile likelihood for non-stationary panel data models with fixed effects 

Geert Dhaene, Koen Jochmans

## To cite this version:

Geert Dhaene, Koen Jochmans. An Adjusted profile likelihood for non-stationary panel data models with fixed effects. 2011. hal-01073732

## HAL Id: hal-01073732 https://sciencespo.hal.science/hal-01073732

Preprint submitted on 30 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# AN ADJUSTED PROFILE LIKELIHOOD FOR NON-STATIONARY PANEL DATA MODELS WITH FIXED EFFECTS 

Geert Dhaene* Koen Jochmans ${ }^{\dagger}$<br>K.U. Leuven<br>CORE

First version: March 2007; This version: March 22, 2011


#### Abstract

We calculate the bias of the profile score for the autoregressive parameters $\rho$ and covariate slopes $\beta$ in the linear model for $N \times T$ panel data with $p$ lags of the dependent variable, exogenous covariates, fixed effects, and unrestricted initial observations. The bias is a vector of multivariate polynomials in $\rho$ with coefficients that depend only on $T$. We center the profile score and, on integration, obtain an adjusted profile likelihood. When $p=1$, the adjusted profile likelihood coincides with Lancaster's (2002) marginal posterior. More generally, it is an integrated likelihood, in the sense of Arellano and Bonhomme (2009), with fixed effects integrated out using a new data-independent prior. It appears that $\rho$ and $\beta$ are identified as the unique point where the large $N$ adjusted profile likelihood reaches a local maximum (or a flat inflection point, as a limiting case) inside or on an ellipsoid centered at the maximum likelihood estimator. We prove this when $p=1$ and report numerical calculations that support it when $p>1$. The global maximum of the adjusted profile likelihood lies at infinity for any $N$.


JEL classification: C13, C22, C23
Keywords: adjusted likelihood, bias correction, dynamic panel data, fixed effects, non-stationarity.

## Introduction

Maximum likelihood estimates of dynamic linear fixed effect models from short panels are known to be severely biased and inconsistent because of the incidental parameter problem first described by Neyman and Scott (1948); see Nerlove $(1967,1971)$ and Nickell (1981), for example. The failure of maximum likelihood has shifted interest towards GMM-based estimation in these models. An overview and discussion is provided by Arellano (2003, Chapter 6).

In an interesting paper, Lancaster (2002) argued for a return to likelihood-based inference and suggested a Bayesian resolution of the incidental parameter problem in the non-stationary AR(1) model. ${ }^{1}$ The fixed effects are first orthogonalized (in the information sense) to the

[^0]common parameters and then integrated from the likelihood using uniform priors. The first order condition for the posterior mode is unbiased. Unfortunately, orthogonalization is not possible when the $\mathrm{AR}(1)$ model is augmented with covariates, as Lancaster showed, or when the autoregressive order is greater than one, as we show. Lancaster noted, however, that incomplete orthogonalization suffices in the $\operatorname{AR}(1)$ model with covariates. Arellano and Bonhomme (2009) considered integrating possibly non-orthogonalized fixed effects from the likelihood using more general priors, generalizing Lancaster (2002). Their characterization of bias-reducing priors in a general class of models (including nonlinear ones) further shows that orthogonality is sufficient but not necessary for being able to improve on maximum likelihood.

In this paper, we deal with the incidental parameter problem in the non-stationary $\operatorname{AR}(p)$ model with covariates by adjusting the profile likelihood. Our approach is frequentist and does not require orthogonalization or finding suitable priors. The adjustment is based on a bias calculation, and subsequent centering, of the profile score, similar in spirit to McCullagh and Tibshirani (1990). ${ }^{2}$ An intermediate result of our analysis is a simple unbiased estimating equation for the $\operatorname{AR}(p)$ model. The adjustment term to the profile log-likelihood that results from it is a multivariate polynomial in the autoregressive parameters with coefficients depending only on the number of time periods. Thus, the adjustment to the profile likelihood is independent of the data and the true parameter values.

When $p=1$, the adjusted profile likelihood coincides with the marginal posterior obtained by Lancaster (2002). Similarly, when $p>1$, it coincides with the marginal posterior obtained by integrating out suitably reparameterized fixed effects using a uniform prior. This view relates the adjusted likelihood to Lancaster's approach, although the transformed fixed effects are no longer information orthogonal when $p>1$ or when covariates are present. We show that the adjusted profile likelihood is also an integrated likelihood in the sense of Arellano and Bonhomme (2009). That is, for a specific choice of prior, the integrated likelihood coincides with the adjusted profile likelihood. From the integrated likelihood point of view, what we find is a data-independent prior for the $\mathrm{AR}(p)$ model that fully eliminates the bias from the first order condition. Such a prior was thought not to exist when $p>1$. We also show that centering the profile score is equivalent to inverting a plug-in estimate of the probability limit of the within-group estimator, complementing and generalizing Bun and Carree (2005).

Recently, Chamberlain and Moreira (2009) and Moreira (2009) suggested the use of invariance arguments to eliminate the fixed effects from the likelihood. It would be interesting if a connection could be established between their invariant likelihood and the adjusted profile likelihood, but we doubt that this is possible in the present model. In the simplest case, the

[^1]AR(1) model without covariates, Moreira's (2009) maximal invariant likelihood differs from the adjusted profile likelihood.

Perhaps our main (and unexpected) finding is that one should look for a local maximizer of the adjusted likelihood/marginal posterior because the global maximum is reached at infinity. The reason for this is that the profile log-likelihood is log-quadratic in the parameters while the adjustment term is a polynomial, with positive coefficients, in the autoregressive parameters. Thus, for large enough autoregressive parameter values, the adjustment term dominates, and the adjusted profile log-likelihood must be re-increasing. This finding complements observations by Lancaster (2002), who noticed in Monte Carlo experiments that his marginal posterior was sometimes re-increasing. This phenomenon is not just a small sample problem or an artifact of an unbounded parameter space. The adjusted likelihood/marginal posterior has its global maximum at infinity for any sample size, and it may already be re-increasing in the stationary parameter region and reach its maximum at the boundary. Consequently, identification is not achieved by global maximization, and estimation based solely on the first order condition is incomplete.

To examine identification, we profile out the covariate slopes, which are unique given the autoregressive parameters, so the analysis reduces to that of the model without covariates.

In the $\operatorname{AR}(1)$ model, we show that the autoregressive parameter is identified as the unique local maximizer - or flat inflection point, as a limiting case - of the (large sample) adjusted profile likelihood in the interval where the unadjusted profile score is downward sloping. The (large sample) maximum likelihood estimate is the midpoint of this interval. Identification as a flat inflection point is an instance of first order lack of identification, in Sargan's (1983) terminology. The leading case of interest is a unit root without deterministic trends, which is identified as a flat inflection point regardless of the number of time periods, in line with the rank deficiency of the expected Jacobian associated with the moment conditions of Ahn and Schmidt (1995); see Alvarez and Arellano (2004). For the unit root case the flat inflection point property has been noted independently by Ahn and Thomas (2006).

Identification in the $\operatorname{AR}(p)$ model via the adjusted likelihood is a harder problem. It appears that the autoregressive parameter vector is identified as the unique local maximizer or flat inflection point of the adjusted profile likelihood on a region bounded by a well-identified ellipsoid that is centered at the maximum likelihood estimate. We have no proof of this conjecture, but we report numerical calculations that support it.

The model parameters can be estimated by the local maximizer of the adjusted likelihood over a plug-in estimate of the ellipsoid (or, when no local maximizer exists, by the minimizer of the norm of the adjusted profile score). When the parameters are identifed as local maximizers, the estimator converges at the parametric rate, and its limit distribution is normal and correctly centered.

In Section 1, we present the model and derive the bias of the profile score. The adjusted likelihood is given in Section 2, where we also connect with Lancaster (2002), Arellano and Bonhomme (2009), and Bun and Carree (2005). Global properties of the adjusted likelihood are discussed in Section 3. In Section 4, we define the adjusted likelihood estimator and derive its main asymptotic properties. Section 5 discusses some simulation results. Technical proofs are collected in an appendix.

## 1 Bias of the profile score

### 1.1 Model and profile likelihood

Suppose we observe a scalar variable $y$, the first $p \geq 1$ lags of $y$, and a $q$-vector of covariates $x$ (which may include lags), for $N$ units $i$ and $T$ periods $t$. Assume that $y_{i t}$ is generated by

$$
\begin{equation*}
y_{i t}=y_{i t-}^{\prime} \rho+x_{i t}^{\prime} \beta+\alpha_{i}+\varepsilon_{i t}, \quad i=1, \ldots, N ; t=1, \ldots, T, \tag{1.1}
\end{equation*}
$$

where $y_{i t-}=\left(y_{i t-1}, \ldots, y_{i t-p}\right)^{\prime}$ and the $\varepsilon_{i t}$ are identically distributed with mean 0 and variance $\sigma^{2}$ and are independent across $i$ and $t$ and also of $x_{i^{\prime} t^{\prime}}$ for all $i^{\prime}$ and $t^{\prime}$. Let $y_{i}^{0}=\left(y_{i(1-p)}, \ldots, y_{i 0}\right)^{\prime}$, $X_{i}=\left(x_{i 1}, \ldots, x_{i T}\right)^{\prime}$, and $\varepsilon_{i}=\left(\varepsilon_{i 1}, \ldots, \varepsilon_{i T}\right)^{\prime}$. We make no assumptions about how $\left(y_{i}^{0}, \alpha_{i}, X_{i}\right), i=$ $1, \ldots, N$, are generated, except for the assumption that the covariates are not linearly dependent. The unknown parameters are $\theta=\left(\rho^{\prime}, \beta^{\prime}\right)^{\prime}, \sigma^{2}$, and $\alpha_{1}, \ldots, \alpha_{N}$. Let $\theta_{0}$ and $\sigma_{0}^{2}$ be the true values of $\theta$ and $\sigma^{2}$. Our interest lies in consistently estimating $\theta_{0}$ under large $N$ and fixed $T$ asymptotics. We do not impose the stationarity condition on $\rho_{0}$, i.e., we allow any $\rho_{0} \in \mathbb{R}^{p}$.

Let $z_{i t}=\left(y_{i t-}^{\prime}, x_{i t}^{\prime}\right)^{\prime}, Y_{i-}=\left(y_{i 1-}, \ldots, y_{i T-}\right)^{\prime}, Z_{i}=\left(Y_{i-}, X_{i}\right)$, and $y_{i}=\left(y_{i 1}, \ldots, y_{i T}\right)^{\prime}$, so that $M y_{i}=M Z_{i} \theta+M \varepsilon_{i}$ where $M=I_{T}-T^{-1} \iota \iota^{\prime}$ and $\iota$ is a conformable vector of ones. We assume that $N^{-1} \sum_{i=1}^{N} Z_{i}^{\prime} M Z_{i}$ and its probability limit as $N \rightarrow \infty$ are nonsingular. The Gaussian quasi-log-likelihood, conditional on $y_{1}^{0}, \ldots, y_{N}^{0}$ and normalized by the number of observations, is given by

$$
-\frac{1}{2 N T} \sum_{i=1}^{N} \sum_{t=1}^{T}\left(\log \sigma^{2}+\frac{1}{\sigma^{2}}\left(y_{i t}-z_{i t}^{\prime} \theta-\alpha_{i}\right)^{2}\right)+\mathrm{c},
$$

where, here and later, c is a non-essential constant. Replacing the nuisance parameters $\alpha_{1}, \ldots, \alpha_{N}$ and $\sigma^{2}$ with their MLE for given $\theta$ gives the (normalized) profile log-likelihood,

$$
l(\theta)=-\frac{1}{2} \log \left(\frac{1}{N} \sum_{i=1}^{N}\left(y_{i}-Z_{i} \theta\right)^{\prime} M\left(y_{i}-Z_{i} \theta\right)\right)+\mathrm{c} .
$$

The profile score, $s(\theta)=\nabla_{\theta} l(\theta)$, has elements

$$
\begin{aligned}
& s_{\rho_{j}}(\theta)=\frac{\sum_{i=1}^{N}\left(y_{i}-Z_{i} \theta\right)^{\prime} M y_{i,-j}}{\sum_{i=1}^{N}\left(y_{i}-Z_{i} \theta\right)^{\prime} M\left(y_{i}-Z_{i} \theta\right)}, \quad j=1, \ldots, p, \\
& s_{\beta_{j}}(\theta)=\frac{\sum_{i=1}^{N}\left(y_{i}-Z_{i} \theta\right)^{\prime} M x_{i, j}}{\sum_{i=1}^{N}\left(y_{i}-Z_{i} \theta\right)^{\prime} M\left(y_{i}-Z_{i} \theta\right)}, \quad j=1, \ldots, q,
\end{aligned}
$$

where $y_{i,-j}$ is the $j$ th column of $Y_{i-}$ and $x_{i, j}$ is the $j$ th column of $X_{i}$.

### 1.2 Bias of the profile score

It is well known that the profile score is asymptotically biased, that is,

$$
\operatorname{plim}_{N \rightarrow \infty} s\left(\theta_{0}\right) \neq 0,
$$

where, here and later, probability limits and expectations are taken conditionally, given ( $y_{i}^{0}, \alpha_{i}, X_{i}$ ), $i=1, \ldots, N$. Hence, $\theta_{0} \neq \arg \max _{\theta} \operatorname{plim}_{N \rightarrow \infty} l(\theta)$ and the MLE, solving $s(\theta)=0$, is inconsistent.

The following calculation yields the bias of the profile score, either asymptotically or, when $\varepsilon_{i t} \sim \mathcal{N}\left(0, \sigma^{2}\right)$, for fixed $N$. Rewrite (1.1) as

$$
D y_{i}=C y_{i}^{0}+X_{i} \beta+\iota \alpha_{i}+\varepsilon_{i}, \quad i=1, \ldots, N
$$

where $D=D(\rho)$ and $C=C(\rho)$ are the $T \times T$ and $T \times p$ matrices

$$
D=\left(\begin{array}{ccccccc}
1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
-\rho_{1} & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
-\rho_{p} & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -\rho_{p} & \cdots & -\rho_{1} & 1
\end{array}\right), \quad C=\left(\begin{array}{ccccc}
\rho_{p} & \cdots & \cdots & \cdots & \rho_{1} \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & \rho_{p} \\
0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & 0
\end{array}\right) .
$$

Then

$$
\binom{y_{i}^{0}}{y_{i}}=\xi_{i}+F \varepsilon_{i}, \quad \xi_{i}=\binom{y_{i}^{0}}{D^{-1}\left(C y_{i}^{0}+X_{i} \beta+\iota \alpha_{i}\right)}, \quad F=\binom{0}{D^{-1}}
$$

and $y_{i,-j}=S_{j}\left(\xi_{i}+F \varepsilon_{i}\right)$, where $S_{j}=\left(0_{T \times(p-j)}: I_{T}: 0_{T \times j}\right)$, a selection matrix. Therefore,

$$
\begin{aligned}
\operatorname{plim}_{N \rightarrow \infty} s_{\rho_{j}}\left(\theta_{0}\right) & =\frac{\operatorname{pim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^{N} \varepsilon_{i}^{\prime} M S_{j}\left(\xi_{0 i}+F_{0} \varepsilon_{i}\right)}{\operatorname{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^{N} \varepsilon_{i}^{\prime} M \varepsilon_{i}}=\frac{\mathbb{E}\left(\varepsilon_{i}^{\prime} M S_{j} F_{0} \varepsilon_{i}\right)}{\mathbb{E}\left(\varepsilon_{i}^{\prime} M \varepsilon_{i}\right)} \\
& =\frac{\operatorname{tr} M S_{j} F_{0}}{T-1}, \\
\operatorname{plim}_{N \rightarrow \infty} s_{\beta_{j}}\left(\theta_{0}\right) & =0,
\end{aligned}
$$

where $\xi_{0 i}$ and $F_{0}$ are $\xi_{i}$ and $F$, evaluated at $\theta_{0}$. If, in addition, $\varepsilon_{i t} \sim \mathcal{N}\left(0, \sigma^{2}\right)$, then

$$
\begin{align*}
\mathbb{E}\left[s_{\rho_{j}}\left(\theta_{0}\right)\right] & =\mathbb{E}\left(\frac{\sum_{i=1}^{N} \varepsilon_{i}^{\prime} M S_{j}\left(\xi_{0 i}+F_{0} \varepsilon_{i}\right)}{\sum_{i=1}^{N} \varepsilon_{i}^{\prime} M \varepsilon_{i}}\right)=\mathbb{E}\left(\frac{\sum_{i=1}^{N} \varepsilon_{i}^{\prime} M S_{j} F_{0} M \varepsilon_{i}}{\sum_{i=1}^{N} \varepsilon_{i}^{\prime} M \varepsilon_{i}}\right) \\
& =\frac{\mathbb{E}\left(\varepsilon_{i}^{\prime} M S_{j} F_{0} M \varepsilon_{i}\right)}{\mathbb{E}\left(\varepsilon_{i}^{\prime} M \varepsilon_{i}\right)}=\frac{\operatorname{tr} M S_{j} F_{0}}{T-1}  \tag{1.2}\\
\mathbb{E}\left[s_{\beta_{j}}\left(\theta_{0}\right)\right] & =0
\end{align*}
$$

by well-known properties of the normal distribution and the following geometric argument, which goes back to Fisher (1930) and Geary (1933). Let $v \sim \mathcal{N}\left(0, \sigma^{2} I_{g}\right)$ and let $Q$ be a $g \times h$ matrix such that $Q^{\prime} Q=I_{h}$, so $Q Q^{\prime}$ is idempotent. Write $I_{g}-Q Q^{\prime}$ as $P P^{\prime}$, where $P^{\prime} P=I_{g-h}$. Transform $v$ into $m=P^{\prime} v$, the radius $r=\left(v^{\prime} Q Q^{\prime} v\right)^{1 / 2}$, and the $h-1$ polar angles $a$ of $Q^{\prime} v$. Then the elements of $\left(m^{\prime}, r, a^{\prime}\right)^{\prime}$ are independent. Therefore, for any $g \times g$ matrix $W$, if $A=v^{\prime} Q Q^{\prime} W Q Q^{\prime} v$ and $B=v^{\prime} Q Q^{\prime} v$, then the ratio $A / B$ depends on $v$ only through $a$ and hence is independent of $B$, which implies that $\mathbb{E}(A)=\mathbb{E}(A / B) \mathbb{E}(B)$ and $\mathbb{E}(A / B)=\mathbb{E}(A) / \mathbb{E}(B) .^{3}$ The transition to (1.2) now follows from applying this property to the ratio

$$
\frac{\sum_{i=1}^{N} \varepsilon_{i}^{\prime} M S_{j} F_{0} M \varepsilon_{i}}{\sum_{i=1}^{N} \varepsilon_{i}^{\prime} M \varepsilon_{i}}=\frac{\varepsilon^{\prime}\left(I_{N} \otimes M\right)\left(I_{N} \otimes S_{j} F_{0}\right)\left(I_{N} \otimes M\right) \varepsilon}{\varepsilon^{\prime}\left(I_{N} \otimes M\right) \varepsilon}
$$

with $v=\varepsilon=\left(\varepsilon_{1}^{\prime}, \ldots, \varepsilon_{N}^{\prime}\right)^{\prime}, Q Q^{\prime}=I_{N} \otimes M$, and $W=I_{N} \otimes S_{j} F_{0}$.
To summarize the results so far regarding the bias of the profile score, let

$$
b(\rho)=\left(b_{1}(\rho), \ldots, b_{p+q}(\rho)\right)^{\prime}, \quad b_{j}(\rho)= \begin{cases}\frac{\operatorname{tr} M S_{j} F}{T-1}, & j=1, \ldots, p \\ 0, & j=p+1, \ldots, p+q\end{cases}
$$

