A further look at Modified ML estimation of the panel AR(1) model with fixed effects and arbitrary initial conditions.

Kruiniger, Hugo

Department of Economics and Finance, University of Durham

16 June 2018

A further look at Modified ML estimation of the panel AR(1) model with fixed effects and arbitrary initial conditions.

Hugo Kruiniger*
Durham University

This version: 15 August 2021
Previous version 16 June 2018

Abstract

In this paper we consider two kinds of generalizations of Lancaster's (*Review of Economic Studies*, 2002) Modified ML estimator (MMLE) for the panel AR(1) model with fixed effects and arbitrary initial conditions and possibly covariates when the time dimension, T, is fixed. When the autoregressive parameter $\rho = 1$, the limiting modified profile log-likelihood function for this model has a stationary point of inflection and ρ is first-order underidentified but second-order identified. We show that the generalized MMLEs exist w.p.a.1 and are uniquely defined w.p.1. and consistent for any value of $|\rho| \leq 1$. When $\rho = 1$, the rate of convergence of the MMLEs is $N^{1/4}$, where N is the cross-sectional dimension of the panel. We then develop an asymptotic theory for GMM estimators when one of the parameters is only second-order identified and use this to derive the limiting distributions of the MMLEs. They are generally asymmetric when $\rho = 1$. We also show that Quasi LM tests that are based on the modified profile log-likelihood and use its expected rather than observed Hessian, with an additional modification for $\rho = 1$, and confidence regions that are based on inverting these tests have correct asymptotic size in a uniform sense when $|\rho| \leq 1$. Finally, we investigate the finite sample properties of the MMLEs and the QLM test in a Monte Carlo study.

JEL classification: C11, C13, C23.

Keywords: dynamic panel data, expected Hessian, fixed effects, Generalized Method of Moments (GMM), inflection point, Modified Maximum Likelihood, Quasi LM test, second-order identification, singular information matrix, weak moment conditions.

---

*Address: hugo.kruiniger@durham.ac.uk; Dept. of Economics, 23-26 Old Elvet, Durham DH1 3HY, England. An earlier version of this paper was presented in July 2013 at the International Panel Data Conference in London and in August 2014 at the ESEM in Toulouse. It also appeared in 2014 as an SSRN WP.

1 Introduction

In this paper we reconsider Modified ML estimation (cf. Neyman and Scott, 1948) of the panel AR(1) model with fixed effects (FE) and arbitrary initial conditions and possibly strictly exogenous covariates, when the time dimension of the panel, T, is fixed.

It is well known that the FE ML estimator for the autoregressive parameter ρ that is equal to the LSDV estimator is inconsistent when T is fixed, cf. Nickell (1981).[1] To obtain a consistent FE estimator for ρ (or for $\theta_0 = (\rho\, \sigma^2\, \beta')'$ where σ^2 is the error variance and β is the vector of coefficients of the covariates) based on the likelihood function for the model, Lancaster (2002) proposed a Bayesian approach that involves using a reparametrization of the fixed effects, which aims to achieve information orthogonality (but fails to do so when covariates are present), and integrating the new effects from the likelihood function using a uniform prior density. He defined his estimator for ρ (or for θ_0) as a local rather than a global maximizer of the resulting marginal (or joint) posterior density because this posterior density is improper and has a global maximum at $r = \infty$ for any sample size, cf. Dhaene and Jochmans (2016).[2] Bun and Carree (2005) took a different route and proposed a bias-corrected LSDV estimator for θ_0 with the correction based on formulae for the asymptotic biases of the LSDV estimators for ρ and β. However, a version of their estimator is equal to Lancaster's estimator for θ_0, cf. Dhaene and Jochmans (2016), and both of them can be viewed as a Modified ML estimator (MMLE), cf. Alvarez and Arellano (2004). Bun and Carree (2005) also investigated the finite sample properties of their estimator using various Monte Carlo experiments. They reported non-convergence of their estimator in about 40% of the replications in some experiments where $N = 100$, $T = 6$ and $\rho = 0.8$. The possible non-existence of the MMLE is also related to the fact that the underlying density function is improper. Specifically, when $\rho = 1$, the limiting modified profile log-likelihood function of r has a stationary point of inflection at $r = 1$, cf. Ahn and Thomas (2004), so that the modified profile log-likelihood function may fail to have a local maximum even asymptotically.

In this paper we discuss two kinds of generalizations of Lancaster's MMLEs that exist as N increases with probability approaching one (w.p.a.1) for any $|\rho| \leq 1$.[3] The first type of generalized MMLE minimizes a quadratic form in the modified profile score vector subject to a second-order condition for a maximum of the modified profile likelihood while

[1] FE estimators only use data in differences and are consistent under minimal assumptions.

[2] Lancaster discards the global maxima at $r = \pm\infty$ and only considers local maxima that are stationary points.

[3] Note that w.p.a.1. means with probability approaching one, i.e., w.p.1 asymptotically.

the second type minimizes the norm of the modified profile score for ρ only, subject to a second-order condition for a maximum. The former MMLE depends on a weight matrix.

While Lancaster has only argued that *one* of the local maxima of the posterior density is consistent (if one exists at all), we show that when $|\rho| \leq 1$ the generalized MMLEs are uniquely defined w.p.1. and consistent.

Both types of generalized MMLEs will select a local maximum if one exists. In this case the estimators are equivalent irrespective of the choice of the weight matrix. However, if the modified profile likelihood function of r has no local maximum on the interval $[-1, \infty)$, then these estimators are still consistent but different and the first type of generalized MMLE depends on the choice of the weight matrix.

Dhaene and Jochmans (2016) have shown that their Adjusted Likelihood estimator for the nonstationary panel AR(1) model, which is a constrained version of our second MMLE, is uniquely defined asymptotically. However, they have not demonstrated that their constraints, which depend on the LSDV estimator, guarantee uniqueness of their estimator in finite samples.

We also derive the limiting distributions of the generalized MMLEs. Similar to the cases of the FEMLE of Hsiao et al. (2002) and the REMLE of Chamberlain (1980) and Anderson and Hsiao (1982), if $\rho = 1$, ρ is only second-order identified by their objective functions and as a result the rate of convergence of the MMLEs for ρ is $N^{1/4}$, cf. Ahn and Thomas (2004) and Kruiniger (2013). Our analysis for $\rho = 1$ is closely related to Sargan (1983) for instrumental variable and ML estimators and also to Rotnitzky et al. (2000) for MLEs when a parameter is only second-order identified, although there are some important differences. We view the MMLEs as GMM estimators in order to derive their limiting distributions when $\rho = 1$.[4] Using an appropriate reparametrization of the modified profile likelihood, we find that if $\rho = 1$ and the data are i.i.d. and normal, then the limiting distributions of the MMLEs are generally asymmetric unlike those of the RE- and FEMLE and other MLEs for parameters that are only second-order identified.

Finally, we discuss inference methods related to the modified profile likelihood. Wald tests, some versions of (Quasi) LM tests, and (Quasi) LR tests that are used for testing hypotheses involving ρ and are based on the reparametrized modified profile likelihood do

[4] Madsen (2009) considers the limiting distribution of another GMM estimator for a panel AR(1) model when $\rho = 1$ but, as she points out, her analysis is incomplete. Dovonon and Hall (2018) present a generic version of the limiting distribution theory for GMM estimators when first-order identification fails but second-order identification holds. Unfortunately, their theory is incomplete for the exactly identified case and therefore cannot be used to derive my results, see section 3.2 below.

not uniformly converge to their fixed parameter first-order limiting distributions when ρ is close or equal to one, cf. Rotnitzky et al. (2000) and Bottai (2003). As a consequence these tests do not asymptotically have correct size in a uniform sense when $|\rho| \leq 1$. Similarly to Kruiniger (2016) in the case of (Quasi) LM tests related to the RE- and the FE(Q)MLE, we show that (Q)LM test-statistics that are based on the modified profile log-likelihood and use its *expected* rather than *observed* Hessian, with an additional modification for $\rho = 1$, and confidence regions that are based on inverting these tests have correct asymptotic size in a uniform sense when $|\rho| \leq 1$.

Monte Carlo results show that the QLM tests have correct size and that when the data are i.i.d. and normal and $|\rho| < 1$, the MMLEs for ρ can have a significantly smaller RMSE than the asymptotically efficient REMLE in panels as large as $T = 9$ and $N = 500$. When the data are not i.i.d. and normal, it is generally not possible to rank the Quasi MMLEs, the RE- and the FEQMLE in terms of asymptotic efficiency.

Both types of generalized MMLEs are also useful for estimating other models with parameters that may correspond to stationary points of inflection of the profile likelihood function. Examples of such models are the sample selection model and the stochastic production frontier model for a cross-section of units that are discussed in Lee and Chesher (1986) and models with skew-normal distributions, see e.g. Hallin and Ley (2014).

Dhaene and Jochmans (2016) discuss several alternative approaches to constructing modified (profile) objective functions for the nonstationary panel AR(1) model that yield estimators similar to Lancaster's MMLE. Hahn and Kuersteiner (2002) modified the LSDV estimator to remove bias up to order $O(T^{-1})$. Other FE estimators for dynamic panel models include the first-difference (FD) instrumental variable estimator of Anderson and Hsiao (1981), the FE GMM estimators of Kruiniger (2001), the Maximum Invariant Likelihood estimator of Moreira (2009), the FDMLE of Kruiniger (2008) and the Panel Fully Aggregated Estimator of Han et al. (2015), which is based on X-differencing. The latter two estimators rely on covariance stationarity of the data when $|\rho| < 1$.

The paper is organised as follows. Section 2 presents the panel AR(1) model and the assumptions. Section 3 discusses existence, uniqueness and consistency of the generalized MMLEs as well as their asymptotic distributions. Section 4 discusses inference methods that have correct asymptotic size in a uniform sense. Section 5 studies the finite sample properties of the MMLEs and a (Q)LM test. Finally, section 6 offers some concluding remarks. Derivations and proofs can be found in the appendix.

2 The panel AR(1) model

We consider ML-type estimators for the panel AR(1) model with K strictly exogenous covariates $x_{i,t,k}$, $k = 1, ..., K$:

$$y_{i,t} = \rho y_{i,t-1} + x'_{i,t}\beta + \alpha_i + \varepsilon_{i,t} \text{ with } \beta = (1-\rho)\breve{\beta} \text{ and } \alpha_i = (1-\rho)\mu_i, \qquad (1)$$

for $i = 1, ..., N$ and $t = 1, ..., T$, where $x'_{i,t}$ is the $t - th$ row of the $T \times K$ matrix X_i, α_i is a fixed effect and $\varepsilon_{i,t}$ is an error term. We can also allow for time effects in the model.

Let $y_i = (y_{i,1} \ ... \ y_{i,T})'$, $y_{i,-1} = (y_{i,0} \ ... \ y_{i,T-1})'$, $\varepsilon_i = (\varepsilon_{i,1} \ ... \ \varepsilon_{i,T})'$ and $\overline{x}'_i = T^{-1}\iota' X_i$, with ι equal to a $T-$vector of ones. If we let $v_i = (\rho - 1)y_{i,0} + \alpha_i + \overline{x}'_i\beta$ for $i = 1, ..., N$, then the model in (1) can also be written as $y_i - y_{i,0}\iota = \rho(y_{i,-1} - y_{i,0}\iota) + QX_i\beta + v_i\iota + \varepsilon_i$ for $i = 1, ..., N$, where $Q = I_T - T^{-1}\iota\iota'$ and I_T is an identity matrix with dimension T, cf. Lancaster (2002). We make the following assumption:

Assumption 1 *The variable $y_{i,t}$ is generated by (1) with (i) $T \geq 2$; (ii) $-1 \leq \rho \leq 1$; (iii) $\{(\varepsilon'_i, v_i, (vech(QX_i))')'\}_{i=1}^N$ is a sequence of i.i.d. random vectors with $E(v_i) = 0$, $Var(v_i) = \sigma_v^2 < \infty$ and $E(X'_iQX_i)$ is a finite and positive definite matrix; and (iv) $\varepsilon_i \perp (v_i, (vech(QX_i))')'$, $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2 I_T < \infty$, $i = 1, ..., N$.*

Thus we assume cross-sectional independence, strict exogeneity of the regressors in first-differences, homoskedasticity and no multicollinearity. On the other hand, we allow for ARCH and non-normality of the error terms, the $\varepsilon_{i,t}$.

We require that $T \geq 2$ and $\rho \geq -1$ for identification. In economics the assumption $\rho \geq -1$ can reasonably be expected to hold when the covariates are strictly exogenous. The restrictive parametrization $\alpha_i = (1-\rho)\mu_i$ and $\beta = (1-\rho)\breve{\beta}$ prevents the fixed effects and the means of the individual regressors from turning into trends at $\rho = 1$ and thereby avoids a discontinuity in the data generating process at $\rho = 1$. These restrictions and the restriction $\rho \leq 1$ are only imposed on the DGP but not in estimation.

We are interested in consistent estimation of the common parameters ρ, σ^2 and β under large N, fixed T asymptotics. We will treat the individual effects as nuisance parameters. We will work with a Gaussian homoskedastic (quasi-)likelihood but we note that consistency of the MMLEs (for ρ and β) does not depend on normality or cross-sectional homoskedasticity of the errors.

