Likelihood inference in an Autoregression with fixed effects

Geert Dhaene, Koen Jochmans

Discussion paper 2013-07

Likelihood inference in an autoregression with fixed effects

Geert Dhaene
Koen Jochmans

Sciences Po Economics Discussion Papers

LIKELIHOOD INFERENCE IN AN AUTOREGRESSION WITH FIXED EFFECTS

GEERT DHAENE[1] and KOEN JOCHMANS[2*]

[1] *Department of Economics, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium.*

[2] *Department of Economics, Sciences Po, 28 rue des Saint-Pères, 75007 Paris, France.*

We calculate the bias of the profile score for the regression coefficients in a multistratum autoregressive model with stratum-specific intercepts. The bias is free of incidental parameters. Centering the profile score delivers an unbiased estimating equation and, upon integration, an adjusted profile likelihood. A variety of other approaches to constructing modified profile likelihoods are shown to yield equivalent results. However, the global maximizer of the adjusted likelihood lies at infinity for any sample size, and the adjusted profile score has multiple zeros. We argue that the parameters are local maximizers inside or on an ellipsoid centered at the maximum likelihood estimator.

Keywords: adjusted likelihood, autoregression, incidental parameters, local maximizer, recentered estimating equation.

1. Introduction

In the presence of nuisance parameters, inference based on the profile likelihood can be highly misleading. In an $N \times T$ data array setting with stratum nuisance parameters, the maximum likelihood estimator is often inconsistent as the number of strata, N, tends to infinity. This is the incidental parameter problem (Neyman and Scott, 1948). It arises because profiling out the nuisance parameters from the likelihood introduces a bias into the (profile) score function. One possible solution is to calculate this bias and to subtract it from the profile score, as suggested by Neyman and Scott (1948, Section 5) and McCullagh and Tibshirani (1990). When the bias is free of incidental parameters this yields a fully recentered score function which, in principle, paves the way for consistent estimation under Neyman-Scott asymptotics (Godambe and Thompson, 1974). This is the case in the classic many-normal-means example, but little is known about this possibility in other situations.

In this paper we consider a time series extension of Neyman and Scott's (1948) classic example. The problem is to estimate a pth order autoregressive model, possibly augmented with covariates, from data on N short time series of length T. The model has stratum-specific intercepts (the fixed effects). The distribution of the initial observations is left unrestricted and the $p \times 1$ vector of autoregressive parameters, ρ, may lie outside the stationary region. The bias of the profile score is found to depend only on ρ and T. Hence, adjusting the profile score by subtracting its bias gives a fixed T unbiased estimating equation and, upon integration, an adjusted profile likelihood in the sense of Pace and Salvan (2006).

However, contrary to what standard maximum likelihood theory would suggest, the parameters of interest are *local* maximizers of the adjusted likelihood. The global maximum is reached at infinity. This phenomenon is not a small sample problem or an artifact of an unbounded parameter space. The adjusted likelihood has its global maximum at infinity for any sample size, and may already be re-increasing in the stationary parameter region and reach its maximum at the boundary. Consequently, consistent estimation is not achieved by global maximization of the adjusted likelihood, and solving the adjusted score equation has to be supplemented by a

[*]Corresponding author. E-mail: koen.jochmans@sciences-po.org. Previous versions of this paper circulated as 'An adjusted profile likelihood for non-stationary panel data models with incidental parameters' and 'An adjusted profile likelihood for non-stationary panel data models with fixed effects'. This version: November 13, 2012.

solution selection rule, which we derive. The adjusted likelihood is re-increasing because the initial observations are unrestricted. This difficulty does not arise when stationarity of the initial observations is imposed, as in Cruddas, Reid, and Cox (1989). Further, when the data carry only little information, in a sense that we specify, the Hessian of the adjusted likelihood is zero, implying first-order underidentification (Sargan, 1983) and non-standard asymptotic properties of the resulting point estimates (Rotnitzky, Cox, Bottai, and Robins, 2000).

These features are not unique to our approach of modifying the profile likelihood. We show that several other routes to constructing modified likelihoods yield the same results. When $p = 1$, the adjusted profile likelihood coincides with Lancaster's (2002) marginal posterior, which, in the absence of covariates, is a Bayesian version of a Cox and Reid (1987) approximate conditional likelihood (see Sweeting, 1987). For general p, it is an integrated likelihood in the sense of Kalbfleisch and Sprott (1970) and Arellano and Bonhomme (2009), as well as a penalized likelihood as defined by Bester and Hansen (2009) (see DiCiccio, Martin, Stern, and Young, 1996 and Severini, 1998 for related approaches). The adjusted profile score equation, in turn, is a Woutersen (2002) integrated moment equation and a locally orthogonal Cox and Reid (1987) moment equation, and solving it is equivalent to inverting the probability limit of the least-squares estimator, as proposed by Bun and Carree (2005).

2. Adjusted profile likelihood

2.1. 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 strata i and T periods t. Assume that y_{it} is generated by

$$y_{it} = y_{it-}^{\top}\rho + x_{it}^{\top}\beta + \alpha_i + \varepsilon_{it}, \qquad i = 1, ..., N; t = 1, ..., T, \tag{2.1}$$

where $y_{it-} = (y_{it-1}, ..., y_{it-p})^{\top}$ and the ε_{it} are identically distributed with mean zero and variance σ^2 and are independent across i and t and also of $x_{i't'}$ for all i' and t'. Let $y_i^0 = (y_{i(1-p)}, ..., y_{i0})^{\top}$, $X_i = (x_{i1}, ..., x_{iT})^{\top}$, and $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})^{\top}$. We place no restrictions on how (y_i^0, α_i, X_i), $i = 1, ..., N$, are generated. The unknown parameters are $\theta = (\rho^{\top}, \beta^{\top})^{\top}$, σ^2, and $\alpha_1, ..., \alpha_N$. Let θ_0 and σ_0^2 be the true values of θ and σ^2. Our interest lies in consistently estimating θ_0 under large N and fixed T asymptotics. We do not impose the stationarity condition on ρ_0, i.e., we allow any $\rho_0 \in \mathbb{R}^p$.

Let $z_{it} = (y_{it-}^{\top}, x_{it}^{\top})^{\top}$, $Y_{i-} = (y_{i1-}, ..., y_{iT-})^{\top}$, $Z_i = (Y_{i-}, X_i)$, and $y_i = (y_{i1}, ..., y_{iT})^{\top}$, so that $My_i = MZ_i\theta + M\varepsilon_i$ where $M = I_T - T^{-1}\iota\iota^{\top}$ and ι is a conformable vector of ones. We assume that $N^{-1}\sum_{i=1}^N Z_i^{\top} M Z_i$ and its probability limit as $N \to \infty$ are nonsingular. The Gaussian quasi-log-likelihood, conditional on $y_1^0, ..., y_N^0$ and normalized by the number of observations, is

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

where, here and later, c is a non-essential constant. Profiling out $\alpha_1, ..., \alpha_N$ and σ^2 gives the (normalized) profile log-likelihood for θ,

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

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

$$\begin{aligned} s_{\rho_j}(\theta) &= \frac{\sum_{i=1}^N (y_i - Z_i\theta)^\top M y_{i,-j}}{\sum_{i=1}^N (y_i - Z_i\theta)^\top M (y_i - Z_i\theta)}, \qquad j = 1, ..., p, \\ s_{\beta_j}(\theta) &= \frac{\sum_{i=1}^N (y_i - Z_i\theta)^\top M x_{i,j}}{\sum_{i=1}^N (y_i - Z_i\theta)^\top M (y_i - Z_i\theta)}, \qquad j = 1, ..., q, \end{aligned}$$

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

For the analysis below, rewrite (2.1) as

$$Dy_i = Cy_i^0 + X_i\beta + \iota\alpha_i + \varepsilon_i, \qquad i = 1, ..., N,$$

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

$$D = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -\rho_1 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -\rho_p & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\rho_p & \cdots & -\rho_1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} \rho_p & \cdots & \cdots & \cdots & \rho_1 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \rho_p \\ 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}.$$

Then

$$\begin{pmatrix} y_i^0 \\ y_i \end{pmatrix} = \xi_i + F\varepsilon_i, \qquad \xi_i = \begin{pmatrix} y_i^0 \\ D^{-1}\left(Cy_i^0 + X_i\beta + \iota\alpha_i\right) \end{pmatrix}, \qquad F = \begin{pmatrix} 0 \\ D^{-1} \end{pmatrix}, \tag{2.2}$$

and $y_{i,-j} = S_j(\xi_i + F\varepsilon_i)$, where $S_j = (0_{T\times(p-j)}, I_T, 0_{T\times j})$, a selection matrix.

2.2. Bias of the profile score

The profile score is asymptotically biased, i.e., $\text{plim}_{N\to\infty} s(\theta_0) \neq 0$. Hence, the maximum likelihood estimator, solving $s(\theta) = 0$, is inconsistent. (Throughout, probability limits and expectations are taken conditionally, given (y_i^0, α_i, X_i), $i = 1, ..., N$.) The profile score bias is a polynomial in ρ_0. For $k = (k_1, ..., k_p)^\top \in \mathbb{N}^p$, let $\rho^k = \prod_{j=1}^p \rho_j^{k_j}$. Also, let $\tau = (1, ..., p)^\top$,

$$\varphi_t = \sum_{\tau^\top k = t} \frac{(\iota^\top k)!}{k_1! \cdots k_p!}\rho^k, \qquad t = 1, \ldots, T-1, \tag{2.3}$$

and set $\varphi_0 = 0$.

LEMMA 1. *The asymptotic bias of the profile score is* $\text{plim}_{N\to\infty} s(\theta_0) = b(\rho_0)$, *where* $b(\rho) = (b_1(\rho), \ldots, b_{p+q}(\rho))^\top$ *and*

$$\begin{aligned} b_j(\rho) &= -\textstyle\sum_{t=0}^{T-j-1} \frac{T-j-t}{T(T-1)}\varphi_t, & j &= 1, \ldots, p, \\ b_j(\rho) &= 0, & j &= p+1, \ldots, p+q. \end{aligned}$$

In addition, if $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_0^2)$, *then* $\mathbb{E}[s(\theta_0)] = b(\rho_0)$.

The bias of the profile score, $b(\rho_0)$, depends only on ρ_0 and T. It is independent of the initial observations, the fixed effects, and the covariates. This is in sharp contrast with the bias of the maximum likelihood estimator, which was first derived by Nickell (1981) for the first order autoregressive model under the assumption of stationarity of the initial observations. This bias depends on the initial observations, the fixed effects, and the covariate values. Note, also, that the Nickell bias concerns a probability limit as $N \to \infty$ whereas here, when the errors are normal, $\mathbb{E}[s(\theta_0)] = b(\rho_0)$ is a finite sample result holding for fixed N and T and may therefore be of independent interest in a time series setting.

2.3. 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., $\mathrm{plim}_{N\to\infty} s_{\mathrm{a}}(\theta_0) = 0$. Hence, $s_{\mathrm{a}}(\theta) = 0$ is a bias-adjusted estimating equation. The question arises whether there is a corresponding adjustment to the profile likelihood. The differential equation $\nabla_\theta a(\rho) = b(\rho)$ has a solution indeed.

LEMMA 2. *Up to an arbitrary constant of integration, the solution to $\nabla_\theta a(\rho) = b(\rho)$ is given by*

$$a(\rho) = \sum_{S \in \mathcal{S}} a_S(\rho), \qquad a_S(\rho) = -\sum_{t=|S|}^{T-1} \frac{T-t}{T(T-1)} \sum_{k \in \mathcal{K}_S : \tau^\top k = t} \frac{(\iota^\top k - 1)!}{k_1! \cdots k_p!} \rho_S^{k_S},$$

where $\mathcal{S}$ is the collection of the non-empty subsets of $\{1, ..., p\}$; $|S|$ is the sum of the elements of S; $\mathcal{K}_S = \{k \in \mathbb{N}^p | k_j > 0$ if and only if $j \in S\}$; and $\rho_S = (\rho_j)_{j\in S}$ and $k_S = (k_j)_{j \in S}$ are subvectors of ρ and k determined by S.

It follows that $s_{\mathrm{a}}(\theta) = 0$ is an estimating equation associated with the function

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

which we call an adjusted profile log-likelihood. Every subvector ρ_S of ρ contributes to $l_{\mathrm{a}}(\theta)$ an adjustment term, $-a_S(\rho)$, which takes the form of a multivariate polynomial in ρ_j, $j \in S$, with positive coefficients that are independent of p.

3. Connections with the literature

Lancaster (2002) studied the first-order autoregressive model, with and without covariates, from a Bayesian perspective. With $p = 1$, we have $\varphi_t = \rho^t$ and

$$b_1(\rho) = -\sum_{t=1}^{T-1} \frac{T-t}{T(T-1)} \rho^{t-1}, \qquad a(\rho) = -\sum_{t=1}^{T-1} \frac{T-t}{T(T-1)t} \rho^t.$$

With independent uniform priors on the reparameterized effects $\eta_i = \alpha_i e^{-(T-1)a(\rho)}$ and on θ and $\log \sigma^2$, Lancaster's posterior for $\vartheta = (\theta^\top, \sigma^2)^\top$ is

$$f(\vartheta|\text{data}) \propto \sigma^{-N(T-1)-2} \exp\left(-N(T-1)a(\rho) - Q^2(\theta)\sigma^{-2}/2\right),$$

where $Q^2(\theta) = \sum_{i=1}^N (y_i - Z_i\theta)^\top M(y_i - Z_i\theta) \propto e^{-2l(\theta)}$. Integrating over σ^2 gives

$$f(\theta|\text{data}) \propto e^{-N(T-1)a(\rho)}(Q^2(\theta))^{-N(T-1)/2}$$

and, hence,

$$f(\theta|\text{data}) \propto e^{N(T-1)l_{\mathrm{a}}(\theta)}. \tag{3.1}$$

Thus, the posterior and the adjusted likelihood are equivalent. More generally, for any p and q, independent uniform priors on $\eta_1, ..., \eta_N, \theta, \log\sigma^2$, with $\eta_i = \alpha_i e^{-(T-1)a(\rho)}$ and $a(\rho)$ as in Lemma 2, yield a posterior $f(\theta|\text{data})$ that is related to $l_{\mathrm{a}}(\theta)$ as in (3.1).

Lancaster's choice of a prior on the reparameterized effects η_i that is independent of ϑ is motivated by a first-order autoregression without covariates, where η_i is orthogonal to ϑ and the posterior $f(\theta|\text{data})$ (hence also $e^{l_{\mathrm{a}}(\theta)}$) has an interpretation as a Cox and Reid (1987) approximate conditional likelihood; see also Sweeting (1987). Orthogonalization to a multidimensional parameter is generally not possible (see, e.g., Severini, 2000, pp. 340–342). Here, orthogonalization is not possible when the model is augmented with covariates, as shown by Lancaster, or when the autoregressive order, p, is greater than one, as we show in Appendix A. From a bias correction perspective, however, orthogonality is sufficient but not necessary. In the present model, 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.