Then $\operatorname{plim}_{N \rightarrow \infty} s\left(\theta_{0}\right)=b\left(\rho_{0}\right)$; if, in addition, $\varepsilon_{i t} \sim \mathcal{N}\left(0, \sigma_{0}^{2}\right)$, then $\mathbb{E}\left[s\left(\theta_{0}\right)\right]=b\left(\rho_{0}\right)$.
Two remarks are worth making. First, the bias of the profile score, $b\left(\rho_{0}\right)$, depends only on $\rho_{0}$ and $T$. It is independent of the initial observations, the fixed effects, and the covariates (recall that the expectations are conditioning on these). This is in sharp contrast with the bias of the ML (the within-group) estimator, which was first derived by Nickell (1981) for the first order autoregressive model under the assumption of stationarity of the initial observations. The bias of the MLE depends on the initial observations, the fixed effects, and the covariate values; it is smaller when the initial observations are more outlying relative to the stationary distributions. Second, the Nickell bias concerns a probability limit as $N \rightarrow \infty$ whereas here, when the errors are normal, $\mathbb{E}\left[s\left(\theta_{0}\right)\right]=b\left(\rho_{0}\right)$ is a finite sample result holding for fixed $N$ and $T$ and may therefore be of independent interest in a time series setting.

[^2]For $k=\left(k_{1}, \ldots, k_{p}\right)^{\prime} \in \mathbb{N}^{p}$, let $\rho^{k}=\prod_{j=1}^{p} \rho_{j}^{k_{j}}$, and let $\tau=(1, \ldots, p)^{\prime}$. It is shown in the Appendix that

$$
D^{-1}=\left(\begin{array}{cccc}
1 & 0 & \cdots & 0  \tag{1.3}\\
\varphi_{1} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
\varphi_{T-1} & \cdots & \varphi_{1} & 1
\end{array}\right), \quad \varphi_{t}=\sum_{\tau^{\prime} k=t} \frac{\left(\iota^{\prime} k\right)!}{k_{1}!\cdots k_{p}!} \rho^{k}
$$

where the summation is over all $k \in \mathbb{N}^{p}$ satisfying $\tau^{\prime} k=t$. Therefore, since

$$
S_{j} F=\left(\begin{array}{cc}
0 & 0 \\
D_{j}^{-1} & 0
\end{array}\right)
$$

where $D_{j}^{-1}$ is the leading $(T-j) \times(T-j)$ block of $D^{-1}$, we obtain

$$
b_{j}(\rho)=-\frac{\iota^{\prime} D_{j}^{-1} \iota}{T(T-1)}=-\sum_{t=0}^{T-j-1} \frac{T-j-t}{T(T-1)} \varphi_{t}, \quad j=1, \ldots, p
$$

where $\varphi_{0}=1$. Hence, for $j=1, \ldots, p, b_{j}(\rho)$ is a multivariate polynomial of degree $T-j-1$ in $\rho_{1}, \ldots, \rho_{p}$ with non-positive coefficients.

## 2 Adjusted profile likelihood

### 2.1 Centered profile score and adjusted profile likelihood

By construction, the centered (or adjusted) profile score,

$$
s_{\mathrm{A}}(\theta)=s(\theta)-b(\rho),
$$

is asymptotically unbiased, i.e., $\operatorname{plim}_{N \rightarrow \infty} s_{\mathrm{A}}\left(\theta_{0}\right)=0$. In the Appendix, we show that the differential equation $\nabla_{\theta} a(\rho)=b(\rho)$ has a solution which, up to an arbitrary constant of integration, is

$$
\begin{equation*}
a(\rho)=\sum_{S \in \mathcal{S}} a_{S}(\rho), \quad a_{S}(\rho)=-\sum_{t=|S|}^{T-1} \frac{T-t}{T(T-1)} \sum_{k \in \mathcal{K}_{S}: \tau^{\prime} k=t} \frac{\left(\iota^{\prime} k-1\right)!}{k_{1}!\cdots k_{p}!} \rho_{S}^{k_{S}} \tag{2.1}
\end{equation*}
$$

where $\mathcal{S}$ is the collection of the non-empty subsets of $\{1, \ldots, p\} ;|S|$ is the sum of the elements of $S ; \mathcal{K}_{S}=\left\{k \in \mathbb{N}^{p} \mid k_{j}>0\right.$ iff $\left.j \in S\right\} ;$ and $\rho_{S}=\left(\rho_{j}\right)_{j \in S}$ and $k_{S}=\left(k_{j}\right)_{j \in S}$ are subvectors of $\rho$ and $k$ determined by $S$.

We call the function

$$
l_{\mathrm{A}}(\theta)=l(\theta)-a(\rho),
$$

which has derivative $\nabla_{\theta} l_{\mathrm{A}}(\theta)=s_{\mathrm{A}}(\theta)$, an adjusted profile log-likelihood. Every subvector $\rho_{S}$ of $\rho$ contributes to $l_{\mathrm{A}}(\theta)$ an adjustment term, $-a_{S}(\rho)$, which takes the form of a multivariate polynomial in $\rho_{j}, j \in S$, with positive coefficients that are independent of $p .^{4,5}$

### 2.2 Connections with the literature

Lancaster (2002) studied the $\mathrm{AR}(1)$ model, with and without covariates, from a Bayesian perspective. It turns out that $e^{l_{\mathrm{A}}(\theta)}$ coincides, up to a non-essential normalization, with Lancaster's posterior density for $\theta$, thereby giving it a frequentist interpretation. With $p=1$, we have $\varphi_{t}=\rho^{t}$ and

$$
b_{1}(\rho)=-\sum_{t=1}^{T-1} \frac{T-t}{T(T-1)} \rho^{t-1}, \quad a(\rho)=-\sum_{t=1}^{T-1} \frac{T-t}{T(T-1) t} \rho^{t} .
$$

With independent uniform priors on the reparameterized effects $\eta_{i}=\alpha_{i} e^{-(T-1) a(\rho)}$ and on $\theta$ and $\log \sigma^{2}$, Lancaster's posterior density for $\vartheta=\left(\theta^{\prime}, \sigma^{2}\right)^{\prime}$ is

$$
f(\vartheta \mid d a t a) \propto \sigma^{-N(T-1)-2} \exp \left(-N(T-1) a(\rho)-Q^{2}(\theta) \sigma^{-2} / 2\right)
$$

where $Q^{2}(\theta)=\sum_{i=1}^{N}\left(y_{i}-Z_{i} \theta\right)^{\prime} M\left(y_{i}-Z_{i} \theta\right) \propto e^{-2 l(\theta)}$. Integrating over $\sigma^{2}$ gives $f(\theta \mid$ data $) \propto$ $e^{-N(T-1) a(\rho)}\left(Q^{2}(\theta)\right)^{-N(T-1) / 2}$ and, hence,

$$
\begin{equation*}
f(\theta \mid \text { data }) \propto e^{N(T-1) l_{\mathrm{A}}(\theta)} \tag{2.2}
\end{equation*}
$$

Thus, the posterior density and the adjusted likelihood for $\theta$ are equivalent as a basis for inference. More generally, for any $p$ and $q$, independent uniform priors on $\eta_{1}, \ldots, \eta_{N}, \theta, \log \sigma^{2}$, with $\eta_{i}=$ $\alpha_{i} e^{-(T-1) a(\rho)}$ and $a(\rho)$ as in (2.1), yield a posterior $f\left(\theta \mid\right.$ data) that is related to $l_{\mathrm{A}}(\theta)$ as in (2.2).

Lancaster's choice of independent uniform priors on $\eta_{i}$ has its foundation in the $\operatorname{AR}(1)$ model without covariates, where $\eta_{i}$ is information orthogonal to $\vartheta$ and the posterior density (hence also $e^{l_{\mathrm{A}}(\theta)}$ ) has an interpretation as a Cox and Reid (1987) approximate conditional likelihood (see also Arellano, 2003, pp. 103-106). Information orthogonalization to a multidimensional parameter is generally not possible (see, e.g., Severini, 2000, pp. 340-342). Here, orthogonalization is not possible when the $\mathrm{AR}(1)$ model is augmented with covariates, as shown by Lancaster, or when the autoregressive order, $p$, is greater than one, as shown in the Appendix. From a bias correction perspective, however, information orthogonality is sufficient but not necessary. For any $p$ and $q$, $s_{\mathrm{A}}(\theta)=0$ is an unbiased estimating equation, and the bias calculation underlying it is immune to the non-existence of orthogonalized fixed effects.

[^3]Arellano and Bonhomme's (2009) approach shares the integration step with Lancaster (2002) but allows non-uniform priors on fixed effects or, equivalently, non-orthogonalized fixed effects. Of interest are bias-reducing priors, i.e., weighting schemes that deliver an integrated likelihood whose score equation has bias $o\left(T^{-1}\right)$ as opposed to the standard $O\left(T^{-1}\right)$. The linear dynamic model (with general $p, q$ ) illustrates an interesting result of Arellano and Bonhomme that generalizes the scope of uniform integration to situations where orthogonalization is impossible. For a given prior $\pi_{i}\left(\alpha_{i} \mid \vartheta\right)$, the (normalized) log integrated likelihood is

$$
l^{I}(\vartheta)=\frac{1}{N T} \sum_{i=1}^{N} \log \int \sigma^{-T / 2} \exp \left(-\frac{1}{2 \sigma^{2}} \sum_{t=1}^{T}\left(y_{i t}-z_{i t}^{\prime} \theta-\alpha_{i}\right)^{2}\right) \pi_{i}\left(\alpha_{i} \mid \vartheta\right) \mathrm{d} \alpha_{i}+\mathrm{c} .
$$

Choosing $\pi_{i}\left(\alpha_{i} \mid \vartheta\right) \propto e^{-(T-1) a(\rho)}$ yields

$$
l^{I}(\vartheta)=-\frac{T-1}{2 T} \log \sigma^{2}-\frac{T-1}{T} a(\rho)-\frac{Q^{2}(\theta)}{2 N T \sigma^{2}}+\mathrm{c}
$$

Profiling out $\sigma^{2}$ gives $\sigma^{2}(\theta)=\arg \max _{\sigma^{2}} l^{I}(\vartheta)=Q^{2}(\theta) /(N(T-1))$ and

$$
l^{I}(\theta)=\max _{\sigma^{2}} l^{I}(\vartheta)=\frac{T-1}{T} l_{\mathrm{A}}(\theta)+\mathrm{c},
$$

so $l^{I}(\theta)$ and $l_{\mathrm{A}}(\theta)$ are equivalent. Because $a(\rho)$ does not depend on true parameter values, $\pi_{i}\left(\alpha_{i} \mid \vartheta\right) \propto e^{-(T-1) a(\rho)}$ is a data-independent bias-reducing (in fact, bias-eliminating) prior in the sense of Arellano and Bonhomme. ${ }^{6}$ Now, $\pi_{i}\left(\alpha_{i} \mid \vartheta\right) \propto e^{-(T-1) a(\rho)}$ is equivalent to $\pi_{i}\left(\eta_{i} \mid \vartheta\right) \propto 1$, i.e., to a uniform prior on $\eta_{i}=\alpha_{i} e^{-(T-1) a(\rho)}$, leading to the same $l^{I}(\vartheta)$. Arellano and Bonhomme (2009, Eq. (11)) give a necessary and sufficient condition for a uniform prior to be bias-reducing. With $\ell_{i}\left(\vartheta, \eta_{i}\right)$ denoting $i$ 's log-likelihood contribution in a parametrization $\eta_{i}$, the condition is that

$$
\begin{equation*}
\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \nabla_{\eta_{i}}\left(A_{i}^{-1} B_{i}\right)=o(1) \quad \text { as } T \rightarrow \infty \tag{2.3}
\end{equation*}
$$

where $A_{i}=A_{i}\left(\vartheta, \eta_{i}\right)=-\mathbb{E}_{\vartheta, \eta_{i}} \nabla_{\eta_{i} \eta_{i}} \ell_{i}\left(\vartheta, \eta_{i}\right), B_{i}=B_{i}\left(\vartheta, \eta_{i}\right)=\mathbb{E}_{\vartheta, \eta_{i}} \nabla_{\vartheta \eta_{i}} \ell_{i}\left(\vartheta, \eta_{i}\right)$, and $\nabla_{\eta_{i}}\left(A_{i}^{-1} B_{i}\right)$ is evaluated at the true parameter values. When $\eta_{i}$ and $\vartheta$ are information orthogonal, $B_{i}=0$ and (2.3) holds. However, condition (2.3) is considerably weaker than parameter orthogonality. In the present model, when $p>1$ or $q>0$, and thus no orthogonalization is possible, it follows from our analysis and Arellano and Bonhomme (2009) that (2.3) must hold for $\eta_{i}=\alpha_{i} e^{-(T-1) a(\rho)}$. Indeed, as we show in the Appendix,

$$
\begin{equation*}
\nabla_{\eta_{i}}\left(A_{i}^{-1} B_{i}\right)=0 \tag{2.4}
\end{equation*}
$$

${ }^{6}$ Our focus is on estimating $\theta_{0}$ but, noting the degrees-of-freedom correction in $\sigma^{2}(\theta)$, the estimating equation derived from $l^{I}(\vartheta)$ is also unbiased for $\sigma_{0}^{2}$.
because $A_{i}^{-1} B_{i}$ is free of $\eta_{i}$.
Finally, the centered profile score relates directly to Bun and Carree (2005). Observe that $s(\theta)=\sum_{i=1}^{N} Z_{i}^{\prime} M\left(y_{i}-Z_{i} \theta\right) / Q^{2}(\theta)$. Further, $M y_{i}=M Z_{i} \widehat{\theta}+M \widehat{\varepsilon}_{i}$ where $\widehat{\theta}$ is the within-group estimator, with residuals $\widehat{\varepsilon}_{i}$ satisfying $\sum_{i=1}^{N} Z_{i}^{\prime} M \widehat{\varepsilon}_{i}=0$. Therefore, solving $s_{\mathrm{A}}(\theta)=0$ is equivalent to solving

$$
\begin{equation*}
\widehat{\theta}-\theta=\left(\sum_{i=1}^{N} Z_{i}^{\prime} M Z_{i}\right)^{-1} b(\rho) Q^{2}(\theta) \tag{2.5}
\end{equation*}
$$

In the $\mathrm{AR}(1)$ model, (2.5) corresponds to Bun and Carree's (2005) proposal for bias-correcting the within-group estimator.

## 3 Global properties of the adjusted profile likelihood when $N$ is large

At this point it is tempting to anticipate that $\theta_{0}$ maximizes $\operatorname{plim}_{N \rightarrow \infty} l_{\mathrm{A}}(\theta)$. However, as the following analysis shows, $-a(\rho)$ dominates $\operatorname{plim}_{N \rightarrow \infty} l(\theta)$ as $\|\rho\| \rightarrow \infty$ in almost all directions and, as a result, $\operatorname{plim}_{N \rightarrow \infty} l_{\mathrm{A}}(\theta)$ is unbounded from above.