3 Modified ML estimation of the panel AR(1) model

Conditional on $y_{i,0}$ and X_i, $i = 1, ..., N$ and normalized by N, the Gaussian FE log-likelihood function for the model in (1) is, up to an additive constant, given by:

$$-\frac{T}{2}\log s^2 - \frac{1}{2s^2}\frac{1}{N}\sum_{i=1}^{N}(y_i - ry_{i,-1} - X_i b - a_i\iota)'(y_i - ry_{i,-1} - X_i b - a_i\iota). \tag{2}$$

To obtain a consistent FE estimator for θ_0 based on (2), Lancaster (2002) proposed a Bayesian approach that involves using a reparametrization of the fixed effects, which aims to achieve information orthogonality (but fails to do so when covariates are present), and integrating the new effects from the likelihood function using a uniform prior density. He defines his estimator for θ_0 as a local maximum of the joint posterior density. Letting $\theta = (r\ s^2\ b)'$, his joint posterior log-density for the model in (1), normalized by N, which can be interpreted as a (normalized) modified profile log-likelihood function, is given by:

$$\begin{aligned} \widetilde{l}_N(\theta) &= \widetilde{l}_N(r, s^2, b) = (T-1)\xi(r) - \frac{T-1}{2}\log s^2 \qquad (3)\\ &\quad -\frac{1}{2s^2}\frac{1}{N}\sum_{i=1}^{N}(y_i - ry_{i,-1} - X_i b)'Q(y_i - ry_{i,-1} - X_i b), \\ \text{where } \xi(r) &= \frac{1}{T(T-1)}\sum_{t=1}^{T-1}\frac{(T-t)}{t}r^t, \qquad (4) \end{aligned}$$

and the corresponding modified profile likelihood equations are given by:

$$\begin{aligned} \Psi_\rho(\theta) &= (T-1)\xi'(r) + \frac{1}{s^2}\frac{1}{N}\sum_{i=1}^{N}(y_i - ry_{i,-1} - X_i b)'Qy_{i,-1} = 0, \qquad (5)\\ \Psi_{\sigma^2}(\theta) &= -\frac{T-1}{2s^2} + \frac{1}{2s^4}\frac{1}{N}\sum_{i=1}^{N}(y_i - ry_{i,-1} - X_i b)'Q(y_i - ry_{i,-1} - X_i b) = 0, \\ \Psi_\beta(\theta) &= \frac{1}{s^2}\frac{1}{N}\sum_{i=1}^{N}X_i'Q(y_i - ry_{i,-1} - X_i b) = 0. \end{aligned}$$

Note that the joint posterior density is not proper.

Let $\widehat{\theta}_{LAN}$ denote Lancaster's estimator for θ_0 and let Θ_N be the set of roots of $\frac{\partial \widetilde{l}_N}{\partial \theta} = 0$ corresponding to local maxima of $\widetilde{l}_N$ on Ω which is an open subset of $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^K$. Thus $\widehat{\theta}_{LAN} \in \Theta_N$ unless Θ_N is empty, in which case (we will say that) $\widehat{\theta}_{LAN}$ does not exist. In that case Lancaster effectively puts $\widehat{\theta}_{LAN} = \mathbf{0}$, see his consistency proof. This 'trick' ensures that $\widehat{\theta}_{LAN}$ always exists so that one can consider whether $\widehat{\theta}_{LAN}$ is a consistent estimator for θ_0. Note that none of the roots of $\frac{\partial \widetilde{l}_N}{\partial \theta} = 0$ correspond to the global maxima that can occur at $r = \infty$ and, if T is odd, at $r = -\infty$.

Lancaster showed that $\widetilde{l}_N(\theta)$ converges uniformly in probability to a nonstochastic differentiable function of θ, say $\widetilde{l}(\theta)$, and that $\frac{\partial \widetilde{l}(\theta)}{\partial \theta}|_{\theta_0} = 0$. Next we derive necessary and sufficient conditions for negative definiteness of the Hessian of $\widetilde{l}(\theta)$ at θ_0, viz.:

$$MH = \begin{pmatrix} (T-1)\xi''(\rho) - tr(\Phi' Q \Phi) - \frac{\Sigma_{zqz}}{\sigma^2} & \frac{(T-1)\xi'(\rho)}{\sigma^2} & -\frac{\Sigma'_{xqz}}{\sigma^2} \\ \frac{(T-1)\xi'(\rho)}{\sigma^2} & -\frac{T-1}{2\sigma^4} & 0 \\ -\frac{\Sigma_{xqz}}{\sigma^2} & 0 & -\frac{\Sigma_{xqx}}{\sigma^2} \end{pmatrix}, \tag{6}$$

where $\Sigma_{zqz} = \text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} \widetilde{Z}_i Q \widetilde{Z}_i$, $\Sigma_{xqx} = \text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} X_i' Q X_i$ and $\Sigma_{xqz} = \text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} X_i' Q \widetilde{Z}_i$ with $\widetilde{Z}_i = \varphi v_i + \Phi Q X_i \beta$,

$$\Phi = \Phi(\rho) = \begin{pmatrix} 0 & . & . & 0 & 0 & 0 \\ 1 & 0 & & & 0 & 0 \\ \rho & 1 & 0 & & & 0 \\ . & \rho & 1 & 0 & & . \\ . & & \rho & 1 & 0 & . \\ \rho^{T-2} & . & . & \rho & 1 & 0 \end{pmatrix} \text{ and } \varphi = \varphi(\rho) = \begin{pmatrix} 1 \\ \rho \\ \rho^2 \\ \vdots \\ \rho^{T-2} \\ \rho^{T-1} \end{pmatrix}. \tag{7}$$

It follows from lemma 4.1 in Dhaene and Jochmans (2016) that if $T = 2$ and $\Sigma_{zqz} > 0$ (so that $\rho \neq 1$) or if $T > 2$ and $\rho \neq 1$, then MH is negative definite so that $\widetilde{l}(\theta)$ has a local maximum at θ_0.[5] Kruiniger (2001) had already shown that if $\rho = 1$ and $T \geq 2$, then MH is singular. Moreover, Ahn and Thomas (2004) have shown that $\widetilde{l}(\theta)$ actually has a stationary point of inflection when $\rho = 1$ rather than a local maximum. This property is related to the fact that the posterior density is not proper. Later on, in the context of Theorem 1 below, we will show that if $\rho = 1$, $\widetilde{l}_N$ may not have any local maximum on $\widetilde{\Omega} = [-1, \infty) \times (0, \infty) \times \mathbb{R}^K$ asymptotically, so that $\widehat{\theta}_{LAN}$ is inconsistent.[6] $\widehat{\theta}_{LAN}$ has two more drawbacks. Firstly, $\widetilde{l}_N(\theta)$ may not have any local maximum in small samples, in which case $\widehat{\theta}_{LAN}$ does not exist. This may happen when ρ is close or equal to unity. Secondly, Lancaster did not rule out that $\widetilde{l}_N(\theta)$ and $\widetilde{l}(\theta)$ have multiple local maxima on Ω and he did not explain how to find the consistent estimator if that were the case.

[5]Their lemma 4.1 implies that $\xi''(\rho) - (T-1)^{-1} tr(\Phi' Q \Phi) + 2(\xi'(\rho))^2 \leq 0$ with equality if and only if $T = 2$ or $\rho = 1$.

[6]Lancaster's model is $y_i = \rho y_{i,-1} + X_i \beta + \alpha_i \iota + \varepsilon_i$ without the restrictions $\beta = (1-\rho)\check{\beta}$ and $\alpha_i = (1-\rho)\mu_i$. Therefore, if $\rho = 1$ and $\beta \neq 0$, then the probability limit of the Hessian of *his* modified log-likelihood function at θ_0 is still negative definite and his estimator is consistent. However, if $\rho = 1$, $\beta = 0$ and $\alpha_i = 0$ for $i = 1, ..., N$, then his estimator is inconsistent.

3.1 Generalized Modified ML estimators

We will now introduce two generalizations of $\widehat{\theta}_{LAN}$. We have assumed that $|\rho| \leq 1$. Under this assumption we will be able to show below that $\widetilde{l}_N(\theta)$ can have one local maximum on $\widetilde{\Omega}$ at most. To ensure that the MMLE for θ_0 is also defined in most cases where $\Theta_N \cap \widetilde{\Omega} = \varnothing$, we will generalize its definition as follows:

$$\widehat{\theta}_W = \arg\min_{\theta\in\widetilde{\Omega}} \left(\frac{\partial \widetilde{l}_N(\theta)}{\partial\theta}\right)' W_N \left(\frac{\partial \widetilde{l}_N(\theta)}{\partial\theta}\right) \text{ s.t. } x'\left(\frac{\partial^2 \widetilde{l}_N(\theta)}{\partial\theta\partial\theta'}\right) x \leq 0 \ \forall x \in \mathbb{R}^{2+K}, \tag{8}$$

where W_N is a positive definite (PD) symmetric weight matrix and $\text{plim}_{N\to\infty} W_N = W$ where W is PD. Thus our MMLE is defined as the minimizer of a quadratic form in the modified profile score vector, $\frac{\partial \widetilde{l}_N}{\partial\theta}$, subject to the Hessian of $\widetilde{l}_N$ being negative semi-definite. If $\widetilde{l}_N(\theta)$ has a local maximum, then our MMLE for θ_0 does not depend on W_N and is equal to $\widehat{\theta}_{LAN}$. Theorem 1 below asserts that $\widehat{\theta}_W$ exists w.p.a.1, is uniquely defined (given W_N) w.p.1 and is consistent for any $\theta_0 \in \widetilde{\Omega}$.

Note that among the likelihood equations in (5) only the one for r is modified. Hence, when solving $\Psi_\beta(\theta) = 0$ for b we obtain the unique solution $\widehat{\beta}(r) = (\sum_{i=1}^N X_i'QX_i)^{-1} \times \sum_{i=1}^N X_i'Q(y_i - ry_{i,-1})$ and when solving $\Psi_{\sigma^2}(\theta) = 0$ for s^2 we obtain the unique solution $\widehat{\sigma}^2(r,b) = (T-1)^{-1}N^{-1}\sum_{i=1}^N (y_i - ry_{i,-1} - X_ib)'Q(y_i - ry_{i,-1} - X_ib)$. Let $\widehat{\theta}(r) = (r, \widehat{\sigma}^2(r, \widehat{\beta}(r)), \widehat{\beta}(r))'$, then the (normalized) modified profile log-likelihood function of r, $\widetilde{l}_N^c(r)$, is defined by the equality $\widetilde{l}_N^c(r) = \widetilde{l}_N(\widehat{\theta}(r))$, i.e. $\widetilde{l}_N^c(r) = \widetilde{l}_N(r, \widehat{\sigma}^2(r, \widehat{\beta}(r)), \widehat{\beta}(r))$.

An alternative MMLE for θ_0, which is based on $\widetilde{l}_N^c(r)$, is given by $\widehat{\theta}_C$ with [7]

$$\widehat{\rho}_C = \arg\min_{r\in[-1,\infty)} \left(\frac{\partial \widetilde{l}_N^c(r)}{\partial r}\right)^2 \text{ s.t. } \frac{\partial^2 \widetilde{l}_N^c(r)}{\partial r^2} \leq 0, \tag{9}$$
$$\widehat{\sigma}_C^2 = \widehat{\sigma}^2(\widehat{\rho}_C, \widehat{\beta}(\widehat{\rho}_C)) \quad \text{and} \quad \widehat{\beta}_C = \widehat{\beta}(\widehat{\rho}_C).$$

The Adjusted Likelihood estimator of Dhaene and Jochmans (2016), viz. $\widehat{\theta}_{ADJ}$, is a constrained version of $\widehat{\theta}_C$.[8] However, using their constraint is not required for uniqueness of this MMLE and would also not guarantee its uniqueness in finite samples if the modified profile likelihood would have multiple local maxima. Theorem 1 below asserts that $\widehat{\theta}_C$ exists w.p.a.1, is uniquely defined w.p.1 and is consistent for any $\theta_0 \in \widetilde{\Omega}$.

[7] One can also define a class of MMLEs where only s^2 is profiled out but not b.

[8] $\widehat{\rho}_{ADJ} = \arg\min_{r\in\mathcal{E}} \left(\frac{\partial \widetilde{l}_N^c(r)}{\partial r}\right)^2$ s.t $\frac{\partial^2 \widetilde{l}_N^c(r)}{\partial r^2} \leq 0$, where $\mathcal{E}$ is a certain interval centered at the LSDV estimator $\widehat{\rho}_{ML}$. $\widehat{\sigma}_{ADJ}^2 = \widehat{\sigma}^2(\widehat{\rho}_{ADJ}, \widehat{\beta}(\widehat{\rho}_{ADJ}))$ and $\widehat{\beta}_{ADJ} = \widehat{\beta}(\widehat{\rho}_{ADJ})$.

There is no W_N such that the $\widehat{\theta}_W$ estimator equals the $\widehat{\theta}_C$ estimator: if $\frac{\partial \widetilde{l}^c_N(r)}{\partial r}|_{\widehat{\rho}_C} = 0$, then $\frac{\partial \widetilde{l}_N(\theta)}{\partial \theta}|_{\widehat{\theta}_W} = 0$ and both estimates of θ are equal but if $\frac{\partial \widetilde{l}^c_N(r)}{\partial r}|_{\widehat{\rho}_C} \neq 0$, then $\frac{\partial \widetilde{l}_N(\theta)}{\partial \theta}|_{\widehat{\theta}_W} \neq 0$ and the two estimates of θ are unequal although the value of $\widehat{\theta}_W$ will be close to that of $\widehat{\theta}_C$ for W_N that give relatively little weight to $\frac{\partial \widetilde{l}_N(\theta)}{\partial r}$.

We can also consider a variation on $\widehat{\theta}_W$ that is given by (8) with the first element of $\frac{\partial \widetilde{l}_N(\theta)}{\partial \theta}$ replaced by $\frac{\partial \widetilde{l}^c_N(r)}{\partial r}$. We call this MMLE $\widehat{\theta}_F$.