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 $o(T^{-1})$ as opposed to the standard $O(T^{-1})$. The present 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 $\pi_i(\alpha_i|\vartheta)$, the (normalized) log integrated likelihood is

$$l_{\mathrm{int}}(\vartheta) = \frac{1}{NT}\sum_{i=1}^N \log \int \sigma^{-T/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^T (y_{it} - z_{it}^\top\theta - \alpha_i)^2\right)\pi_i(\alpha_i|\vartheta)\;\mathrm{d}\alpha_i + \mathrm{c}.$$

Choosing $\pi_i(\alpha_i|\vartheta) \propto e^{-(T-1)a(\rho)}$ yields

$$l_{\mathrm{int}}(\vartheta) = -\frac{T-1}{2T}\log\sigma^2 - \frac{T-1}{T}a(\rho) - \frac{Q^2(\theta)}{2NT\sigma^2} + \mathrm{c}.$$

Profiling out σ^2 gives $\sigma^2(\theta) = \arg\max_{\sigma^2} l_{\mathrm{int}}(\vartheta) = Q^2(\theta)/(N(T-1))$ and

$$l_{\mathrm{int}}(\theta) = \max_{\sigma^2} l_{\mathrm{int}}(\vartheta) = \frac{T-1}{T} l_{\mathrm{a}}(\theta) + \mathrm{c},$$

so $l_{\mathrm{int}}(\theta)$ and $l_{\mathrm{a}}(\theta)$ are equivalent. Because $a(\rho)$ does not depend on true parameter values, $\pi_i(\alpha_i|\vartheta) \propto e^{-(T-1)a(\rho)}$ is a data-independent bias-reducing (in fact, bias-eliminating) prior in the sense of Arellano and Bonhomme. Now, $\pi_i(\alpha_i|\vartheta) \propto e^{-(T-1)a(\rho)}$ is equivalent to $\pi_i(\eta_i|\vartheta) \propto 1$, i.e., to a uniform prior on $\eta_i = \alpha_i e^{-(T-1)a(\rho)}$, leading to the same $l_{\mathrm{int}}(\vartheta)$. Arellano and Bonhomme (2009, Eq. (11)) give a necessary and sufficient condition for a uniform prior to be bias-reducing. With $\ell_i(\vartheta, \eta_i) = T^{-1}\sum_{t=1}^T \ell_{it}(\vartheta, \eta_i)$ denoting i's (normalized) log-likelihood contribution in a parametrization η_i, the condition is that

$$\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\nabla_{\eta_i}(A_i^{-1}B_i)=o(1)\qquad\text{as } T\to\infty,\tag{3.2}$$

where $A_i = A_i(\vartheta,\eta_i) = -\mathbb{E}_{\vartheta,\eta_i}\nabla_{\eta_i\eta_i}\ell_i(\vartheta,\eta_i)$, $B_i = B_i(\vartheta,\eta_i) = \mathbb{E}_{\vartheta,\eta_i}\nabla_{\vartheta\eta_i}\ell_i(\vartheta,\eta_i)$, and $\nabla_{\eta_i}(A_i^{-1}B_i)$ is evaluated at the true parameter values. When η_i and ϑ are orthogonal, $B_i=0$ and (3.2) holds. However, Condition (3.2) 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 (3.2) must hold for $\eta_i=\alpha_i e^{-(T-1)a(\rho)}$. Indeed, as we show in Appendix A,

$$\nabla_{\eta_i}(A_i^{-1}B_i)=0\tag{3.3}$$

because $A_i^{-1}B_i$ is free of η_i.

Woutersen (2002) derived a likelihood-based moment condition in which parameters of interest and fixed effects are orthogonal by construction even though orthogonality in the information matrix may not be possible. With $\ell_i=\ell_i(\vartheta,\alpha_i)=\sum_{t=1}^{T}\ell_{it}(\vartheta,\alpha_i)$ a generic log-likelihood for stratum i, let

$$g_i = g_i(\vartheta,\alpha_i) = \nabla_{\vartheta}\ell_i-\nabla_{\alpha_i}\ell_i\frac{\mathbb{E}_{\vartheta,\alpha_i}\nabla_{\alpha_i\vartheta}\ell_i}{\mathbb{E}_{\vartheta,\alpha_i}\nabla_{\alpha_i\alpha_i}\ell_i}.\tag{3.4}$$

Then $\mathbb{E}_{\vartheta,\alpha_i}g_i=0$ and parameter orthogonality holds in the sense that $\mathbb{E}_{\vartheta,\alpha_i}\nabla_{\alpha_i}g_i=0$ (under regularity conditions). Woutersen's (2002) integrated moment estimator of ϑ minimizes $g_{\mathrm{int}}^{\top}g_{\mathrm{int}}$ where $g_{\mathrm{int}}=(NT)^{-1}\sum_{i=1}^{N}g_{\mathrm{int}i}$ and

$$g_{\mathrm{int}i}=g_{\mathrm{int}i}(\vartheta)=\left[g_i-\frac{1}{2}\frac{\nabla_{\alpha_i\alpha_i}g_i}{\nabla_{\alpha_i\alpha_i}\ell_i}+\frac{1}{2}\frac{\nabla_{\alpha_i\alpha_i\alpha_i}\ell_i}{\nabla_{\alpha_i\alpha_i}\ell_i}\nabla_{\alpha_i}g_i\right]_{\alpha_i=\widehat{\alpha}_i(\vartheta)},$$

with $\widehat{\alpha}_i(\vartheta)=\arg\max_{\alpha_i}\ell_i$. The function $g_{\mathrm{int}i}$ is the Laplace approximation to $\int g_i e^{\ell_i}\mathrm{d}\alpha_i/\int e^{\ell_i}\mathrm{d}\alpha_i$, that is, to g_i with α_i integrated out using likelihood weights. Arellano (2003) obtained the same $g_{\mathrm{int}i}$ as a locally orthogonal Cox and Reid (1987) moment function. Woutersen and Voia (2004) calculated g_{int} for the present model with $p=1$. For any p and q, the integrated moment condition essentially coincides with the adjusted profile score. In Appendix A, it is shown that

$$g_{\mathrm{int}i}(\theta,\sigma^2)=\begin{pmatrix}\sigma^{-2}Z_i^{\top}M(y_i-Z_i\theta)-(T-1)\,b(\rho)\\ \sigma^{-4}(y_i-Z_i\theta)^{\top}M(y_i-Z_i\theta)/2-\sigma^{-2}\,(T-1)\,/2\end{pmatrix}.\tag{3.5}$$

On profiling out σ^2 from the minimand $g_{\mathrm{int}}^{\top}g_{\mathrm{int}}$, we obtain

$$g_{\mathrm{int}}(\theta)=\frac{T-1}{T}(s(\theta)-b(\rho))=\frac{T-1}{T}s_{\mathrm{a}}(\theta).$$

Thus, Woutersen's (2002) estimator of θ minimizes the norm of the adjusted profile score.

The adjusted likelihood can also be viewed as a penalized log-likelihood in the sense of Bester and Hansen (2009). With $\ell=\sum_{i=1}^{N}\sum_{t=1}^{T}\ell_{it}$, $\ell_{it}=\ell_{it}(\vartheta,\alpha_i)$, again denoting a generic log-likelihood, let $\pi_i=\pi_i(\vartheta,\alpha_i)$ be a function satisfying

$$\nabla_{\alpha_i}\pi_i\quad\xrightarrow{p}\quad\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\nabla_{\alpha_i\alpha_i}\ell_{it}\sum_{t=1}^{T}\psi_{it}\right]+\frac{1}{2}\mathbb{E}\left[\nabla_{\alpha_i\alpha_i\alpha_i}\ell_{it}\right]\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_{it}\sum_{t=1}^{T}\psi_{it}\right],\tag{3.6}$$

$$\nabla_{\vartheta}\pi_i\quad\xrightarrow{p}\quad\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\nabla_{\alpha_i\vartheta}\ell_{it}\sum_{t=1}^{T}\psi_{it}\right]+\frac{1}{2}\mathbb{E}\left[\nabla_{\alpha_i\alpha_i\vartheta}\ell_{it}\right]\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_{it}\sum_{t=1}^{T}\psi_{it}\right],\tag{3.7}$$

where $\psi_{it} = -\mathbb{E}\left[\nabla_{\alpha_i\alpha_i}\ell_{it}\right]^{-1}\nabla_{\alpha_i}\ell_{it}$. Then $\ell_\pi = \ell - \sum_{i=1}^{n}\pi_i$ is a penalized log-likelihood. Bester and Hansen (2009) provide a function that satisfies (3.6)–(3.7) in a general class of fixed-effect models and show that it leads to ℓ_π whose first-order condition has bias $o(T^{-1})$. In the present model, (3.6)–(3.7) can be solved exactly, i.e., for finite T, thus allowing a full recentering of the score. With $\ell_{it} = -\frac{1}{2}[\log\sigma^2 + (y_{it} - z_{it}^\top\theta - \alpha_i)^2/\sigma^2] + \mathrm{c}$, the relevant differential equations are

$$\nabla_{\alpha_i}\pi_i = 0, \qquad \nabla_\theta\pi_i = (T-1)b(\rho), \qquad \nabla_{\sigma^2}\pi_i = -\frac{1}{2\sigma^2},$$

which yields $\pi_i = -\frac{1}{2}\log\sigma^2 + (T-1)a(\rho) + \mathrm{c}$. Therefore,

$$\ell_\pi = \ell + \frac{N}{2}\log\sigma^2 - N(T-1)a(\rho) + \mathrm{c} \tag{3.8}$$

and $l_\pi(\theta) = \max_{\alpha_1,\ldots,\alpha_N,\sigma^2}\ell_\pi = N(T-1)l_{\mathrm{a}}(\theta) + \mathrm{c}$. Thus, the (normalized) profile penalized log-likelihood and the adjusted log-likelihood coincide. Bester and Hansen (2009) derived the exact solution to (3.6)–(3.7) for the case $p = 1$ and noted the equivalence between the penalized log-likelihood and Lancaster's (2002) posterior. Bester and Hansen's approach is to adjust the likelihood *before* profiling out the incidental parameters, while we adjust it *after* doing so. In the present model, the two approaches coincide.

Finally, the adjusted profile score is also related to Bun and Carree (2005). Note that $s(\theta) = \sum_{i=1}^{N} Z_i^\top M(y_i - Z_i\theta)/Q^2(\theta)$ and $My_i = MZ_i\widehat{\theta} + M\widehat{\varepsilon}_i$ where $\widehat{\theta}$ is the maximum likelihood estimator, with residuals $\widehat{\varepsilon}_i$ satisfying $\sum_{i=1}^{N} Z_i^\top M\widehat{\varepsilon}_i = 0$. Therefore, solving $s_{\mathrm{a}}(\theta) = 0$ is equivalent to solving

$$\widehat{\theta} - \theta = \left(\sum_{i=1}^{N} Z_i^\top M Z_i\right)^{-1} b(\rho)Q^2(\theta). \tag{3.9}$$

When $p = 1$, (3.9) corresponds to Bun and Carree's (2005) proposal for bias-correcting the maximum likelihood estimate.

4. Global properties of the adjusted profile likelihood

At this point it is tempting to anticipate that θ_0 maximizes $\mathrm{plim}_{N\to\infty} l_{\mathrm{a}}(\theta)$. However, as shown below, $-a(\rho)$ dominates $\mathrm{plim}_{N\to\infty} l(\theta)$ as $\|\rho\| \to \infty$ in almost all directions and $\mathrm{plim}_{N\to\infty} l_{\mathrm{a}}(\theta)$ is unbounded from above.

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

$$\begin{aligned} L_{\mathrm{a}}(\theta) &= L(\theta) - a(\rho), & L(\theta) &= \mathrm{plim}_{N\to\infty} l(\theta), \\ S_{\mathrm{a}}(\theta) &= S(\theta) - b(\rho), & S(\theta) &= \mathrm{plim}_{N\to\infty} s(\theta), \\ H_{\mathrm{a}}(\theta) &= H(\theta) - c(\rho), & H(\theta) &= \mathrm{plim}_{N\to\infty} h(\theta). \end{aligned}$$

Using $M(y_i - Z_i\theta) = -MZ_i(\theta - \theta_0) + M\varepsilon_i$, we have

$$L(\theta) = -\frac{1}{2}\log\left(\mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_i^\top M\varepsilon_i - 2(\theta-\theta_0)^\top Z_i^\top M\varepsilon_i + (\theta-\theta_0)^\top Z_i^\top M Z_i(\theta-\theta_0)\right)\right) + \mathrm{c}.$$

Let $b_0 = b(\rho_0) = S(\theta_0)$ and note that

$$\mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} Z_i^\top M\varepsilon_i = \left(\mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^\top M\varepsilon_i\right) b_0 = \sigma_0^2\,(T-1)\,b_0.$$

Hence, defining $V_0 = V(\theta_0)$ by

$$\mathrm{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} Z_i^\top M Z_i = \left(\mathrm{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^\top M \varepsilon_i\right) V_0 = \sigma_0^2\,(T-1)\,V_0,$$

we can write

$$L(\theta) = -\frac{1}{2}\log\left(1 - 2(\theta-\theta_0)^\top b_0 + (\theta-\theta_0)^\top V_0(\theta-\theta_0)\right) + \mathrm{c}$$

by absorbing the term $-\frac{1}{2}\log\left(\sigma_0^2\,(T-1)\right)$ into c. As $N\to\infty$, the maximum likelihood estimator of θ converges in probability to $\theta_{\mathrm{ml}} = \arg\max_\theta L(\theta) = \theta_0 + V_0^{-1}b_0$ and has asymptotic bias $V_0^{-1}b_0$. This expression generalizes the fixed T bias calculations in Nickell (1981) and Bun and Carree (2005). Note that $(\theta_0-\theta_{\mathrm{ml}})^\top V_0(\theta_0-\theta_{\mathrm{ml}}) = b_0^\top V_0^{-1} b_0$. Furthermore,

$$\begin{aligned}
L(\theta) &= -\frac{1}{2}\log\left(1 - b_0^\top V_0^{-1}b_0 + (\theta-\theta_{\mathrm{ml}})^\top V_0(\theta-\theta_{\mathrm{ml}})\right) + \mathrm{c},\\
S(\theta) &= -\frac{V_0(\theta-\theta_{\mathrm{ml}})}{1 - b_0^\top V_0^{-1}b_0 + (\theta-\theta_{\mathrm{ml}})^\top V_0(\theta-\theta_{\mathrm{ml}})},\\
H(\theta) &= -\frac{V_0}{1 - b_0^\top V_0^{-1}b_0 + (\theta-\theta_{\mathrm{ml}})^\top V_0(\theta-\theta_{\mathrm{ml}})} + 2S(\theta)S(\theta)^\top.
\end{aligned}$$

Note that $L(\cdot)$ and $H(\cdot)$ are even and $S(\cdot)$ is odd about θ_{ml} and that $H(\theta_0) = 2b_0b_0^\top - V_0$ and $H_{\mathrm{a}}(\theta_0) = 2b_0b_0^\top - V_0 - c_0$, where $c_0 = c(\rho_0)$. Since $L(\theta)$ is log-quadratic in θ 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 $\rho = kr$ with r in the positive orthant of $\mathbb{R}^p$ and let $k\to\infty$, the term $-a(\rho)$ dominates and $L_{\mathrm{a}}(\theta)\to\infty$.

It follows that $\theta_0 \neq \arg\max_\theta L_{\mathrm{a}}(\theta)$ and θ_0 has to be identified as a functional of $L_{\mathrm{a}}(\theta)$ other than its global maximizer (as in standard maximum likelihood theory). Because $S_{\mathrm{a}}(\theta_0) = 0$, we need to select θ_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 θ_0 is a local maximizer of $L_{\mathrm{a}}(\theta)$ (which will often be the case, as shown below), $L_{\mathrm{a}}(\theta)$, being smooth and unbounded, must also have at least one local minimum. Because $l(\theta)$ is log-quadratic for any $N\geq 1$ and $a(\rho)$ does not depend on the data, $l_{\mathrm{a}}(\theta)$, too, is re-increasing, regardless of the sample size. Therefore, an estimation strategy based on solving $s_{\mathrm{a}}(\theta) = 0$ has to be complemented by a solution selection rule.