Let $h(\theta)=\nabla_{\theta^{\prime}} s(\theta), c(\rho)=\nabla_{\theta^{\prime}} b(\rho)$, and

$$
\begin{array}{ll}
L_{\mathrm{A}}(\theta)=L(\theta)-a(\rho), & L(\theta)=\operatorname{pim}_{N \rightarrow \infty} l(\theta) \\
S_{\mathrm{A}}(\theta)=S(\theta)-b(\rho), & S(\theta)=\operatorname{pim}_{N \rightarrow \infty} s(\theta) \\
H_{\mathrm{A}}(\theta)=H(\theta)-c(\rho), & H(\theta)=\operatorname{plim}_{N \rightarrow \infty} h(\theta)
\end{array}
$$

Using $M\left(y_{i}-Z_{i} \theta\right)=-M Z_{i}\left(\theta-\theta_{0}\right)+M \varepsilon_{i}$, we have

$$
L(\theta)=-\frac{1}{2} \log \left(\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N}\left(\varepsilon_{i}^{\prime} M \varepsilon_{i}-2\left(\theta-\theta_{0}\right)^{\prime} Z_{i}^{\prime} M \varepsilon_{i}+\left(\theta-\theta_{0}\right)^{\prime} Z_{i}^{\prime} M Z_{i}\left(\theta-\theta_{0}\right)\right)\right)+\mathrm{c}
$$

Let $b_{0}=b\left(\rho_{0}\right)=S\left(\theta_{0}\right)$ and note that

$$
\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} Z_{i}^{\prime} M \varepsilon_{i}=\left(\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i}^{\prime} M \varepsilon_{i}\right) b_{0}=\sigma_{0}^{2}(T-1) b_{0}
$$

Hence, defining $V_{0}=V\left(\theta_{0}\right)$ by

$$
\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} Z_{i}^{\prime} M Z_{i}=\left(\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i}^{\prime} M \varepsilon_{i}\right) V_{0}=\sigma_{0}^{2}(T-1) V_{0}
$$

we can write

$$
L(\theta)=-\frac{1}{2} \log \left(1-2\left(\theta-\theta_{0}\right)^{\prime} b_{0}+\left(\theta-\theta_{0}\right)^{\prime} V_{0}\left(\theta-\theta_{0}\right)\right)+\mathrm{c}
$$

by absorbing the term $-\frac{1}{2} \log \left(\sigma_{0}^{2}(T-1)\right)$ into c. As $N \rightarrow \infty$, the ML estimator of $\theta$ converges in probability to $\theta_{M L}=\arg \max _{\theta} L(\theta)=\theta_{0}+V_{0}^{-1} b_{0}$ and has asymptotic bias $V_{0}^{-1} b_{0}$. This expression
generalizes the bias calculations in Nickell (1981). Note that $\left(\theta_{0}-\theta_{M L}\right)^{\prime} V_{0}\left(\theta_{0}-\theta_{M L}\right)=b_{0}^{\prime} V_{0}^{-1} b_{0}$. Furthermore,

$$
\begin{aligned}
L(\theta) & =-\frac{1}{2} \log \left(1-b_{0}^{\prime} V_{0}^{-1} b_{0}+\left(\theta-\theta_{M L}\right)^{\prime} V_{0}\left(\theta-\theta_{M L}\right)\right)+\mathrm{c} \\
S(\theta) & =-\frac{V_{0}\left(\theta-\theta_{M L}\right)}{1-b_{0}^{\prime} V_{0}^{-1} b_{0}+\left(\theta-\theta_{M L}\right)^{\prime} V_{0}\left(\theta-\theta_{M L}\right)} \\
H(\theta) & =-\frac{V_{0}}{1-b_{0}^{\prime} V_{0}^{-1} b_{0}+\left(\theta-\theta_{M L}\right)^{\prime} V_{0}\left(\theta-\theta_{M L}\right)}+2 S(\theta) S(\theta)^{\prime}
\end{aligned}
$$

Note that $L(\cdot)$ and $H(\cdot)$ are even and $S(\cdot)$ is odd about $\theta_{M L}$ and that $H\left(\theta_{0}\right)=2 b_{0} b_{0}^{\prime}-V_{0}$ and $H_{\mathrm{A}}\left(\theta_{0}\right)=2 b_{0} b_{0}^{\prime}-V_{0}-c_{0}$, where $c_{0}=c\left(\rho_{0}\right)$. Since $L(\theta)$ is log-quadratic in $\theta$ and $a(\rho)$ is a multivariate polynomial with negative coefficients, $L_{\mathrm{A}}(\theta)=L(\theta)-a(\rho)$ is unbounded from above. For example, if we put $\rho=k r$ with $r$ in the positive orthant of $\mathbb{R}^{p}$ and let $k \rightarrow \infty$, then the term $-a(\rho)$ dominates and $L_{\mathrm{A}}(\theta) \rightarrow \infty .^{7}$ Hence, $\theta_{0} \neq \arg \max _{\theta} L_{\mathrm{A}}(\theta)$ and $\theta_{0}$ has to be identified as a functional of $L_{\mathrm{A}}(\theta)$ other than its maximizer. Because $S_{\mathrm{A}}\left(\theta_{0}\right)=0$, we need to select $\theta_{0}$ from the set of stationary points of $L_{\mathrm{A}}(\theta)$, that is, from the set of zeros of $S_{\mathrm{A}}(\theta)$. In general, this set is not a singleton. Indeed, whenever $\theta_{0}$ is a local maximizer of $L_{\mathrm{A}}(\theta)$ (which will often be the case), $L_{\mathrm{A}}(\theta)$, being smooth and unbounded, must also have at least one local minimum. A moment-based estimation strategy based solely on solving $s_{\mathrm{A}}(\theta)=0$ is, therefore, incomplete.

### 3.1 AR(1) without covariates

Consider the $\operatorname{AR}(1)$ model without covariates, i.e., $p=1$ and $q=0$. Let $\zeta_{0}^{2}=\left(V_{0}-b_{0}^{2}\right) / V_{0}^{2}$. Then,

$$
\begin{aligned}
L(\rho) & =-\frac{1}{2} \log \left(\zeta_{0}^{2}+\left(\rho-\rho_{M L}\right)^{2}\right)+\mathrm{c} \\
S(\rho) & =-\frac{\rho-\rho_{M L}}{\zeta_{0}^{2}+\left(\rho-\rho_{M L}\right)^{2}}, \quad H(\rho)=-\frac{\zeta_{0}^{2}-\left(\rho-\rho_{M L}\right)^{2}}{\left(\zeta_{0}^{2}+\left(\rho-\rho_{M L}\right)^{2}\right)^{2}}
\end{aligned}
$$

by absorbing $-\frac{1}{2} \log V_{0}$ into c. Note that $\zeta_{0}^{2}=-1 / H\left(\rho_{M L}\right)$. Recall that $S(\rho)$ is odd about $\rho_{M L}=\rho_{0}+b_{0} / V_{0}$. The zeros of $H(\rho)$ are $\underline{\rho}=\rho_{M L}-\zeta_{0}$ and $\bar{\rho}=\rho_{M L}+\zeta_{0}$, so $S(\rho)$ decreases on $[\underline{\rho}, \bar{\rho}]$ and increases elsewhere. Note that all of $\underline{\rho}, \bar{\rho}, \rho_{M L}$, and $\zeta_{0}$ are identified by $S(\cdot)$. Further, $\rho_{M L}$ and $\zeta_{0}$ act as location and scale parameters of $S(\cdot)$. For any given $\rho_{0}, \rho_{M L}$ and $\zeta_{0}$ are determined by $V_{0}$. As $V_{0}$ increases, $\left|b_{0} / V_{0}\right|$ and $\zeta_{0}$ decrease, that is, the bias of $\rho_{M L}$ decreases in absolute value, the length of $[\underline{\rho}, \bar{\rho}]$ shrinks, and $S(\rho)$ becomes steeper on $(\underline{\rho}, \bar{\rho})$.

[^4]There is a sharp lower bound on $V_{0}$. From $y_{i,-1}=S_{1}\left(\xi_{0 i}+F_{0} \varepsilon_{i}\right)$ and the independence between $\xi_{0 i}$ and $\varepsilon_{i}$, we obtain

$$
V_{0}=\frac{\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} y_{i,-1}^{\prime} M y_{i,-1}}{\sigma_{0}^{2}(T-1)}=V_{0}^{L B}+V_{\xi \xi}
$$

where

$$
V_{0}^{L B}=\frac{\operatorname{tr} F_{0}^{\prime} S_{1}^{\prime} M S_{1} F_{0}}{T-1}, \quad V_{\xi \xi}=\frac{\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \xi_{0 i}^{\prime} S_{1}^{\prime} M S_{1} \xi_{0 i}}{\sigma_{0}^{2}(T-1)}
$$

So $V_{0} \geq V_{0}^{L B}$ and this lower bound implies an upper bound on $\left|b_{0} / V_{0}\right|$ and on the length of $[\underline{\rho}, \bar{\rho}]$, and a lower bound on the steepness of $S(\rho)$ on $(\underline{\rho}, \bar{\rho})$.

Lemma 1. $V_{0}^{L B}$ is given by

$$
V_{0}^{L B}=\frac{1}{T-1}\left(\sum_{j=0}^{T-2}(T-j-1) \rho_{0}^{2 j}-\frac{1}{T} \sum_{j=0}^{T-2}\left(\sum_{k=0}^{j} \rho_{0}^{k}\right)^{2}\right)
$$

and satisfies (i) $V_{0}^{L B} \geq 2 b_{0}^{2}$; (ii) $V_{0}^{L B} \geq 2 b_{0}^{2}-c_{0}$ with equality if and only if $T=2$ or $\rho_{0}=1$.
By Lemma $1, H\left(\rho_{0}\right)=2 b_{0}^{2}-V_{0} \leq 0$ and, hence,

$$
\left(\bar{\rho}-\rho_{M L}\right)^{2}=\frac{V_{0}-b_{0}^{2}}{V_{0}^{2}} \geq \frac{b_{0}^{2}}{V_{0}^{2}}=\left(\rho_{0}-\rho_{M L}\right)^{2} .
$$

Therefore, $\rho_{0} \in[\underline{\rho}, \bar{\rho}]$. Since $S(\rho)$ is a rational function that vanishes at $\pm \infty$ and $b(\rho)$ is a polynomial, $S_{\mathrm{A}}(\rho)$ has finitely many zeros. Thus, because $S_{\mathrm{A}}\left(\rho_{0}\right)=0$ and, by Lemma 1 , $H_{\mathrm{A}}\left(\rho_{0}\right)=2 b_{0}^{2}-V_{0}-c_{0} \leq 0$, it follows that $L_{\mathrm{A}}(\rho)$ has a local maximum or a flat inflection point at $\rho_{0}$. Our main result for the $\operatorname{AR}(1)$ model without covariates is the uniqueness of such a point in $[\underline{\rho}, \bar{\rho}]$, thereby identifying $\rho_{0}$ as a functional of $L_{\mathrm{A}}(\rho)$.

Theorem 1. $\rho_{0}$ is the unique point in $[\underline{\rho}, \bar{\rho}]$ where $L_{\mathrm{A}}(\rho)$ has a local maximum or a flat inflection point.

Equivalently, $\rho_{0}$ is the unique point in $[\underline{\rho}, \bar{\rho}]$ where $b(\rho)$ approaches $S(\rho)$ from below.
$L_{\mathrm{A}}(\rho)$ has a flat inflection point at $\rho_{0}$ if and only if $V_{0}=V_{0}^{L B}=2 b_{0}^{2}-c_{0}$. The latter equality holds if and only if $T=2$ or $\rho_{0}=1$. The former holds if and only if $V_{\xi \xi}=0$, which requires $M S_{1} \xi_{0 i}$ to be negligibly small for almost all $i$. The elements of $S_{1} \xi_{0 i}$ are $\rho_{0}^{j-1} y_{i}^{0}+\alpha_{i} \sum_{k=1}^{j-1} \rho_{0}^{k-1}$, $j=1, \ldots, T$, so $M S_{1} \xi_{0 i}=0$ if and only if $y_{i}^{0}\left(1-\rho_{0}\right)=\alpha_{i}$. Thus, when $\rho_{0}=1$ and there are no deterministic trends, i.e., $\alpha_{i}=0, L_{\mathrm{A}}(\rho)$ has a flat inflection point at $\rho_{0}$ for any $T$. When $\rho_{0} \neq 1, V_{0}=V_{0}^{L B}=2 b_{0}^{2}-c_{0}$ only when $T=2$ and a very strong condition holds on the initial observations and the fixed effects, which is unlikely to hold in situations where a fixed effect modeling approach is called for. Thus, when $\rho_{0} \neq 1$, except in quite special circumstances, $\rho_{0}$ is
the unique point in $[\underline{\rho}, \bar{\rho}]$ where $L_{\mathrm{A}}(\rho)$ attains a strict local maximum. Note that, when $\rho_{0}$ is a local maximizer of $L_{\mathrm{A}}(\rho)$, it need not be the global maximizer on $[\underline{\rho}, \bar{\rho}]$, which may instead be $\bar{\rho}$. To see why this may happen, interpret the situation where $L_{\mathrm{A}}(\rho)$ has a flat inflection point at $\rho_{0}$ as a limiting case of the property that $L_{\mathrm{A}}(\rho)$ is re-increasing.

Figure 1 illustrates how $\rho_{0}$ is identified by $L_{\mathrm{A}}(\rho)$ for two cases, each with $T=4$. The figures on the left correspond to the case $\rho_{0}=.5$ with $V_{0}=V_{0}^{L B}+V_{\xi \xi}$ and $V_{\xi \xi}$ corresponding to stationary initial observations. Those on the right correspond to the unit root case without deterministic trends, i.e., $\rho_{0}=1$ and $V_{0}=V_{0}^{L B}$. In each case, the bottom figures show $S(\rho)$ (solid line) and $b(\rho)$ (dashed line); the top figures show $L(\rho)$ (solid line), $-a(\rho)$ (dashed line), and $L_{\mathrm{A}}(\rho)=L(\rho)-a(\rho)$ (thick line). In all the figures, vertical lines indicate $\underline{\rho}, \rho_{0}$, and $\bar{\rho}$, from left to right. In the case of $\rho_{0}=.5, \rho_{0}$ is the unique local maximizer of $L_{\mathrm{A}}(\bar{\rho})$ on $[\underline{\rho}, \bar{\rho}]$. Note that there is a second solution of $S_{\mathrm{A}}(\rho)=0$ on $[\underline{\rho}, \bar{\rho}]$, which corresponds to a local minimum of $L_{\mathrm{A}}(\rho)$. In the unit root case, $\rho_{0}$ is the unique flat inflection point of $L_{\mathrm{A}}(\rho)$ on $[\underline{\rho}, \bar{\rho}]$.

Figure 1: Identification


Left: $\rho_{0}=0.5$. Right: $\rho_{0}=1$. Bottom: $S(\rho)$ (solid), $b(\rho)$ (dashed). Top: $L(\rho)$ (solid), $-a(\rho)$ (dashed), $L_{\mathrm{A}}(\rho)$ (thick). Vertical lines at $\underline{\rho}, \rho_{0}$, and $\bar{\rho}$.

Note, finally, that the asymptotic bias of the MLE has the same sign as $b_{0}$ because $\rho_{M L}=$
$\rho_{0}+b_{0} / V_{0}$. The proof of Theorem 1 , as a by-product, shows that if $T$ is even, then $b_{0}<0$; and, if $T$ is odd, then $b(\rho)$ decreases and has a unique zero at some point $\rho_{u} \in[-2,-1)$, so $b_{0}$ has the same sign as $\rho_{u}-\rho_{0}$.

### 3.2 AR(1) with covariates

In the $\operatorname{AR}(1)$ model with covariates, profiling out $\beta$ yields a profile likelihood of $\rho$ with essentially the same properties as in the $\operatorname{AR}(1)$ model without covariates. Let $\beta(\rho)=\arg \max _{\beta} L_{\mathrm{A}}(\rho, \beta)$. Clearly, $\beta(\rho)=\arg \max _{\beta} L(\rho, \beta)=\arg \min _{\beta}\left(\theta-\theta_{M L}\right)^{\prime} V_{0}\left(\theta-\theta_{M L}\right)$. Partition $V_{0}, V_{0}^{-1}$, and $b_{0}$ as

$$
V_{0}=\left(\begin{array}{cc}
V_{0 \rho \rho} & V_{0 \rho \beta} \\
V_{0 \beta \rho} & V_{0 \beta \beta}
\end{array}\right), \quad V_{0}^{-1}=\left(\begin{array}{cc}
V_{0}^{\rho \rho} & V_{0}^{\rho \beta} \\
V_{0}^{\beta \rho} & V_{0}^{\beta \beta}
\end{array}\right), \quad b_{0}=\binom{b_{0 \rho}}{0},
$$

where $V_{0 \rho \rho}, V_{0}^{\rho \rho}$, and $b_{0 \rho}$ are scalars. Using $V_{0}^{\rho \rho}=\left(V_{0 \rho \rho}-V_{0 \rho \beta} V_{0 \beta \beta}^{-1} V_{0 \beta \rho}\right)^{-1}$, we have

$$
\begin{aligned}
V_{0 \beta \beta}\left(\beta(\rho)-\beta_{M L}\right) & =-V_{0 \beta \rho}\left(\rho-\rho_{M L}\right) \\
\min _{\beta}\left(\theta-\theta_{M L}\right)^{\prime} V_{0}\left(\theta-\theta_{M L}\right) & =\left(\rho-\rho_{M L}\right)^{2} / V_{0}^{\rho \rho} \\
1-b_{0}^{\prime} V_{0}^{-1} b_{0} & =1-b_{0 \rho}^{2} V_{0}^{\rho \rho} .
\end{aligned}
$$

The first of these equations, together with $V_{0}\left(\theta_{0}-\theta_{M L}\right)=-b_{0}$, yields $\beta\left(\rho_{0}\right)=\beta_{0}$, so $\beta_{0}$ is identified whenever $\rho_{0}$ is. Profiling out $\beta$ from $L(\rho, \beta)$ gives the limiting profile log-likelihood of $\rho$ as

$$
L(\rho)=L(\rho, \beta(\rho))=-\frac{1}{2} \log \left(\zeta_{0}^{2}+\left(\rho-\rho_{M L}\right)^{2}\right)+\mathrm{c}
$$

(slightly abusing notation), where $\zeta_{0}^{2}$ is redefined as $\zeta_{0}^{2}=\left(1-b_{0 \rho}^{2} V_{0}^{\rho \rho}\right) V_{0}^{\rho \rho}$ and $\frac{1}{2} \log V_{0}^{\rho \rho}$ is absorbed into c.