In the appendix we show that $\widetilde{l}^c_N(r)$ converges uniformly in probability to a nonstochastic differentiable function of r, say $\widetilde{l}^c(r)$, that $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_\rho = 0$ and that $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho \leq 0$, with equality holding if $\rho = 1$ or if $T = 2$ and $\sigma_v^2 = \beta = 0$ (i.e., $\Sigma_{zqz} = 0$). Thus, similar to $\widetilde{l}(\theta)$, $\widetilde{l}^c(r)$ has a local maximum at ρ when $\rho \neq 1$ and, in case $T = 2$, $\Sigma_{zqz} > 0$. In the appendix we also show that $\widetilde{l}^c(r)$ has a stationary point of inflection at ρ when $\rho = 1$. To simplify the exposition we assume in the remainder of this paper that if $T = 2$ and $\rho \neq 1$, then either $\sigma_v^2 > 0$ or $\beta \neq 0$ so that $\Sigma_{zqz} > 0$.

Note that $\widehat{\theta}_C$ would only fail to exist in the extremely unlikely case that $\frac{\partial^2 \widetilde{l}^c_N(r)}{\partial r^2} > 0$ on the entire interval $[-1, \infty)$. Similarly, $\widehat{\theta}_W$ and $\widehat{\theta}_F$ would only fail to exist in the extremely unlikely case that for no $\theta \in \widetilde{\Omega}$, $x' \left(\frac{\partial^2 \widetilde{l}_N(\theta)}{\partial \theta \partial \theta'} \right) x \leq 0 \; \forall x \in \mathbb{R}^{2+K}$. [9] The second-order conditions $\frac{\partial^2 \widetilde{l}^c_N(r)}{\partial r^2} \leq 0$ and $x' \left(\frac{\partial^2 \widetilde{l}_N(\theta)}{\partial \theta \partial \theta'} \right) x \leq 0 \; \forall x \in \mathbb{R}^{2+K}$ are a crucial part of the definitions of $\widehat{\theta}_C$, $\widehat{\theta}_W$ and $\widehat{\theta}_F$ because $\widetilde{l}^c_N(r)$ and $\widetilde{l}_N(r)$ may attain a minimum on $[-1, \infty)$ and $\widetilde{\Omega}$, respectively, see lemma 1 in the appendix.

The next theorem asserts uniqueness and consistency of $\widehat{\theta}_W$, $\widehat{\theta}_F$ and $\widehat{\theta}_C$:

Theorem 1 *Let Assumption 1 hold. Then the Modified MLEs $\widehat{\theta}_W$, $\widehat{\theta}_F$ and $\widehat{\theta}_C$ for θ_0 are uniquely defined w.p.1 when they exist, exist w.p.a.1 and are consistent.*

If $-1 \leq \rho < 1$, $\lim_{N\to\infty} \Pr(\Theta_N \cap \widetilde{\Omega} = \varnothing) = 0$, i.e., $\widehat{\theta}_{LAN}$ exists w.p.a.1. In this case $\widehat{\theta}_{LAN}$ is also unique w.p.1. (if it exists) and consistent. However, if $\rho = 1$, $\lim_{N\to\infty} \Pr(\Theta_N \cap \widetilde{\Omega} = \varnothing) > 0$ by lemma 4 in the appendix (and $\theta_0 \neq \mathbf{0}$), i.e., $\widehat{\theta}_{LAN}$ may not exist even asymptotically, which implies that $\widehat{\theta}_{LAN}$ is inconsistent.

When $-1 \leq \rho < 1$, the first-order, fixed parameter asymptotic distributions of $\widehat{\theta}_W$, $\widehat{\theta}_F$, $\widehat{\theta}_C$ and $\widehat{\theta}_{LAN}$ are the same and given by (cf. Kruiniger, 2001):

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, (MH)^{-1} MIM (MH)^{-1}\right), \tag{10}$$

[9] One could ensure that $\widehat{\theta}_W$, $\widehat{\theta}_F$ and $\widehat{\theta}_C$ are always defined by replacing them by $\widehat{\theta}(\widehat{\rho}_{ML} + \frac{3}{T+1})$ in these improbable cases, where $-\frac{3}{T+1}$ is the asymptotic bias of $\widehat{\rho}_{ML}$ when $\rho = 1$. The rationale for this proposed solution is that the non-existence problem most likely only occurs (if ever) when the sample size is very small and ρ is close or equal to unity.

where MH is given in (6) and under normality of the ε_i MIM (Modified Information Matrix) equals:[10]

$$MIM = \begin{pmatrix} tr(Q\Phi Q\Phi) + \frac{\sigma^2 tr(\Phi' Q\Phi) + \Sigma_{zqz}}{\sigma^2} & -\frac{(T-1)\xi'(\rho)}{\sigma^2} & \frac{\Sigma'_{xqz}}{\sigma^2} \\ -\frac{(T-1)\xi'(\rho)}{\sigma^2} & \frac{T-1}{2\sigma^4} & 0 \\ \frac{\Sigma_{xqz}}{\sigma^2} & 0 & \frac{\Sigma_{xqx}}{\sigma^2} \end{pmatrix}. \quad (11)$$

It can easily be checked that $tr(Q\Phi Q\Phi) \neq -(T-1)\xi''(\rho)$ and hence $MH \neq -MIM$.

If $T = 2$, $\widehat{\rho}_{LAN}$ is equal to the FEMLE for ρ that has been proposed by Hsiao et al. (2002), henceforth $\widehat{\rho}_{FEML}$, but if $T > 2$, the data are i.i.d. and normal and $|\rho| < 1$, $\widehat{\rho}_{LAN}$ is asymptotically less efficient than $\widehat{\rho}_{FEML}$, see Ahn and Thomas (2004); when the data are not i.i.d and normal, $\widehat{\rho}_{LAN}$ may be asymptotically more efficient than $\widehat{\rho}_{FEML}$.

If $\rho = 1$, $\det(MIM) \neq 0$ but $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho = 0$ and $\det(MH) = 0$. Thus ρ and θ are first-order underidentified when $\rho = 1$. Although we cannot directly apply the results of Rotnitzky et al. (2000), who developed an asymptotic theory for MLEs when the information matrix is singular, to $\widehat{\theta}_W$, $\widehat{\theta}_F$ and $\widehat{\theta}_C$ when $\rho = 1$, because they are *Modified* MLEs and $\det(MIM) \neq 0$, arguments similar to theirs suggest that these MMLEs have a slower than $\sqrt{N}$ rate of convergence and that their limiting distributions are non-standard. When deriving their limiting distributions for $\rho = 1$ below, we will view the MMLEs as GMM estimators. If ρ is close to 1, $\det(MH)$ and $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho$ are close to zero and the MMLEs will have a "weak moment conditions" problem, cf. Kruiniger (2013).

3.2 The limiting distributions of $\widehat{\theta}_C$ and $\widehat{\theta}_F$ when $\rho = 1$

W.p.a.1 $\widehat{\rho}_C$ is a solution of the first-order condition (f.o.c.) $G^c_N(r) \equiv \frac{\partial^2 \widetilde{l}^c_N(r)}{\partial r^2} \frac{\partial \widetilde{l}^c_N(r)}{\partial r} = 0$. Using a Taylor expansion of $G^c_N(\widehat{\rho}_C)$ around $r = 1$, we show in the appendix that when $\rho = 1$, $N^{1/4}(\widehat{\rho}_C - 1) = O_p(1)$, i.e., the rate of convergence of $\widehat{\rho}_C$ is at least $N^{1/4}$. This quartic root rate of convergence reflects the fact that $\frac{\partial^2 \widetilde{l}^c(1)}{\partial r^2} = 0$ and $\frac{\partial^3 \widetilde{l}^c(1)}{\partial r^3} = \frac{T(T-1)(T+1)}{12} \neq 0$, which means that ρ is second-order identified when $\rho = 1$, and is in line with results in Sargan (1983), Rotnitzky et al. (2000), Ahn and Thomas (2004), Madsen (2009), Dovonon and Renault (2013) and Kruiniger (2013) who also study estimation when a parameter is only second-order identified. Note that this rate is faster than the $N^{1/6}$-rate of the MLEs of the parameters that correspond to the inflection point of the likelihood functions of the sample selection model and the stochastic production frontier model for a cross-section that are discussed in Lee and Chesher (1986) and the models with skew-normal distributions that are discussed in Hallin and Ley (2014).

[10]To derive (11) we have used that if $\varepsilon_i|(v_i, QX_i) \sim N(0, \sigma^2 I_T)$, then for any constant $T \times T$ matrices M_1 and M_2, $E(\varepsilon_i' M_1 \varepsilon_i \varepsilon_i' M_2 \varepsilon_i) = \sigma^4(tr(M_1)tr(M_2) + tr(M_1 M_2 + M_1' M_2))$.

Next we discuss the derivation of the limiting distribution of $\widehat{\theta}_C$ when $\rho = 1$. Let $M_N^c(r) = N\left(\frac{\partial \widetilde{l}_N^c(r)}{\partial r}\right)^2$. Analogously to Sargan (1983) and Rotnitzky et al. (2000) consider the following Taylor expansion of $M_N^c(r)$ around $r = 1$:

$$M_N^c(r) = M_N^c(1) + \sum_{j=1}^{4} \frac{1}{j!}\frac{\partial^j M_N^c(1)}{\partial r^j}(r-1)^j + P_{3,N}(N^{1/4}(r-1)), \tag{12}$$

where $P_{3,N}(N^{1/4}(r-1))$ is a polynomial in $N^{1/4}(r-1)$ with coefficients that are $o_p(1)$. Let $\widehat{\rho} = \widehat{\rho}_C$. Substituting $\widehat{\rho}$ for r in (12) we obtain

$$\begin{aligned} M_N^c(\widehat{\rho}) \quad = \quad & N\left(\frac{\partial \widetilde{l}_N^c(1)}{\partial r}\right)^2 + \frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3} N^{1/2} \frac{\partial \widetilde{l}_N^c(1)}{\partial r} N^{1/2}(\widehat{\rho}-1)^2 + \\ & \frac{1}{4}\left(\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3}\right)^2 N(\widehat{\rho}-1)^4 + R_{1,N}^c(N^{1/4}(\widehat{\rho}-1)), \end{aligned} \tag{13}$$

where $R_{1,N}^c(N^{1/4}(\widehat{\rho}-1)) = o_p(1)$.

Let $Z_{1,N} = \left(-\frac{1}{2}\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3}\right)^{-1} N^{1/2}\left(\frac{\partial \widetilde{l}_N^c(1)}{\partial r}\right)$. In the proof of Theorem 2 we show that $Z_{1,N} = O_p(1)$ and that there exists a sequence $\{U_N\}$ with $U_N = O_p(N^{-1/2})$ such that if $Z_{1,N} + U_N > 0$, then $M_N^c(r)$ has two local minima attained at values $\widetilde{\rho}$ such that $N^{1/2}(\widetilde{\rho}-1)^2 = Z_{1,N} + o_p(1)$, whereas if $Z_{1,N}+U_N < 0$, then $M_N^c(r)$ has one local minimum attained at $r = \widehat{\rho}$ with $N^{1/2}(\widehat{\rho}-1)^2 = o_p(1)$. Furthermore, when $Z_{1,N} + U_N > 0$, the sign of $N^{1/4}(\widehat{\rho}-1)$ is determined by the remainder $R_{1,N}^c(N^{1/4}(\widehat{\rho}-1))$.

To obtain the limiting distribution of $\widehat{\theta}_C$ when $\rho = 1$ we use the following new parametrization (indicated by the subscript n), cf. Kruiniger (2013): $\theta_n = (r_n, s_n^2, b_n')'$ where $r_n = r$, $s_n^2 = s^2/r$ and $b_n = b$. Noting that we can express the elements of θ as functions of the elements of θ_n, viz. $\theta = \theta(\theta_n) = (r_n, s_n^2 r_n, b_n')'$, the reparameterized modified log-likelihood function is given by $\widetilde{l}_{N,n}(\theta_n) = \widetilde{l}_N(\theta(\theta_n))$. Similarly to Lancaster (2002), it can be shown that $\widetilde{l}_{N,n}(\theta_n)$ converges uniformly in probability to a nonstochastic continuous function of θ_n, i.e. $\widetilde{l}_n(\theta_n) = \widetilde{l}(\theta(\theta_n))$. The reparametrization is such that the elements of the first row and the first column of the Hessian of $\widetilde{l}_n(\theta_n)$ at $\theta_{0,n} = (\rho_n, \sigma_n^2, \beta_n')' = \theta_* \equiv (1, \sigma^2, 0')'$ are equal to zero. Note that if $\rho = 1$, then $\theta_0 = \theta_{0,n} = \theta_*$ for some σ^2.