4.1. First-order autoregression without covariates

In the first-order autoregressive model without covariates ($p=1, q=0$), let $\zeta_0^2 = \left(V_0 - b_0^2\right)/V_0^2$. Then,

$$\begin{aligned}
L(\rho) &= -\frac{1}{2}\log\left(\zeta_0^2 + (\rho-\rho_{\mathrm{ml}})^2\right) + \mathrm{c},\\
S(\rho) &= -\frac{\rho-\rho_{\mathrm{ml}}}{\zeta_0^2 + (\rho-\rho_{\mathrm{ml}})^2}, \qquad H(\rho) = -\frac{\zeta_0^2 - (\rho-\rho_{\mathrm{ml}})^2}{\left(\zeta_0^2 + (\rho-\rho_{\mathrm{ml}})^2\right)^2},
\end{aligned}$$

by absorbing $-\frac{1}{2}\log V_0$ into c. Note that $\zeta_0^2 = -1/H(\rho_{\mathrm{ml}})$. Recall that $S(\rho)$ is odd about $\rho_{\mathrm{ml}} = \rho_0 + b_0/V_0$. The zeros of $H(\rho)$ are $\underline{\rho} = \rho_{\mathrm{ml}} - \zeta_0$ and $\overline{\rho} = \rho_{\mathrm{ml}} + \zeta_0$, so $S(\rho)$ decreases on $[\underline{\rho}, \overline{\rho}]$ and increases elsewhere. All of $\underline{\rho}$, $\overline{\rho}$, ρ_{ml}, and ζ_0 are identified by $S(\cdot)$, and ρ_{ml} and ζ_0 act as location and scale parameters of $S(\cdot)$. For any given ρ_0, ρ_{ml} and ζ_0 are determined by V_0. As V_0 increases, $|b_0/V_0|$ and ζ_0 decrease, that is, the bias of ρ_{ml} decreases in absolute value, the length of $[\underline{\rho}, \overline{\rho}]$ shrinks, and $S(\rho)$ becomes steeper on $[\underline{\rho}, \overline{\rho}]$.

There is a sharp lower bound on V_0. With ξ_{0i} and F_0 denoting ξ_i and F evaluated at ρ_0, we have $y_{i,-1} = S_1(\xi_{0i} + F_0\varepsilon_i)$. From the independence between ξ_{0i} and ε_i, we obtain

$$V_0 = \frac{\mathrm{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^N y_{i,-1}^\top M y_{i,-1}}{\sigma_0^2 (T-1)} = V_0^{LB} + V_{\xi\xi},$$

where

$$V_0^{LB} = \frac{\mathrm{tr} F_0^\top S_1^\top M S_1 F_0}{T-1}, \qquad V_{\xi\xi} = \frac{\mathrm{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \xi_{0i}^\top S_1^\top M S_1 \xi_{0i}}{\sigma_0^2 (T-1)}.$$

So $V_0 \geq V_0^{LB}$ and this lower bound implies an upper bound on $|b_0/V_0|$ and on the length of $[\underline{\rho}, \overline{\rho}]$, and a lower bound on the steepness of $S(\rho)$ on $[\underline{\rho}, \overline{\rho}]$.

Lemma 3. *V_0^{LB} is given by*

$$V_0^{LB} = \frac{1}{T-1}\left(\sum_{j=0}^{T-2}(T-j-1)\,\rho_0^{2j} - \frac{1}{T}\sum_{j=0}^{T-2}\left(\sum_{k=0}^{j}\rho_0^k\right)^2\right)$$

and satisfies (i) $V_0^{LB} \geq 2b_0^2$; (ii) $V_0^{LB} \geq 2b_0^2 - c_0$ with equality if and only if $T = 2$ or $\rho_0 = 1$.

By Lemma 3, $H(\rho_0) = 2b_0^2 - V_0 \leq 0$ and, hence,

$$(\overline{\rho} - \rho_{\mathrm{ml}})^2 = \frac{V_0 - b_0^2}{V_0^2} \geq \frac{b_0^2}{V_0^2} = (\rho_0 - \rho_{\mathrm{ml}})^2.$$

Therefore, $\rho_0 \in [\underline{\rho}, \overline{\rho}]$. 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 $S_{\mathrm{a}}(\rho_0) = 0$ and, by Lemma 3, $H_{\mathrm{a}}(\rho_0) = 2b_0^2 - V_0 - c_0 \leq 0$, it follows that $L_{\mathrm{a}}(\rho)$ has a local maximum or a flat inflection point at ρ_0. Our main result for a first-order autoregression without covariates is the uniqueness of such a point in $[\underline{\rho}, \overline{\rho}]$, thereby identifying ρ_0 as a functional of $L_{\mathrm{a}}(\rho)$. Equivalently, ρ_0 is the unique point in $[\underline{\rho}, \overline{\rho}]$ where $b(\rho)$ approaches $S(\rho)$ from below.

Theorem 1. *ρ_0 is the unique point in $[\underline{\rho}, \overline{\rho}]$ where $L_{\mathrm{a}}(\rho)$ has a local maximum or a flat inflection point.*

$L_{\mathrm{a}}(\rho)$ has a flat inflection point at ρ_0 if and only if $V_0 = V_0^{LB} = 2b_0^2 - c_0$. 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 $MS_1\xi_{0i}$ to be negligibly small for almost all i. The elements of $S_1\xi_{0i}$ are $\rho_0^{j-1}y_i^0 + \alpha_i\sum_{k=1}^{j-1}\rho_0^{k-1}$, $j = 1, ..., T$, so $MS_1\xi_{0i} = 0$ if and only if $y_i^0(1-\rho_0) = \alpha_i$. The following corollary has been independently obtained by Ahn and Thomas (2006).

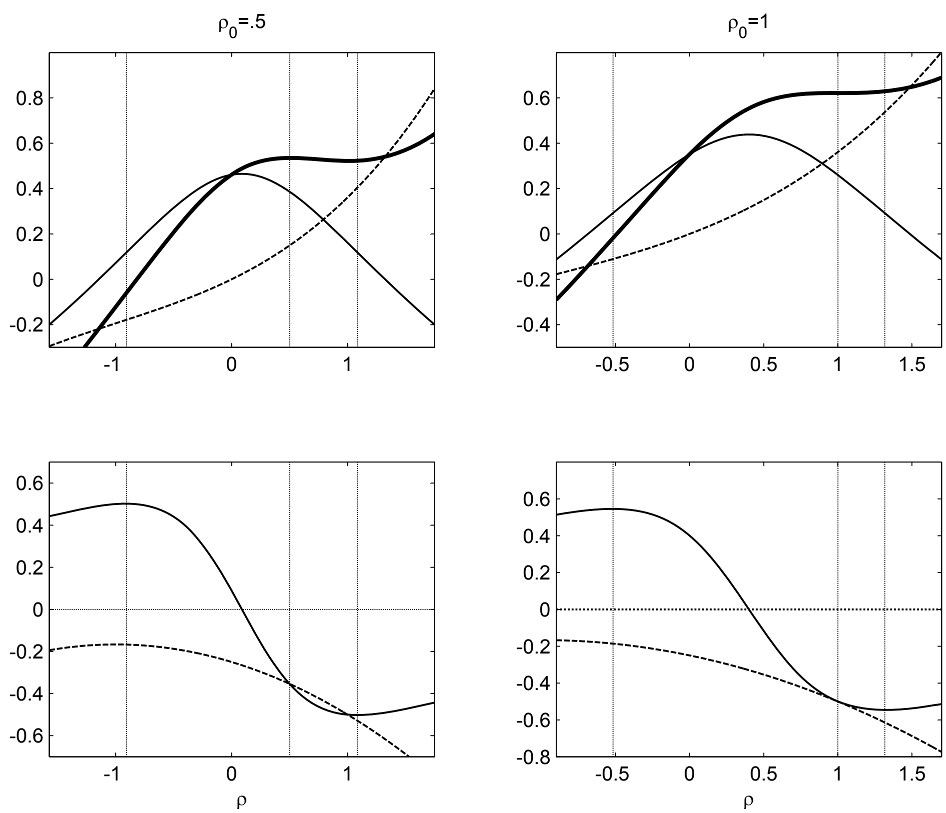
Corollary 1. *When $\rho_0 = 1$ and $\alpha_i = 0$, $L_{\mathrm{a}}(\rho)$ has a flat inflection point at ρ_0 for any T.*

When $\rho_0 \neq 1$, $V_0 = V_0^{LB} = 2b_0^2 - c_0$ 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, ρ_0 is the unique point in $[\underline{\rho}, \overline{\rho}]$ where $L_{\mathrm{a}}(\rho)$ attains a strict local maximum. Note that, when ρ_0 is a local maximizer of $L_{\mathrm{a}}(\rho)$, it need not be the global maximizer on $[\underline{\rho}, \overline{\rho}]$, which may instead be $\overline{\rho}$. To see why this may happen, interpret the situation where $L_{\mathrm{a}}(\rho)$ has a flat inflection point at ρ_0 as a limiting case of the property that $L_{\mathrm{a}}(\rho)$ is re-increasing.

Figure 1 illustrates how ρ_0 is identified by $L_{\mathrm{a}}(\rho)$ for two cases, each with $T = 4$. The plots on the left correspond to the case $\rho_0 = .5$ with $V_0 = V_0^{LB} + V_{\xi\xi}$ 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 $V_0 = V_0^{LB}$. In each case, the bottom figures show $S(\rho)$ (solid line) and $b(\rho)$ (dashed line); the top plots show $L(\rho)$ (solid line), $-a(\rho)$ (dashed line), and $L_{\mathrm{a}}(\rho) = L(\rho) - a(\rho)$ (thick line). In all the plots, vertical lines indicate $\underline{\rho}$, ρ_0, and $\overline{\rho}$, from left to right. In the case of $\rho_0 = .5$, ρ_0 is the unique local maximizer of $L_{\mathrm{a}}(\rho)$ on $[\underline{\rho}, \overline{\rho}]$. Note that there is a second solution of $S_{\mathrm{a}}(\rho) = 0$ on $[\underline{\rho}, \overline{\rho}]$, which corresponds to a local minimum of $L_{\mathrm{a}}(\rho)$. In the unit root case, ρ_0 is the unique flat inflection point of $L_{\mathrm{a}}(\rho)$ on $[\underline{\rho}, \overline{\rho}]$.

The asymptotic bias of the maximum likelihood estimator has the same sign as b_0 because $\rho_{\mathrm{ml}} = \rho_0 + b_0/V_0$. 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 $\rho_u - \rho_0$.

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 $\underline{\rho}$, ρ_0, and $\overline{\rho}$.

4.2. First-order autoregression with covariates

In the first-order autoregressive model with covariates ($p = 1, q \geq 1$), profiling out β yields a profile likelihood of ρ with essentially the same properties as in the model without covariates. Let $\beta(\rho) = \arg\max_\beta L_{\mathrm{a}}(\rho, \beta) =$

$\arg\max_{\beta} L(\rho,\beta) = \arg\min_{\beta}(\theta-\theta_{\rm ml})^\top V_0(\theta-\theta_{\rm ml})$. Partition V_0, V_0^{-1}, and b_0 as

$$V_0 = \begin{pmatrix} V_{0\rho\rho} & V_{0\rho\beta} \\ V_{0\beta\rho} & V_{0\beta\beta} \end{pmatrix}, \qquad V_0^{-1} = \begin{pmatrix} V_0^{\rho\rho} & V_0^{\rho\beta} \\ V_0^{\beta\rho} & V_0^{\beta\beta} \end{pmatrix}, \qquad b_0 = \begin{pmatrix} b_{0\rho} \\ 0 \end{pmatrix}.$$

With $V_0^{\rho\rho} = (V_{0\rho\rho} - V_{0\rho\beta}V_{0\beta\beta}^{-1}V_{0\beta\rho})^{-1}$, we have

$$\begin{aligned} V_{0\beta\beta}\,(\beta(\rho)-\beta_{\rm ml}) &= -V_{0\beta\rho}(\rho-\rho_{\rm ml}), \\ \min_{\beta}(\theta-\theta_{\rm ml})^\top V_0(\theta-\theta_{\rm ml}) &= (\rho-\rho_{\rm ml})^2/V_0^{\rho\rho}, \\ 1-b_0^\top V_0^{-1}b_0 &= 1-b_{0\rho}^2V_0^{\rho\rho}. \end{aligned}$$

The first of these equations, together with $V_0(\theta_0-\theta_{\rm ml}) = -b_0$, yields $\beta(\rho_0) = \beta_0$, so β_0 is identified whenever ρ_0 is. Profiling out β from $L(\rho,\beta)$ gives the limiting profile log-likelihood of ρ as

$$L(\rho) = L(\rho,\beta(\rho)) = -\frac{1}{2}\log\left(\zeta_0^2 + (\rho-\rho_{\rm ml})^2\right) + {\rm c}$$

(slightly abusing notation), where ζ_0^2 is redefined as $\zeta_0^2 = \left(1-b_{0\rho}^2V_0^{\rho\rho}\right)V_0^{\rho\rho}$ and $\frac{1}{2}\log V_0^{\rho\rho}$ is absorbed into c.

LEMMA 4. $(V_0^{\rho\rho})^{-1} \geq V_0^{LB}$, *with* V_0^{LB} *as defined earlier and given in Lemma 3.*

We can now invoke the result for the model without covariates. Let $\underline{\rho} = \rho_{\rm ml} - \zeta_0$ and $\overline{\rho} = \rho_{\rm ml} + \zeta_0$, with ζ_0 redefined as indicated.

THEOREM 2. ρ_0 *is the unique point in* $[\underline{\rho},\overline{\rho}]$ *where* $L_{\rm a}(\rho) = L(\rho) - a(\rho)$ *has a local maximum or a flat inflection point.*

By the proof of Lemma 4, the conditions under which ρ_0 is a flat inflection point of $L_{\rm a}(\rho)$ are the same as before. The presence of covariates does not affect the sign of the asymptotic bias of the maximum likelihood estimator of ρ. It also follows from the proof of Lemma 4 that the inclusion of covariates in the model cannot increase $V_0^{\rho\rho}$, so the magnitude of $\rho_{\rm ml} - \rho_0 = V_0^{\rho\rho}b_{0\rho}$ can only decrease relative to the model without covariates.

4.3. pth-order autoregression

Consider first an autoregression with $p > 1$ and without covariates, i.e., $q = 0$. Then

$$\begin{aligned} L(\rho) &= -\frac{1}{2}\log\left(1+(\rho-\rho_{\rm ml})^\top W_0(\rho-\rho_{\rm ml})\right) + {\rm c}, \qquad W_0 = \frac{V_0}{1-b_0^\top V_0^{-1}b_0}, \\ S(\rho) &= -\frac{W_0(\rho-\rho_{\rm ml})}{1+(\rho-\rho_{\rm ml})^\top W_0(\rho-\rho_{\rm ml})}, \\ H(\rho) &= -\frac{W_0}{1+(\rho-\rho_{\rm ml})^\top W_0(\rho-\rho_{\rm ml})} + 2S(\rho)S(\rho)^\top, \end{aligned}$$

where $-\frac{1}{2}\log(1-b_0^\top V_0^{-1}b_0)$ is absorbed into c. Because $W_0 = -H(\rho_{\rm ml})$, W_0 is identified by $L(\cdot)$.