Lemma 2. $\left(V_{0}^{\rho \rho}\right)^{-1} \geq V_{0}^{L B}$, with $V_{0}^{L B}$ as defined earlier and given in Lemma 1.
In view of Lemma 2, we can invoke the result for the $\mathrm{AR}(1)$ model without covariates. If $\underline{\rho}=\rho_{M L}-\zeta_{0}$ and $\bar{\rho}=\rho_{M L}+\zeta_{0}$ with $\zeta_{0}$ redefined as indicated, then $\rho_{0}$ is the unique point in $[\underline{\rho}, \bar{\rho}]$ where $L_{\mathrm{A}}(\rho)=L(\rho)-a(\rho)$ has a local maximum or a flat inflection point. By the proof of Lemma 2, the conditions under which $\rho_{0}$ is a flat inflection point of $L_{\mathrm{A}}(\rho)$ are the same as in the $\operatorname{AR}(1)$ model without covariates. Note, also, that the presence of covariates does not affect the sign of the asymptotic bias of the MLE of $\rho$. Furthermore, it follows from the proof of Lemma 2 that the inclusion of covariates in the model cannot increase $V_{0}^{\rho \rho}$, so the magnitude of $\rho_{M L}-\rho_{0}=V_{0}^{\rho \rho} b_{0 \rho}$ can only decrease relative to the model without covariates. This generalizes an observation of Phillips and Sul (2007) in the stationary AR(1) model, and contrasts with Nickell (1981, p. 1424).

## 3.3 $\operatorname{AR}(p)$

Consider first the $\operatorname{AR}(p)$ model with $p>1$ and without covariates, i.e., $q=0$. Then

$$
\begin{aligned}
L(\rho) & =-\frac{1}{2} \log \left(1+\left(\rho-\rho_{M L}\right)^{\prime} W_{0}\left(\rho-\rho_{M L}\right)\right)+\mathrm{c}, \quad W_{0}=\frac{V_{0}}{1-b_{0}^{\prime} V_{0}^{-1} b_{0}}, \\
S(\rho) & =-\frac{W_{0}\left(\rho-\rho_{M L}\right)}{1+\left(\rho-\rho_{M L}\right)^{\prime} W_{0}\left(\rho-\rho_{M L}\right)}, \\
H(\rho) & =-\frac{W_{0}}{1+\left(\rho-\rho_{M L}\right)^{\prime} W_{0}\left(\rho-\rho_{M L}\right)}+2 S(\rho) S(\rho)^{\prime},
\end{aligned}
$$

where $-\frac{1}{2} \log \left(1-b_{0}^{\prime} V_{0}^{-1} b_{0}\right)$ is absorbed into c. Note that $W_{0}=-H\left(\rho_{M L}\right)$, so $W_{0}$ is identified by $L(\cdot)$.

As in the $\mathrm{AR}(1)$ case, there is a lower bound on $V_{0}$. Recalling that $Y_{i-}=\left(y_{i,-1}, \ldots, y_{i,-p}\right)$ and $y_{i,-j}=S_{j}\left(\xi_{0 i}+F_{0} \varepsilon_{i}\right)$, where $\xi_{0 i}$ and $\varepsilon_{i}$ are independent, we have

$$
V_{0}=\frac{\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} Y_{i-}^{\prime} M Y_{i-}}{\sigma_{0}^{2}(T-1)}=V_{0}^{L B}+V_{\xi \xi}
$$

where $V_{0}^{L B}$ and $V_{\xi \xi}$ have elements

$$
\left(V_{0}^{L B}\right)_{j k}=\frac{\operatorname{tr} F_{0}^{\prime} S_{j}^{\prime} M S_{k} F_{0}}{T-1}, \quad\left(V_{\xi \xi}\right)_{j k}=\frac{\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \xi_{0 i}^{\prime} S_{j}^{\prime} M S_{k} \xi_{0 i}}{\sigma_{0}^{2}(T-1)}
$$

for $1 \leq j, k \leq p$. Hence, $V_{0}-V_{0}^{L B}$ is positive semi-definite, which we write as $V_{0} \geq V_{0}^{L B}$. When $p \geq T$, while $V_{0}$ is nonsingular by assumption, $\operatorname{rank}\left(V_{0}^{L B}\right) \leq T-1$ because $S_{j} F_{0}=0$ for $j \geq T$, which implies that $\left(V_{0}^{L B}\right)_{j k}=0$ whenever $j \geq T$ or $k \geq T$. Thus, when $p \geq T$, although $V_{0}$ can be arbitrarily close to $V_{0}^{L B}, V_{0} \neq V_{0}^{L B}$. Further, when $p \geq T, b_{j}(\rho)=0$ for $j \geq T$ because the sum defining $b_{j}(\rho)$ is empty, and $c_{i j}(\rho)=0$ for $i+j \geq T$. Hence, when $p \geq T, V_{0}^{L B}-2 b_{0} b_{0}^{\prime}$ and $V_{0}^{L B}-2 b_{0} b_{0}^{\prime}+c_{0}$ have only zeros beyond their leading $(T-1) \times(T-1)$ blocks.

A proof of generalizations of (i)-(ii) of Lemma 1 and Theorem 1 to the $\operatorname{AR}(p)$ model would be desirable but is more difficult than in the $\mathrm{AR}(1)$ case and our attempts failed. ${ }^{8}$ Numerical computations, however, suggest that

$$
\begin{gather*}
V_{0}^{L B} \geq 2 b_{0} b_{0}^{\prime}, \quad V_{0}^{L B} \geq 2 b_{0} b_{0}^{\prime}-c_{0},  \tag{3.1}\\
\operatorname{rank}\left(V_{0}^{L B}-2 b_{0} b_{0}^{\prime}+c_{0}\right)= \begin{cases}\min (p, T-2) & \text { if } \sum_{j=1}^{p} \rho_{0 j} \neq 1 \text { or } T<p+2, \\
p-1 & \text { else. }\end{cases} \tag{3.2}
\end{gather*}
$$

[^5]Specifically, we computed the eigenvalues of $V_{0}^{L B}-2 b_{0} b_{0}^{\prime}$ and $V_{0}^{L B}-2 b_{0} b_{0}^{\prime}+c_{0}$ for $p=2,3,4$; $T=2, \ldots, 7$; and all $\rho_{0}$ in a subset of $\mathbb{R}^{p}$ chosen as follows. For $p=4$, we put a square grid on the Cartesian product of the two triangles defined by

$$
\begin{array}{ll}
-1 \leq \gamma_{2} \leq 1, & \gamma_{2}-1 \leq \gamma_{1} \leq 1-\gamma_{2} \\
-1 \leq \gamma_{4} \leq 1, & \gamma_{4}-1 \leq \gamma_{3} \leq 1-\gamma_{4} \tag{3.3}
\end{array}
$$

which is the stationary region of the lag polynomial $\gamma(L)=\left(1-\gamma_{1} L-\gamma_{2} L^{2}\right)\left(1-\gamma_{3} L-\gamma_{4} L^{2}\right)$. For each point on this grid and for each of the values $m=1,2,4, \rho_{0}$ was calculated by equating the coefficients on both sides of $m-\rho_{01} L-\rho_{02} L^{2}-\rho_{03} L^{3}-\rho_{04} L^{4}=m \gamma(L)$. For $m=1$, the stationary region is covered, while for larger $m$ a larger region is covered, though less densely. In addition to (3.3) we set $\gamma_{4}=0$ for $p=3$, and $\gamma_{3}=\gamma_{4}=0$ for $p=2$. The grid points on the region defined by (3.3) were spaced at intervals of .002 when $p=2, .02$ when $p=3$, and .1 when $p=4$. We found that, uniformly over this numerical design, the eigenvalues of $V_{0}^{L B}-2 b_{0} b_{0}^{\prime}$ and $V_{0}^{L B}-2 b_{0} b_{0}^{\prime}+c_{0}$ are non-negative and the rank of $V_{0}^{L B}-2 b_{0} b_{0}^{\prime}+c_{0}$ is as given by (3.2). These findings, while obviously not a proof, support (3.1) and (3.2), and we shall proceed under the assumption that (3.1) and (3.2) hold. ${ }^{9}$

Because $V_{0} \geq V_{0}^{L B}$, (3.1) implies that $V_{0} \geq 2 b_{0} b_{0}^{\prime}$ and that $H_{\mathrm{A}}\left(\rho_{0}\right)=2 b_{0} b_{0}^{\prime}-V_{0}-c_{0} \leq 0$. Pre- and postmultiplication of $V_{0} \geq 2 b_{0} b_{0}^{\prime}$ by $b_{0}^{\prime} V_{0}^{-1}$ and $V_{0}^{-1} b_{0}$ gives $b_{0}^{\prime} V_{0}^{-1} b_{0} \leq \frac{1}{2} \leq 1-b_{0}^{\prime} V_{0}^{-1} b_{0}$. Recalling that $\left(\rho_{0}-\rho_{M L}\right)^{\prime} V_{0}\left(\rho_{0}-\rho_{M L}\right)=b_{0}^{\prime} V_{0}^{-1} b_{0}$, we see that

$$
\left(\rho_{0}-\rho_{M L}\right)^{\prime} W_{0}\left(\rho_{0}-\rho_{M L}\right) \leq 1
$$

Therefore, if (3.1) and (3.2) hold, $\rho_{0}$ is a point in the ellipsoidal disk $\mathcal{E}=\left\{\rho:\left(\rho-\rho_{M L}\right)^{\prime} W_{0}(\rho-\right.$ $\left.\left.\rho_{M L}\right) \leq 1\right\}$ where $L_{\mathrm{A}}(\rho)$ has a local maximum or a flat inflection point. We approached the question of uniqueness of such a point numerically. For the same numerical design as above and with $V_{0}=V_{0}^{L B}$, we applied the Newton-Raphson algorithm to find a stationary point of $L_{\mathrm{A}}(\rho)$, starting at $\rho_{M L}$ and using the Moore-Penrose inverse of $H_{\mathrm{A}}(\rho)$ whenever $H_{\mathrm{A}}(\rho)$ is singular. Uniformly over this design, the algorithm was found to converge to $\rho_{0}$, thus supporting the conjecture that $\rho_{0}$ is the unique point in $\mathcal{E}$ where $L_{\mathrm{A}}(\rho)$ has a local maximum or a flat inflection point. ${ }^{10}$

In the $\operatorname{AR}(p)$ model with covariates, $\beta$ can be profiled out of $L_{\mathrm{A}}(\theta)$ just as in the $\operatorname{AR}(1)$ model. Here, again, $\beta_{0}=\beta\left(\rho_{0}\right)$. Lemma 2 continues to hold for $p>1$. Hence, if $\rho_{0}$ is identified in the model without covariates in the way we suggested, then it is identified in the model with covariates in exactly the same way, now with $\mathcal{E}$ defined through $W_{0}=\left(1-b_{0 \rho}^{\prime} V_{0}^{\rho \rho} b_{0 \rho}\right)^{-1} V_{0 \rho \rho}$, in obvious notation.

[^6]
## 4 Estimation

For a given $\rho$, we define

$$
\widehat{\beta}_{\mathrm{AL}}(\rho)=\arg \max _{\beta} l_{\mathrm{A}}(\rho, \beta)=\left(\sum_{i=1}^{N} X_{i}^{\prime} M X_{i}\right)^{-1} \sum_{i=1}^{N} X_{i}^{\prime} M\left(y_{i}-Y_{i-} \rho\right)
$$

as the adjusted likelihood (AL) estimator of $\beta_{0}$. (Note that $\widehat{\beta}_{\mathrm{AL}}(\rho)$ coincides with the MLE of $\beta_{0}$ for a given $\rho$.) The unadjusted and adjusted profile log-likelihoods for $\rho$ are $l(\rho)=l\left(\rho, \widehat{\beta}_{\text {AL }}(\rho)\right)$ and $l_{\mathrm{A}}(\rho)=l(\rho)-a(\rho)$. Let $s(\rho), s_{\mathrm{A}}(\rho), h(\rho)$, and $h_{\mathrm{A}}(\rho)$ be the corresponding profile scores and Hessians. Let $\widehat{W}=-h\left(\widehat{\rho}_{M L}\right)$, where $\widehat{\rho}_{M L}$ is the MLE of $\rho_{0}$, and let $\widehat{\mathcal{E}}=\left\{\rho:\left(\rho-\widehat{\rho}_{M L}\right)^{\widehat{W}}(\rho-\right.$ $\left.\left.\widehat{\rho}_{M L}\right) \leq 1\right\}$. We define the AL estimator of $\rho_{0}$ as

$$
\widehat{\rho}_{\mathrm{AL}}=\arg \min _{\rho \in \widehat{\mathcal{E}}} s_{\mathrm{A}}^{\prime}(\rho) s_{\mathrm{A}}(\rho) \quad \text { s.t. } \quad h_{\mathrm{A}}(\rho) \leq 0
$$

that is, as the strict local maximizer of $l_{\mathrm{A}}(\rho)$ on the interior of $\widehat{\mathcal{E}}$ if such a maximizer exists and otherwise as the minimizer of the norm of $s_{\mathrm{A}}(\rho)$ on $\widehat{\mathcal{E}}$. The AL estimator of $\beta_{0}$, then, is $\widehat{\beta}_{\mathrm{AL}}=\widehat{\beta}_{\mathrm{AL}}\left(\widehat{\rho}_{\mathrm{AL}}\right)$.

Let $N \rightarrow \infty$. Then, $l_{\mathrm{A}}(\rho)$ converges to $L_{\mathrm{A}}(\rho)$ uniformly in $\rho$ since $-a(\rho)$ is nonstochastic and $\sup _{\rho}|l(\rho)-L(\rho)|=o_{p}(1)$. Further, $\widehat{\rho}_{M L} \xrightarrow{p} \rho_{M L}, \widehat{W} \xrightarrow{p}-H\left(\rho_{M L}\right)=W_{0}$, and $\widehat{\mathcal{E}} \xrightarrow{p} \mathcal{E}$ in the sense that $\operatorname{Pr}[\rho \in \widehat{\mathcal{E}}] \rightarrow 1_{\{\rho \in \mathcal{E}\}}$ for any $\rho$ not on the boundary of $\mathcal{E}$. It follows that $\widehat{\theta}_{\mathrm{AL}}=\left(\widehat{\rho}_{\mathrm{AL}}^{\prime}, \widehat{\beta}_{\mathrm{AL}}^{\prime}\right)^{\prime} \xrightarrow{p} \theta_{0}$. When $H_{\mathrm{A}}\left(\theta_{0}\right)$ is nonsingular, by the usual argument involving a Taylor series expansion of $s_{\mathrm{A}}(\theta)$ around $\theta_{0}$,

$$
\begin{equation*}
\sqrt{N}\left(\widehat{\theta}_{\mathrm{AL}}-\theta_{0}\right) \xrightarrow{d} \mathcal{N}(0, \Omega), \quad \Omega=H_{\mathrm{A}}\left(\theta_{0}\right)^{-1}\left(V_{0}-b_{0} b_{0}^{\prime}\right) H_{\mathrm{A}}\left(\theta_{0}\right)^{-1} . \tag{4.1}
\end{equation*}
$$

The asymptotic variance can be estimated in the usual way. We have not investigated the limit distribution of $\widehat{\theta}_{\mathrm{AL}}$ in the situation where $H_{\mathrm{A}}\left(\theta_{0}\right)$ is singular. Presumably this could be done along the lines of Rotnitzky, Cox, Bottai, and Robins (2000).