We also need to introduce some additional notation. Let $\widehat{\theta} = \widehat{\theta}_C$ and $\widehat{\theta}_n = \widehat{\theta}_{n,C} = (\widehat{\rho}_C,\ \widehat{\sigma}_{n,C}^2,\ \widehat{\beta}_C')'$ with $\widehat{\sigma}_{n,C}^2 = \widehat{\sigma}_C^2/\widehat{\rho}_C$. Furthermore, let $Z_{2,N} = N^{1/2}(\widehat{\sigma}^2(1,\widehat{\beta}(1)) - \sigma^2)$, $Z_{3,N} = N^{1/2}(\widehat{\beta}-\beta)$ and $Z_N = (Z_{1,N}, Z_{2,N}, Z_{3,N}')'$. Then we have the following results:

Theorem 2 *Let Assumption 1 hold, $\varepsilon_i \sim N(0, \sigma^2 I)$, $i = 1, ..., N$, and $\rho = 1$. Then*

(i) $Z_N \xrightarrow{d} Z = (Z_1, Z_2, Z_3')' \sim N(0, \Sigma_Z)$, *where* $E(Z_1 Z_2) = 0$, $E(Z_1 Z_3) = 0$, $E(Z_2 Z_3) = 0$, $Var(Z_1) = 48T^{-2}((T-1)(T+1))^{-1}$, $Var(Z_2) = 2\sigma^4(T-1)^{-1}$ *and* $Var(Z_3) = \sigma^2(\Sigma_{xqx})^{-1}$;

(ii) letting $K_+ = \sigma^2(T+1)/6$ *and* $B^c = \mathbf{1}(R^c > 0)$ *with the r.v.* R^c *defined in (31),*

$$\begin{bmatrix} N^{1/4}(\widehat{\rho}_C - 1) \\ N^{1/2}(\widehat{\sigma}^2_{n,C} - \sigma^2) \\ N^{1/2}(\widehat{\beta}_C - \beta) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (-1)^{B^c} Z_1^{1/2} \\ Z_2 + K_+ Z_1 \\ Z_3 \end{bmatrix} \mathbf{1}\{Z_1 > 0\} + \begin{bmatrix} 0 \\ Z_2 \\ Z_3 \end{bmatrix} \mathbf{1}\{Z_1 \leq 0\}.$$

C*omments:* In the proof of Theorem 2 we show that the sign of $N^{1/4}(\widehat{\rho}_C - 1)$ depends on $\frac{\partial^5 \widetilde{l}^c_N(1)}{\partial r^5}$, whereas it follows from Kruiniger (2013) and corollary 1 in Rotnitzky et al. (2000) that the sign of $N^{1/4}(\widehat{\rho}_{FEML} - 1)$ only depends on the second and third derivatives of the FE log-likelihood. The latter is generally true for MLEs of parameters that are only second-order identified, cf. Rotnitzky et al. (2000);

Relaxing the assumption of normality of the ε_i affects Σ_Z and the conditional distribution of B^c given Z but otherwise does not change Theorem 2;

The limiting distribution of $\widehat{\rho}_C$ is asymmetric unlike that of $\widehat{\rho}_{FEML}$ and other MLEs of parameters that are only second-order identified, cf. Rotnitzky et al. (2000);

From $\widehat{\theta}_C = \theta(\widehat{\theta}_{n,C})$ we have $\widehat{\sigma}^2_C = \widehat{\sigma}^2_{n,C}\widehat{\rho}_C$. Hence the rate of convergence of $\widehat{\sigma}^2_C$ is also $N^{1/4}$ and $N^{1/4}(\widehat{\sigma}^2_C - \sigma^2) = N^{1/4}(\widehat{\rho}_C - 1)\sigma^2 + o_p(1)$;

Finally, the following result implies the sign of the asymptotic bias of $\widehat{\rho}_C$ and $\widehat{\sigma}^2_C$:

Corollary 1 *Let Assumption 1 hold, $\varepsilon_i \sim N(0, \sigma^2 I)$, $i = 1, ..., N$, and $\rho = 1$. Then if $T \geq 4$, $E((-1)^{B^c} Z_1^{1/2} | Z_1 > 0) > 0$ whereas if $T = 2$ or $T = 3$, $E((-1)^{B^c} Z_1^{1/2} | Z_1 > 0) < 0$.*

We now consider the minimum rate of convergence of $\widehat{\rho} = \widehat{\rho}_F$ and the limiting distribution of $\widehat{\theta}_F$ when $\rho = 1$. Details of the derivations of these properties of $\widehat{\rho}_F$ and $\widehat{\theta}_F$ are given in the appendix. There we show that $N^{1/4}(\widehat{\rho} - 1) = O_p(1)$, cf. Lemma 5.

Let $\Psi_{N,n}(\theta_n) = (\frac{\partial \widetilde{l}^c_N(r)}{\partial r}, s_n^2 r \frac{\partial \widetilde{l}_{N,n}(\theta_n)}{\partial s_n^2}, s_n^2 r \frac{\partial \widetilde{l}_{N,n}(\theta_n)}{\partial b'})'$, $\widehat{\underline{\omega}}_n = ((\widehat{\sigma}^2_{n,F} - \sigma^2), \widehat{\beta}_F')'$ and $\underline{w}_n = (s_n^2, b')'$. Then we have the following results:

Theorem 3 *Let Assumption 1 hold, $\varepsilon_i \sim N(0, \sigma^2 I)$, $i = 1, ..., N$, $\rho = 1$, and let W_N be a PD matrix. Then*

$$\begin{bmatrix} N^{1/4}(\widehat{\rho}_F - 1) \\ N^{1/2}\widehat{\underline{\omega}}_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (-1)^B Z_1^{1/2} \\ \underline{\omega}_+ \end{bmatrix} \mathbf{1}\{Z_1 > 0\} + \begin{bmatrix} 0 \\ \underline{\omega}_+ + K_- Z_1 \end{bmatrix} \mathbf{1}\{Z_1 \leq 0\},$$

where $(Z_1,\ \underline{\omega}_+')' \sim N(0,\Sigma_\omega)$, $B = \mathbf{1}(R>0)$ *and the r.v.* R*, the matrix* Σ_ω *and the constant vector* K_- *are implicitly defined in the proof.*

C*omments:* In the proof of Theorem 3 we see that the sign of $N^{1/4}(\widehat{\rho}_F - 1)$ depends on $\frac{\partial^5 \widetilde{l}_N^c(1)}{\partial r^5}$ in line with the results in Kruiniger (2013) for Quasi MLEs of second-order identified parameters but in contrast to the results for MLEs in Rotnitzky et al. (2000);

Relaxing the assumption of normality of the ε_i affects Σ_ω and the conditional distributions of B and R given $(Z_1,\ \underline{\omega}_+')'$ but otherwise does not fundamentally change the results in Theorem 3;

Like $\widehat{\rho}_C$ and $\widehat{\sigma}_C^2$, when $\rho = 1$, $\widehat{\rho}_F$ and $\widehat{\sigma}_F^2$ converge at a rate of at least $N^{1/4}$ to ρ and σ^2, whereas $\widehat{\beta}_F$ converges at a rate of $N^{1/2}$ to β just like $\widehat{\beta}_C$;

For any W, $(\widehat{\rho}_F - 1)^2$ is first-order asymptotically equivalent to $(\widehat{\rho}_C - 1)^2$ and hence the RMSEs of $\widehat{\rho}_F$ and $\widehat{\rho}_C$ are asymptotically the same. However, the limiting distribution of B and hence that of $N^{1/4}(\widehat{\rho}_F - 1)$ depends on W. The limiting distributions of $\widehat{\sigma}_F^2$ and $\widehat{\beta}_F$ also depend on W and are different from those of $\widehat{\sigma}_C^2$ and $\widehat{\beta}_C$ unless $W_N = diag(W_{N,1,1}, \underline{W}_{N,2,2})$ where $W_{N,1,1}$ is a scalar. In the latter case $\underline{\omega}_+ + K_- Z_1 = (Z_2, Z_3')'$ and $K_- = (-K_+, 0)'$. If in addition $W_{N,1,1} = \infty$ while the elements of $\underline{W}_{N,2,2}$ are finite, then the limiting distributions of $N^{1/4}(\widehat{\rho}_F - 1)$ and $N^{1/4}(\widehat{\rho}_C - 1)$ are also the same;

The results in Theorem 3 can easily be reinterpreted to obtain a version for the generic possibly overidentified case. Treating $\Psi_{N,n}(\theta_n)$ as generic moment functions and ρ and $\underline{\omega}_n$ as generic parameters, with $\underline{\omega}_n$ a vector and ρ a scalar that is only second-order identified, by following the logic of the proofs of Lemma 5 and Theorem 3 we would still obtain Theorem 3 but with $Z_1 = -2(\Psi_{n,\rho\rho}' W^{1/2} M_\omega W^{1/2} \Psi_{n,\rho\rho})^{-1} \times (\Psi_{n,\rho\rho}' W^{1/2} M_\omega W^{1/2} \Psi_n)$, $\underline{\omega}_+ = \mathcal{M}(\Psi_n + \frac{1}{2}\Psi_{n,\rho\rho} Z_1)$ and $K_- = -\frac{1}{2}\mathcal{M}\Psi_{n,\rho\rho}$, where $M_\omega = I - W^{1/2}\Psi_{n,\omega}(\Psi_{n,\omega}' W \Psi_{n,\omega})^{-1}\Psi_{n,\omega}' W^{1/2}$, $\mathcal{M} = -(\Psi_{n,\omega}' W \Psi_{n,\omega})^{-1}\Psi_{n,\omega}' W$ and Ψ_n, $\Psi_{n,\omega}$ and $\Psi_{n,\rho\rho}$ are defined in the proof of Theorem 3. In the exact identified case R would still be defined similarly as in the proof of Theorem 3 and in particular the sign of $N^{1/4}(\widehat{\rho} - 1)$ would still depend on $\text{plim}_{N\to\infty} \frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4}$. In the overidentified case R would be defined as a generic version of R_2 in the proof of Theorem 3 and in particular the sign of $N^{1/4}(\widehat{\rho} - 1)$ would depend on $\text{plim}_{N\to\infty} \frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3}$ but not on $\text{plim}_{N\to\infty} \frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4}$. [11]

[11] Dovonon and Hall (2018) have also derived the limiting distribution of the GMM estimator of ρ and $\underline{\omega}_n$ with ρ a scalar that is only second-order identified, but unfortunately their distributional result for $N^{1/4}(\widehat{\rho} - 1)$ in the exact identified case is incorrect because the order of the expansion of the objective function that they used to study the distribution of B is too low which resulted in an expression for R (their formula (18)) that is actually equal to zero, see the proof of my Theorem 3.

It can be expected that the MMLEs also have non-standard asymptotic properties close to the singularity point, θ_*. Rotnitzky et al. (2000) informally discuss a richness of possibilities for the MLEs close to the singularity point and one can expect several possibilities for the MMLEs too. To save space we don't explore them here. Nonetheless they are a warning of the care needed in conducting inference close to θ_*. Finally, we note that the local-to-unity asymptotic behaviour of various GMM estimators for the panel AR(1) model discussed in Kruiniger (2009) is unrelated to second-order identification.

4 Modified likelihood based inference

Wald tests, some versions of (Quasi) LM tests, and (Quasi) LR tests that are used for testing hypotheses involving ρ and are based on the reparametrized modified likelihood do not asymptotically have correct size in a uniform sense when $|\rho| \leq 1$, cf. Rotnitzky et al. (2000) and especially Bottai (2003), who discusses why these tests do not have correct size in the single parameter case. Generalizing the testing approach proposed in Bottai (2003) that has correct size to a multiple parameter setting, Kruiniger (2016) has shown that (Quasi) LM tests that are related to the RE- and the FE(Q)MLE and standardised by using (a sandwich formula involving) the *expected* rather than the *observed* Hessian do asymptotically have correct size in a uniform sense when $|\rho| \leq 1$. However, the situation is somewhat special in the case of the QLM tests that are used for testing hypotheses involving ρ and are based on the reparametrized modified likelihood. In this case the singularity point, θ_*, corresponds to an inflection point rather than a maximum. As a result in small samples the (normalized) reparametrized modified log-likelihood, $\widetilde{l}_{N,n}(\theta_n)$, may not even have a local maximum when ρ is close to one. Nevertheless, the expected Hessian of $\widetilde{l}_{N,n}(\theta_{0,n})$, viz. $\mathcal{H}(\underline{\theta}_{0,n})$, where $\underline{\theta}_{0,n} = (\theta_{0,n}' \; \overline{\sigma}^2_{v,n})'$ with $\overline{\sigma}^2_{v,n} = \overline{\sigma}^2_v/\sigma^2 - (1-\rho)$ and $\overline{\sigma}^2_v = (1-\rho)^2\sigma^2_v$, is still negative definite close to the singularity point $\underline{\theta}_* = (\theta_*' \; 0)'$.[12] [13] We will now introduce the QLM test-statistic $QLM(\theta_{0,n})$ for testing $H_0 : A\theta_{0,n} = a$, where A is a $J \times \dim(\theta)$ constant matrix of rank J and J is the number of restrictions, which include a restriction on ρ with $-1 \leq \rho < 1$. Let $\mathcal{J}_i(\theta_{0,n}) = \frac{\partial \widetilde{l}_{n,i}(\theta_{0,n})}{\partial \theta_n} \frac{\partial \widetilde{l}_{n,i}(\theta_{0,n})}{\partial \theta_n'}$ and $\mathcal{J}(\theta_{0,n}) = N^{-1}\sum_{i=1}^N \mathcal{J}_i(\theta_{0,n})$, where $\widetilde{l}_{n,i}(\theta_n)$ is the contribution to the reparametrized modified log-likelihood, $N \times \widetilde{l}_{N,n}(\theta_n)$, by individual i. Then $QLM(\theta_{0,n})$ is given by

[12] Note that $\mathcal{H}(\underline{\theta}_{0,n}) = E_{\underline{\theta}_{0,n}}(\partial^2\widetilde{l}_{N,n}(\theta_{0,n})/\partial\theta_n\partial\theta_n')$ depends on $\underline{\theta}_{0,n} = (\theta_{0,n}' \; \overline{\sigma}^2_{v,n})'$, whereas the observed Hessian $\partial^2\widetilde{l}_{N,n}(\theta_{0,n})/\partial\theta_n\partial\theta_n'$ only depends on $\theta_{0,n}$.