As in the $p = 1$ case, there is a lower bound on V_0. Recalling that $Y_{i-} = (y_{i,-1},\ldots,y_{i,-p})$ and $y_{i,-j} = S_j(\xi_{0i} + F_0\varepsilon_i)$, where ξ_{0i} and ε_i are independent, we have

$$V_0 = \frac{{\rm plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N Y_{i-}^\top MY_{i-}}{\sigma_0^2\,(T-1)} = V_0^{LB} + V_{\xi\xi}$$

where V_0^{LB} and $V_{\xi\xi}$ have elements

$$\left(V_0^{LB}\right)_{jk} = \frac{\mathrm{tr} F_0^\top S_j^\top M S_k F_0}{T-1}, \qquad \left(V_{\xi\xi}\right)_{jk} = \frac{\mathrm{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \xi_{0i}^\top S_j^\top M S_k \xi_{0i}}{\sigma_0^2\,(T-1)},$$

for $1 \le j,k \le p$. Hence, $V_0 - V_0^{LB}$ is positive semi-definite, which we write as $V_0 \ge V_0^{LB}$. When $p \ge T$, while V_0 is nonsingular by assumption, $\mathrm{rank}(V_0^{LB}) \le T-1$ because $S_j F_0 = 0$ for $j \ge T$, which implies that $\left(V_0^{LB}\right)_{jk} = 0$ whenever $j \ge T$ or $k \ge T$. Thus, when $p \ge T$, although V_0 can be arbitrarily close to V_0^{LB}, $V_0 \ne V_0^{LB}$. Further, when $p \ge T$, $b_j(\rho) = 0$ for $j \ge T$ because the sum defining $b_j(\rho)$ is empty, and $c_{ij}(\rho) = 0$ for $i+j \ge T$. Hence, when $p \ge T$, $V_0^{LB} - 2b_0 b_0^\top$ and $V_0^{LB} - 2b_0 b_0^\top + c_0$ have only zeros beyond their leading $(T-1)\times(T-1)$ blocks.

A proof of generalizations of (i)–(ii) of Lemma 3 and Theorem 1 to the $p > 1$ would be desirable but is more difficult.[1] We resorted to numerical computations, which suggest that

$$V_0^{LB} \ge 2b_0 b_0^\top, \qquad V_0^{LB} \ge 2b_0 b_0^\top - c_0, \tag{4.1}$$

$$\mathrm{rank}(V_0^{LB} - 2b_0 b_0^\top + c_0) = \begin{cases} \min(p, T-2) & \text{if } \sum_{j=1}^p \rho_{0j} \ne 1 \text{ or } T < p+2, \\ p-1 & \text{else.} \end{cases} \tag{4.2}$$

Specifically, we computed the eigenvalues of $V_0^{LB} - 2b_0 b_0^\top$ and $V_0^{LB} - 2b_0 b_0^\top + c_0$ for $p = 2,3,4$; $T = 2,...,7$; and all ρ_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

$$\begin{array}{ll} -1 \le \gamma_2 \le 1, & \gamma_2 - 1 \le \gamma_1 \le 1 - \gamma_2, \\ -1 \le \gamma_4 \le 1, & \gamma_4 - 1 \le \gamma_3 \le 1 - \gamma_4, \end{array} \tag{4.3}$$

which is the stationary region of the lag polynomial $\gamma(L) = (1 - \gamma_1 L - \gamma_2 L^2)(1 - \gamma_3 L - \gamma_4 L^2)$. For each point on this grid and for each of the values $m = 1,2,4$, ρ_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 (4.3) 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 (4.3) 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 $V_0^{LB} - 2b_0 b_0^\top$ and $V_0^{LB} - 2b_0 b_0^\top + c_0$ are non-negative and the rank of $V_0^{LB} - 2b_0 b_0^\top + c_0$ is as given by (4.2). These findings, while obviously not a proof, support (4.1) and (4.2), and we shall proceed under the assumption that (4.1) and (4.2) hold.[2]

Because $V_0 \ge V_0^{LB}$, (4.1) implies that $V_0 \ge 2b_0 b_0^\top$ and that $H_{\mathrm{a}}(\rho_0) = 2b_0 b_0^\top - V_0 - c_0 \le 0$. Pre- and postmultiplication of $V_0 \ge 2b_0 b_0^\top$ by $b_0^\top V_0^{-1}$ and $V_0^{-1} b_0$ gives $b_0^\top V_0^{-1} b_0 \le \frac{1}{2} \le 1 - b_0^\top V_0^{-1} b_0$. Recalling that $(\rho_0 - \rho_{\mathrm{ml}})^\top V_0 (\rho_0 - \rho_{\mathrm{ml}}) = b_0^\top V_0^{-1} b_0$, we have

$$(\rho_0 - \rho_{\mathrm{ml}})^\top W_0 (\rho_0 - \rho_{\mathrm{ml}}) \le 1.$$

[1] A major difficulty is the rapidly increasing complexity of φ_t as p increases. For example, $\varphi_t = \sum_{k=0}^{\lfloor t/2 \rfloor} \frac{(t-k)!}{(t-2k)!k!} \rho_1^{t-2k} \rho_2^k$ when $p = 2$. In comparison, $\varphi_t = \rho_1^t$ when $p = 1$.

[2] 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.

Therefore, if (4.1) and (4.2) hold, ρ_0 is a point in the ellipsoidal disk $\mathcal{E} = \{\rho : (\rho - \rho_{\rm ml})^\top W_0 (\rho - \rho_{\rm ml}) \leq 1\}$ where $L_{\rm a}(\rho)$ 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 $V_0 = V_0^{LB}$, we applied the Newton-Raphson algorithm to find a stationary point of $L_{\rm a}(\rho)$, starting at $\rho_{\rm ml}$ and using the Moore-Penrose inverse of $H_{\rm a}(\rho)$ whenever $H_{\rm a}(\rho)$ is singular. Uniformly over this design, the algorithm was found to converge to ρ_0, thus supporting the conjecture that ρ_0 is the unique point in $\mathcal{E}$ where $L_{\rm a}(\rho)$ has a local maximum or a flat inflection point.[3]

In the model with covariates, just as before, β can be profiled out of $L_{\rm a}(\theta)$. Here, again, $\beta_0 = \beta(\rho_0)$. Lemma 4 continues to hold for $p > 1$. Hence, if ρ_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 $W_0 = (1 - b_{0\rho}^\top V_0^{\rho\rho} b_{0\rho})^{-1} V_{0\rho\rho}$, in obvious notation.

5. Estimation and inference

For a given ρ, define

$$\widehat{\beta}(\rho) = \arg\max_\beta l_{\rm a}(\rho, \beta) = \arg\max_\beta l(\rho, \beta) = \left(\sum_{i=1}^N X_i^\top M X_i \right)^{-1} \sum_{i=1}^N X_i^\top M \left(y_i - Y_{i-} \rho \right).$$

The unadjusted and adjusted profile log-likelihoods for ρ are $l(\rho) = l(\rho, \widehat{\beta}(\rho))$ and $l_{\rm a}(\rho) = l(\rho) - a(\rho)$. Let $s(\rho)$, $s_{\rm a}(\rho)$, $h(\rho)$, and $h_{\rm a}(\rho)$ be the corresponding profile scores and Hessians. Let $\widehat{W} = -h(\widehat{\rho}_{\rm ml})$, where $\widehat{\rho}_{\rm ml}$ is the maximum likelihood estimator of ρ_0, and let $\widehat{\mathcal{E}} = \{\rho : (\rho - \widehat{\rho}_{\rm ml})^\top \widehat{W} (\rho - \widehat{\rho}_{\rm ml}) \leq 1\}$. We define the adjusted likelihood estimator of ρ_0 as

$$\widehat{\rho}_{\rm al} = \arg\min_{\rho \in \widehat{\mathcal{E}}} s_{\rm a}^\top(\rho) s_{\rm a}(\rho) \quad \text{s.t.} \quad h_{\rm a}(\rho) \leq 0,$$

that is, as the strict local maximizer of $l_{\rm a}(\rho)$ on the interior of $\widehat{\mathcal{E}}$ if such a maximizer exists and otherwise as the minimizer of the norm of $s_{\rm a}(\rho)$ on $\widehat{\mathcal{E}}$. The adjusted likelihood estimator of β_0, then, is $\widehat{\beta}_{\rm al} = \widehat{\beta}(\widehat{\rho}_{\rm al})$.

Let $N \to \infty$. Then $l_{\rm a}(\rho)$ converges to $L_{\rm a}(\rho)$ uniformly in ρ since $-a(\rho)$ is nonstochastic and $\sup_\rho |l(\rho) - L(\rho)| = o_p(1)$. Further, $\widehat{\rho}_{\rm ml} \xrightarrow{p} \rho_{\rm ml}$, $\widehat{W} \xrightarrow{p} -H(\rho_{\rm ml}) = W_0$, and $\widehat{\mathcal{E}} \xrightarrow{p} \mathcal{E}$ in the sense that $\Pr[\rho \in \widehat{\mathcal{E}}] \to 1_{\{\rho \in \mathcal{E}\}}$ for any ρ not on the boundary of $\mathcal{E}$. It follows that $\widehat{\theta}_{\rm al} = (\widehat{\rho}_{\rm al}^\top, \widehat{\beta}_{\rm al}^\top)^\top \xrightarrow{p} \theta_0$.

When $H_{\rm a}(\theta_0)$ is nonsingular, by a Taylor series expansion of $s_{\rm a}(\theta)$ around θ_0,

$$\sqrt{N}(\widehat{\theta}_{\rm al} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega), \qquad \Omega = H_{\rm a}(\theta_0)^{-1} (V_0 - b_0 b_0^\top) H_{\rm a}(\theta_0)^{-1}. \tag{5.1}$$

The asymptotic variance can be estimated in the usual way.[4]

We have not investigated the limit distribution of $\widehat{\theta}_{\rm al}$ in the situation where $H_{\rm a}(\theta_0)$ is singular (which includes all cases where ρ_0 is a flat inflection point). Presumably this could be done by using arguments along the lines of Rotnitzky, Cox, Bottai, and Robins (2000).

[3] 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 $H_{\rm a}(\rho)$ was underestimated.

[4] Note that the information equality does not hold, although this could be rectified by rescaling the adjusted profile score; see McCullagh and Tibshirani (1990).

We used simulations to examine the finite sample properties of the adjusted likelihood estimator in first- and second-order autoregressions without covariates and in a first-order autoregressive model with one stationary covariate. We compared the estimator with the one-step Arellano and Bond (1991) estimator, which leaves the initial observations unrestricted,[5] and with two estimators that are consistent only under rectangular array asymptotics (see Li, Lindsay, and Waterman, 2003 and Sartori, 2003). The latter two estimators have an asymptotic bias that is $O(T^{-2})$. The first of these estimators corrects the maximum likelihood estimate. For the first-order autoregression without covariates, we used the estimator of Hahn and Kuersteiner (2002), which is targeted to this setup. For the other models, we used its extension to possibly nonlinear models as proposed by Hahn and Kuersteiner (2011). The second large T estimator considered is the penalized likelihood estimator of Bester and Hansen (2009).[6]

In all the designs, we set $N = 100$, generated ε_{it} and α_i as $\mathcal{N}(0,1)$ variates, and chose ρ_0 in the interior of the stationary region, which implies that y_{it} is eventually stationary as $t \to \infty$. We varied the information content of the data through the initial observations. Let $\mu_i = \lim_{t\to\infty} \mathbb{E}(y_{it-}|\alpha_i)$ and $\Sigma_i = \lim_{t\to\infty} \mathrm{Var}(y_{it-}|\alpha_i)$, so, if y_i^0 was drawn from the stationary distribution, we would just have $\mu_i = \mathbb{E}(y_i^0|\alpha_i)$ and $\Sigma_i = \mathrm{Var}(y_i^0|\alpha_i)$. Let $G_i G_i^\top = \Sigma_i$ be the Cholesky factorization of Σ_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 ψ standard deviations away from the stationary mean. So ψ controls the outlyingness of the initial observations relative to the stationary distributions. All else being equal, V_0 increases in ψ and $V_0 \to V_0^{LB}$ as $\psi \to 0$, so the data carry less information as ψ gets smaller. The effect of strong inlying observations (small ψ) 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 ψ is vanishingly small as ρ_0 moves to the boundary of the stationary region. We set $\psi = 0, 1, 2$ when $p = 1$ and $\psi = .3, 1, 2$ when $p = 2$.

In the models without a covariate, μ_i and Σ_i follow immediately from α_i and ρ_0. In the model with a covariate, x_{it} was generated by $x_{it} = \delta\alpha_i + \gamma x_{it-1} + u_{it}$ with $u_{it} \sim \mathcal{N}(0, \sigma_u^2)$ and x_{i0} drawn from the stationary distribution. Here,

$$\mu_i = \frac{\alpha_i}{1-\rho_0}\left(1 + \frac{\delta\beta_0}{1-\gamma}\right), \qquad \Sigma_i = \frac{1}{1-\rho_0^2}\left(1 + \frac{\beta_0^2}{1-\gamma^2}\left(\frac{1+\gamma\rho_0}{1-\gamma\rho_0}\right)\sigma_u^2\right).$$

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.

Tables 1–5 in Appendix B present Monte Carlo estimates, based on $10,000$ replications, of the bias and the standard deviation (std) of the estimators considered, as well as the coverage rates of the corresponding asymptotic and bootstrap 95% confidence intervals ($\mathrm{ci}^{\mathrm{a}}_{.95}$ and $\mathrm{ci}^{\mathrm{b}}_{.95}$). Bootstrap confidence intervals were computed using the percentile method with 39 bootstrap samples formed by randomly drawing N strata with replacement from $\{1, ..., N\}$.

In the first-order autoregression with $\rho_0 = .5$ (upper panel in Table 1), both $\widehat{\rho}_{\mathrm{al}}$ and $\widehat{\rho}_{\mathrm{ab}}$ perform well. The adjusted likelihood estimator has smaller standard deviation and is virtually unbiased, except when $\psi = 0$

[5] The Arellano and Bond (1991) estimator is a generalized method of moments estimator based on the moments $\mathbb{E}\left[x_{ij}(\varepsilon_{it} - \varepsilon_{it-1})\right] = 0, j = 1, ..., T; t = 2, ..., T$; and $\mathbb{E}\left[y_{it-j}(\varepsilon_{it} - \varepsilon_{it-1})\right] = 0, j = 2, ..., t; t = 2, ..., T$.

[6] The estimators of Hahn and Kuersteiner (2011) and Bester and Hansen (2009) require a bandwidth choice. We set the bandwidth equal to unity, following the suggestion of Bester and Hansen (2009, p. 134).

and $T = 2$. Both estimators deliver 95% confidence intervals with broadly correct coverage, although the coverage errors are somewhat larger for $\widehat{\rho}_{\mathrm{ab}}$, where they also increase in T. The latter observation is in line with the theoretical results of Alvarez and Arellano (2003). The estimator of Hahn and Kuersteiner (2002) outperforms the Bester and Hansen (2009) estimator, although both exhibit substantial bias for small T and their performance is sensitive to ψ.