## 5 Simulations

We used simulations to examine the finite sample properties of the AL estimator in $A R(1)$ and $\mathrm{AR}(2)$ models without covariates and in an $\operatorname{AR}(1)$ model with a single stationary and exogenous regressor. We compared the AL with the one-step GMM estimator of Arellano and Bond (1991), which leaves the initial observations unrestricted. ${ }^{11}$ In all the designs, we set $N=100$, generated

[^7]$\varepsilon_{i t}$ and $\alpha_{i}$ as $\mathcal{N}(0,1)$ variates, ${ }^{12}$ and chose $\rho_{0}$ in the interior of the stationary region, which implies that $y_{i t}$ is eventually stationary as $t \rightarrow \infty$. We varied the information content of the data through the initial observations. Let $\mu_{i}=\lim _{t \rightarrow \infty} \mathbb{E}\left(y_{i t-} \mid \alpha_{i}\right)$ and $\Sigma_{i}=\lim _{t \rightarrow \infty} \operatorname{Var}\left(y_{i t-} \mid \alpha_{i}\right)$, so, if $y_{i}^{0}$ was drawn from the stationary distribution, we would just have $\mu_{i}=\mathbb{E}\left(y_{i}^{0} \mid \alpha_{i}\right)$ and $\Sigma_{i}=\operatorname{Var}\left(y_{i}^{0} \mid \alpha_{i}\right)$. Let $G_{i} G_{i}^{\prime}=\Sigma_{i}$ be the Cholesky factorization of $\Sigma_{i}$. We set $y_{i}^{0}=\mu_{i}+\psi G_{i} \iota$ for some chosen scalar $\psi \geq 0$, which is a $p$-variate version of setting the initial observations $\psi$ standard deviations away from the stationary mean. So $\psi$ controls the outlyingness of the initial observations relative to the stationary distributions. All else being equal, $V_{0}$ increases in $\psi$ and $V_{0} \rightarrow V_{0}^{L B}$ as $\psi \rightarrow 0$, so the data carry less information as $\psi$ gets smaller. The effect of strong inlying observations (small $\psi$ ) on the informativeness of the data is stronger when $T$ is small because it takes time to revert to the stationary distribution. The effect of $\psi$ is vanishingly small as $\rho_{0}$ moves to the boundary of the stationary region. ${ }^{13}$ We set $\psi=0,1,2$ in the $\operatorname{AR}(1)$ models and $\psi=.3,1,2$ in the $\mathrm{AR}(2)$ model. ${ }^{14}$

In the models without a covariate, $\mu_{i}$ and $\Sigma_{i}$ follow immediately from $\alpha_{i}$ and $\rho_{0}$. In the $\operatorname{AR}(1)$ model with a covariate, $x_{i t}$ was generated by a stationary $\operatorname{AR}(1), x_{i t}=\delta \alpha_{i}+\gamma x_{i t-1}+u_{i t}$ with $u_{i t} \sim \mathcal{N}\left(0, \sigma_{u}^{2}\right)$ and $x_{i 0}$ drawn from the stationary distribution. Here,

$$
\mu_{i}=\frac{\alpha_{i}}{1-\rho_{0}}\left(1+\frac{\delta \beta_{0}}{1-\gamma}\right), \quad \Sigma_{i}=\frac{1}{1-\rho_{0}^{2}}\left(1+\frac{\beta_{0}^{2}}{1-\gamma^{2}}\left(\frac{1+\gamma \rho_{0}}{1-\gamma \rho_{0}}\right) \sigma_{u}^{2}\right)
$$

We set $\delta=\gamma=\sigma_{u}=.5$ and $\beta_{0}=1-\rho_{0}$, inducing dependence between the covariate and the fixed effect, and keeping the long-run multiplier of $x$ on $y$ constant at unity across designs, as in Kiviet (1995).

We varied $\psi, T$, and $\rho_{0}$. Tables $1-3$ present Monte Carlo estimates, based on 10,000 replications at each design point, of the bias and standard deviation (STD) of the AL and GMM estimators and of the coverage rates of the asymptotic and bootstrap $95 \%$ confidence intervals $\left(\mathrm{CI}_{.95}^{A}\right.$ and $\left.\mathrm{CI}_{.95}^{B}\right) .{ }^{15}$

In the $\mathrm{AR}(1)$ model (Table 1), for all designs $\widehat{\rho}_{\text {AL }}$ has less bias and less standard deviation than $\widehat{\rho}_{\text {GMM }}$. When $\rho_{0}=.5$, the AL estimator is virtually unbiased except when $\psi=0$ and $T=2$, in which case $\rho_{0}$ is weakly identified. ${ }^{16}$ Both estimators deliver $95 \%$ confidence intervals with broadly correct coverage, although the coverage errors are somewhat larger for $\widehat{\rho}_{\text {GMM }}$,

[^8]Table 1: Simulations for the $\operatorname{AR}(1)$ model, $N=100$

|  | bias |  |  |  |  |  | STD |  | $\mathrm{CI}_{.95}^{A}$ |  |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| $\psi$ | $T$ | $\rho_{0}$ | $\widehat{\rho}_{\text {AL }}$ | $\widehat{\rho}_{\mathrm{GMM}}$ | $\widehat{\rho}_{\mathrm{AL}}$ | $\widehat{\rho}_{\mathrm{GMM}}$ | $\widehat{\rho}_{\mathrm{AL}}$ | $\widehat{\rho}_{\mathrm{GMM}}$ | $\widehat{\rho}_{\mathrm{AL}}$ | $\widehat{\rho}_{\mathrm{GMM}}^{B}$ |
| 0 | 2 | .5 | -.146 | - | .267 | - | .880 | .916 | .924 | .976 |
| 1 | 2 | .5 | .032 | - | .269 | - | .944 | .934 | .944 | .965 |
| 2 | 2 | .5 | .029 | - | .173 | - | .967 | .941 | .942 | .944 |
| 0 | 4 | .5 | .006 | -.043 | .142 | .150 | .954 | .921 | .947 | .911 |
| 1 | 4 | .5 | .014 | -.057 | .124 | .163 | .965 | .927 | .944 | .906 |
| 2 | 4 | .5 | .002 | -.015 | .064 | .083 | .968 | .940 | .943 | .934 |
| 0 | 8 | .5 | .000 | -.028 | .056 | .058 | .966 | .914 | .946 | .860 |
| 1 | 8 | .5 | .000 | -.039 | .048 | .070 | .960 | .906 | .944 | .842 |
| 2 | 8 | .5 | -.001 | -.022 | .036 | .052 | .953 | .925 | .946 | .887 |
| 0 | 24 | .5 | .000 | -.018 | .021 | .022 | .944 | .871 | .944 | .726 |
| 1 | 24 | .5 | -.001 | -.022 | .020 | .025 | .946 | .859 | .945 | .700 |
| 2 | 24 | .5 | .000 | -.020 | .018 | .023 | .950 | .868 | .947 | .721 |
| 0 | 2 | .95 | -.144 | - | .266 | - | .883 | .919 | .927 | .978 |
| 1 | 2 | .95 | -.118 | - | .268 | - | .888 | .924 | .933 | .979 |
| 2 | 2 | .95 | -.055 | - | .267 | - | .913 | .928 | .942 | .979 |
| 0 | 4 | .95 | -.087 | -.680 | .124 | .465 | .894 | .667 | .903 | .615 |
| 1 | 4 | .95 | -.063 | -.696 | .124 | .464 | .907 | .688 | .933 | .627 |
| 2 | 4 | .95 | -.016 | -.389 | .123 | .426 | .936 | .815 | .946 | .756 |
| 0 | 8 | .95 | -.043 | -.345 | .064 | .176 | .888 | .438 | .902 | .155 |
| 1 | 8 | .95 | -.025 | -.398 | .063 | .192 | .914 | .403 | .942 | .141 |
| 2 | 8 | .95 | .003 | -.225 | .063 | .156 | .952 | .655 | .948 | .350 |
| 0 | 24 | .95 | -.006 | -.090 | .024 | .033 | .923 | .157 | .940 | .006 |
| 1 | 24 | .95 | .000 | -.127 | .024 | .042 | .941 | .063 | .940 | .004 |
| 2 | 24 | .95 | .001 | -.091 | .018 | .038 | .961 | .166 | .940 | .022 |

'-' indicates non-existence of the corresponding population moment.

Table 2: Simulations for the $\operatorname{AR}(2)$ model, $N=100$

| $\psi$ | T | $\rho_{0}$ | bias |  | STD |  | $\mathrm{CI}_{.95}^{A}$ |  | $\mathrm{CI}_{.95}^{B}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | $\widehat{\rho}_{\text {AL }}$ | $\widehat{\rho}_{\text {GMM }}$ | $\widehat{\rho}_{\text {AL }}$ | $\widehat{\rho}_{\text {GMM }}$ | $\widehat{\rho}_{\text {AL }}$ | $\widehat{\rho}_{\text {GMM }}$ | $\widehat{\rho}_{\text {AL }}$ | $\widehat{\rho}_{\text {GMM }}$ |
| . 3 | 2 | . 6 | -. 141 | - | . 267 | - | . 882 | . 920 | . 930 | . 977 |
|  |  | . 2 | -. 129 | - | . 748 | - | . 959 | . 945 | . 940 | . 983 |
| 1 | 2 | . 6 | -. 141 | - | . 267 | - | . 881 | . 916 | . 930 | . 977 |
|  |  | . 2 | $-.124$ | - | . 316 | - | . 909 | . 920 | . 920 | . 977 |
| 2 | 2 | . 6 | -. 147 | - | . 266 | - | . 880 | . 916 | . 927 | . 977 |
|  |  | . 2 | -. 130 | - | . 258 | - | . 890 | . 918 | . 922 | . 978 |
| . 3 | 4 | . 6 | -. 071 | -. 279 | . 123 | . 280 | . 900 | . 782 | . 914 | . 698 |
|  |  | . 2 | $-.033$ | -. 125 | . 098 | . 133 | . 936 | . 808 | . 932 | . 763 |
| 1 | 4 | . 6 | -. 001 | -. 043 | . 123 | . 120 | . 945 | . 921 | . 945 | . 895 |
|  |  | . 2 | -. 002 | -. 023 | . 094 | . 089 | . 959 | . 931 | . 950 | . 925 |
| 2 | 4 | . 6 | . 008 | -. 011 | . 082 | . 064 | . 967 | . 939 | . 946 | . 933 |
|  |  | . 2 | . 003 | -. 007 | . 072 | . 064 | . 965 | . 943 | . 949 | . 942 |
| . 3 | 8 | . 6 | -. 016 | -. 102 | . 065 | . 088 | . 941 | . 773 | . 941 | . 536 |
|  |  | . 2 | -. 009 | $-.053$ | . 056 | . 056 | . 949 | . 833 | . 943 | . 695 |
| 1 | 8 | . 6 | . 006 | -. 033 | . 063 | . 055 | . 965 | . 897 | . 948 | . 816 |
|  |  | . 2 | . 002 | -. 015 | . 052 | . 045 | . 963 | . 929 | . 947 | . 910 |
| 2 | 8 | . 6 | . 000 | -. 011 | . 037 | . 037 | . 953 | . 932 | . 947 | . 913 |
|  |  | . 2 | . 000 | -. 003 | . 037 | . 036 | . 944 | . 943 | . 945 | . 942 |
| . 3 | 24 | . 6 | . 000 | $-.028$ | . 024 | . 025 | . 956 | . 795 | . 949 | . 607 |
|  |  | . 2 | . 000 | -. 018 | . 024 | . 023 | . 950 | . 869 | . 945 | . 755 |
| 1 | 24 | . 6 | . 000 | $-.020$ | . 022 | . 023 | . 951 | . 859 | . 950 | . 732 |
|  |  | . 2 | . 000 | -. 011 | . 022 | . 022 | . 945 | . 914 | . 946 | . 865 |
| 2 | 24 | . 6 | . 000 | -. 011 | . 020 | . 020 | . 950 | . 913 | . 951 | . 856 |
|  |  | . 2 | . 000 | -. 003 | . 019 | . 019 | . 946 | . 944 | . 948 | . 939 |
| . 3 | 2 | 1 | -. 144 | - | . 265 | - | . 882 | . 918 | . 934 | . 977 |
|  |  | -. 2 | $-.083$ | - | . 640 | - | . 969 | . 968 | . 945 | . 992 |
| 1 | 2 | 1 | $-.145$ | - | . 265 | - | . 880 | . 913 | . 927 | . 979 |
|  |  | -. 2 | $-.073$ | - | . 234 | - | . 928 | . 921 | . 931 | . 978 |
| 2 | 2 | 1 | $-.147$ | - | . 266 | - | . 881 | . 918 | . 930 | . 977 |
|  |  | -. 2 | $-.078$ | - | . 165 | - | . 901 | . 918 | . 919 | . 979 |
| . 3 | 4 | 1 | $-.070$ | -. 235 | . 118 | . 256 | . 902 | . 808 | . 906 | . 721 |
|  |  | -. 2 | $.004$ | $-.021$ | . 087 | . 084 | . 966 | . 934 | . 945 | . 937 |
| 1 | 4 | 1 | . 009 | -. 030 | . 115 | . 098 | . 951 | . 930 | . 947 | . 908 |
|  |  | -. 2 | $.006$ | -. 006 | . 081 | . 078 | . 965 | . 939 | . 944 | . 943 |
| 2 | 4 | 1 | . 003 | -. 008 | . 059 | . 051 | . 966 | . 940 | . 945 | . 936 |
|  |  | -. 2 | . 001 | $-.003$ | . 060 | . 059 | . 949 | . 941 | . 945 | . 943 |
| . 3 | 8 | 1 | -. 009 | -. 068 | . 059 | . 072 | . 953 | . 830 | . 945 | . 653 |
|  |  | -. 2 | . 002 | -. 015 | . 051 | . 045 | . 965 | . 926 | . 945 | . 911 |
| 1 | 8 | 1 | . 002 | -. 024 | . 051 | . 048 | . 966 | . 916 | . 947 | . 856 |
|  |  | -. 2 | . 001 | -. 001 | . 042 | . 041 | . 950 | . 945 | . 944 | . 946 |
| 2 | 8 | 1 | $-.001$ | -. 009 | . 033 | . 034 | . 948 | . 936 | . 946 | . 922 |
|  |  | -. 2 | . 001 | . 002 | . 033 | . 034 | . 944 | . 945 | . 947 | . 947 |
| . 3 | 24 | 1 | $-.001$ | -. 018 | . 022 | . 023 | . 950 | . 872 | . 950 | . 770 |
|  |  | -. 2 | . 000 | -. 008 | . 022 | . 022 | . 947 | . 927 | . 946 | . 906 |
| 1 | 24 | 1 | $-.001$ | $-.015$ | . 021 | . 022 | . 948 | . 898 | . 948 | . 818 |
|  |  | -. 2 | . 000 | -. 004 | . 021 | . 021 | . 947 | . 940 | . 947 | . 935 |
| 2 | 24 | 1 | -. 001 | -. 010 | . 019 | . 020 | . 950 | . 920 | . 947 | . 879 |
|  |  | -. 2 | . 000 | . 000 | . 019 | . 019 | . 946 | . 947 | . 948 | . 949 |

'-' indicates non-existence of the corresponding population moment.

Table 3: Simulations for the $\operatorname{AR}(1)$ model with a covariate, $N=100$