[13] This reparametrization is the same as the one used in Kruiniger (2013) for the FE(Q)MLE.

$$QLM(\theta_{0,n}) = N \times \frac{\partial \widetilde{l}'_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n} \mathcal{H}^{-1}(\underline{\widetilde{\theta}}_n) A' \times \tag{14}$$
$$(A\mathcal{H}^{-1}(\underline{\widetilde{\theta}}_n)\mathcal{J}(\widetilde{\theta}_n)\mathcal{H}^{-1}(\underline{\widetilde{\theta}}_n)A')^{-1} A\mathcal{H}^{-1}(\underline{\widetilde{\theta}}_n) \frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n},$$

where $\underline{\widetilde{\theta}}_n$ is a restricted estimate of $\underline{\theta}_{0,n}$. $\overline{\sigma}^2_{v,n}$ can be estimated by the restricted FE(Q)MLE. Under H_0, $QLM(\theta_{0,n}) \sim \chi^2(J)$. When using $QLM(\theta_{0,n})$ to test $H_0 : -1 \leq \rho = a < 1$, $A = (1\ 0\ \mathbf{0}')$ and $\frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n} = A' \frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \rho}$. To test hypotheses that include the restriction $\rho = 1$, one should use a different Quasi LM test, cf. Bottai (2003). In this case one should replace $QLM(\theta_{0,n})$ given in (14) by

$$QLM(\theta_{0,n}) = N \times \widetilde{S}'(\widetilde{\theta}_n)\widetilde{\mathcal{H}}^{-1}(\underline{\widetilde{\theta}}_n)A' \times \tag{15}$$
$$(A\widetilde{\mathcal{H}}^{-1}(\underline{\widetilde{\theta}}_n)\widetilde{\mathcal{J}}(\widetilde{\theta}_n)\widetilde{\mathcal{H}}^{-1}(\underline{\widetilde{\theta}}_n)A')^{-1}A\widetilde{\mathcal{H}}^{-1}(\underline{\widetilde{\theta}}_n)\widetilde{S}(\widetilde{\theta}_n),$$

with

$$\begin{aligned}
\widetilde{S}(\widetilde{\theta}_n) &= N^{-1}\sum\nolimits_{i=1}^{N} S_i, \quad \widetilde{\mathcal{J}}(\widetilde{\theta}_n) = N^{-1}\sum\nolimits_{i=1}^{N}(S_iS_i'), \\
S_i &= (S_{i,1}, S'_{i,2})', \quad S_{i,1} = \frac{1}{2}\frac{\partial^2 \widetilde{l}_{n,i}}{\partial r_n^2}|_{\widetilde{\theta}_n}, \quad S_{i,2} = \frac{\partial \widetilde{l}_{n,i}}{\partial d_n}|_{\widetilde{\theta}_n}, \\
\widetilde{\mathcal{H}}_{1,1} &= \frac{2}{4!}E_{\underline{\widetilde{\theta}}_n}(\frac{\partial^4 \widetilde{l}_{N,n}}{\partial r_n^4}|_{\underline{\widetilde{\theta}}_n}), \quad \widetilde{\mathcal{H}}'_{1,2} = \widetilde{\mathcal{H}}_{2,1} = \frac{1}{2!}E_{\underline{\widetilde{\theta}}_n}(\frac{\partial^3 \widetilde{l}_{N,n}}{\partial r_n^2 \partial d_n}|_{\underline{\widetilde{\theta}}_n}), \\
\widetilde{\mathcal{H}}_{2,2} &= \frac{2}{2!}E_{\underline{\widetilde{\theta}}_n}(\frac{\partial^2 \widetilde{l}_{N,n}}{\partial d_n \partial d_n'}|_{\underline{\widetilde{\theta}}_n}), \quad \widetilde{\mathcal{H}}(\underline{\widetilde{\theta}}_n) = \begin{bmatrix} \widetilde{\mathcal{H}}_{1,1} & \widetilde{\mathcal{H}}_{1,2} \\ \widetilde{\mathcal{H}}_{2,1} & \widetilde{\mathcal{H}}_{2,2} \end{bmatrix},
\end{aligned}$$

where we have partitioned θ_n as $\theta_n = (r_n, d_n')'$ and used $\widetilde{l}_{N,n}$ and $\widetilde{l}_{n,i}$ as short for $\widetilde{l}_{N,n}(\theta_n)$ and $\widetilde{l}_{n,i}(\theta_n)$, respectively. When using $QLM(\theta_{0,n})$ to test $H_0 : \rho = 1$, $A = (1\ 0\ \mathbf{0}')$ and $\widetilde{S}(\widetilde{\theta}_n) = A'(N^{-1}\sum_{i=1}^{N} S_{i,1})$. It can be shown that $QLM(\theta_{0,n})$ given by (14) and (15) is continuous at $\theta_{0,n} = \theta_*$ for any $\sigma^2 > 0$ by using de l'Hôpital's rule twice.

Theorem 4 *The Quasi LM test based on (14) or (15) for testing $H_0 : A\theta_{0,n} = a$, which includes a restriction on ρ with $|\rho| \leq 1$, has correct asymptotic size in a uniform sense.*

Confidence sets (CSs) that are obtained by inverting the tests based on (14) and (15) have correct asymptotic size in a uniform sense. Other tests (and CSs) for ρ that have correct asymptotic size include (CSs based on) the GMM LM test(-statistic)s of Newey and West (1987) that exploit the moments conditions of the System GMM and the nonlinear Ahn-Schmidt (AS) GMM estimator, respectively, see Kruiniger (2009) for the System version and Bun and Kleibergen (2017) for the AS version of the test, and identication-robust test(-statistics)s such as the GMM AR test of Stock and Wright (2000) and the KLM and GMM-CLR tests of Kleibergen (2005) that exploit System and AS moments conditions, cf. Bun and Kleibergen (2017). Kruiniger (2016) has shown that the Quasi LM test for testing an hypothesis about ρ shares the optimal power properties of the KLM test in a worst case scenario. To test $H_0 : \rho = 1$ one could also use a Wald test based on $\sqrt{N}(\widehat{\rho}_C - 1)^2$. Under H_0 $\sqrt{N}(\widehat{\rho}_C - 1)^2 \xrightarrow{d} Z_1\mathbf{1}\{Z_1 > 0\}$, cf. Theorem 2. Recall that $Z_{1,N} = \left(-\frac{1}{2}\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3}\right)^{-1} N^{1/2}\left(\frac{\partial\widetilde{l}_N^c(1)}{\partial r}\right) \xrightarrow{d} Z_1$, with $\mathrm{plim}_{N\to\infty}\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3} = \frac{\partial^3\widetilde{l}^c(1)}{\partial r^3} = \frac{T(T-1)(T+1)}{12}$ and $\frac{\partial\widetilde{l}_N^c(1)}{\partial r}$ given in (29). When the data are heterogeneous and/or non-normal, one can bootstrap the distribution of $N^{1/2}\left(\frac{\partial\widetilde{l}_N^c(1)}{\partial r}\right)$ or estimate the averages of the second and the fourth moments of the $\varepsilon_{i,t}$ by using that under H_0 $\varepsilon_i = y_i - y_{i,-1}$ for $i = 1, ..., N$. To test $H_0 : \rho = 1$ one could also use any other panel unit root test, e.g. the test of Harris and Tzavalis (1999) that is based on the bias-corrected LSDV estimator for ρ, i.e., $\widehat{\rho}_{ML} + \frac{3}{T+1}$, where $-\frac{3}{T+1}$ is the asymptotic bias of $\widehat{\rho}_{ML}$ when $\rho = 1$. The rate of convergence of $\widehat{\rho}_{ML}$ is $N^{1/2}$ which is faster than $N^{1/4}$, the rate of $\widehat{\rho}_C$. Hence if N is large enough inference based on $\widehat{\rho}_{ML}$ is better in terms of power and size. Finally, to test a hypothesis that only involves β, one can use a Wald test based on $\widehat{\beta}_C$.

5 The finite sample performance of the Modified ML estimators and the Quasi LM test

In this section we compare through Monte Carlo simulations the finite sample properties of three estimators in various panel AR(1) models without covariates: $\widehat{\rho}_C$; the REMLE for ρ that has been proposed by both Chamberlain (1980) and Anderson and Hsiao (1982), henceforth $\widehat{\rho}_{REML}$; and the FEMLE for ρ (i.e., $\widehat{\rho}_{FEML}$) that has been proposed by Hsiao et al. (2002). We study how the properties of these estimators are affected if we change (1) the distributions of the $v_i = y_{i,0} - \mu_i$ or (2) the ratio of the variances of the error components, i.e. σ_μ^2/σ^2. We conducted the simulation experiments for $(T, N) = (4, 100)$, $(9, 100)$, $(4, 500)$ or $(9, 500)$ and $\rho = 0.5$, 0.8, 0.9, 0.95, 0.98 or 1.

In all simulation experiments the error components have been drawn from normal distributions with zero means. We assumed that $\sigma_\mu^2 = 0$, 1 or 25. For the $\varepsilon_{i,t}$ we assumed homoskedasticity and no autocorrelation: $E(\varepsilon_i\varepsilon_i') = \sigma^2 I$ with $\sigma^2 = 1$.

In order to assess how the assumptions with respect to $y_{i,0} - \mu_i$, $i = 1, ..., N$, affect the properties of the estimators, we conducted two different sets of experiments, which are identified by a capital: in one set, labeled NS, the initial observations are non-stationary, i.e., $y_{i,0}-\mu_i=0$, $i=1,...,N$, whereas in the other set, labeled S, the initial observations are drawn from stationary distributions when $|\rho|<1$, i.e., $(y_{i,0}-\mu_i)\sim N(0,\sigma^2_{i,0}/(1-\rho^2))$ with $\sigma^2_{i,0}=\sigma^2$, although $y_{i,0}-\mu_i=0$, $i=1,...,N$, when $\rho=1$.

Note that all four estimators suffer from a weak moment conditions problem when ρ is close to one, cf. Kruiniger (2013).

In the cases of the RE- and FEMLE $(1-\rho)\mu_i+\varepsilon_i$ is decomposed as $(1-\rho)\pi y_{i,0}-(1-\rho)v_i+\varepsilon_i=(1-\rho)\pi y_{i,0}+u_i$ with $\pi=1$ for the FE case. In the experiments we imposed homoskedasticity on their likelihood functions and added the restrictions $\sigma^2>0$ and $(T-1)(1-\rho)^2\sigma^2_v+\sigma^2>0$ to ensure that the estimates of $E(u_iu_i')$ were PD.

We allowed for time effects by subtracting cross-sectional averages from the data.

We computed $\widehat{\rho}_C$ by maximizing $\widetilde{l}_N(\theta)$ subject to $-1\leq r\leq 1.4$. (We also tried using $-1\leq r\leq 2$ but never found a maximum between 1.4 and 2.) If no local maximum was found, we computed $\widehat{\rho}_C$ by solving (9) s.t. $-1\leq r\leq 1.4$ using grid search.

Tables 1-6 report the simulation results in terms of the biases and root mean squared errors (RMSEs) of the estimators and the relative frequencies that $\widehat{\rho}_{LAN}$ did not exist (NM). The tables differ with respect to the dimensions of the panel and the assumptions made about the $y_{i,0}-\mu_i$, $i=1,...,N$. Inspection of the results leads to the following conclusions:[14]

1. In almost all experiments (the exception is design NS with $N=100$ and $\rho=.0.5$) $\widehat{\rho}_{REML}$ is superior in terms of RMSE for 'smaller' values of ρ (i.e., values closer to 0), $\widehat{\rho}_{FEML}$ is superior for 'larger' values of ρ (i.e., values closer to 1), while $\widehat{\rho}_C$ is superior on an interval of 'intermediate' values of ρ, which includes $\rho=0.8$ when $T=4$ and $N=100$, and $\rho=0.9$ when $T=4$ and $N=500$. In most experiments $\widehat{\rho}_{REML}$ is superior when $\rho=0.5$, while $\widehat{\rho}_{FEML}$ is superior when ρ is near/equals 1. When ρ is near 1, the bias of $\widehat{\rho}_C$ is larger than the biases of $\widehat{\rho}_{FEML}$ and $\widehat{\rho}_{REML}$.

2. When T or N increases, the values of the bounds of the interval for ρ on which $\widehat{\rho}_C$ is superior increase. When $T=9$ and $N=500$, $\widehat{\rho}_C$ is superior around $\rho=0.95$.

[14] Dhaene and Jochmans (2016) report simulations results on the finite sample properties of their Adjusted Likelihood estimator ($\widehat{\rho}_{ADJ}$), the bias corrected LSDV estimator of Hahn and Kuersteiner (2002) ($\widehat{\rho}_{HK}$) and the RE GMM estimator of Arellano and Bond (1991) ($\widehat{\rho}_{AB}$). Some of their simulation experiments are equal to some of our experiments. The results for these experiments show that $\widehat{\rho}_{ADJ}$ and $\widehat{\rho}_C$ are very similar and that $\widehat{\rho}_{HK}$ has a large bias when T is small. $\widehat{\rho}_{AB}$ has poor properties when ρ is close to 1 due to weak instruments.

Furthermore, when $\rho = 0.50$ and $T = 9$ or $N = 500$, $\widehat{\rho}_{FEML}$ is often the most efficient estimator after $\widehat{\rho}_{REML}$.

3. When σ_μ^2/σ^2 increases, the RMSE of $\widehat{\rho}_{REML}$ increases and hence the value of the lowerbound of the interval of values of ρ on which $\widehat{\rho}_C$ is superior decreases.

4. When $Var(y_{i,0} - \mu_i)/\sigma^2$ decreases, the bias and the RMSE of $\widehat{\rho}_C$ and the RMSE of $\widehat{\rho}_{REML}$ increase and the value of the upperbound of the interval of values of ρ on which $\widehat{\rho}_C$ is superior decreases.

5. The bias of $\widehat{\rho}_U$ is about the same as the bias of $\widehat{\rho}_C$, also when ρ is (close to) one. Moreover, the sign of the bias of $\widehat{\rho}_C$ is the opposite of the sign that is implied by corollary 1. This suggests that the biases of $\widehat{\rho}_C$ and $\widehat{\rho}_U$ are mainly caused by other factors than the random sign of $N^{1/4}(\widehat{\rho} - 1)$ when $Z_1 > 0$.