When ρ_0 is increased to .95 (lower panel in Table 1), the performance of all estimators tends to worsen. $\widehat{\rho}_{\mathrm{ab}}$ deteriorates the most, showing a substantial bias, large dispersion, and confidence intervals with much lower coverage. $\widehat{\rho}_{\mathrm{al}}$ continues to have little bias and provides confidence intervals with approximately correct coverage, with the bootstrap-based confidence intervals being slightly better. In most designs, both $\widehat{\rho}_{\mathrm{hk}}$ and $\widehat{\rho}_{\mathrm{bh}}$ outperform $\widehat{\rho}_{\mathrm{ab}}$ in terms of bias and standard deviation. Their confidence intervals, however, are not reliable.

In the second-order autoregression (Tables 2 and 3), both $\widehat{\rho}_{\mathrm{al}}$ and $\widehat{\rho}_{\mathrm{ab}}$ perform well in terms of bias, 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 ψ are small. In most designs, $\widehat{\rho}_{\mathrm{al}}$ has smaller standard deviation than $\widehat{\rho}_{\mathrm{ab}}$, with the difference decreasing in T and ψ. For both estimators, the confidence intervals have very reasonable coverage. As before, $\widehat{\rho}_{\mathrm{hk}}$ and $\widehat{\rho}_{\mathrm{bh}}$ still show a substantial bias for most of the designs considered. Together with their small standard deviation for most values of T, this again leads to their confidence intervals being too narrow.

In the model with a covariate (Tables 4 and 5), the coefficient on x_{it}, β_0, is generally estimated with small bias by all estimators. Regarding the estimation of ρ_0, the tables show a similar pattern as in the model without a covariate. $\widehat{\rho}_{\mathrm{al}}$ has little bias and well-behaved confidence intervals for all designs, especially when computed by bootstrapping. The same holds for $\widehat{\rho}_{\mathrm{ab}}$ only when $\rho_0 = .5$. $\widehat{\rho}_{\mathrm{hk}}$ and $\widehat{\rho}_{\mathrm{bh}}$ only start to perform reasonably when $T \geq 16$, although the coverage errors of their confidence intervals remain large for all values of T considered.

Appendix A: Proofs

Proof of Lemma 1. Using (2.2),

$$\begin{aligned}
\operatorname{plim}_{N\to\infty} s_{\rho_j}(\theta_0) &= \frac{\operatorname{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \varepsilon_i^\top M S_j(\xi_{0i}+F_0\varepsilon_i)}{\operatorname{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \varepsilon_i^\top M\varepsilon_i} = \frac{\mathbb{E}(\varepsilon_i^\top M S_j F_0 \varepsilon_i)}{\mathbb{E}(\varepsilon_i^\top M\varepsilon_i)} \\
&= \frac{\operatorname{tr} M S_j F_0}{T-1}, \\
\operatorname{plim}_{N\to\infty} s_{\beta_j}(\theta_0) &= 0,
\end{aligned}$$

where ξ_{0i} and F_0 are ξ_i and F, evaluated at θ_0. If, in addition, the disturbances are normal variates, i.e., $\varepsilon_{it} \sim \mathcal{N}(0,\sigma^2)$, then

$$\begin{aligned}
\mathbb{E}[s_{\rho_j}(\theta_0)] &= \mathbb{E}\left(\frac{\sum_{i=1}^N \varepsilon_i^\top M S_j(\xi_{0i}+F_0\varepsilon_i)}{\sum_{i=1}^N \varepsilon_i^\top M\varepsilon_i}\right) = \mathbb{E}\left(\frac{\sum_{i=1}^N \varepsilon_i^\top M S_j F_0 M\varepsilon_i}{\sum_{i=1}^N \varepsilon_i^\top M\varepsilon_i}\right) \\
&= \frac{\mathbb{E}(\varepsilon_i^\top M S_j F_0 M\varepsilon_i)}{\mathbb{E}(\varepsilon_i^\top M\varepsilon_i)} = \frac{\operatorname{tr} M S_j F_0}{T-1}, \qquad \text{(A.2)} \\
\mathbb{E}[s_{\beta_j}(\theta_0)] &= 0,
\end{aligned}$$

by well-known properties of the normal distribution and the following geometric argument, which goes back to Fisher (1930) and Geary (1933). Let $v \sim \mathcal{N}(0,\sigma^2 I_g)$ and let Q be a $g\times h$ matrix such that $Q^\top Q = I_h$, so $QQ^\top$ is idempotent. Write $I_g - QQ^\top$ as $PP^\top$, where $P^\top P = I_{g-h}$. Transform v into $m = P^\top v$, the radius $r = (v^\top QQ^\top v)^{1/2}$, and the

$h-1$ polar angles a of $Q^\top v$. Then the elements of $(m^\top, r, a^\top)^\top$ are independent. Therefore, for any $g \times g$ matrix W, if $A = v^\top QQ^\top WQQ^\top v$ and $B = v^\top QQ^\top v$, 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 $\mathbb{E}(A/B) = \mathbb{E}(A)/\mathbb{E}(B)$.[7] The transition to (A.2) now follows from applying this property to the ratio

$$\frac{\sum_{i=1}^N \varepsilon_i^\top MS_jF_0M\varepsilon_i}{\sum_{i=1}^N \varepsilon_i^\top M\varepsilon_i} = \frac{\varepsilon^\top (I_N \otimes M)(I_N \otimes S_jF_0)(I_N \otimes M)\varepsilon}{\varepsilon^\top (I_N \otimes M)\varepsilon}$$

with $v = \varepsilon = (\varepsilon_1^\top, ..., \varepsilon_N^\top)^\top$, $QQ^\top = I_N \otimes M$, and $W = I_N \otimes S_jF_0$. The proof is completed by writing $\mathrm{tr} MS_jF_0$ in terms of the φ_t. Note that

$$S_jF = \begin{pmatrix} 0 & 0 \\ D_j^{-1} & 0 \end{pmatrix},$$

where D_j^{-1} is the leading $(T-j)\times(T-j)$ block of D^{-1}. For arbitrary $\rho_1, ..., \rho_{T-1}$, D and its inverse are

$$D = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\rho_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -\rho_{T-1} & \cdots & -\rho_1 & 1 \end{pmatrix}, \qquad D^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \phi_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \phi_{T-1} & \cdots & \phi_1 & 1 \end{pmatrix},$$

where $\phi_1, ..., \phi_{T-1}$ are recursively obtained as $\phi_1 = \rho_1$ and $\phi_j = \rho_j + \sum_{k=1}^{j-1} \phi_k \rho_{j-k}$, $j = 2, ..., T-1$. Recursive substitution gives

$$\phi_j = \sum_{k_1+2k_2+\cdots+jk_j=j} \frac{(k_1+...+k_p)!}{k_1!\cdots k_p!}\rho_1^{k_1}\rho_2^{k_2}\cdots\rho_j^{k_j}.$$

Putting $\rho_{p+1} = \cdots = \rho_{T-1} = 0$ gives $\phi_j = \varphi_j$. Therefore,

$$\frac{\mathrm{tr} MS_jF_0}{T-1} = -\frac{\iota^\top D_j^{-1}\iota}{T(T-1)} = -\sum_{t=0}^{T-j-1} \frac{T-j-t}{T(T-1)}\varphi_t, \quad j = 1, \ldots, p,$$

which equals $b_j(\rho)$. □

Proof of Lemma 2. For $j = 1, ..., p$, let $\mathcal{S}_j = \{S \in \mathcal{S} | j \in S\}$. 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

$$B_{j,S}(\rho) = -\sum_{t=0}^{T-j-1} \frac{T-j-t}{T(T-1)} \sum_{k\in\mathcal{K}_{j,S}:\tau^\top k=t} \frac{(\iota^\top k)!}{k_1!\cdots(k_j+1)!\cdots k_p!}\rho_j\rho_S^{k_S}$$

and $\mathcal{K}_{j,S} = \{k \in \mathbb{N}^p |$ for all $j' \neq j$, $k_{j'} > 0$ if ond only if $j' \in S\} \supset \mathcal{K}_S$. A change of variable from k_j+1 to k_j gives

$$B_{j,S}(\rho) = -\sum_{t=|S|-j}^{T-j-1} \frac{T-j-t}{T(T-1)} \sum_{k\in\mathcal{K}_S:\tau^\top k=t+j} \frac{(\iota^\top k-1)!}{k_1!\cdots k_p!}\rho_S^{k_S},$$

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 $\tau^\top k = t+j$. 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). 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. □

[7] For a discussion and historical perspective on this device, see Conniffe and Spencer (2001).

Proof of Equation (3.3). In the parameterization $\eta_i = \alpha_i e^{-(T-1)a(\rho)}$, we have

$$\begin{aligned}\ell_i(\vartheta,\eta_i) &= -\frac{1}{2}\log\sigma^2 - \frac{1}{2T\sigma^2}\sum_{t=1}^{T}(y_{it} - z_{it}^\top\theta - \eta_i e^{(T-1)a(\rho)})^2 + \mathrm{c},\\ \nabla_{\eta_i}\ell_i(\vartheta,\eta_i) &= \frac{e^{(T-1)a(\rho)}}{T\sigma^2}(y_i - Z_i\theta - \eta_i e^{(T-1)a(\rho)}\iota)^\top\iota,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\vartheta,\eta_i}\nabla_{\eta_i\eta_i}\ell_i(\vartheta,\eta_i) &= -\sigma^{-2}e^{2(T-1)a(\rho)}, \qquad \mathbb{E}_{\vartheta,\eta_i}\nabla_{\sigma^2\eta_i}\ell_i(\vartheta,\eta_i) = 0,\\ \mathbb{E}_{\vartheta,\eta_i}\nabla_{\theta\eta_i}\ell_i(\vartheta,\eta_i) &= -\sigma^{-2}e^{(T-1)a(\rho)}\Big(\eta_i(T-1)b(\rho)e^{(T-1)a(\rho)} + \mathbb{E}_{\vartheta,\eta_i}Z_i^\top\iota/T\Big).\end{aligned}$$

The jth column of Y_{i-} is $y_{i,-j} = S_j(\xi_i + F\varepsilon_i)$, so the jth element of $\mathbb{E}_{\vartheta,\eta_i}Y_{i-}^\top\iota$ is

$$\mathbb{E}_{\vartheta,\eta_i}y_{i,-j}^\top\iota = \iota^\top S_j\xi_i = \iota^\top D_j^{-1}\iota\eta_i e^{(T-1)a(\rho)} + Tm_j, \qquad m_j = \iota^\top S_j\begin{pmatrix} y_i^0 \\ D^{-1}\left(Cy_i^0 + X_i\beta\right)\end{pmatrix}/T.$$

Hence,

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

where $m = (m_1,...,m_p,\iota^\top X_i/T)^\top$ is free of η_i. Consequently,

$$A_i^{-1}B_i = -e^{-(T-1)a(\rho)}\begin{pmatrix} m \\ 0\end{pmatrix}$$

and $\nabla_{\eta_i}(A_i^{-1}B_i) = 0$. □

Proof that no orthogonalization exists when $p > 1$. In the original parameterization, if $l_i(\vartheta,\alpha_i)$ is i's log-likelihood contribution, we have

$$\begin{aligned}\mathbb{E}_{\vartheta,\alpha_i}\nabla_{\alpha_i\alpha_i}l_i(\vartheta,\alpha_i) &= -\sigma^{-2}, \qquad \mathbb{E}_{\vartheta,\alpha_i}\nabla_{\sigma^2\alpha_i}l_i(\vartheta,\alpha_i) = 0,\\ \mathbb{E}_{\vartheta,\alpha_i}\nabla_{\theta\alpha_i}l_i(\vartheta,\alpha_i) &= -\sigma^{-2}\mathbb{E}_{\vartheta,\alpha_i}Z_i^\top\iota/T,\end{aligned}$$

and so, by the preceeding proof,

$$A_i^{-1}B_i = -\begin{pmatrix}\mathbb{E}_{\vartheta,\alpha_i}Z_i^\top\iota/T \\ 0\end{pmatrix} = -\begin{pmatrix}-(T-1)b(\rho)\alpha_i + m \\ 0\end{pmatrix}.$$

Suppose some reparameterized fixed effect, say ζ_i, is orthogonal to ϑ. Then $\alpha_i = \alpha_i(\vartheta,\zeta_i)$ must satisfy the differential equation $\nabla_\vartheta\alpha_i = A_i^{-1}B_i$, that is,

$$\nabla_{\rho_j}\alpha_i = (T-1)b_j(\rho)\alpha_i - m_j, \qquad j = 1,\ldots,p, \tag{A.3}$$

$$\nabla_{\beta_j}\alpha_i = -m_{p+j}, \qquad j = 1,\ldots,q, \tag{A.4}$$

and $\nabla_{\sigma^2}\alpha_i = 0$. We show that these equations are inconsistent. Suppose $q > 0$. Then (A.3) implies $\nabla_{\rho_j\beta_{j'}}\alpha_i = -\nabla_{\beta_{j'}}m_j$, which is generally non-zero, while (A.4) implies $\nabla_{\rho_j\beta_{j'}}\alpha_i = 0$, so the equations are inconsistent. Suppose $q = 0$. Then

$$Tm_j = \iota^\top S_j\begin{pmatrix} I_p \\ D^{-1}C\end{pmatrix}y_i^0, \qquad j = 1,\ldots,p,$$

and, because $\nabla_{\rho_{j'}}b_j(\rho) = \nabla_{\rho_{j'}\rho_j}a(\rho) = \nabla_{\rho_j}b_{j'}(\rho)$, (A.3) will be inconsistent if $\nabla_{\rho_{j'}}m_j \neq \nabla_{\rho_j}m_{j'}$ for some j,j'. Take $j = p$ and $j' = p-1$. The first element of y_i^0 appears in Tm_p and Tm_{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

$$\nabla_{\rho_{p-1}}\gamma_p = \rho_p\sum_{t=0}^{T-p-1}\nabla_{\rho_{p-1}}\varphi_t = \rho_p\sum_{t=1}^{T-p}\nabla_{\rho_p}\varphi_t, \qquad \nabla_{\rho_p}\gamma_{p-1} = \rho_p\sum_{t=1}^{T-p}\nabla_{\rho_p}\varphi_t + \sum_{k=0}^{T-p}\varphi_t,$$

using $\varphi_0 = 1$ and $\nabla_{\rho_{p-1}}\varphi_t = \nabla_{\rho_p}\varphi_{t+1}$. The latter follows from differentiating φ_t and a change of variable from $k_{p-1}-1$ to k_{p-1}, giving

$$\nabla_{\rho_{p-1}}\varphi_t = \sum_{\tau^\top k = t-p+1}\frac{(\iota^\top k + 1)!}{k_1!\cdots k_p!}\rho^k,$$

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.3) is inconsistent. □

Proof of Equation (3.5). By the preceeding proof,

$$\frac{\mathbb{E}_{\vartheta,\alpha_i}\nabla_{\alpha_i\vartheta}\ell_i}{\mathbb{E}_{\vartheta,\alpha_i}\nabla_{\alpha_i\alpha_i}\ell_i} = \begin{pmatrix} \mathbb{E}_{\vartheta,\alpha_i}Z_i^\top\iota/T \\ 0 \end{pmatrix}$$

and so

$$g_i = \begin{pmatrix} \sigma^{-2}(Z_i^\top - \mathbb{E}_{\vartheta,\alpha_i}Z_i^\top\iota\iota^\top/T)\,(y_i - Z_i\theta - \iota\alpha_i) \\ \sigma^{-4}\,(y_i - Z_i\theta - \iota\alpha_i)^\top\,(y_i - Z_i\theta - \iota\alpha_i)\,/2 - \sigma^{-2}T/2 \end{pmatrix}.$$