|  |  |  | bias |  | STD |  | $\mathrm{CI}_{.95}^{A}$ |  | $\mathrm{CI}_{.95}^{B}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\psi$ | $T$ | $\theta_{0}$ | $\widehat{\theta}_{\text {AL }}$ | $\widehat{\theta}_{\text {GMM }}$ | $\widehat{\theta}_{\text {AL }}$ | $\widehat{\theta}_{\text {GMM }}$ | $\widehat{\theta}_{\text {AL }}$ | $\widehat{\theta}_{\text {GMM }}$ | $\widehat{\theta}_{\text {AL }}$ | $\widehat{\theta}_{\text {GMM }}$ |
| 0 | 2 | . 5 | -. 092 | -. 051 | . 266 | - | . 900 | . 882 | . 943 | . 945 |
|  |  | . 5 | -. 027 | -. 015 | . 249 | - | . 965 | . 943 | . 945 | . 962 |
| 1 | 2 | . 5 | . 019 | . 001 | . 269 | - | . 939 | . 910 | . 947 | . 945 |
|  |  | . 5 | -. 001 | . 000 | . 258 | - | . 968 | . 950 | . 939 | . 949 |
| 2 | 2 | . 5 | . 030 | . 002 | . 182 | - | . 966 | . 940 | . 949 | . 945 |
|  |  | . 5 | -. 010 | -. 002 | . 261 | - | . 949 | . 941 | . 942 | . 941 |
| 0 | 4 | . 5 | . 014 | -. 093 | . 141 | . 119 | . 957 | . 845 | . 945 | . 737 |
|  |  | . 5 | . 004 | -. 008 | . 127 | . 124 | . 959 | . 941 | . 947 | . 946 |
| 1 | 4 | . 5 | . 012 | -. 069 | . 119 | . 102 | . 968 | . 883 | . 948 | . 802 |
|  |  | . 5 | -. 001 | . 010 | . 126 | . 123 | . 951 | . 943 | . 949 | . 947 |
| 2 | 4 | . 5 | . 001 | -. 024 | . 063 | . 062 | . 968 | . 921 | . 944 | . 890 |
|  |  | . 5 | . 000 | . 009 | . 129 | . 128 | . 944 | . 940 | . 941 | . 941 |
| 0 | 8 | . 5 | -. 001 | -. 053 | . 051 | . 050 | . 965 | . 813 | . 948 | . 582 |
|  |  | . 5 | . 001 | . 004 | . 074 | . 075 | . 943 | . 943 | . 946 | . 945 |
| 1 | 8 | . 5 | -. 001 | -. 047 | . 045 | . 047 | . 959 | . 837 | . 945 | . 627 |
|  |  | . 5 | . 001 | . 011 | . 075 | . 076 | . 940 | . 938 | . 943 | . 937 |
| 2 | 8 | . 5 | -. 001 | -. 026 | . 035 | . 036 | . 950 | . 884 | . 939 | . 754 |
|  |  | . 5 | . 001 | . 011 | . 075 | . 075 | . 948 | . 945 | . 949 | . 943 |
| 0 | 24 | . 5 | -. 001 | -. 032 | . 020 | . 020 | . 948 | . 646 | . 943 | . 323 |
|  |  | . 5 | . 000 | . 007 | . 038 | . 038 | . 950 | . 945 | . 950 | . 941 |
| 1 | 24 | . 5 | . 000 | -. 031 | . 019 | . 020 | . 946 | . 652 | . 945 | . 343 |
|  |  | . 5 | . 000 | . 009 | . 038 | . 038 | . 950 | . 942 | . 952 | . 937 |
| 2 | 24 | . 5 | . 000 | -. 026 | . 017 | . 018 | . 947 | . 700 | . 948 | . 416 |
|  |  | . 5 | . 000 | . 009 | . 038 | . 038 | . 949 | . 943 | . 952 | . 935 |
| 0 | 2 | . 95 | -. 145 | -. 904 | . 266 | - | . 879 | . 809 | . 930 | . 926 |
|  |  | . 05 | -. 007 | -. 030 | . 234 | - | . 974 | . 977 | . 947 | . 991 |
| 1 | 2 | . 95 | -. 124 | -. 975 | . 266 | - | . 891 | . 811 | . 937 | . 923 |
|  |  | . 05 | . 004 | . 046 | . 240 | - | . 974 | . 977 | . 941 | . 989 |
| 2 | 2 | . 95 | -. 068 | -. 534 | . 264 | - | . 910 | . 842 | . 948 | . 936 |
|  |  | . 05 | . 007 | . 063 | . 246 | - | . 972 | . 968 | . 943 | . 980 |
| 0 | 4 | . 95 | -. 086 | -. 641 | . 124 | . 253 | . 892 | . 267 | . 906 | . 123 |
|  |  | . 05 | -. 002 | -. 013 | . 121 | . 115 | . 973 | . 944 | . 948 | . 955 |
| 1 |  | . 95 | -. 067 | -. 595 | . 123 | . 252 | . 906 | . 312 | . 930 | . 148 |
|  |  | . 05 | . 004 | . 041 | . 122 | . 114 | . 973 | . 938 | . 948 | . 943 |
| 2 | 4 | . 95 | -. 020 | -. 338 | . 122 | . 207 | . 939 | . 561 | . 950 | . 346 |
|  |  | . 05 | . 003 | . 056 | . 128 | . 120 | . 971 | . 916 | . 945 | . 917 |
| 0 | 8 | . 95 | -. 043 | -. 347 | . 064 | . 100 | . 888 | . 036 | . 903 | . 001 |
|  |  | . 05 | . 000 | -. 006 | . 073 | . 075 | . 970 | . 943 | . 948 | . 946 |
| 1 | 8 | . 95 | -. 026 | -. 306 | . 064 | . 098 | . 911 | . 067 | . 942 | . 004 |
|  |  | . 05 | . 003 | . 028 | . 075 | . 074 | . 972 | . 926 | . 944 | . 933 |
| 2 | 8 | . 95 | . 003 | -. 155 | . 064 | . 068 | . 947 | . 317 | . 939 | . 056 |
|  |  | . 05 | . 000 | . 032 | . 075 | . 072 | . 967 | . 928 | . 951 | . 918 |
| 0 | 24 | . 95 | -. 007 | -. 116 | . 024 | . 022 | . 920 | . 000 | . 943 | . 000 |
|  |  | . 05 | . 000 | -. 001 | . 038 | . 039 | . 970 | . 946 | . 950 | . 949 |
| 1 | 24 | . 95 | . 000 | -. 101 | . 024 | . 021 | . 942 | . 000 | . 944 | . 000 |
|  |  | . 05 | . 000 | . 009 | . 038 | . 039 | . 964 | . 942 | . 950 | . 942 |
| 2 | 24 | . 95 | . 001 | -. 055 | . 017 | . 015 | . 963 | . 024 | . 946 | . 000 |
|  |  | . 05 | . 000 | . 010 | . 038 | . 038 | . 950 | . 941 | . 951 | . 937 |

'-' indicates non-existence of the corresponding population moment.
[21]
where they increase in $T$. When $\rho_{0}$ is increased to $.95, \widehat{\rho}_{\text {GMM }}$ has a large negative bias and its standard deviation increases considerably. The AL estimator also deteriorates, but much less. The confidence intervals based on $\widehat{\rho}_{\text {AL }}$ continue to provide good coverage, with the bootstrap being slightly superior, while those based on $\widehat{\rho}_{\text {GMM }}$ have large coverage errors for all $\psi$, especially when $T \geq 8$.

Turning to the $\mathrm{AR}(2)$ model (Table 2), the AL and GMM estimators perform well in terms of bias (for the chosen values of $\rho_{0}$ ), although there is a non-negligible bias when $T=2$ and also when $T=4$ and the initial observations are strong inlyers. As $N=100$, the probability that the adjusted likelihood has no local maximum in the relevant region is fairly large when both $T$ and $\psi$ are small. In most designs, $\widehat{\rho}_{\text {AL }}$ has smaller standard deviation than $\widehat{\rho}_{\text {GMM }}$, with the difference decreasing in $T$ and $\psi$. For both estimators, the confidence intervals have very reasonable coverage, with those based on $\widehat{\rho}_{\text {AL }}$ being slightly more accurate in most designs.

The results for the $\mathrm{AR}(1)$ model with a covariate (Table 3) follow a similar pattern as in the $\operatorname{AR}(1)$ model without covariates. The coefficient on the exogenous variable, $\beta_{0}$, is generally estimated with small bias by both estimators. When $\left(\rho_{0}, \beta_{0}\right)=(.5, .5)$, the biases and standard deviations are fairly small and, here, the AL estimator does not uniformly dominate GMM. When $T=2, \widehat{\theta}_{\text {GMM }}$ is less biased than $\widehat{\theta}_{\text {AL }}$ while the opposite holds when $T \geq 4$. When $\left(\rho_{0}, \beta_{0}\right)$ is changed to $(.95, .05)$, the bias of $\widehat{\beta}_{\mathrm{GMM}}$ increases only moderately, but $\widehat{\rho}_{\mathrm{GMM}}$ is now heavily biased. As a result, GMM-based confidence intervals for $\rho_{0}$ tend to have poor coverage while those for $\beta_{0}$ have broadly correct coverage. Compared to GMM, the AL estimator is much less biased and yields confidence intervals with approximately correct coverage, especially when computed by bootstrapping.

## Appendix

Proof of equations (1.3). For arbitrary $\rho_{1}, \ldots, \rho_{T-1}, D$ and its inverse are

$$
D=\left(\begin{array}{cccc}
1 & 0 & \cdots & 0 \\
-\rho_{1} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
-\rho_{T-1} & \cdots & -\rho_{1} & 1
\end{array}\right), \quad D^{-1}=\left(\begin{array}{cccc}
1 & 0 & \cdots & 0 \\
\phi_{1} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
\phi_{T-1} & \cdots & \phi_{1} & 1
\end{array}\right)
$$

where $\phi_{1}, \ldots, \phi_{T-1}$ are recursively obtained as $\phi_{1}=\rho_{1}$ and $\phi_{j}=\rho_{j}+\sum_{k=1}^{j-1} \phi_{k} \rho_{j-k}, j=2, \ldots, T-1$. Recursive substitution gives

$$
\phi_{j}=\sum_{k_{1}+2 k_{2}+\cdots+j k_{j}=j} \frac{\left(k_{1}+\ldots+k_{p}\right)!}{k_{1}!\cdots k_{p}!} \rho_{1}^{k_{1}} \rho_{2}^{k_{2}} \cdots \rho_{j}^{k_{j}}
$$

equivariant under translations of $\rho_{0}$ and, hence, the bias, standard deviation, and coverage rates are independent of $\rho_{0}$.

Putting $\rho_{p+1}=\cdots=\rho_{T-1}=0$ gives $\phi_{j}=\varphi_{j}$.

Proof of equations (2.1). For $j=1, \ldots, p$, let $\mathcal{S}_{j}=\{S \in \mathcal{S} \mid j \in S\}$. Group terms by $S \in \mathcal{S}_{j}$ to write

$$
\int b_{j}(\rho) d \rho_{j}=\sum_{S \in \mathcal{S}_{j}} B_{j, S}(\rho)+\mathrm{c}
$$

where

$$
B_{j, S}(\rho)=-\sum_{t=0}^{T-j-1} \frac{T-j-t}{T(T-1)} \sum_{k \in \mathcal{K}_{j, S}: \tau^{\prime} k=t} \frac{\left(\iota^{\prime} k\right)!}{k_{1}!\cdots\left(k_{j}+1\right)!\cdots k_{p}!} \rho_{j} \rho_{S}^{k_{S}}
$$

and $\mathcal{K}_{j, S}=\left\{k \in \mathbb{N}^{p} \mid\right.$ for all $j^{\prime} \neq j, k_{j^{\prime}}>0$ iff $\left.j^{\prime} \in S\right\} \supset \mathcal{K}_{S}$. A change of variable from $k_{j}+1$ to $k_{j}$ gives

$$
B_{j, S}(\rho)=-\sum_{t=|S|-j}^{T-j-1} \frac{T-j-t}{T(T-1)} \sum_{k \in \mathcal{K}_{S}: \tau^{\prime} k=t+j} \frac{\left(\iota^{\prime} k-1\right)!}{k_{1}!\cdots k_{p}!} \rho_{S}^{k_{S}},
$$

where the lower limit in the first sum changed from 0 to $|S|-j$ because, when $t<|S|-j$, no $k \in \mathcal{K}_{S}$ satisfies $\tau^{\prime} k=t+j$. A further change of variable from $t+j$ to $t$ gives $B_{j, S}(\rho)=a_{S}(\rho)$, with $a_{S}(\rho)$ as defined in (2.1). Therefore,

$$
b_{j}(\rho)=\nabla_{\rho_{j}} \sum_{S \in \mathcal{S}_{j}} a_{S}(\rho)=\nabla_{\rho_{j}} \sum_{S \in \mathcal{S}} a_{S}(\rho)=\nabla_{\rho_{j}} a(\rho)
$$

which completes the proof.
Proof of (2.4). In the parameterization $\eta_{i}=\alpha_{i} e^{-(T-1) a(\rho)}$, we have

$$
\begin{aligned}
\ell_{i}\left(\vartheta, \eta_{i}\right) & =-\frac{1}{2} \log \sigma^{2}-\frac{1}{2 T \sigma^{2}} \sum_{t=1}^{T}\left(y_{i t}-z_{i t}^{\prime} \theta-\eta_{i} e^{(T-1) a(\rho)}\right)^{2}+\mathrm{c} \\
\nabla_{\eta_{i}} \ell_{i}\left(\vartheta, \eta_{i}\right) & =\frac{e^{(T-1) a(\rho)}}{T \sigma^{2}}\left(y_{i}-Z_{i} \theta-\eta_{i} e^{(T-1) a(\rho)} \iota\right)^{\prime} \iota
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}_{\vartheta, \eta_{i}} \nabla_{\eta_{i} \eta_{i}} \ell_{i}\left(\vartheta, \eta_{i}\right) & =-\sigma^{-2} e^{2(T-1) a(\rho)}, \quad \mathbb{E}_{\vartheta, \eta_{i}} \nabla_{\sigma^{2} \eta_{i}} \ell_{i}\left(\vartheta, \eta_{i}\right)=0 \\
\mathbb{E}_{\vartheta, \eta_{i}} \nabla_{\theta \eta_{i}} \ell_{i}\left(\vartheta, \eta_{i}\right) & =-\sigma^{-2} e^{(T-1) a(\rho)}\left(\eta_{i}(T-1) b(\rho) e^{(T-1) a(\rho)}+\mathbb{E}_{\vartheta, \eta_{i}} Z_{i}^{\prime} \iota / T\right)
\end{aligned}
$$

The $j$ th column of $Y_{i-}$ is $y_{i,-j}=S_{j}\left(\xi_{i}+F \varepsilon_{i}\right)$, so the $j$ th element of $\mathbb{E}_{\vartheta, \eta_{i}} Y_{i-}^{\prime} \iota$ is

$$
\mathbb{E}_{\vartheta, \eta_{i}} y_{i,-j}^{\prime} \iota=\iota^{\prime} S_{j} \xi_{i}=\iota^{\prime} D_{j}^{-1} \iota \eta_{i} e^{(T-1) a(\rho)}+m_{j}, \quad m_{j}=\iota^{\prime} S_{j}\binom{y_{i}^{0}}{D^{-1}\left(C y_{i}^{0}+X_{i} \beta\right)}
$$

Hence,

$$
\mathbb{E}_{\vartheta, \eta_{i}} Z_{i-\iota}^{\prime} \iota / T=-\eta_{i}(T-1) b(\rho) e^{(T-1) a(\rho)}+m
$$

where $m=\left(m_{1}, \ldots, m_{p}, \iota^{\prime} X_{i} / T\right)^{\prime}$ is free of $\eta_{i}$. Consequently,

$$
A_{i}^{-1} B_{i}=-e^{-(T-1) a(\rho)}\binom{m}{0}
$$

and $\nabla_{\eta_{i}}\left(A_{i}^{-1} B_{i}\right)=0$.
Proof that no orthogonalization exists when $p>1$. In the original parameterization, if $l_{i}\left(\vartheta, \alpha_{i}\right)$ is $i$ 's log-likelihood contribution, we have

$$
\begin{aligned}
\mathbb{E}_{\vartheta, \alpha_{i}} \nabla_{\alpha_{i} \alpha_{i}} l_{i}\left(\vartheta, \alpha_{i}\right) & =-\sigma^{-2}, \quad \mathbb{E}_{\vartheta, \alpha_{i}} \nabla_{\sigma^{2} \alpha_{i}} l_{i}\left(\vartheta, \alpha_{i}\right)=0, \\
\mathbb{E}_{\vartheta, \alpha_{i}} \nabla_{\theta \alpha_{i}} l_{i}\left(\vartheta, \alpha_{i}\right) & =-\sigma^{-2} \mathbb{E}_{\vartheta, \alpha_{i}} Z_{i}^{\prime} \iota / T,
\end{aligned}
$$

and so, by the preceeding proof,

$$
A_{i}^{-1} B_{i}=-\binom{\mathbb{E}_{\vartheta, \alpha_{i}} Z_{i}^{\prime} \iota / T}{0}=-\binom{-(T-1) b(\rho) \alpha_{i}+m}{0}
$$

Suppose some reparameterized fixed effect, say $\zeta_{i}$, is information orthogonal to $\vartheta$. Then $\alpha_{i}=$ $\alpha_{i}\left(\vartheta, \zeta_{i}\right)$ must satisfy the differential equation $\nabla_{\vartheta} \alpha_{i}=A_{i}^{-1} B_{i}$, that is,

$$
\begin{array}{ll}
\nabla_{\rho_{j}} \alpha_{i}=(T-1) b_{j}(\rho) \alpha_{i}-m_{j}, & j=1, \ldots, p \\
\nabla_{\beta_{j}} \alpha_{i}=-m_{p+j}, & j=1, \ldots, q \tag{A.2}
\end{array}
$$

and $\nabla_{\sigma^{2}} \alpha_{i}=0$. We show that these equations are inconsistent. Suppose $q>0$. Then (A.1) implies $\nabla_{\rho_{j} \beta_{j^{\prime}}} \alpha_{i}=-\nabla_{\beta_{j^{\prime}}} m_{j}$, which is generally non-zero, while (A.2) implies $\nabla_{\rho_{j} \beta_{j^{\prime}}} \alpha_{i}=0$, so the equations are inconsistent. Suppose $q=0$. Then

$$
m_{j}=\iota^{\prime} S_{j}\binom{I_{p}}{D^{-1} C} y_{i}^{0}, \quad j=1, \ldots, p
$$

and, because $\nabla_{\rho_{j^{\prime}}} b_{j}(\rho)=\nabla_{\rho_{j^{\prime} \rho_{j}}} a(\rho)=\nabla_{\rho_{j}} b_{j^{\prime}}(\rho)$, (A.1) will be inconsistent if $\nabla_{\rho_{j^{\prime}}} m_{j} \neq \nabla_{\rho_{j}} m_{j^{\prime}}$ for some $j, j^{\prime}$. Take $j=p$ and $j^{\prime}=p-1$. The first element of $y_{i}^{0}$ appears in $m_{p}$ and $m_{p-1}$ with coefficients $\gamma_{p}=1+\rho_{p} \sum_{t=0}^{T-p-1} \varphi_{t}$ and $\gamma_{p-1}=\rho_{p} \sum_{t=0}^{T-p} \varphi_{t}$, respectively. Differentiating gives

$$
\nabla_{\rho_{p-1}} \gamma_{p}=\rho_{p} \sum_{t=0}^{T-p-1} \nabla_{\rho_{p-1}} \varphi_{t}=\rho_{p} \sum_{t=1}^{T-p} \nabla_{\rho_{p}} \varphi_{t}, \quad \nabla_{\rho_{p}} \gamma_{p-1}=\rho_{p} \sum_{t=1}^{T-p} \nabla_{\rho_{p}} \varphi_{t}+\sum_{k=0}^{T-p} \varphi_{t}
$$

using $\varphi_{0}=1$ and $\nabla_{\rho_{p-1}} \varphi_{t}=\nabla_{\rho_{p}} \varphi_{t+1}$. The latter follows from differentiating $\varphi_{t}$ and a change of variable from $k_{p-1}-1$ to $k_{p-1}$, giving

$$
\nabla_{\rho_{p-1}} \varphi_{t}=\sum_{\tau^{\prime} k=t-p+1} \frac{\left(\iota^{\prime} k+1\right)!}{k_{1}!\cdots k_{p}!} \rho^{k}
$$

which is invariant under a unit shift of $p$ and $t$. Therefore, $\nabla_{\rho_{p-1}} \gamma_{p} \neq \nabla_{\rho_{p}} \gamma_{p-1}$, and (A.1) is inconsistent.