6. When $T = 4$ and $N = 100$, $NM > 0.35$ for $\rho \geq 0.8$; when $T = 4$ and $N = 500$, $NM > 0.29$ for $\rho \geq 0.8$; when $T = 9$ and $N = 100$, $NM > 0.35$ for $\rho \geq 0.9$; and when $T = 9$ and $N = 500$, $NM > 0.25$ for $\rho \geq 0.9$. Generally, the higher the value of ρ, the higher the value of NM. When $\rho = 1$, $NM \approx 0.50$ for all panels considered, which supports the idea that even asymptotically $\widehat{\rho}_{LAN}$ may not exist when $\rho = 1$. If the value of $Var(y_{i,0} - \mu_i)/\sigma^2$ decreases, the value of NM increases. Under design NS, when $T = 4$, $N = 100$ and $\rho = 0.5$, we still have $NM > 0.3$.

We have also investigated the size and power properties of the modified likelihood based QLM-test for testing $H_0 : \rho = a$, that is, $QLM(\rho)$. To this end, we conducted three types of Monte Carlo experiments. The designs of two of them, labelled S-Normal and NS-Normal, were similar to designs S and NS described above. The designs of the third kind of experiments, labelled S-ChiSq., were also similar to S with one difference: the $\varepsilon_{i,t}$ were i.i.d. $(\chi^2(1) - 1)/\sqrt{2}$ instead of i.i.d. $N(0, 1)$ so that $(y_{i,0} - \mu_i) \sim (\chi^2(1) - 1)/\sqrt{2(1 - \rho^2)}$ instead of $N(0, 1/(1 - \rho^2))$. In all experiments $\mu_i \sim N(0, 1)$. We used various true values for ρ including 0.5, 0.9, 0.95 and 0.99. The results for the power of $QLM(\rho)$ were based on testing $H_0 : \rho = 0.8$. In all experiments $T = 9$ and $N \in \{100, 500\}$.

$QLM(\rho)$ depends on $\mathcal{H}(\widetilde{\underline{\theta}}_n)$, i.e., an estimate of the expected Hessian that is based on the restricted estimate $\widetilde{\underline{\theta}}_n$. One of the parameters in $\mathcal{H}(\underline{\theta}_{0,n})$ is $\overline{\sigma}^2_{v,n}$. However, the latter is not estimated by a MMLE. Instead we used the restricted FE(Q)MLE for $\overline{\sigma}^2_{v,n}$.

Tables 7 and 8 report the simulation results for the size and the power of $QLM(\rho)$, respectively. Table 7 shows that the empirical size of the test is very close to the nominal size of 5% in all experiments, including those where ρ is close to one. Finally, table 8 shows that the power properties of $QLM(\rho)$ do not change much across the three types of experiments and also that its power is still high when (true) $\rho = 0.99$ despite weak identification in that case.

6 Concluding remarks

Alvarez and Arellano (2004) and Juodis (2013) have extended the MMLE of Lancaster to panel AR(1) models that allow for time-series heteroskedasticity. Their estimators suffer from the same problems as Lancaster's MMLE, namely a weak moment conditions problem if the parameter values are close to the unit root *and* time-series homoskedasticity, cf. Alvarez and Arellano (2004) and Kruiniger (2013); the related problem of possible non-existence; and the possibility of non-uniqueness of local maxima of the modified profile likelihood function. The non-existence problem can be solved by generalizing their estimators in a similar way as Lancaster's estimator has been generalized to (8) or (9). However, it is unclear whether the modified profile likelihood function has at most one local maximum even when the parameter space for ρ is restricted to $[-1, 1]$.[15] If uniqueness would not hold, then one could select a local maximum that is (plausible and) closest to the value of $\widehat{\rho}_{FEML}$ (or $\widehat{\rho}_{REML}$), which is a consistent estimator, as the MMLE.[16]

Alvarez and Arellano (2004) and Dhaene and Jochmans (2016) have also extended the MMLE of Lancaster to panel AR(p) models, while Juodis (2013) has also extended the MMLE of Lancaster to panel VARX(1) models. Comments similar to those made in the previous paragraph apply to these extensions. The MMLEs discussed in section 3 are inconsistent for models with endogenous or predetermined covariates. However, in some cases these models can be replaced by VAR models.

It seems reasonable to expect that the aforementioned extensions of the MMLEs to more general models may also outperform the RE- and FEMLEs for those models in panels of realistic dimensions for some parts of the parameter space. However, a comprehensive Monte Carlo study of their finite sample properties is left for future research.

Finally, we note that Bester and Hansen (2007) and Arellano and Bonhomme (2009) have proposed priors that result in first-order unbiased Bayesian estimators for ρ in a version of model (1) that does not include the K exogenous covariates.

[15] The modified profile likelihood equation for ρ is a polynomial in r. If the model has no covariates, then the coefficients of this polynomial are functions of ρ, σ_v^2 and T variance parameters instead of one.

[16] Note that this method of selecting the MMLE is also "sensible" in finite samples.

A Proofs and derivations

The asymptotic bias of the LSDV estimator for ρ, $\widehat{\rho}_{ML}$:

The LSDV estimators for ρ and β, $\widehat{\rho}_{ML}$ and $\widehat{\beta}_{ML}$, satisfy the profile likelihood equations for ρ and β :

$$\begin{aligned} \sum_{i=1}^{N} y_{i,-1}' Q(y_i - \widehat{\rho}_{ML} y_{i,-1} - X_i \widehat{\beta}_{ML}) &= 0 \text{ and} \\ \sum_{i=1}^{N} X_i' Q(y_i - \widehat{\rho}_{ML} y_{i,-1} - X_i \widehat{\beta}_{ML}) &= 0. \end{aligned} \tag{16}$$

Let $r_{xy_{-1}}^2 = (\sum_{i=1}^N y_{i-1}' Q y_{i,-1})^{-1} \sum_{i=1}^N (y_{i-1}' Q X_i)(\sum_{i=1}^N X_i' Q X_i)^{-1} \sum_{i=1}^N (X_i' Q y_{i,-1})$, $s_y^2 = (T-1)^{-1} N^{-1} \sum_{i=1}^N y_{i,-1}' Q y_{i,-1}$ and $\rho_{ML} = \text{plim}_{N\to\infty} \widehat{\rho}_{ML}$. Using that $y_{i,-1} - \mu_i \iota - \iota \overline{x}_i' \breve{\beta} = \varphi v_i + \Phi Q X_i \beta + \Phi \varepsilon_i = \widetilde{Z}_i + \Phi \varepsilon_i$ and $Q\iota = 0$, it can be shown that the asymptotic bias of $\widehat{\rho}_{ML}$ is given by (cf. e.g. Bun and Carree, 2005):

$$\rho_{ML} - \rho = -\frac{\sigma^2 h(\rho)}{(1 - \rho_{xy_{-1}}^2)\sigma_y^2}, \tag{17}$$

where $\rho_{xy_{-1}}^2 = \text{plim}_{N\to\infty} r_{xy_{-1}}^2$, $\sigma_y^2 = \text{plim}_{N\to\infty} s_y^2$ and $h(\rho) = -(T-1)^{-1} tr(Q\Phi) = \frac{1}{T(T-1)} \sum_{t=1}^{T-1} (T-t)\rho^{t-1} = \xi'(\rho)$. Note that $h(\rho) = \frac{T-1-T\rho+\rho^T}{T(T-1)(1-\rho)^2}$, when $\rho \neq 1$, and $h(1) = \frac{1}{2}$. Assumption 1 implies that $\sigma_y^2 = \frac{\sigma^2}{T-1} tr(\Phi' Q \Phi) + \frac{1}{T-1} E(\widetilde{Z}_i Q \widetilde{Z}_i)$ and $\frac{\sigma^2}{T-1} tr(\Phi' Q \Phi) > 0$ and hence $\sigma_y^2 > 0$. We also have $\rho_{xy_{-1}}^2 < 1$. Furthermore, if $|\rho| \leq 1$, $h(\rho) > 0$ and hence $\rho_{ML} - \rho < 0$ (cf. e.g. Bun and Carree, 2005).

It can also be shown that if $\rho = 1$, then $\rho_{ML} - \rho = -\frac{3}{T+1}$. Note that $E(\widetilde{Z}_i Q \widetilde{Z}_i) = \sigma_v^2 \varphi' Q \varphi + 2E(v_i \varphi' Q \Phi Q X_i)\beta + \beta' E(X_i' Q \Phi Q \Phi Q X_i)\beta$. Let $f(\rho) = \frac{1}{T-1} tr(\Phi' Q \Phi)$ and $g(\rho) = \frac{1}{T-1} \varphi' Q \varphi$. Below we show that $f(1) = \frac{1}{6}(T+1)$. Furthermore, $g(1) = 0$ and when $\rho = 1$, we also have $\beta = \rho_{xy_{-1}}^2 = 0$. We conclude that when $\rho = 1$, then $E(\widetilde{Z}_i Q \widetilde{Z}_i) = 0$, $\sigma_y^2 = \frac{\sigma^2}{6}(T+1)$ and $\rho_{ML} - \rho = -\frac{3}{T+1}$ (cf. Harris and Tzavalis, 1999).

Proof of the claim that $f(1) = \frac{1}{6}(T+1)$:

We have $f(\rho) = (T-1)^{-1} tr(\Phi' Q \Phi) = (T-1)^{-1}(tr\Phi'\Phi - T^{-1}\iota'\Phi\Phi'\iota) = (T-1)^{-1} \times (\sum_{t=0}^{T-2} \sum_{s=0}^{t} \rho^{2s} - T^{-1} \sum_{t=0}^{T-2} (\sum_{s=0}^{t} \rho^s)^2)$. It follows that $f(1) = (T-1)^{-1}(\sum_{t=0}^{T-2}(t+1) - T^{-1} \sum_{t=1}^{T-1} t^2) = (T-1)^{-1}(\frac{1}{2}(T-1)T - \frac{1}{6}(T-1)(2T-1)) = \frac{1}{6}(T+1)$. $\square$

Some results related to $\widetilde{l}_N^c(r)$ and $\frac{\partial \widetilde{l}_N^c(r)}{\partial r}$:

By the envelope theorem we have $\frac{\partial \widetilde{l}_N^c(r)}{\partial r} = \Psi_\rho(r, \widehat{\sigma}^2(r, \widehat{\beta}(r)), \widehat{\beta}(r))$, i.e.,

$$\frac{\partial \widetilde{l}_N^c(r)}{\partial r} = (T-1)\xi'(r) + \widehat{\sigma}^{-2}(r, \widehat{\beta}(r))N^{-1}\sum_{i=1}^{N}(y_i - ry_{i,-1} - X_i\widehat{\beta}(r))'Qy_{i,-1}. \tag{18}$$

Let $\widehat{\sigma}^2_{ML} = (T-1)^{-1}N^{-1}\sum_{i=1}^N[(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML})'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML})]$. Next we show that the first-order condition for a local maximum of $\widetilde{l}_N^c(r)$ can be written as

$$\frac{\partial \widetilde{l}_N^c(r)}{\partial r} = (T-1)\xi'(r) - \frac{(T-1)(r-\widehat{\rho}_{ML})}{\widehat{\sigma}^2_{ML}/(s_y^2(1-r^2_{xy_{-1}})) + (r-\widehat{\rho}_{ML})^2} = 0. \tag{19}$$

Derivation of (19): Using $\sum_{i=1}^N X_i'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML}) = 0$ from (16) and $\sum_{i=1}^N X_i'Q(y_i - ry_{i,-1} - X_ib) = 0$ from (5), we obtain

$$\widehat{\beta}_{ML} - b = (\sum_{i=1}^N X_i'QX_i)^{-1}\sum_{i=1}^N(X_i'Qy_{i,-1})(r-\widehat{\rho}_{ML}). \tag{20}$$

Next, using $\sum_{i=1}^N y_{i,-1}'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML}) = 0$ from (16), we obtain

$$\sum_{i=1}^N y_{i,-1}'Q(y_i - ry_{i,-1} - X_ib) = \sum_{i=1}^N y_{i,-1}'Q(y_{i,-1}(\widehat{\rho}_{ML} - r) + X_i(\widehat{\beta}_{ML} - b)) =$$
$$(\widehat{\rho}_{ML} - r)(\sum_{i=1}^N(y_{i-1}'Qy_{i,-1}) - \sum_{i=1}^N(y_{i-1}'QX_i)(\sum_{i=1}^N X_i'QX_i)^{-1}\sum_{i=1}^N(X_i'Qy_{i,-1})).$$

Hence

$$(T-1)^{-1}N^{-1}\sum_{i=1}^N(y_i - ry_{i,-1} - X_i\widehat{\beta}(r))'Qy_{i,-1} = (\widehat{\rho}_{ML} - r)s_y^2(1-r^2_{xy_{-1}}). \tag{21}$$

Using $\sum_{i=1}^N y_{i,-1}'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML}) = 0$ and $\sum_{i=1}^N X_i'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML}) = 0$ from (16), we obtain

$$\sum_{i=1}^N[(y_i - ry_{i,-1} - X_ib)'Q(y_i - ry_{i,-1} - X_ib)] =$$

$$\sum_{i=1}^{N}[(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML})'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML})+$$
$$((\widehat{\rho}_{ML} - r)y_{i,-1} + X_i(\widehat{\beta}_{ML} - b))'Q(y_{i,-1}(\widehat{\rho}_{ML} - r) + X_i(\widehat{\beta}_{ML} - b))].$$

In addition, by using (20) once more, we obtain

$$\sum_{i=1}^{N}[((\widehat{\rho}_{ML} - r)y_{i,-1} + X_i(\widehat{\beta}_{ML} - b))'Q(y_{i,-1}(\widehat{\rho}_{ML} - r) + X_i(\widehat{\beta}_{ML} - b))] =$$
$$(\widehat{\rho}_{ML} - r)^2[\sum_{i=1}^{N} y'_{i,-1}Qy_{i,-1} - \sum_{i=1}^{N}(y'_{i-1}QX_i)(\sum_{i=1}^{N} X'_iQX_i)^{-1}\sum_{i=1}^{N}(X'_iQy_{i,-1})].$$

Hence

$$\widehat{\sigma}^2(r, \widehat{\beta}(r)) = (T-1)^{-1}N^{-1}\sum_{i=1}^{N}(y_i - ry_{i,-1} - X_i\widehat{\beta}(r))'Q(y_i - ry_{i,-1} - X_i\widehat{\beta}(r)) =$$
$$\widehat{\sigma}^2_{ML} + (\widehat{\rho}_{ML} - r)^2 s_y^2(1 - r^2_{xy_{-1}}). \tag{22}$$

Finally, combining (18) with (21) and (22) yields (19).