Recalling $\mathbb{E}_{\vartheta,\alpha_i}Z_i^\top\iota/T = -(T-1)b(\rho)\alpha_i + m$, we have

$$\nabla_{\alpha_i\alpha_i}g_i = \begin{pmatrix} -2\sigma^{-2}T(T-1)b(\rho) \\ \sigma^{-4}T \end{pmatrix}, \qquad \nabla_{\alpha_i\alpha_i}\ell_i = -\sigma^{-2}T,$$

and therefore

$$g_i - \frac{1}{2}\frac{\nabla_{\alpha_i\alpha_i}g_i}{\nabla_{\alpha_i\alpha_i}\ell_i} = \begin{pmatrix} \sigma^{-2}(Z_i^\top - \mathbb{E}_{\vartheta,\alpha_i}Z_i^\top\iota\iota^\top/T)\,(y_i - Z_i\theta - \iota\alpha_i) - (T-1)b(\rho) \\ \sigma^{-4}(y_i - Z_i\theta - \iota\alpha_i)^\top(y_i - Z_i\theta - \iota\alpha_i)/2 - \sigma^{-2}(T-1)/2 \end{pmatrix}.$$

Evaluating at $\alpha_i = \widehat{\alpha}_i(\vartheta) = \iota^\top(y_i - Z_i\theta)/T$ and noting that $\nabla_{\alpha_i\alpha_i\alpha_i}\ell_i = 0$ gives (3.5). $\square$

Proof of Lemma 3. Let $A = S_1F_0$ and $B = \nabla_{\rho_0}A$. Then

$$b_0 = -\frac{\iota^\top A\iota}{T(T-1)}, \qquad c_0 = -\frac{\iota^\top B\iota}{T(T-1)}, \qquad V_0^{LB} = \frac{\mathrm{tr}A^\top MA}{T-1} = \frac{T\mathrm{tr}AA^\top - \iota^\top AA^\top\iota}{T\,(T-1)}.$$

Hence, $V_0^{LB} \geq 2b_0^2$ and $V_0^{LB} \geq 2b_0^2 - c_0$ if and only if

$$T\mathrm{tr}AA^\top - \iota^\top AA^\top\iota - \frac{2\left(\iota^\top A\iota\right)^2}{T\,(T-1)} \geq 0, \tag{A.5}$$

$$T\mathrm{tr}AA^\top - \iota^\top AA^\top\iota - \frac{2\left(\iota^\top A\iota\right)^2}{T\,(T-1)} - \iota^\top B\iota \geq 0. \tag{A.6}$$

The matrix $A = A_T$ is

$$\begin{aligned} A &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad T = 2, \\ A &= \begin{pmatrix} A_{T-1} & 0 \\ a_T^\top & 0 \end{pmatrix}, \qquad a_T = (\rho^{T-2}, \rho^{T-3}, ..., 1)^\top, \qquad T > 2, \end{aligned}$$

where the subscript on ρ is omitted. By recursion, it can be deduced that

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

yielding V_0^{LB} as stated in the lemma. Now let $r > 0$ and use the equalities just obtained to see that if (A.6) holds for $\rho = r$, then (A.5) holds for $\rho = r$ and (A.5) and (A.6) hold for $\rho = -r$, with strict inequalities for $T \geq 3$. Hence, we only need to show that (A.6) holds for $\rho \geq 0$, with equality if and only if $T = 2$ or $\rho = 1$. Write (A.6) as $Q_T \geq 0$. Because $Q_2 = 0$, to show that (A.6) 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 ΔQ_T as

$$\begin{aligned} \Delta Q_T &= \Delta\left(T\mathrm{tr}AA^\top - \iota^\top AA^\top\iota - 2\frac{\left(\iota^\top A\iota\right)^2}{T\,(T-1)} - \iota^\top B\iota\right)_T \\ &= \left\{\left(\mathrm{tr}AA^\top\right)_{T+1} - 2\frac{\left(\iota^\top A\iota\right)_{T+1}^2}{T\,(T+1)}\right\} + \left\{2\frac{\left(\iota^\top A\iota\right)_T^2}{T\,(T-1)} - \Delta\left(\iota^\top B\iota\right)_T\right\} \\ &\quad + \left\{T\Delta\left(\mathrm{tr}AA^\top\right)_T - \Delta\left(\iota^\top AA^\top\iota\right)_T\right\} \end{aligned}$$

and denote the quantities in braces as τ_1, τ_2, and τ_3. Using $T(T+1)/2=\sum_{i=0}^{T-1}(T-i)$, we have

$$\begin{aligned}\tau_1 &= \sum_{j=0}^{T-1}(T-j)\rho^{2j}-\frac{2}{T(T+1)}\left(\sum_{j=0}^{T-1}(T-j)\rho^j\right)^2\\ &= \frac{2}{T(T+1)}\left(\sum_{i=0}^{T-1}\sum_{j=0}^{T-1}(T-i)(T-j)\rho^{2j}-\sum_{i=0}^{T-1}\sum_{j=0}^{T-1}(T-i)(T-j)\rho^{i+j}\right)\\ &= \frac{2}{T(T+1)}u^\top Ru,\end{aligned}$$

where $u=(T,T-1,...,1)^\top$ and

$$R=\begin{pmatrix}1 & 1 & \cdots & 1\\ \rho^2 & \rho^2 & \cdots & \rho^2\\ \vdots & \vdots & \ddots & \vdots\\ \rho^{2T-2} & \rho^{2T-2} & \cdots & \rho^{2T-2}\end{pmatrix}-\begin{pmatrix}1 & \rho & \cdots & \rho^{T-1}\\ \rho & \rho^2 & \cdots & \rho^T\\ \vdots & \vdots & \ddots & \vdots\\ \rho^{T-1} & \rho^T & \cdots & \rho^{2T-2}\end{pmatrix}.$$

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

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

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,

$$\tau_3=T\sum_{j=0}^{T-1}\rho^{2j}-\left(\sum_{j=0}^{T-1}\rho^j\right)^2=\sum_{i=0}^{T-1}\sum_{j=0}^{T-1}\rho^{2j}-\sum_{i=0}^{T-1}\sum_{j=0}^{T-1}\rho^{i+j}=\iota^\top R\iota\geq 0.$$

Use

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

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 τ_2 is a polynomial of degree $2T-4$ in ρ. When $T=2$ or $\rho=1$, $\tau_2=0$. When $\rho\neq 1$,

$$\left(\iota^\top A\iota\right)_T=\frac{T-1-T\rho+\rho^T}{(1-\rho)^2},\qquad \Delta\left(\iota^\top B\iota\right)_T=\frac{1-\rho^T-T\rho^{T-1}+T\rho^T}{(1-\rho)^2},$$

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 $T>2$,

$$\lim_{\rho\to 1}\tau_2(1-\rho)^{-2}=\frac{1}{72}T(T-1)(T-2)(T+1)>0$$

and, therefore, $\tau_2=(1-\rho)^2P(T,\rho)$, where $P(T,\rho)$ is a polynomial of degree $2T-6$. If all coefficients of $P(T,\rho)=\sum_{j=0}^{2T-6}p_j\rho^j$ are positive, we conclude that $\tau_2\geq 0$. Write $\tau_2=\sum_{j=0}^{2T-4}q_j\rho^j$, where q_j is found as

$$q_j=\begin{cases}\frac{d}{6}\left\{(j+1)(j(j-1)+6(T-1)(T-j-1))-3jT(T-1)\right\}, & j\leq T-2,\\ \frac{d}{6}(2T-j-1)(2T-j-2)(2T-j-3), & T-1\leq j.\end{cases}$$

Equating the coefficients of τ_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 $0 \le k \le 2T-6$, 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 $\min(T-2, 2T-6)$, we obtain

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

and, hence, $p_k > 0$ because either $k < T-2$, implying $2T-k-2 > T$, or $k = T-2 \le 2T-6$, implying $T \ge 4$ and $k \ge 2$. Therefore, $\tau_2 \ge 0$. This establishes $Q_T \ge 0$, that is, (A.6). 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 $T \ge 2$ and $\rho \ne 1$, 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 ρ_0 is equivalent to $b(\rho)$ approaching $S(\rho)$ from below as ρ approaches ρ_0 from the left. We will write this as $b(\rho) \uparrow S(\rho)$ at ρ_0, and show that $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \overline{\rho}]$ at most once. From

$$\nabla_\rho H(\rho) = \frac{2(\rho-\rho_{\mathrm{ml}})\left(3\zeta_0^2 - (\rho-\rho_{\mathrm{ml}})^2\right)}{\left(\zeta_0^2 + (\rho-\rho_{\mathrm{ml}})^2\right)^3}$$

it follows that $S(\rho)$ is strictly concave on $[\underline{\rho}, \rho_{\mathrm{ml}}]$ and strictly convex on $[\rho_{\mathrm{ml}}, \overline{\rho}]$. Because $\varphi_t = \rho^t$, $b(\rho)$ and its first two derivatives are

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

For $\rho \ne 1$,

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

When $T \le 3$, $b(\rho)$ is linear and so, given that $S(\rho)$ is concave-convex on $[\underline{\rho}, \overline{\rho}]$, $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \overline{\rho}]$ at most once. Suppose $T \ge 4$. 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 $[\underline{\rho}, \overline{\rho}]$ can only be on $(\rho_{\mathrm{ml}}, \overline{\rho}]$, where $S(\rho)$ is strictly convex and is approached from below by $b(\rho)$ at most once. Now suppose T is odd and $T \ge 5$. Then,

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

so $b(\rho)$ is strictly convex on $(-\infty, \rho_v]$ and strictly concave on $[\rho_v, \infty)$ for some $\rho_v \in (-1, -\frac{1}{2})$ and decreases on $\mathbb{R}$. Define ρ_u by $b(\rho_u) = 0$, that is, by $T(1-\rho_u) = 1 - \rho_u^T$, $\rho_u \in \mathbb{R}_-$. Since $T \ge 5$, we have $-2 < \rho_u < -1$. Thus, $b(\rho)$ is negative and strictly convex on $(\rho_u, \rho_v]$, with $-2 < \rho_u < -1 < \rho_v < -\frac{1}{2}$. Let $R = [\rho_u, \rho_v] \cap [\rho_{\mathrm{ml}}, \overline{\rho}]$. If R is empty, then $\rho_v < \rho_{\mathrm{ml}}$ or $\overline{\rho} < \rho_u$; in either case, by the concavity-convexity of $S(\rho)$, $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \overline{\rho}]$ at most once. If R is non-empty, to show that $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \overline{\rho}]$ 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) $V_0^{LB} \ge \frac{T-1}{T}$ if $\rho_0 \le 0$; (ii) $V_0^{LB} \ge \frac{1}{2}$ if $\rho_0 > 0$. By (ii), $\rho_{\mathrm{ml}} = \rho_0 + b_0/V_0 \ge \rho_0 + 2b_0 > -\frac{1}{2}$ if $0 < \rho_0 \le 1$ because $b(0) = -\frac{1}{T}$, $b(1) = -\frac{1}{2}$, and $b(\rho)$ is concave on $[0,1]$. Further, $\rho_{\mathrm{ml}} > 0$ if $\rho_0 > 1$ because, then, $\frac{b_0}{V_0} > \frac{1}{2b_0} > -1$. Hence, R is empty if $\rho_0 > 0$. Now suppose $\rho_0 \le 0$. Define ρ_w by $S(\rho_w) = b(\rho_v)$, $\rho_w \in [\rho_{\mathrm{ml}}, \overline{\rho}]$; and ρ_w' by $S(\rho_w') = b(0) = -\frac{1}{T}$, $\rho_w' \in [\rho_{\mathrm{ml}}, \overline{\rho}]$. Then $\rho_w - \rho_{\mathrm{ml}} < \rho_w' - \rho_{\mathrm{ml}} = \frac{1}{2}(T - \sqrt{T^2 - 4\zeta_0^2})$. By (i), $\zeta_0^2 = \frac{V_0 - b_0^2}{V_0^2} \le \frac{1}{V_0} \le \frac{T}{T-1} \le \frac{5}{4}$. Since $H(\rho)$ increases on $[\rho_{\mathrm{ml}}, \overline{\rho}]$ and $H(\rho_w')$ decreases in T and increases in ζ_0^2,

$$\begin{aligned} H(\rho_w) &= -\frac{\zeta_0^2}{\left(\zeta_0^2 + (\rho_w - \rho_{\mathrm{ml}})^2\right)^2} + 2S^2(\rho_w) < -\frac{\zeta_0^2}{\left(\zeta_0^2 + (\rho_w' - \rho_{\mathrm{ml}})^2\right)^2} + \frac{2}{T^2} \\ &\le -\frac{5/4}{\left(\frac{5}{4} + \frac{1}{4}\left(5 - \sqrt{20}\right)^2\right)^2} + \frac{2}{25} < -\frac{1}{2} \end{aligned}$$

and so, $H(\rho) < -\frac{1}{2}$ for $\rho \in [\rho_{\rm ml}, \rho_w]$. On the other hand, $T(1-\rho_u) = 1-\rho_u^T$ implies $\frac{1-\rho_u}{\rho_u} = \frac{1-\rho_u^{T-1}}{T-1}$ and, therefore,

$$c(\rho_u) = -\frac{-T + T\rho_u^{T-1}}{T(T-1)(1-\rho_u)^2} = \frac{1}{\rho_u(1-\rho_u)} > -\frac{1}{2}.$$

So, $c(\rho) > -\frac{1}{2}$ for $\rho \in [\rho_u, \rho_v]$ and $H(\rho) < c(\rho)$ for $\rho \in R$. We conclude that $b(\rho) \uparrow S(\rho)$ on $[\underline{\rho}, \overline{\rho}]$ at most once, provided (i) and (ii) hold, which we now show. Write $V_0^{LB} = \frac{1}{T(T-1)}\sum_{j=0}^{2T-4} v_j \rho_0^j$, where

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

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

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

Hence, V_0^{LB} decreases in ρ_0 on $\mathbb{R}_-$ and (i) follows because $V_0^{LB} = \frac{T-1}{T}$ when $\rho_0 = 0$. When $0 < \rho_0 < 1$, a sufficient condition for $V_0^{LB} \geq \frac{1}{2}$ is that $d_k \geq 0$ for $0 \leq k \leq T-2$, where $d_k = \sum_{j=0}^{k}(v_{2j} + v_{2j+1}) - \frac{T(T-1)}{2}$. We have

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

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

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

Define f_k by $d_k = \frac{1}{2}(k+1)f_k$. Then, $f_0 = (T-2)(T-5) \geq 0$, $f_1 = \frac{1}{2}\left(3T^2 - 25T + 54\right) > 0$, and, for $k \geq 2$,

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

Hence, $V_0^{LB} \geq \frac{1}{2}$ when $0 < \rho_0 < 1$. When $\rho_0 \geq 1$, it also holds that $V_0^{LB} \geq \frac{1}{2}$ because then $b_0 < b(1) = -\frac{1}{2}$ and $V_0^{LB} \geq 2b_0^2$. Therefore, (ii) holds. □