Proof of Lemma 1. Let $A=S_{1} F_{0}$ and $B=\nabla_{\rho_{0}} A$. Then

$$
b_{0}=-\frac{\iota^{\prime} A \iota}{T(T-1)}, \quad c_{0}=-\frac{\iota^{\prime} B \iota}{T(T-1)}, \quad V_{0}^{L B}=\frac{\operatorname{tr} A^{\prime} M A}{T-1}=\frac{T \operatorname{tr} A A^{\prime}-\iota^{\prime} A A^{\prime} \iota}{T(T-1)} .
$$

Hence, $V_{0}^{L B} \geq 2 b_{0}^{2}$ and $V_{0}^{L B} \geq 2 b_{0}^{2}-c_{0}$ if and only if

$$
\begin{align*}
T \operatorname{tr} A A^{\prime}-\iota^{\prime} A A^{\prime} \iota-\frac{2\left(\iota^{\prime} A \iota\right)^{2}}{T(T-1)} & \geq 0  \tag{A.3}\\
T \operatorname{tr} A A^{\prime}-\iota^{\prime} A A^{\prime} \iota-\frac{2\left(\iota^{\prime} A \iota\right)^{2}}{T(T-1)}-\iota^{\prime} B \iota & \geq 0 \tag{A.4}
\end{align*}
$$

The matrix $A=A_{T}$ is

$$
\begin{aligned}
& A=\left(\begin{array}{ll}
0 & 0 \\
1 & 0
\end{array}\right), \quad T=2 \\
& A=\left(\begin{array}{cc}
A_{T-1} & 0 \\
a_{T}^{\prime} & 0
\end{array}\right), \quad a_{T}=\left(\rho^{T-2}, \rho^{T-3}, \ldots, 1\right)^{\prime}, \quad T>2
\end{aligned}
$$

where the subscript on $\rho$ is omitted. By recursion, it can be deduced that

$$
\begin{aligned}
& \iota^{\prime} A \iota=\sum_{j=0}^{T-2}(T-j-1) \rho^{j}, \quad \iota^{\prime} B \iota=\sum_{j=1}^{T-2} j(T-j-1) \rho^{j-1}, \\
& \operatorname{tr} A A^{\prime}=\sum_{j=0}^{T-2}(T-j-1) \rho^{2 j}, \quad \iota^{\prime} A A^{\prime} \iota=\sum_{j=0}^{T-2}\left(\sum_{k=0}^{j} \rho^{k}\right)^{2},
\end{aligned}
$$

yielding $V_{0}^{L B}$ as stated in the lemma. Now let $r>0$ and use the equalities just obtained to see that if (A.4) holds for $\rho=r$, then (A.3) holds for $\rho=r$ and (A.3) and (A.4) hold for $\rho=-r$, with strict inequalities for $T \geq 3$. Hence, we only need to show that (A.4) holds for $\rho \geq 0$, with equality if and only if $T=2$ or $\rho=1$. Write (A.4) as $Q_{T} \geq 0$. Because $Q_{2}=0$, to show that (A.4) holds, it suffices to show that $\Delta Q_{T} \geq 0$ for $T \geq 2$, where $\Delta(\cdot)_{T}=(\cdot)_{T+1}-(\cdot)_{T}$. Write $\Delta Q_{T}$ as

$$
\begin{aligned}
\Delta Q_{T}= & \Delta\left(T \operatorname{tr} A A^{\prime}-\iota^{\prime} A A^{\prime} \iota-2 \frac{\left(\iota^{\prime} A \iota\right)^{2}}{T(T-1)}-\iota^{\prime} B \iota\right)_{T} \\
= & \left\{\left(\operatorname{tr} A A^{\prime}\right)_{T+1}-2 \frac{\left(\iota^{\prime} A \iota\right)_{T+1}^{2}}{T(T+1)}\right\}+\left\{2 \frac{\left(\iota^{\prime} A \iota\right)_{T}^{2}}{T(T-1)}-\Delta\left(\iota^{\prime} B \iota\right)_{T}\right\} \\
& +\left\{T \Delta\left(\operatorname{tr} A A^{\prime}\right)_{T}-\Delta\left(\iota^{\prime} A A^{\prime} \iota\right)_{T}\right\}
\end{aligned}
$$

and denote the quantities in braces as $\tau_{1}, \tau_{2}$, and $\tau_{3}$. Using $T(T+1) / 2=\sum_{i=0}^{T-1}(T-i)$, we have

$$
\begin{aligned}
\tau_{1} & =\sum_{j=0}^{T-1}(T-j) \rho^{2 j}-\frac{2}{T(T+1)}\left(\sum_{j=0}^{T-1}(T-j) \rho^{j}\right)^{2} \\
& =\frac{2}{T(T+1)}\left(\sum_{i=0}^{T-1} \sum_{j=0}^{T-1}(T-i)(T-j) \rho^{2 j}-\sum_{i=0}^{T-1} \sum_{j=0}^{T-1}(T-i)(T-j) \rho^{i+j}\right) \\
& =\frac{2}{T(T+1)} u^{\prime} R u
\end{aligned}
$$

where $u=(T, T-1, \ldots, 1)^{\prime}$ and

$$
R=\left(\begin{array}{cccc}
1 & 1 & \cdots & 1 \\
\rho^{2} & \rho^{2} & \cdots & \rho^{2} \\
\vdots & \vdots & \ddots & \vdots \\
\rho^{2 T-2} & \rho^{2 T-2} & \cdots & \rho^{2 T-2}
\end{array}\right)-\left(\begin{array}{cccc}
1 & \rho & \cdots & \rho^{T-1} \\
\rho & \rho^{2} & \cdots & \rho^{T} \\
\vdots & \vdots & \ddots & \vdots \\
\rho^{T-1} & \rho^{T} & \cdots & \rho^{2 T-2}
\end{array}\right)
$$

Consider the principal minors of $R$. Those of order 1 are 0 ; those of order 2 are

$$
\operatorname{det}\left(\begin{array}{cc}
0 & \rho^{2 i}-\rho^{i+j} \\
\rho^{2 j}-\rho^{i+j} & 0
\end{array}\right)=\rho^{i+j}\left(\rho^{j}-\rho^{i}\right)^{2} \geq 0, \quad 0<i<j<T
$$

given $\rho \geq 0$; and those of order greater than 2 are 0 because $R$ is the sum of two matrices of rank 1 and, hence, $\operatorname{rank}(R) \leq 2$. Therefore, $R$ is positive semi-definite and $\tau_{1} \geq 0$. Furthermore,

$$
\tau_{3}=T \sum_{j=0}^{T-1} \rho^{2 j}-\left(\sum_{j=0}^{T-1} \rho^{j}\right)^{2}=\sum_{i=0}^{T-1} \sum_{j=0}^{T-1} \rho^{2 j}-\sum_{i=0}^{T-1} \sum_{j=0}^{T-1} \rho^{i+j}=\iota^{\prime} R \iota \geq 0
$$

Use

$$
\begin{aligned}
\Delta\left(\iota^{\prime} B \iota\right)_{T} & =\sum_{j=1}^{T-1} j(T-j) \rho^{j-1}-\sum_{j=1}^{T-2} j(T-j-1) \rho^{j-1}=\sum_{j=0}^{T-2}(j+1) \rho^{j} \\
& =\frac{2}{T(T-1)} \sum_{i=0}^{T-2} \sum_{j=0}^{T-2}(T-i-1)(j+1) \rho^{j}
\end{aligned}
$$

to write

$$
\tau_{2}=d\left(\sum_{i=0}^{T-2} \sum_{j=0}^{T-2}(T-i-1)(T-j-1) \rho^{i+j}-\sum_{i=0}^{T-2} \sum_{j=0}^{T-2}(T-i-1)(j+1) \rho^{j}\right),
$$

where $d=\frac{2}{T(T-1)}$. Note that $\tau_{2}$ is a polynomial of degree $2 T-4$ in $\rho$. When $T=2$ or $\rho=1$, $\tau_{2}=0$. When $\rho \neq 1$,

$$
\left(\iota^{\prime} A \iota\right)_{T}=\frac{T-1-T \rho+\rho^{T}}{(1-\rho)^{2}}, \quad \Delta\left(\iota^{\prime} B \iota\right)_{T}=\frac{1-\rho^{T}-T \rho^{T-1}+T \rho^{T}}{(1-\rho)^{2}}
$$

and so

$$
\tau_{2}=\frac{d\left(T-1-T \rho+\rho^{T}\right)^{2}}{(1-\rho)^{4}}-\frac{1-\rho^{T}-T \rho^{T-1}+T \rho^{T}}{(1-\rho)^{2}}
$$

For $T>2$,

$$
\lim _{\rho \rightarrow 1} \tau_{2}(1-\rho)^{-2}=\frac{1}{72} T(T-1)(T-2)(T+1)>0
$$

and, therefore, $\tau_{2}=(1-\rho)^{2} P(T, \rho)$, where $P(T, \rho)$ is a polynomial of degree $2 T-6$. If all coefficients of $P(T, \rho)=\sum_{j=0}^{2 T-6} p_{j} \rho^{j}$ are positive, we conclude that $\tau_{2} \geq 0$. Write $\tau_{2}=$ $\sum_{j=0}^{2 T-4} q_{j} \rho^{j}$, where $q_{j}$ is found as

$$
q_{j}= \begin{cases}\frac{d}{6}\{(j+1)(j(j-1)+6(T-1)(T-j-1))-3 j T(T-1)\}, & j \leq T-2, \\ \frac{d}{6}(2 T-j-1)(2 T-j-2)(2 T-j-3), & T-1 \leq j\end{cases}
$$

Equating the coefficients of $\tau_{2}$ and $(1-\rho)^{2} P(T, \rho)$ gives $p_{k}=\sum_{j=0}^{k}(k+1-j) q_{j}$. To show that $p_{k}>0$ for $0 \leq k \leq 2 T-6$, we only need to show that $p_{k}>0$ for $k$ up to $T-2$ because for larger $k, q_{k}>0$ and so $p_{k}$ increases in $k$. For $k$ up to $\min (T-2,2 T-6)$, we obtain

$$
p_{k}=\frac{d}{12}(k+1)(k+2)(k+3)\left((T-1)(2 T-k-2)+\frac{k}{10}(k-1)-T(T-1)\right)
$$

and, hence, $p_{k}>0$ because either $k<T-2$, implying $2 T-k-2>T$, or $k=T-2 \leq 2 T-6$, implying $T \geq 4$ and $k \geq 2$. Therefore, $\tau_{2} \geq 0$. This establishes $Q_{T} \geq 0$, that is, (A.4). Recall that $Q_{2}=0$ and note that $\rho=1$ implies $\tau_{1}=\tau_{2}=\tau_{3}=0$ and, hence, $Q_{T}=0$. Therefore, $Q_{T}=0$ if $T=2$ or $\rho=1$. If $T \geq 2$ and $\rho \neq 1$, then $\Delta Q_{T}>0$ because $\tau_{3}>0$ when $T=2$ and $\tau_{2}>0$ when $T>2$. Therefore, $Q_{T}=0$ only if $T=2$ or $\rho=1$.

Proof of Theorem 1. $L_{\mathrm{A}}(\rho)$ having a local maximum or a flat inflection point at $\rho_{0}$ is equivalent to $b(\rho)$ approaching $S(\rho)$ from below as $\rho$ approaches $\rho_{0}$ from the left. We will write this as $b(\rho) \uparrow S(\rho)$ at $\rho_{0}$, and show that $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \bar{\rho}]$ at most once. From

$$
\nabla_{\rho} H(\rho)=\frac{2\left(\rho-\rho_{M L}\right)\left(3 \zeta_{0}^{2}-\left(\rho-\rho_{M L}\right)^{2}\right)}{\left(\zeta_{0}^{2}+\left(\rho-\rho_{M L}\right)^{2}\right)^{3}}
$$

it follows that $S(\rho)$ is strictly concave on $\left[\rho, \rho_{M L}\right]$ and strictly convex on $\left[\rho_{M L}, \bar{\rho}\right]$. Because $\varphi_{t}=\rho^{t}, b(\rho)$ and its first two derivatives are

$$
\begin{aligned}
b(\rho) & =-\sum_{t=0}^{T-2} \frac{T-1-t}{T(T-1)} \rho^{t} \\
c(\rho) & =-\sum_{t=1}^{T-2} \frac{t(T-1-t)}{T(T-1)} \rho^{t-1}, \quad d(\rho)=-\sum_{t=2}^{T-2} \frac{t(t-1)(T-1-t)}{T(T-1)} \rho^{t-2}
\end{aligned}
$$

For $\rho \neq 1$,

$$
\begin{aligned}
b(\rho) & =-\frac{T-1-T \rho+\rho^{T}}{T(T-1)(1-\rho)^{2}}, \quad c(\rho)=-\frac{T-2-T \rho+T \rho^{T-1}-(T-2) \rho^{T}}{T(T-1)(1-\rho)^{3}} \\
d(\rho) & =-\frac{2 T-6-2 T \rho+T(T-1) \rho^{T-2}-2 T(T-3) \rho^{T-1}+(T-2)(T-3) \rho^{T}}{T(T-1)(1-\rho)^{4}} .
\end{aligned}
$$

When $T \leq 3, b(\rho)$ is linear and so, given that $S(\rho)$ is concave-convex on $[\underline{\rho}, \bar{\rho}], b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \bar{\rho}]$ at most once. Suppose $T \geq 4$. Then, $b(\rho)$ is a polynomial of degree 2 or higher with negative coefficients, so $b(\rho)$ is negative, decreasing, and strictly concave, on $\mathbb{R}_{+}$. Further, by Descartes' rule of signs, $c(\rho)$ has one zero on $\mathbb{R}_{-}$when $T$ is even and none when $T$ is odd, and $d(\rho)$ has no zeros on $\mathbb{R}_{-}$when $T$ is even and one when $T$ is odd. Suppose $T$ is even. Then $c(-1)=0$ and $b(-1)=-\frac{1}{2(T-1)}<0$, so $b(\rho)$ is negative and strictly concave on $\mathbb{R}$, and, hence, its intersection with $S(\rho)$ on $[\underline{\rho}, \bar{\rho}]$ can only be on $\left(\rho_{M L}, \bar{\rho}\right]$, where $S(\rho)$ is strictly convex and is approached from below by $b(\rho)$ at most once. Now suppose $T$ is odd and $T \geq 5$. Then,

$$
d(-1)=\frac{T-3}{4 T}>0, \quad d\left(-\frac{1}{2}\right)=-\frac{2^{4-T}(T-2)\left(2^{T}-3 T+1\right)}{27 T(T-1)}<0
$$

so $b(\rho)$ is strictly convex on $\left(-\infty, \rho_{v}\right]$ and strictly concave on $\left[\rho_{v}, \infty\right)$ for some $\rho_{v} \in\left(-1,-\frac{1}{2}\right)$ and decreases on $\mathbb{R}$. Define $\rho_{u}$ by $b\left(\rho_{u}\right)=0$, that is, by $T\left(1-\rho_{u}\right)=1-\rho_{u}^{T}, \rho_{u} \in \mathbb{R}_{-}$. Since $T \geq 5$, we have $-2<\rho_{u}<-1$. Thus, $b(\rho)$ is negative and strictly convex on $\left(\rho_{u}, \rho_{v}\right]$, with $-2<\rho_{u}<-1<\rho_{v}<-\frac{1}{2}$. Let $R=\left[\rho_{u}, \rho_{v}\right] \cap\left[\rho_{M L}, \bar{\rho}\right]$. If $R$ is empty, then $\rho_{v}<\rho_{M L}$ or $\bar{\rho}<\rho_{u}$; in either case, by the concavity-convexity of $S(\rho), b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \bar{\rho}]$ at most once. If $R$ is non-empty, to show that $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \bar{\rho}]$ at most once, it suffices to show that $S(\rho)$ decreases faster than $b(\rho)$ on $R$, i.e., $H(\rho)<c(\rho)$ for $\rho \in R$. We will show below that (i) $V_{0}^{L B} \geq \frac{T-1}{T}$ if $\rho_{0} \leq 0$; (ii) $V_{0}^{L B} \geq \frac{1}{2}$ if $\rho_{0}>0$. By (ii), $\rho_{M L}=\rho_{0}+b_{0} / V_{0} \geq \rho_{0}+2 b_{0}>-\frac{1}{2}$ if $0<\rho_{0} \leq 1$ because $b(0)=-\frac{1}{T}, b(1)=-\frac{1}{2}$, and $b(\rho)$ is concave on [0, 1]. Further, $\rho_{M L}>0$ if $\rho_{0}>1$ because, then, $\frac{b_{0}}{V_{0}}>\frac{1}{2 b_{0}}>-1$. Hence, $R$ is empty if $\rho_{0}>0$. Now suppose $\rho_{0} \leq 0$. Define $\rho_{w}$ by $S\left(\rho_{w}\right)=b\left(\rho_{v}\right), \rho_{w} \in\left[\rho_{M L}, \bar{\rho}\right]$; and $\rho_{w}^{\prime}$ by $S\left(\rho_{w}^{\prime}\right)=b(0)=-\frac{1}{T}, \rho_{w}^{\prime} \in\left[\rho_{M L}, \bar{\rho}\right]$. Then $\rho_{w}-\rho_{M L}<\rho_{w}^{\prime}-\rho_{M L}=\frac{1}{2}\left(T-\sqrt{T^{2}-4 \zeta_{0}^{2}}\right)$. By (i), $\zeta_{0}^{2}=\frac{V_{0}-b_{0}^{2}}{V_{0}^{2}} \leq \frac{1}{V_{0}} \leq \frac{T}{T-1} \leq \frac{5}{4}$. Since $H(\rho)$ increases on $\left[\rho_{M L}, \bar{\rho}\right]$ and $H\left(\rho_{w}^{\prime}\right)$ decreases in $T$ and increases in $\zeta_{0}^{2}$,