Next we show that $\sigma^2_{ML} = \text{plim}_{N\to\infty}\,\widehat{\sigma}^2_{ML} > 0$.

Proof of the claim that $\sigma^2_{ML} > 0$:

Using $Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML}) = Q(\varepsilon_i + (\rho - \widehat{\rho}_{ML})y_{i,-1} + X_i(\beta - \widehat{\beta}_{ML}))$ and $Qy_{i,-1} = Q(\widetilde{Z}_i + \Phi\varepsilon_i)$, where $\widetilde{Z}_i = \varphi v_i + \Phi QX_i\beta$, we obtain $\widehat{\sigma}^2_{ML} = (T-1)^{-1}N^{-1}\sum_{i=1}^{N}[(\varepsilon_i + (\rho - \widehat{\rho}_{ML})(\widetilde{Z}_i + \Phi\varepsilon_i) + X_i(\beta - \widehat{\beta}_{ML}))'Q(\varepsilon_i + (\rho - \widehat{\rho}_{ML})(\widetilde{Z}_i + \Phi\varepsilon_i) + X_i(\beta - \widehat{\beta}_{ML}))]$.

Assumption 1 implies that $\varepsilon_i|(v_i, QX_i) \sim i.i.d.N(0, \sigma^2 I_T)$, $i = 1, ..., N$, with $\sigma^2 > 0$. It follows that $\sigma^2_{ML} = \text{plim}_{N\to\infty}\,\widehat{\sigma}^2_{ML} \geq \text{plim}_{N\to\infty}(T-1)^{-1}N^{-1}\sum_{i=1}^{N}[(\varepsilon_i + (\rho - \widehat{\rho}_{ML})\Phi\varepsilon_i)'Q\times (\varepsilon_i + (\rho - \widehat{\rho}_{ML})\Phi\varepsilon_i)] = \sigma^2(T-1)^{-1}tr((I + (\rho - \rho_{ML})\Phi)'Q(I + (\rho - \rho_{ML})\Phi)) > 0$. $\square$

Proof of the claim that $\widetilde{l}^c_N(r)$ converges uniformly in probability to $\widetilde{l}^c(r)$:

We have $\widetilde{l}^c_N(r) = \widetilde{l}_N(r, \widehat{\sigma}^2(r, \widehat{\beta}(r)), \widehat{\beta}(r)) = (T-1)\xi(r) - \frac{T-1}{2}\log(\widehat{\sigma}^2(r, \widehat{\beta}(r))) - \frac{T-1}{2}$ and from (22), $\widehat{\sigma}^2(r, \widehat{\beta}(r)) = \widehat{\sigma}^2_{ML} + (\widehat{\rho}_{ML} - r)^2 s_y^2(1 - r^2_{xy_{-1}})$. Note that $-\log(\widehat{\sigma}^2(r, \widehat{\beta}(r)))$ is a concave function of r. Then it follows from pointwise convergence of $\log(\widehat{\sigma}^2(r, \widehat{\beta}(r)))$ to the function $\log(\sigma^2(r)) \equiv \log(\sigma^2_{ML} + (\rho_{ML} - r)^2\sigma_y^2(1 - \rho^2_{xy_{-1}}))$ that $\text{plim}_{N\to\infty}\sup_{r\in[-1,\infty)}\left|\widetilde{l}^c_N(r) - \widetilde{l}^c(r)\right| = 0$, see e.g. Newey and McFadden (1994, section 2.6). $\square$

Some results related to $\widetilde{l}^c(r)$, $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_\rho = 0$ and $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho$:

The first-order condition for a local maximum of $\widetilde{l}^c(r)$ can be written as:

$$\frac{\partial \widetilde{l}^c(r)}{\partial r} = (T-1)\xi'(r) - \frac{(T-1)(r-\rho_{ML})}{\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}})) + (r-\rho_{ML})^2} = 0. \tag{23}$$

The second-order condition for a local maximum of $\widetilde{l}^c(r)$ is given by:

$$\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2} = (T-1)\xi''(r) - \frac{(T-1)(\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}})) - (r-\rho_{ML})^2)}{(\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}})) + (r-\rho_{ML})^2)^2} < 0.$$

Below we show that

$$\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}})) = -\left(\frac{\sigma^2\xi'(\rho)}{\sigma^2_y(1-\rho^2_{xy_{-1}})}\right)^2 + \sigma^2/(\sigma^2_y(1-\rho^2_{xy_{-1}})). \tag{24}$$

Then it is easily verified that $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_\rho = 0$ and

$$\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho = (T-1)\xi''(\rho) + (T-1)(2\left(\xi'(\rho)\right)^2 - \sigma^2_y(1-\rho^2_{xy_{-1}})/\sigma^2).$$

Note that $(T-1)\sigma^2_y = \sigma^2 tr(\Phi' Q \Phi) + E(\widetilde{Z}_i Q \widetilde{Z}_i)$. Let $\sigma^2_x = \text{plim}_{N\to\infty} \frac{1}{(T-1)N}\sum_{i=1}^N X_i' Q X_i$ and $\sigma^2_{xy_{-1}} = \text{plim}_{N\to\infty} \frac{1}{(T-1)N}\sum_{i=1}^N X_i' Q y_{i,-1}$. Using $Qy_{i,-1} = Q(\widetilde{Z}_i + \Phi\varepsilon_i)$ it is easily seen that $\sigma^2_y(1-\rho^2_{xy_{-1}})/\sigma^2 = (\sigma^2_y - \sigma'_{xy_{-1}}\sigma_x^{-2}\sigma_{xy_{-1}})/\sigma^2 \geq (T-1)^{-1} tr(\Phi' Q \Phi)$, with equality holding if $\rho = 1$ or $\sigma^2_v = \beta = 0$ (i.e. if $\Sigma_{zqz} = 0$). We also have $\xi''(\rho) - (T-1)^{-1}\times tr(\Phi(\rho)' Q \Phi(\rho)) + 2(\xi'(\rho))^2 \leq 0$, with equality holding if $\rho = 1$ or $T = 2$. It follows that $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho \leq 0$, with equality holding if $\rho = 1$ or if $T = 2$ and $\sigma^2_v = \beta = 0$. Thus $\widetilde{l}^c(r)$ has a local maximum at ρ when $\rho \neq 1$ and, in case $T = 2$, $\Sigma_{zqz} > 0$. Below we show that $\widetilde{l}^c(r)$ has a stationary point of inflection at ρ when $\rho = 1$.

Derivation of (24): Given that the equality $y_i - r y_{i,-1} - X_i b = (\rho - r) y_{i,-1} + X_i(\beta - b) + \alpha_i \iota + \varepsilon_i$ holds for any r and b, including for $r = \widehat{\rho}_{ML}$ and $b = \widehat{\beta}_{ML}$, we can rewrite $\widehat{\sigma}^2_{ML}$ as

$$\widehat{\sigma}^2_{ML} = (T-1)^{-1}N^{-1}\sum_{i=1}^N [((\rho-\widehat{\rho}_{ML})y_{i,-1} + X_i(\beta - \widehat{\beta}_{ML}) + \varepsilon_i)'\times$$
$$Q((\rho-\widehat{\rho}_{ML})y_{i,-1} + X_i(\beta - \widehat{\beta}_{ML}) + \varepsilon_i)]. \tag{25}$$

Let $\beta_{ML} = \text{plim}_{N\to\infty}\widehat{\beta}_{ML}$, $\sigma_x^2 = \text{plim}_{N\to\infty}(T-1)^{-1}N^{-1}\sum_{i=1}^N X_i'QX_i$, and $\sigma_{xy_{-1}}^2 = \text{plim}_{N\to\infty}(T-1)^{-1}N^{-1}\sum_{i=1}^N X_i'Qy_{i,-1}$. Then combining $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N X_i'Q(y_i - \widehat{\rho}_{ML}y_{i,-1} - X_i\widehat{\beta}_{ML}) = 0$ from (16) with $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N X_i'Q(y_i - \rho y_{i,-1} - X_i\beta) = p\lim_{N\to\infty} N^{-1}\sum_{i=1}^N X_i'Q\varepsilon_i = 0$ gives

$$\beta_{ML} - \beta = \sigma_x^{-2}\sigma_{xy_{-1}}(\rho - \rho_{ML}). \tag{26}$$

Using (25) and (26) and recalling that $\xi'(\rho) = h(\rho) = -(T-1)^{-1}tr(Q\Phi)$, we obtain

$$\begin{aligned}\sigma_{ML}^2 = \ &\text{p}\lim_{N\to\infty}\widehat{\sigma}_{ML}^2 = (\rho-\rho_{ML})^2\sigma_y^2 + 2(\beta-\beta_{ML})'\sigma_{xy_{-1}}(\rho-\rho_{ML}) + \\ &(\beta-\beta_{ML})'\sigma_x^2(\beta-\beta_{ML}) + 2(\rho-\rho_{ML})\sigma^2(T-1)^{-1}tr(Q\Phi) + \sigma^2 = \\ &(\rho-\rho_{ML})^2\sigma_y^2 - \sigma_{xy_{-1}}'\sigma_x^{-2}\sigma_{xy_{-1}}(\rho-\rho_{ML})^2 + 2(\rho-\rho_{ML})\sigma^2(T-1)^{-1}tr(Q\Phi) + \sigma^2 = \\ &(\rho-\rho_{ML})^2\sigma_y^2(1-\rho_{xy_{-1}}^2) - 2(\rho-\rho_{ML})\sigma^2\xi'(\rho) + \sigma^2.\end{aligned}$$

Finally, using $\rho_{ML} - \rho = -\frac{\sigma^2\xi'(\rho)}{\sigma_y^2(1-\rho_{xy_{-1}}^2)}$, we find that

$$\sigma_{ML}^2/(\sigma_y^2(1-\rho_{xy_{-1}}^2)) = -\left(\frac{\sigma^2\xi'(\rho)}{\sigma_y^2(1-\rho_{xy_{-1}}^2)}\right)^2 + \sigma^2/(\sigma_y^2(1-\rho_{xy_{-1}}^2)).$$

Proof of the claim that $\widetilde{l}^c(r)$ has an inflection point at ρ when $\rho = 1$:
We have already seen that $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho=1} = \frac{\partial^2\widetilde{l}^c(r)}{\partial r^2}|_{\rho=1} = 0$. In addition, we have

$$\begin{aligned}\frac{\partial^3\widetilde{l}^c(r)}{\partial r^3} &= (T-1)\xi'''(r) + \frac{6(T-1)(r-\rho_{ML})(\sigma_{ML}^2/(\sigma_y^2(1-\rho_{xy_{-1}}^2)))}{(\sigma_{ML}^2/(\sigma_y^2(1-\rho_{xy_{-1}}^2)) + (r-\rho_{ML})^2)^3} \\ &\quad - \frac{2(T-1)(r-\rho_{ML})^3}{(\sigma_{ML}^2/(\sigma_y^2(1-\rho_{xy_{-1}}^2)) + (r-\rho_{ML})^2)^3},\end{aligned}$$

$\xi'''(1) = \frac{(T-2)(T-3)}{12}$, $f(1) = \frac{T+1}{6}$, $\xi'(1) = \frac{1}{2}$, $\lim_{\rho\to1}\rho_{xy_{-1}}^2 = 0$, $\lim_{\rho\to1}(\rho_{ML}-\rho) = -\frac{\xi'(1)}{f(1)} = -\frac{3}{T+1}$ and $\lim_{\rho\to1}(\sigma_{ML}^2/(\sigma_y^2(1-\rho_{xy_{-1}}^2))) = -\left(\frac{\xi'(1)}{f(1)}\right)^2 + \frac{1}{f(1)} = 3\frac{(2T-1)}{(T+1)^2}$. It follows that $\frac{\partial^3\widetilde{l}^c(r)}{\partial r^3}|_{\rho=1} = (T-1)\xi'''(1) + \frac{(T-1)^2}{2} \neq 0$ (in fact > 0) for $T \geq 2$. $\square$

We now present two lemmata that help to establish uniqueness and consistency of our MMLEs:

Lemma 1 *Let Assumption 1 hold. Then (i) $\widetilde{l}_N(\theta)$ has either no local optima or one local* *maximum, namely* $\widehat{\theta}_W = \widehat{\theta}_C$*, and one local minimum on the set* $\widetilde{\Omega}$ *w.p.1. (ii)* $\widetilde{l}_N^c(r)$ *has either no local optima or one local maximum, namely* $\widehat{\rho}_W = \widehat{\rho}_C$*, and one local minimum on the interval* $[-1,\infty)$ *w.p.1. (iii) The equation* $\frac{\partial \widetilde{l}_N^c(r)}{\partial r} = 0$ *has either no solution on* $[-1,\infty)$ *or two solutions on* $[-1,\infty)$*, namely* $\widehat{\rho}_1$ *and* $\widehat{\rho}_2$ *with* $\widehat{\rho}_1 < \widehat{\rho}_2$ *and* $\widehat{\rho}_1 = \widehat{\rho}_W = \widehat{\rho}_C$*, w.p.1.*