Proof of Lemma 4. Use $y_{i,-1} = S_1(\xi_{0i} + F_0\varepsilon_i)$ to write $Z_i = (y_{i,-1}, X_i) = (S_1F_0\varepsilon_i, 0) + \Xi_i$, where $\Xi_i = (S_1\xi_{0i}, X_i)$ is independent of ε_i. Proceeding as above, we have

$$V_0 = \frac{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} Z_i^\top M Z_i}{\sigma_0^2(T-1)} = \begin{pmatrix} V_0^{LB} & 0 \\ 0 & 0 \end{pmatrix} + V_\Xi,$$

where

$$V_\Xi = \begin{pmatrix} V_{\xi\xi} & V_{\xi X} \\ V_{X\xi} & V_{XX} \end{pmatrix} = \frac{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Xi_i^\top M \Xi_i}{\sigma_0^2(T-1)}$$

is positive semi-definite and V_{XX} is positive definite by assumption. Therefore, $V_{\xi\xi} - V_{\xi X}V_{XX}^{-1}V_{X\xi} \geq 0$ and $(V_0^{\rho\rho})^{-1} = V_0^{LB} + V_{\xi\xi} - V_{\xi X}V_{XX}^{-1}V_{X\xi} \geq V_0^{LB}$. □

Appendix B: Tables

Table 1. Simulation results for the first-order autoregression, $N = 100$ observations.

			bias				std				$\mathrm{ci}^{\mathrm{a}}_{.95}$				$\mathrm{ci}^{\mathrm{b}}_{.95}$			
ψ	T	ρ_0	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$
0	2	.50	−.143	—	−.747	−.996	.266	—	.151	.106	.879	.919	.003	.002	.928	.979	.011	.000
1	2	.50	.027	—	−.375	−.750	.268	—	.140	.093	.942	.934	.352	.000	.948	.967	.321	.000
2	2	.50	.026	—	.112	−.425	.168	—	.112	.075	.969	.943	1.000	.001	.949	.943	.825	.005
0	4	.50	.006	−.043	−.296	−.441	.142	.148	.066	.055	.955	.921	.009	.000	.945	.913	.025	.000
1	4	.50	.014	−.055	−.141	−.328	.124	.166	.067	.055	.964	.925	.474	.000	.942	.906	.493	.001
2	4	.50	.003	−.015	.073	−.184	.064	.084	.056	.046	.968	.937	.941	.021	.939	.932	.775	.045
0	6	.50	.007	−.033	−.147	−.246	.090	.081	.051	.043	.968	.923	.155	.000	.944	.887	.231	.001
1	6	.50	.003	−.047	−.072	−.196	.067	.098	.049	.041	.971	.911	.666	.003	.945	.864	.698	.009
2	6	.50	−.001	−.022	.043	−.123	.045	.064	.042	.035	.958	.924	.873	.064	.937	.905	.834	.101
0	8	.50	.001	−.027	−.085	−.163	.056	.058	.043	.036	.965	.921	.427	.005	.941	.864	.516	.011
1	8	.50	.000	−.040	−.044	−.137	.048	.069	.040	.034	.960	.908	.773	.022	.944	.835	.809	.037
2	8	.50	.000	−.022	.029	−.092	.036	.052	.034	.030	.954	.927	.870	.128	.946	.889	.877	.171
0	16	.50	−.001	−.020	−.022	−.064	.028	.030	.026	.024	.948	.902	.841	.238	.944	.794	.865	.274
1	16	.50	−.001	−.027	−.012	−.059	.026	.035	.025	.023	.947	.880	.904	.278	.946	.760	.918	.321
2	16	.50	−.001	−.022	.009	−.047	.023	.032	.023	.021	.947	.890	.915	.387	.946	.792	.936	.426
0	24	.50	−.001	−.018	−.010	−.039	.021	.022	.020	.019	.944	.868	.902	.461	.944	.728	.916	.496
1	24	.50	−.001	−.023	−.006	−.037	.020	.025	.020	.019	.943	.850	.925	.492	.941	.691	.931	.527
2	24	.50	.000	−.020	.004	−.031	.019	.024	.018	.017	.946	.865	.928	.555	.943	.724	.939	.582
0	2	.95	−.143	—	−.525	−1.000	.265	—	.150	.100	.879	.922	.287	.000	.927	.980	.128	.000
1	2	.95	−.120	—	−.490	−.977	.266	—	.152	.101	.893	.921	.369	.000	.934	.978	.170	.000
2	2	.95	−.059	—	−.385	−.907	.265	—	.152	.101	.918	.930	.638	.000	.947	.975	.346	.000
0	4	.95	−.086	−.676	−.276	−.537	.124	.456	.071	.057	.889	.673	.074	.000	.908	.615	.059	.000
1	4	.95	−.061	−.696	−.239	−.509	.124	.472	.073	.059	.910	.689	.185	.000	.932	.627	.132	.000
2	4	.95	−.015	−.386	−.151	−.443	.123	.418	.070	.055	.939	.812	.636	.000	.948	.755	.450	.000
0	6	.95	−.059	−.468	−.189	−.382	.084	.263	.051	.042	.889	.543	.046	.000	.899	.306	.056	.000
1	6	.95	−.038	−.511	−.157	−.357	.083	.272	.049	.041	.911	.534	.151	.000	.936	.296	.140	.000
2	6	.95	−.003	−.286	−.080	−.297	.082	.233	.046	.038	.944	.749	.718	.000	.949	.520	.604	.000
0	8	.95	−.044	−.346	−.146	−.301	.063	.174	.038	.032	.889	.433	.032	.000	.908	.151	.048	.000
1	8	.95	−.025	−.399	−.115	−.276	.063	.192	.038	.032	.915	.396	.159	.000	.944	.145	.171	.000
2	8	.95	.002	−.226	−.046	−.222	.063	.159	.035	.029	.946	.660	.807	.000	.940	.356	.730	.000
0	16	.95	−.016	−.148	−.076	−.167	.034	.061	.020	.018	.901	.237	.024	.000	.941	.020	.058	.000
1	16	.95	−.005	−.197	−.052	−.147	.034	.076	.020	.018	.929	.145	.217	.000	.942	.015	.272	.000
2	16	.95	.004	−.125	−.006	−.107	.030	.063	.017	.015	.959	.350	.947	.000	.942	.085	.923	.000
0	24	.95	−.006	−.090	−.050	−.115	.024	.033	.015	.013	.920	.157	.038	.000	.942	.006	.084	.000
1	24	.95	.000	−.126	−.032	−.099	.024	.042	.014	.013	.943	.059	.301	.000	.942	.003	.378	.000
2	24	.95	.001	−.091	.001	−.070	.018	.038	.011	.011	.962	.164	.955	.000	.946	.020	.947	.000

'—' indicates non-existence of the moment.

Table 2. Simulation results for the second-order autoregression, $N = 100$ observations.

			bias				std				$\mathrm{ci}^{\mathrm{a}}_{.95}$				$\mathrm{ci}^{\mathrm{b}}_{.95}$			
ψ	T	ρ_0	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$	$\widehat{\rho}_{\mathrm{al}}$	$\widehat{\rho}_{\mathrm{ab}}$	$\widehat{\rho}_{\mathrm{hk}}$	$\widehat{\rho}_{\mathrm{bh}}$
.3	2	.60	−.147	—	−.999	−.999	.264	—	.102	.102	.885	.919	.000	.000	.930	.980	.000	.000
		.20	−.118	—	−.881	−.881	.714	—	.515	.515	.955	.942	.577	.577	.934	.984	.634	.634
1	2	.60	−.147	—	−1.000	−1.000	.266	—	.103	.103	.873	.923	.000	.000	.925	.980	.000	.000
		.20	−.128	—	−.885	−.885	.315	—	.182	.182	.909	.927	.002	.002	.922	.979	.011	.011
2	2	.60	−.144	—	−1.001	−1.001	.267	—	.100	.100	.878	.926	.000	.000	.924	.982	.000	.000
		.20	−.126	—	−.888	−.888	.261	—	.120	.120	.890	.928	.000	.000	.917	.981	.000	.000
.3	4	.60	−.069	−.263	−.339	−.405	.123	.274	.070	.068	.902	.787	.005	.000	.919	.710	.007	.001
		.20	−.031	−.116	−.454	−.424	.099	.134	.079	.076	.940	.811	.000	.000	.932	.769	.000	.001
1	4	.60	.001	−.040	−.250	−.312	.122	.120	.065	.062	.952	.926	.045	.002	.952	.906	.056	.005
		.20	−.001	−.020	−.386	−.362	.093	.090	.072	.068	.963	.937	.000	.000	.948	.927	.002	.003
2	4	.60	.008	−.011	−.132	−.177	.081	.064	.050	.048	.969	.937	.297	.058	.944	.933	.316	.087
		.20	.004	−.006	−.256	−.241	.072	.064	.059	.056	.965	.942	.005	.007	.941	.938	.020	.024
.3	6	.60	−.031	−.148	−.196	−.237	.082	.137	.054	.053	.926	.777	.071	.007	.931	.591	.087	.019
		.20	−.016	−.072	−.336	−.322	.069	.077	.053	.052	.949	.833	.000	.000	.945	.713	.001	.001
1	6	.60	.009	−.039	−.131	−.167	.082	.073	.051	.050	.962	.910	.314	.096	.948	.850	.326	.128
		.20	.005	−.017	−.269	−.256	.066	.058	.051	.049	.964	.930	.000	.001	.944	.915	.002	.003
2	6	.60	.002	−.011	−.052	−.078	.048	.046	.041	.040	.959	.933	.768	.522	.946	.922	.765	.555
		.20	.000	−.005	−.169	−.158	.046	.044	.042	.041	.952	.943	.025	.036	.942	.943	.051	.063
.3	8	.60	−.015	−.098	−.136	−.162	.064	.088	.046	.045	.943	.782	.180	.065	.944	.555	.217	.093
		.20	−.006	−.050	−.264	−.256	.056	.057	.043	.042	.958	.846	.000	.000	.945	.712	.000	.000
1	8	.60	.008	−.032	−.086	−.108	.063	.054	.042	.042	.964	.905	.507	.288	.950	.835	.532	.325
		.20	.004	−.014	−.208	−.200	.052	.045	.040	.039	.970	.936	.001	.001	.949	.911	.005	.006
2	8	.60	.001	−.011	−.028	−.044	.037	.037	.034	.034	.958	.933	.877	.757	.945	.919	.873	.771
		.20	.000	−.003	−.129	−.121	.036	.036	.034	.034	.950	.949	.046	.062	.952	.948	.085	.104
.3	16	.60	.003	−.041	−.052	−.059	.037	.036	.029	.029	.965	.794	.588	.494	.945	.574	.619	.538
		.20	.001	−.025	−.140	−.138	.035	.031	.027	.026	.963	.862	.001	.001	.944	.744	.003	.003
1	16	.60	.000	−.024	−.033	−.039	.031	.031	.028	.028	.957	.869	.773	.703	.940	.753	.784	.723
		.20	.000	−.011	−.111	−.108	.029	.028	.026	.026	.948	.927	.013	.014	.940	.890	.026	.030
2	16	.60	.000	−.010	−.007	−.012	.025	.026	.025	.025	.947	.925	.944	.921	.945	.887	.937	.921
		.20	−.001	−.003	−.071	−.068	.024	.024	.024	.024	.944	.940	.160	.184	.946	.939	.215	.240
.3	24	.60	.000	−.028	−.029	−.032	.024	.025	.023	.023	.948	.793	.758	.711	.947	.603	.776	.737
		.20	.000	−.018	−.091	−.090	.024	.023	.021	.021	.949	.876	.010	.012	.942	.758	.026	.028
1	24	.60	.000	−.020	−.018	−.021	.023	.023	.022	.022	.946	.848	.855	.828	.946	.724	.864	.837
		.20	−.001	−.011	−.076	−.075	.022	.022	.021	.021	.949	.914	.043	.047	.947	.866	.078	.085
2	24	.60	.000	−.011	−.004	−.007	.020	.020	.020	.020	.951	.913	.948	.940	.952	.858	.946	.941
		.20	.000	−.003	−.051	−.050	.019	.019	.019	.019	.947	.945	.243	.265	.948	.941	.315	.334

'—' indicates non-existence of the moment.

Table 3. Simulation results for the second-order autoregression, $N = 100$ observations (continued).

			bias				std				$\text{ci}^{\text{a}}_{.95}$				$\text{ci}^{\text{b}}_{.95}$			
ψ	T	ρ_0	$\widehat{\rho}_{\text{al}}$	$\widehat{\rho}_{\text{ab}}$	$\widehat{\rho}_{\text{hk}}$	$\widehat{\rho}_{\text{bh}}$	$\widehat{\rho}_{\text{al}}$	$\widehat{\rho}_{\text{ab}}$	$\widehat{\rho}_{\text{hk}}$	$\widehat{\rho}_{\text{bh}}$	$\widehat{\rho}_{\text{al}}$	$\widehat{\rho}_{\text{ab}}$	$\widehat{\rho}_{\text{hk}}$	$\widehat{\rho}_{\text{bh}}$	$\widehat{\rho}_{\text{al}}$	$\widehat{\rho}_{\text{ab}}$	$\widehat{\rho}_{\text{hk}}$	$\widehat{\rho}_{\text{bh}}$
.3	2	1	−.146	—	−1.000	−1.000	.267	—	.102	.102	.875	.923	.000	.000	.924	.977	.000	.000
		−.2	−.082	—	−.523	−.523	.607	—	.444	.444	.966	.962	.772	.772	.943	.990	.806	.806
1	2	1	−.147	—	−1.002	−1.002	.272	—	.103	.103	.875	.916	.000	.000	.927	.979	.000	.000
		−.2	−.077	—	−.520	−.520	.233	—	.150	.150	.921	.922	.065	.065	.927	.976	.128	.128
2	2	1	−.148	—	−1.000	−1.000	.266	—	.102	.102	.883	.924	.000	.000	.927	.980	.000	.000
		−.2	−.077	—	−.518	−.518	.166	—	.088	.088	.905	.923	.000	.000	.918	.978	.001	.001
.3	4	1	−.068	−.224	−.325	−.393	.117	.252	.068	.066	.906	.807	.004	.000	.908	.723	.007	.001
		−.2	.006	−.018	−.318	−.264	.088	.084	.078	.075	.968	.941	.017	.059	.945	.943	.046	.112
1	4	1	.010	−.031	−.210	−.271	.114	.098	.059	.058	.955	.927	.084	.003	.948	.909	.085	.008
		−.2	.009	−.003	−.301	−.250	.081	.078	.074	.071	.969	.940	.015	.055	.947	.948	.041	.099
2	4	1	.002	−.009	−.092	−.132	.060	.051	.043	.042	.970	.939	.454	.126	.947	.936	.442	.152
		−.2	.002	−.002	−.220	−.184	.061	.061	.060	.058	.949	.939	.036	.102	.940	.943	.086	.169
.3	6	1	−.030	−.112	−.163	−.208	.078	.113	.053	.052	.926	.817	.160	.024	.927	.644	.169	.037
		−.2	.002	−.017	−.263	−.230	.062	.057	.053	.052	.965	.929	.002	.009	.946	.914	.009	.022
1	6	1	.007	−.026	−.085	−.124	.072	.061	.047	.046	.964	.923	.604	.260	.947	.878	.582	.273
		−.2	.004	−.001	−.204	−.174	.052	.051	.047	.046	.962	.949	.014	.045	.949	.952	.038	.083
2	6	1	.000	−.009	−.020	−.046	.039	.039	.037	.036	.950	.934	.917	.757	.940	.926	.906	.763
		−.2	.002	.002	−.122	−.101	.041	.042	.040	.039	.945	.941	.134	.271	.949	.945	.207	.354
.3	8	1	−.008	−.064	−.091	−.121	.059	.071	.045	.044	.952	.850	.489	.229	.946	.675	.488	.254
		−.2	.001	−.016	−.217	−.196	.051	.046	.041	.041	.964	.926	.001	.004	.938	.900	.004	.010
1	8	1	.004	−.022	−.044	−.070	.051	.047	.040	.039	.970	.918	.818	.600	.949	.866	.798	.599
		−.2	.000	−.002	−.159	−.140	.041	.041	.038	.037	.953	.945	.016	.042	.947	.946	.042	.084
2	8	1	−.001	−.009	−.005	−.023	.033	.034	.032	.032	.947	.934	.948	.890	.950	.920	.940	.877
		−.2	.001	.002	−.091	−.076	.034	.035	.033	.032	.941	.943	.212	.346	.946	.944	.284	.426
.3	16	1	.000	−.027	−.019	−.027	.030	.032	.028	.028	.955	.859	.902	.844	.948	.727	.894	.834
		−.2	.001	−.008	−.110	−.105	.029	.028	.025	.025	.954	.938	.009	.016	.943	.917	.026	.039
1	16	1	−.002	−.019	−.009	−.017	.027	.029	.026	.026	.948	.895	.939	.905	.944	.807	.929	.895
		−.2	.000	−.002	−.086	−.081	.026	.026	.024	.024	.951	.948	.068	.103	.947	.944	.116	.157
2	16	1	.000	−.009	.004	−.002	.024	.025	.024	.024	.944	.926	.944	.949	.948	.896	.946	.947
		−.2	.000	.001	−.054	−.049	.023	.023	.022	.022	.942	.943	.327	.410	.947	.946	.405	.483
.3	24	1	−.001	−.019	−.005	−.009	.022	.024	.022	.022	.945	.869	.939	.925	.942	.764	.936	.920
		−.2	.000	−.007	−.072	−.070	.022	.022	.020	.020	.948	.933	.055	.073	.949	.911	.104	.123
1	24	1	−.001	−.015	−.002	−.005	.021	.022	.021	.021	.948	.897	.952	.946	.948	.816	.946	.942
		−.2	.000	−.004	−.060	−.058	.021	.021	.020	.020	.949	.944	.146	.182	.951	.938	.209	.243
2	24	1	.000	−.009	.004	.001	.019	.020	.019	.019	.947	.918	.946	.950	.947	.877	.940	.945
		−.2	.000	.000	−.041	−.039	.018	.019	.018	.018	.948	.948	.395	.443	.949	.951	.474	.519