$$
\begin{aligned}
H\left(\rho_{w}\right) & =-\frac{\zeta_{0}^{2}}{\left(\zeta_{0}^{2}+\left(\rho_{w}-\rho_{M L}\right)^{2}\right)^{2}}+2 S^{2}\left(\rho_{w}\right)<-\frac{\zeta_{0}^{2}}{\left(\zeta_{0}^{2}+\left(\rho_{w}^{\prime}-\rho_{M L}\right)^{2}\right)^{2}}+\frac{2}{T^{2}} \\
& \leq-\frac{5 / 4}{\left(\frac{5}{4}+\frac{1}{4}(5-\sqrt{20})^{2}\right)^{2}}+\frac{2}{25}<-\frac{1}{2}
\end{aligned}
$$

and so, $H(\rho)<-\frac{1}{2}$ for $\rho \in\left[\rho_{M L}, \rho_{w}\right]$. On the other hand, $T\left(1-\rho_{u}\right)=1-\rho_{u}^{T}$ implies $\frac{1-\rho_{u}}{\rho_{u}}=$ $\frac{1-\rho_{u}^{T-1}}{T-1}$ and, therefore,

$$
c\left(\rho_{u}\right)=-\frac{-T+T \rho_{u}^{T-1}}{T(T-1)\left(1-\rho_{u}\right)^{2}}=\frac{1}{\rho_{u}\left(1-\rho_{u}\right)}>-\frac{1}{2} .
$$

So, $c(\rho)>-\frac{1}{2}$ for $\rho \in\left[\rho_{u}, \rho_{v}\right]$ and $H(\rho)<c(\rho)$ for $\rho \in R$. We conclude that $b(\rho) \uparrow S(\rho)$ on $[\rho, \bar{\rho}]$ at most once, provided (i) and (ii) hold, which we now show. Write $V_{0}^{L B}=\frac{1}{T(T-1)} \sum_{j=0}^{2 T-4} v_{j} \rho_{0}^{j}$, where

$$
\begin{aligned}
v_{2 j}= & T(T-j-1) \\
& -\left\{(2 j+1)(T-j-1)-j(j+1)+(2 j-T+1)(2 j-T+2) 1_{\{2 j \geq T\}}\right\}, \\
v_{2 j+1}= & -\left\{(2 j+2)(T-j-2)-j(j+1)+(2 j-T+2)(2 j-T+3) 1_{\{2 j+1 \geq T\}}\right\},
\end{aligned}
$$

$\operatorname{using}\left(\sum_{k=0}^{j} \rho^{k}\right)^{2}=\sum_{k=0}^{j}(k+1) \rho^{k}+\sum_{k=1}^{j}(j-k+1) \rho^{j+k}$. Clearly, $v_{2 j+1}<0$. Further, $v_{2 j}>0$ because

$$
v_{2 j}= \begin{cases}(T-2 j-1)(T-j-1)+j(j+1) & \text { if } 0 \leq 2 j<T \\ (T-j-1)(j+1) & \text { if } T \leq 2 j \leq 2 T-4\end{cases}
$$

Hence, $V_{0}^{L B}$ decreases in $\rho_{0}$ on $\mathbb{R}_{-}$and (i) follows because $V_{0}^{L B}=\frac{T-1}{T}$ when $\rho_{0}=0$. When $0<\rho_{0}<1$, a sufficient condition for $V_{0}^{L B} \geq \frac{1}{2}$ is that $d_{k} \geq 0$ for $0 \leq k \leq T-2$, where $d_{k}=\sum_{j=0}^{k}\left(v_{2 j}+v_{2 j+1}\right)-\frac{T(T-1)}{2}$. We have

$$
v_{2 j}+v_{2 j+1}= \begin{cases}(T-2 j-1)(T-j-1)-(T-2 j-2)(2 j+2) & \text { if } 2 j+1<T \\ (2 j-T+3)(T-j-1) & \text { if } 2 j+1 \geq T\end{cases}
$$

Only when $2 j+1<T$ is it possible that $v_{2 j}+v_{2 j+1}<0$, so it suffices to show that $d_{k} \geq 0$ for $2 k+1<T$. We obtain, for $2 k+1<T$,

$$
d_{k}=\frac{1}{2}(k+1)\left(2 T^{2}-5 T k+4 k^{2}-8 T+13 k+10\right)-\frac{T(T-1)}{2} .
$$

Define $f_{k}$ by $d_{k}=\frac{1}{2}(k+1) f_{k}$. Then, $f_{0}=(T-2)(T-5) \geq 0, f_{1}=\frac{1}{2}\left(3 T^{2}-25 T+54\right)>0$, and, for $k \geq 2$,

$$
\begin{aligned}
f_{k} & >\frac{5}{3} T^{2}-5 T k+4 k^{2}-8 T+13 k+10 \\
& =\frac{1}{3}((T-2 k-2)(5 T-6 k-16)+k(T-5)+2(T-1))>0 .
\end{aligned}
$$

Hence, $V_{0}^{L B} \geq \frac{1}{2}$ when $0<\rho_{0}<1$. When $\rho_{0} \geq 1$, it also holds that $V_{0}^{L B} \geq \frac{1}{2}$ because then $b_{0}<b(1)=-\frac{1}{2}$ and $V_{0}^{L B} \geq 2 b_{0}^{2}$. Therefore, (ii) holds.

Proof of Lemma 2. Use $y_{i,-1}=S_{1}\left(\xi_{0 i}+F_{0} \varepsilon_{i}\right)$ to write $Z_{i}=\left(y_{i,-1}, X_{i}\right)=\left(S_{1} F_{0} \varepsilon_{i}, 0\right)+\Xi_{i}$, where $\Xi_{i}=\left(S_{1} \xi_{0 i}, X_{i}\right)$ is independent of $\varepsilon_{i}$. Proceeding as above, we have

$$
V_{0}=\frac{\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} Z_{i}^{\prime} M Z_{i}}{\sigma_{0}^{2}(T-1)}=\left(\begin{array}{cc}
V_{0}^{L B} & 0 \\
0 & 0
\end{array}\right)+V_{\Xi}
$$

where

$$
V_{\Xi}=\left(\begin{array}{cc}
V_{\xi \xi} & V_{\xi X} \\
V_{X \xi} & V_{X X}
\end{array}\right)=\frac{\operatorname{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \Xi_{i}^{\prime} M \Xi_{i}}{\sigma_{0}^{2}(T-1)}
$$

is positive semi-definite and $V_{X X}$ is positive definite by assumption. Therefore, $V_{\xi \xi}-V_{\xi X} V_{X X}^{-1} V_{X \xi} \geq$ 0 and $\left(V_{0}^{\rho \rho}\right)^{-1}=V_{0}^{L B}+V_{\xi \xi}-V_{\xi X} V_{X X}^{-1} V_{X \xi} \geq V_{0}^{L B}$.

## References

Ahn, S. C. and P. Schmidt. (1995). Efficient estimation of models for dynamic panel data. Journal of Econometrics, 68:5-27.
Ahn, S. C. and G. M. Thomas. (2006). Likelihood based inference for dynamic panel data models. Unpublished manuscript.
Alvarez, J. and M. Arellano. (2004). Robust likelihood estimation of dynamic panel data models. Unpublished manuscript.
Arellano, M. (2003). Panel Data Econometrics. Advanced Texts in Econometrics. Oxford University Press.
Arellano, M. and S. Bond. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies, 58:277-297.
Arellano, M. and S. Bonhomme. (2009) Robust priors in nonlinear panel data models. Econometrica, 77:489-536.
Bun, M. J. and M. A. Carree. (2005). Bias-corrected estimation in dynamic panel data models. Journal of Business and Economic Statistics, 23:200-210.
Chamberlain, G. and M. J. Moreira. (2009). Decision theory applied to a linear panel data model. Econometrica, 77:107-133.
Conniffe, D. and J. E. Spencer. (2001). When moments of ratios are ratios of moments. The Statistician, 50:161-168.
Cox, D. R. and N. Reid. (1987). Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society, Series B, 49:1-39.
Cruddas, A., Reid, N. and D. Cox. (1989). A time series illustration of approximate conditional likelihood. Biometrika, 76:231-237.
Dhaene, G. and K. Jochmans. (2007). An adjusted profile likelihood for non-stationary panel data models with incidental parameters. Unpublished manuscript.
Fisher, R. (1930). The moments of the distribution for normal samples of measures of departure from normality. Proceedings of the Royal Society of London, Series A, 130:16-28.
Geary, R. (1933). A general expression for the moments of certain symmetrical functions of normal samples. Biometrika, 25:184-186.

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, 109:107150.

Kiefer, N. M. (1980). Estimation of fixed effects models for time series of cross-sections with arbitrary intertemporal covariance. Journal of Econometrics, 14:195-202.
Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics, 68:53-78.
Kruiniger, H. (2008). Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model. Journal of Econometrics, 144:447-464.
Lancaster, T. (2002). Orthogonal parameters and panel data. Review of Economic Studies, 69:647-666.
McCullagh, P. and R. Tibshirani. (1990). A simple method for the adjustment of profile likelihoods. Journal of the Royal Statistical Society, Series B, 52:325-344.
Moreira, M. J. (2009). A maximum likelihood method for the incidental parameter problem. Annals of Statistics, 37:3660-3696.
Nerlove, M. (1967). Experimental evidence on the estimation of dynamic economic relations from a time series of cross-sections. Economic Studies Quarterly, 18:42-74.
Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross-sections. Econometrica, 39:359-382.
Neyman, J. and E. L. Scott. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16:1-32.
Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica, 49:1417-1426.
Phillips, P. C. B. and D. Sul. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. Journal of Econometrics, 137:162-188.
Rotnitzky, A., D. R. Cox, M. Bottai, and J. Robins. (2000). Likelihood-based inference with singular information matrix. Bernoulli, 6:243-284.
Sargan, J. D. (1983). Identification and lack of identification. Econometrica, 51:1605-1633.
Severini, T. A. (2000). Likelihood Methods in Statistics. Oxford Statistical Science Series. Oxford University Press.


[^0]:    *Address: K.U. Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium. Tel. +3216 326798; Fax +32 16 326796; E-mail: geert.dhaene@econ.kuleuven.be.
    ${ }^{\dagger}$ Address: U.C. Louvain, Center for Operations Research and Econometrics, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Tel. +3210 474329; E-mail: koen.jochmans@uclouvain.be.
    ${ }^{1}$ Throughout, non-stationarity will refer to initial observations being left unrestricted (and being conditioned upon) and the possibility that the autoregressive parameter vector lies outside the stationary region. Likelihoodbased estimation of $\operatorname{AR}(1)$ models under restrictions on the initial observations has been considered by Kiefer (1980), Cruddas, Reid, and Cox (1989), Hsiao, Pesaran, and Tahmiscioglu (2002), and Kruiniger (2008).

[^1]:    ${ }^{2}$ Our bias calculation is related to Alvarez and Arellano's (2004) in the case of time series homoskedasticity. We calculate the bias of the profile score for the regression coefficients, while they calculated the bias of the conditional score for all common parameters given maximum likelihood estimates of the fixed effects.

[^2]:    ${ }^{3}$ For a discussion and historical perspective on this device, see Conniffe and Spencer (2001).

[^3]:    ${ }^{4}$ The adjustment, $-a(\rho)$, is robust to cross-sectional heteroskedasticity ( $\sigma_{i}^{2}$ instead of $\sigma^{2}$, say). We investigated this in an older version of this paper (Dhaene and Jochmans, 2007), where we also derived the adjustment for the $\mathrm{AR}(1)$ model when there are incidental time trends. In the latter case, the adjustment is still a polynomial.
    ${ }^{5}$ Unbalanced panel data can be accommodated by writing $l(\theta)$ as a weighted average of $K$ balanced subpanel profile log-likelihoods, say $l_{k}(\theta)$, and adjusting each $l_{k}(\theta)$. Missing data, with observations missing completely at random, fit into this scheme by forming complete subseries from each incomplete time series.

[^4]:    ${ }^{7}$ It has already been observed that, in finite samples, a bias-adjusted likelihood may be re-increasing. Moreover, Lancaster (2002) presented graphs of a re-increasing marginal posterior log-density for $\rho$ in an AR(1) model with time dummies. Arellano and Bonhomme (2009), using a slightly different prior, presented a graph of a re-increasing integrated likelihood for an $\mathrm{AR}(1)$ model. Because $l(\theta)$ is log-quadratic for any $N \geq 1$ and $a(\rho)$ does not depend on the data, $l_{\mathrm{A}}(\theta)$ is re-increasing regardless of the sample size.

[^5]:    ${ }^{8} \mathrm{~A}$ major difficulty is the rapidly increasing complexity of $\varphi_{t}$ as $p$ increases. For example, $\varphi_{t}=$ $\sum_{k=0}^{\lfloor t / 2\rfloor} \frac{(t-k)!}{(t-2 k)!k!} \rho_{1}^{t-2 k} \rho_{2}^{k}$ when $p=2$. Compare this with $\varphi_{t}=\rho_{1}^{t}$ when $p=1$ and note that $V_{0}^{L B}$, $b_{0}$, and $c_{0}$ depend in non-trivial ways on the $\varphi_{t}$ 's.

[^6]:    ${ }^{9}$ The same computations but with $T=8,9,10$ further supported the conclusions. Here, however, when $m=4$ and $p=3,4$ the computations are numerically less stable because the polynomial terms may be extremely large and their sum numerically imprecise. It is natural to expect, however, that (3.1) and (3.2) hold for all $T$ if they hold for the small values of $T$ considered.
    ${ }^{10}$ Computations with $T=8,9,10$ gave the same results except in certain cases with $m=4$ and $p=3,4$ where the algorithm failed to converge because the rank of $H_{\mathrm{A}}(\rho)$ was underestimated.

[^7]:    ${ }^{11}$ For the GMM estimator we used all the moments that follow from the exogeneity of $x_{i t}$, i.e. $\mathbb{E}\left[x_{i j}\left(\varepsilon_{i t}-\varepsilon_{i t-1}\right)\right]=0, j=1, \ldots, T ; t=2, \ldots, T$; and $\mathbb{E}\left[y_{i t-j}\left(\varepsilon_{i t}-\varepsilon_{i t-1}\right)\right]=0, j=2, \ldots, t ; t=2, \ldots, T$. It should be noted that the GMM estimator remains consistent under heteroskedasticity while the bias calculation underlying the AL estimator requires time series homoskedasticity.

[^8]:    ${ }^{12}$ The AL estimator is invariant with respect to the fixed effects, but the GMM estimator is not.
    ${ }^{13}$ In the $\mathrm{AR}(1)$ model without covariates, for example, $S_{1} \xi_{0 i}=g_{0}\left(y_{i 0}-\mu_{i}\right)+\iota \mu_{i}$ with $g_{0}=\left(1, \rho_{0}, \ldots, \rho_{0}^{T-1}\right)^{\prime}$, hence $V_{\xi \xi}=\frac{g_{0}^{\prime} M g_{0}}{\left(1-\rho_{0}^{2}\right)(T-1)} \psi^{2}$. As $\rho_{0} \uparrow 1, V_{\xi \xi} \rightarrow 0$ for any fixed $\psi$.
    ${ }^{14}$ In the $\operatorname{AR}(2)$ model, when $\psi=0$ the weighting matrix used by GMM is singular for all $T$ and the likelihood function does not depend on $\rho_{2}$ for $T=2$. Therefore, we set $\psi>0$ although it is possible to work with $\psi=0$ for $T \geq 3$ by dropping the earliest observation from the set of instruments.
    ${ }^{15}$ To compute $\mathrm{CI}_{.95}^{B}$, we used the bootstrap percentile method with 39 bootstrap samples formed by randomly drawing $N$ units $i$ with replacement from $\{1, \ldots, N\}$. A small number of bootstrap draws suffices for studying coverage rates via simulation.
    ${ }^{16}$ When $\psi=0$ and $T=2, l(\rho)$ depends on $\rho_{0}$ and $\rho$ only through $\rho_{0}-\rho$, so the distribution of $\widehat{\rho}_{\text {AL }}$ is