'—' indicates non-existence of the moment.

Table 4. Simulation results for the first-order autoregression with a covariate, $N = 100$ observtions.

			bias				std				$\mathrm{ci}^{\mathrm{a}}_{.95}$				$\mathrm{ci}^{\mathrm{b}}_{.95}$			
ψ	T	θ_0	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$
0	2	.50	−.112	−.133	−.961	−.961	.268	—	.103	.103	.890	.863	.000	.000	.935	.942	.000	.000
		.50	−.032	−.039	−.242	−.242	.230	—	.166	.166	.958	.926	.663	.663	.935	.958	.714	.714
1	2	.50	.035	−.001	−.743	−.743	.268	—	.096	.096	.945	.921	.000	.000	.945	.952	.000	.000
		.50	.005	.002	−.075	−.075	.232	—	.178	.178	.972	.953	.916	.916	.948	.954	.928	.928
2	2	.50	.022	.000	−.422	−.422	.167	—	.076	.076	.967	.941	.001	.001	.945	.950	.004	.004
		.50	−.002	−.001	.019	.019	.233	—	.202	.202	.943	.940	.938	.938	.944	.944	.942	.941
0	4	.50	.015	−.097	−.361	−.413	.141	.120	.059	.056	.959	.854	.000	.000	.946	.737	.001	.000
		.50	.006	−.010	−.031	−.048	.124	.120	.127	.121	.964	.944	.906	.909	.949	.949	.943	.936
1	4	.50	.012	−.064	−.260	−.310	.116	.102	.056	.054	.968	.882	.006	.000	.947	.811	.017	.001
		.50	.001	.000	.018	.004	.124	.122	.130	.124	.948	.942	.911	.926	.945	.946	.938	.941
2	4	.50	.001	−.023	−.137	−.175	.060	.061	.046	.044	.966	.927	.161	.027	.949	.897	.223	.054
		.50	−.002	.000	.029	.023	.123	.123	.129	.124	.940	.941	.914	.929	.944	.944	.941	.943
0	6	.50	.004	−.067	−.192	−.230	.082	.071	.042	.042	.969	.831	.007	.000	.945	.635	.014	.001
		.50	.001	−.004	.000	−.010	.092	.091	.096	.094	.949	.945	.927	.933	.947	.945	.946	.944
1	6	.50	.000	−.054	−.150	−.184	.062	.065	.041	.040	.968	.856	.047	.005	.943	.702	.071	.014
		.50	.001	.002	.018	.011	.092	.092	.096	.094	.943	.942	.923	.932	.947	.948	.943	.944
2	6	.50	.000	−.025	−.090	−.114	.042	.044	.034	.034	.958	.906	.259	.080	.943	.830	.309	.123
		.50	.000	.003	.020	.017	.091	.091	.095	.093	.945	.945	.929	.935	.944	.945	.942	.943
0	8	.50	−.001	−.055	−.127	−.152	.052	.052	.035	.035	.966	.812	.054	.008	.945	.573	.074	.020
		.50	.001	−.001	.006	.001	.076	.076	.080	.079	.948	.948	.936	.941	.947	.947	.947	.947
1	8	.50	−.001	−.048	−.105	−.127	.045	.049	.034	.034	.955	.829	.128	.034	.938	.631	.156	.058
		.50	.000	.002	.013	.009	.077	.077	.080	.079	.942	.942	.929	.935	.948	.946	.941	.944
2	8	.50	−.001	−.027	−.069	−.086	.035	.037	.029	.029	.948	.877	.334	.158	.942	.763	.374	.202
		.50	.000	.003	.014	.013	.077	.076	.079	.078	.945	.946	.934	.937	.946	.946	.943	.943
0	16	.50	−.001	−.038	−.052	−.060	.027	.028	.023	.023	.950	.717	.387	.275	.948	.411	.418	.310
		.50	.000	.002	.005	.004	.051	.051	.052	.052	.948	.947	.944	.945	.947	.946	.945	.945
1	16	.50	−.001	−.036	−.048	−.055	.025	.027	.022	.022	.952	.732	.435	.312	.949	.441	.463	.352
		.50	.000	.003	.007	.006	.051	.051	.052	.052	.944	.944	.939	.940	.947	.948	.947	.947
2	16	.50	−.001	−.027	−.038	−.044	.022	.024	.020	.020	.943	.787	.532	.421	.939	.548	.559	.461
		.50	−.001	.003	.006	.006	.051	.051	.052	.052	.945	.946	.944	.944	.950	.950	.948	.948
0	24	.50	.000	−.033	−.033	−.036	.020	.020	.018	.018	.946	.635	.570	.497	.946	.315	.599	.531
		.50	.001	.003	.004	.004	.041	.041	.041	.041	.943	.943	.941	.942	.945	.944	.945	.945
1	24	.50	−.001	−.033	−.031	−.034	.019	.020	.018	.018	.945	.634	.586	.516	.944	.330	.609	.546
		.50	.000	.003	.004	.004	.041	.042	.042	.042	.945	.944	.943	.943	.944	.943	.943	.943
2	24	.50	.000	−.027	−.026	−.029	.018	.018	.016	.016	.947	.692	.647	.578	.948	.407	.665	.606
		.50	.000	.004	.005	.005	.041	.041	.042	.042	.941	.942	.941	.941	.943	.942	.942	.942

'—' indicates non-existence of the moment.

Table 5. Simulation results for the first-order autoregression with a covariate, $N = 100$ observations (continued).

			bias				std				$\mathrm{ci}^{\mathrm{a}}_{.95}$				$\mathrm{ci}^{\mathrm{b}}_{.95}$			
ψ	T	θ_0	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$	$\widehat{\theta}_{\mathrm{al}}$	$\widehat{\theta}_{\mathrm{ab}}$	$\widehat{\theta}_{\mathrm{hk}}$	$\widehat{\theta}_{\mathrm{bh}}$
0	2	.95	−.144	−.942	−1.000	−1.000	.266	—	.102	.102	.879	.804	.000	.000	.929	.919	.000	.000
		.05	−.002	−.029	−.026	−.026	.215	—	.160	.160	.975	.976	.932	.932	.946	.988	.941	.941
1	2	.95	−.122	−.917	−.977	−.977	.267	—	.102	.102	.890	.820	.000	.000	.936	.927	.000	.000
		.05	−.001	.008	.013	.013	.215	—	.163	.163	.977	.979	.938	.938	.947	.988	.943	.943
2	2	.95	−.062	−.382	−.908	−.908	.268	—	.100	.100	.915	.842	.000	.000	.946	.936	.000	.000
		.05	.002	.014	.045	.045	.224	—	.168	.168	.973	.967	.926	.926	.943	.983	.933	.934
0	4	.95	−.086	−.641	−.487	−.537	.124	.253	.060	.058	.890	.270	.000	.000	.909	.125	.000	.000
		.05	−.002	−.015	−.013	−.014	.117	.108	.116	.110	.977	.951	.918	.930	.950	.955	.948	.948
1	4	.95	−.064	−.585	−.461	−.510	.126	.256	.060	.059	.903	.327	.000	.000	.932	.160	.000	.000
		.05	.000	.010	.015	.013	.119	.110	.119	.114	.974	.946	.911	.923	.950	.955	.940	.943
2	4	.95	−.018	−.303	−.398	−.445	.124	.198	.057	.056	.933	.593	.000	.000	.945	.394	.000	.000
		.05	.002	.020	.039	.037	.123	.113	.120	.115	.971	.941	.898	.912	.949	.947	.931	.932
0	6	.95	−.059	−.452	−.352	−.382	.085	.146	.042	.042	.885	.099	.000	.000	.898	.013	.000	.000
		.05	−.002	−.010	−.008	−.009	.091	.087	.091	.089	.971	.947	.924	.931	.946	.949	.945	.945
1	6	.95	−.038	−.407	−.328	−.357	.083	.149	.041	.041	.914	.147	.000	.000	.938	.024	.000	.000
		.05	.000	.008	.010	.009	.091	.086	.091	.089	.973	.943	.925	.932	.947	.946	.943	.942
2	6	.95	−.004	−.202	−.272	−.298	.083	.104	.038	.038	.945	.440	.000	.000	.948	.149	.000	.000
		.05	−.002	.012	.022	.022	.091	.086	.091	.089	.968	.947	.924	.929	.948	.947	.940	.940
0	8	.95	−.044	−.348	−.283	−.301	.064	.099	.032	.033	.888	.033	.000	.000	.902	.002	.000	.000
		.05	.000	−.006	−.005	−.005	.075	.073	.076	.075	.973	.944	.929	.933	.949	.947	.945	.945
1	8	.95	−.025	−.309	−.259	−.277	.063	.100	.032	.032	.912	.066	.000	.000	.942	.004	.000	.000
		.05	.000	.007	.008	.008	.075	.073	.075	.074	.972	.947	0.93	.937	.950	.951	.948	.948
2	8	.95	.003	−.148	−.207	−.222	.063	.066	.029	.029	.948	.340	.000	.000	.944	.066	.000	.000
		.05	.001	.012	.019	.019	.077	.074	.076	.076	.964	.947	.932	.936	.949	.947	.941	.942
0	16	.95	−.016	−.175	−.162	−.166	.034	.039	.018	.018	.904	.001	.000	.000	.936	.000	.000	.000
		.05	.000	−.003	−.003	−.003	.051	.051	.051	.051	.970	.943	.939	.941	.946	.945	.946	.945
1	16	.95	−.004	−.154	−.142	−.146	.033	.038	.018	.018	.945	.004	.000	.000	.945	.000	.000	.000
		.05	.000	.004	.004	.004	.051	.051	.051	.051	.966	.944	.941	.941	.948	.945	.943	.944
2	16	.95	.003	−.076	−.104	−.107	.030	.025	.015	.015	.959	.095	.000	.000	.945	.002	.000	.000
		.05	.000	.005	.008	.008	.051	.050	.051	.051	.952	.946	.941	.942	.948	.947	.946	.946
0	24	.95	−.006	−.116	−.113	−.115	.024	.023	.013	.013	.917	.000	.000	.000	.942	.000	.000	.000
		.05	.000	−.002	−.002	−.002	.041	.041	.041	.041	.969	.944	.943	.943	.946	.947	.948	.947
1	24	.95	.000	−.102	−.097	−.099	.024	.022	.013	.013	.942	.000	.000	.000	.940	.000	.000	.000
		.05	.000	.003	.003	.003	.041	.041	.041	.041	.960	.944	.940	.940	.945	.942	.941	.941
2	24	.95	.001	−.055	−.068	−.069	.018	.015	.011	.011	.962	.024	.000	.000	.943	.000	.000	.000
		.05	.000	.004	.005	.005	.041	.041	.041	.041	.944	.942	.940	.940	.945	.943	.944	.943

‘—’ indicates non-existence of the moment.

References

Ahn, S. C. and G. M. Thomas (2006). Likelihood based inference for dynamic panel data models. Unpublished manuscript.

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

Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Económicas*, 27:423–458.

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.

Bester, C. A. and C. Hansen (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics*, 27:131–148.

Bun, M. J. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics*, 23:200–210.

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 (with discussion). *Journal of the Royal Statistical Society, Series B*, 49:1–39.

Cruddas, A., N. Reid, and D. R. Cox (1989). A time series illustration of approximate conditional likelihood. *Biometrika*, 76:231–237.

DiCiccio, T. J., M. A. Martin, S. E. Stern, and G. A. Young (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society, Series B*, 458:189–203.

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.

Godambe, V. P. and M. E. Thompson (1974). Estimating equations in the presence of a nuisance parameters. *Annals of Statistics*, 2:568–571.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel data model with fixed effects when both n and T are large. *Econometrica*, 70:1639–1657.

Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel data models with fixed effects. *Econometric Theory*, 27:1152–1191.

Kalbfleisch, J. P. and D. A. Sprott (1970). Application of likelihood methods to models involving large numbers of parameters (with discussion). *Journal of the Royal Statistical Society, Series B*, 32:175–208.

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.

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.

Pace, L. and A. Salvan (2006). Adjustments of profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, 136:3554–3564.

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.

Sartori, N. (2003). Modified profile likelihood in models with stratum nuisance parameters. *Biometrika*, 90:533–549.

Severini, T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, 85:403–411.

Severini, T. A. (2000). *Likelihood Methods in Statistics.* Oxford Statistical Science Series. Oxford University Press.

Sweeting, T. J. (1987). Discussion of the paper by Professors Cox and Reid. *Journal of the Royal Statistical Society, Series B*, 49:20–21.

Woutersen, T. (2002). Robustness against incidental parameters. Unpublished manuscript.

Woutersen, T. and M. Voia (2004). Efficient estimation of the dynamic linear model with fixed effects and regressors. Unpublished manuscript.