Lemma 2 *Let Assumption 1 hold. First let* $\rho \neq 1$*. Then (i)* $\widetilde{l}(\theta)$ *has one local maximum and one local minimum but no inflection point on the set* $\widetilde{\Omega}$*. The local maximum is attained at* θ_0*. (ii)* $\widetilde{l}^c(r)$ *has one local maximum and one local minimum but no inflection point on the interval* $[-1,\infty)$*. The local maximum of* $\widetilde{l}^c(r)$ *is attained at* ρ*. (iii) The equation* $\frac{\partial \widetilde{l}^c(r)}{\partial r} = 0$ *has two solutions on* $[-1,\infty)$*:* ρ_1 *and* ρ_2 *with* $\rho_1 < \rho_2$ *and* $\rho_1 = \rho$*.*

Now let $\rho = 1$*. Then (iv)* $\widetilde{l}(\theta)$ *has one stationary point of inflection but no local optima on* $\widetilde{\Omega}$*. The inflection point is attained at* θ_0*. (v)* $\widetilde{l}^c(r)$ *has one stationary point of inflection but no local optima on* $[-1,\infty)$*. The inflection point of* $\widetilde{l}^c(r)$ *is attained at* $\rho = 1$*. (vi) The equation* $\frac{\partial \widetilde{l}^c(r)}{\partial r} = 0$ *has only one solution on* $[-1,\infty)$*:* $\rho_1 = 1$*.*

We first prove the following lemma, which summarizes some useful properties of $\xi'(\rho)$:

Lemma 3 *Let* $\rho \geq -1$*. When* $T \geq 2$*,* $\xi'(\rho) > 0$*,* $\xi'(1) = \frac{1}{2}$*;*

When $T = 2$*,* $\xi'(\rho) = \frac{1}{2}$*;*

When $T = 3$*,* $\xi'(-1) = \frac{1}{6}$ *and* $\xi''(\rho) = \frac{1}{6}$*;*

When $T \geq 4$ *and* T *is even,* $\xi'(-1) = \frac{1}{2(T-1)}$*,* $\xi''(-1) = 0$*,* $\xi''(\rho) > 0$ *when* $\rho > -1$*, and* $\xi'''(\rho) > 0$*;*

When $T \geq 5$ *and* T *is odd,* $\xi'(-1) = \frac{1}{2T}$*,* $\xi''(\rho) > 0$*,* $\xi'''(-1) = -\frac{T-3}{4T} < 0$*,* $\xi'''(-1/2) = \frac{2^{4-T}(2^T-3T+1)}{27T(T-1)} > 0$*, and* $\exists \rho_*$ *with* $-1 < \rho_* < -1/2$ *such that* $\xi'''(\rho_*) = 0$*,* $\xi'''(\rho) < 0$ *for* $\rho < \rho_*$ *and* $\xi'''(\rho) > 0$ *for* $\rho > \rho_*$*.*

Proof of lemma 3: for the proof of most properties see Dhaene and Jochmans (2015). Their proof uses that $\xi'(\rho) = [T(T-1)]^{-1}\sum_{t=1}^{T-1}(T-t)\rho^{t-1} = \frac{T-1-T\rho+\rho^T}{T(T-1)(1-\rho)^2}$ when $\rho \neq 1$, and Descartes' rule of signs. The remaining claims, i.e., $\xi'(\rho) = \frac{1}{2}$ when $\rho = 1$ or $T = 2$, $\xi''(\rho) = \frac{1}{6}$ when $T = 3$, $\xi'(-1) = \frac{1}{2(T-1)}$ when T is even, and $\xi'(-1) = \frac{1}{2T}$ when T is odd, are easily verified. $\square$

Thus when $\rho \geq -1$, we have:

If $T = 2$, then $\xi'(\rho)$ is strictly positive and constant;

If $T = 3$, then $\xi'(\rho)$ is strictly positive and increasing linearly;
If $T \geq 4$ and T is even, then $\xi'(\rho)$ is strictly positive, non-decreasing and strictly convex;
If $T \geq 5$ and T is odd, then $\xi'(\rho)$ is strictly positive, strictly increasing and first strictly concave and then strictly convex.

Proof of lemma 1:

We can write (19) as

$$\xi'(r)\{\widehat{\sigma}^2_{ML}/(s^2_y(1-r^2_{xy_{-1}})) + (r-\widehat{\rho}_{ML})^2\} = (r-\widehat{\rho}_{ML}). \tag{27}$$

Let $\zeta_N(r) = \xi'(r)\{\widehat{\sigma}^2_{ML}/(s^2_y(1-r^2_{xy_{-1}})) + (r-\widehat{\rho}_{ML})^2\}$. Then $\zeta'_N(r) = \xi''(r)\{\widehat{\sigma}^2_{ML}/(s^2_y(1-r^2_{xy_{-1}})) + (r-\widehat{\rho}_{ML})^2\} + 2(r-\widehat{\rho}_{ML})\xi'(r)$ and $\zeta''_N(r) = \xi'''(r)\{\widehat{\sigma}^2_{ML}/(s^2_y(1-r^2_{xy_{-1}})) + (r-\widehat{\rho}_{ML})^2\} + 4(r-\widehat{\rho}_{ML})\xi''(r) + 2\xi'(r)$.

By lemma 3 $\zeta_N(r) > 0$ when $r \geq -1$. Hence any solution r of (27) should satisfy $r > \widehat{\rho}_{ML}$. When $r \geq \max(-1, \widehat{\rho}_{ML})$, we also have by lemma 3 that $\zeta'_N(r) > 0$ and if T is even that $\zeta''_N(r) > 0$ while if T is odd we either have $\zeta''_N(r) > 0$ for all $r \geq \max(-1, \widehat{\rho}_{ML})$ or $\zeta''_N(r) < 0$ for all r on $[\max(-1, \widehat{\rho}_{ML}), \rho_{**})$ and $\zeta''_N(r) > 0$ for all r on (ρ_{**}, ∞) with $\rho_{**} > \max(-1, \rho_{ML})$ and equal to the solution of $\zeta''_N(r) = 0$. It follows that w.p.1. the graph of $\zeta_N(r)$ either does not intersect the line $r - \widehat{\rho}_{ML}$ (this may well happen when ρ is close or equal to unity, see Lancaster for an example) or intersects the line $r - \widehat{\rho}_{ML}$ twice, say at $r = \widehat{\rho}_1$ and $r = \widehat{\rho}_2$ with $\max(-1, \widehat{\rho}_{ML}) < \widehat{\rho}_1 < \widehat{\rho}_2$. Both solutions of (27) would correspond to local optima w.p.1. That is, the possibility that $r - \widehat{\rho}_{ML}$ is a tangent to $\zeta_N(r)$ at $\widehat{\rho}_1$ and/or $\widehat{\rho}_2$ is an event with probability zero, so $\widehat{\rho}_1$ and $\widehat{\rho}_2$ would not correspond to (an) inflection point(s) w.p.1. It is clear that when $|\rho| \leq 1$, $\widetilde{l}^c_N(r)$ and $\widetilde{l}_N(\theta)$ attain at most one local maximum on the interval $[-1, \infty)$ and the set $\widetilde{\Omega}$, respectively. Moreover, if (27) has any solutions, then $\widehat{\rho}_C$ is one of them and $\widehat{\rho}_C = \widehat{\rho}_W \geq \max(-1, \widehat{\rho}_{ML})$. Given that $\widetilde{l}^c_N(r)$ has a global maximum at $r = \infty$ (because $\lim_{r\uparrow\infty} \widetilde{l}^c_N(r) = \infty$), we conclude that if (27) has any solutions, then it has two solutions $\widehat{\rho}_1$ and $\widehat{\rho}_2$ with $\widehat{\rho}_1 < \widehat{\rho}_2$, where $\widehat{\rho}_1 = \widehat{\rho}_C = \widehat{\rho}_W$ corresponds to a local maximum of $\widetilde{l}^c_N(r)$ and $\widehat{\rho}_2$ corresponds to a local minimum of $\widetilde{l}^c_N(r)$, because $\widetilde{l}^c_N(r)$ cannot attain a local maximum at $\widehat{\rho}_2$. Likewise, given that $\lim_{r\uparrow\infty} \widetilde{l}_N(\widehat{\theta}(r)) = \lim_{r\uparrow\infty} \widetilde{l}^c_N(r) = \infty$, we conclude that if (27) has solutions $\widehat{\rho}_1$ and $\widehat{\rho}_2$ with $\widehat{\rho}_1 < \widehat{\rho}_2$, then $\widetilde{l}_N(\theta)$ has two local optima on the set $\widetilde{\Omega}$, say $\widehat{\theta}_1$ and $\widehat{\theta}_2$, where $\widehat{\theta}_1 = \widehat{\theta}(\widehat{\rho}_1) = \widehat{\theta}(\widehat{\rho}_W) = \widehat{\theta}(\widehat{\rho}_C)$ corresponds to a local maximum of $\widetilde{l}_N(\theta)$ and $\widehat{\theta}_2 = \widehat{\theta}(\widehat{\rho}_2)$ corresponds to a local minimum of $\widetilde{l}_N(\theta)$, because $\widetilde{l}_N(\theta)$ cannot attain a local maximum at $\widehat{\theta}(\widehat{\rho}_2)$. $\square$

Proof of lemma 2:

We can write (23) as

$$\xi'(r)\{\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}}))+(r-\rho_{ML})^2\}=(r-\rho_{ML}). \tag{28}$$

Let $\zeta(r)=\xi'(r)\{\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}}))+(r-\rho_{ML})^2\}$. Then $\zeta'(r)=\xi''(r)\{\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}}))+(r-\rho_{ML})^2\}+2(r-\rho_{ML})\xi'(r)$ and $\zeta''(r)=\xi'''(r)\{\sigma^2_{ML}/(\sigma^2_y(1-\rho^2_{xy_{-1}}))+(r-\rho_{ML})^2\}+4(r-\rho_{ML})\xi''(r)+2\xi'(r)$.

By lemma 3 $\zeta(r)>0$ when $r\geq -1$. Hence any solution r of (28) should satisfy $r>\rho_{ML}$. When $r\geq\max(-1,\rho_{ML})$, we also have by lemma 3 that $\zeta'(r)>0$ and if T is even that $\zeta''(r)>0$ while if T is odd we either have $\zeta''(r)>0$ for all $r\geq\max(-1,\rho_{ML})$ or $\zeta''(r)<0$ for all r in $[\max(-1,\rho_{ML}),\rho_{**})$ and $\zeta''(r)>0$ for all r in (ρ_{**},∞) where ρ_{**} satisfies $\rho_{**}>\max(-1,\rho_{ML})$ and $\zeta''(\rho_{**})=0$. It follows that the graph of $\zeta(r)$ intersects the line $r-\rho_{ML}$ at most twice, say at $r=\rho_1$ and $r=\rho_2$ with $\max(-1,\rho_{ML})\leq\rho_1\leq\rho_2$. On the other hand we have $\rho\geq\max(-1,\rho_{ML})$ and using (17), $h(\rho)=\xi'(\rho)$ and (24) it is easily verified that ρ is a solution of (28). We have already seen in the main text that when $\rho\neq 1$, $\widetilde{l}^c(r)$ and $\widetilde{l}(\theta)$ attain a local maximum at ρ and $\theta_0=\text{plim}_{N\to\infty}\widehat{\theta}(\rho)=\text{plim}_{N\to\infty}(\rho,\widehat{\sigma}^2(\rho,\widehat{\beta}(\rho)),\widehat{\beta}(\rho))'$, respectively. Given that $\widetilde{l}^c(r)$ has a global maximum at $r=\infty$ (because $\lim_{r\uparrow\infty}\widetilde{l}^c(r)=\infty$), $\widetilde{l}^c(r)$ cannot attain a local maximum at ρ_2 so we conclude that (28) has two solutions, ρ_1 and ρ_2, where $\rho_1=\rho$ and ρ_2 corresponds to a local minimum of $\widetilde{l}^c(r)$. Similarly, given that $\lim_{r\uparrow\infty}\text{plim}_{N\to\infty}\widetilde{l}(\widehat{\theta}(r))=\lim_{r\uparrow\infty}\widetilde{l}^c(r)=\infty$, $\widetilde{l}(\theta)$ cannot attain a local maximum at $\text{plim}_{N\to\infty}\widehat{\theta}(\rho_2)$ so we conclude that $\text{plim}_{N\to\infty}\widehat{\theta}(\rho_1)=\text{plim}_{N\to\infty}\widehat{\theta}(\rho)=\theta_0$ and that $\text{plim}_{N\to\infty}\widehat{\theta}(\rho_2)$ corresponds to a local minimum of $\widetilde{l}(\theta)$.

When $\rho=1$, we know that $\widetilde{l}(\theta)$ and $\widetilde{l}^c(r)$ have an inflection point at θ_0 and ρ, respectively. When $\rho=1$, we also have that $\rho_{ML}=1-\frac{3}{T+1}$, so that $\max(-1,\rho_{ML})=\rho_{ML}$ and $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho_{ML}}=(T-1)\xi'(\rho_{ML})>0$. Because $\frac{\partial\widetilde{l}^c(r)}{\partial r}$ is continuous for $r\geq\rho_{ML}$, because ρ_1 is the smallest $r>\rho_{ML}$ such that $\frac{\partial\widetilde{l}^c(r)}{\partial r}=0$, and because $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho_{ML}}>0$, we have $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho_1-}>0$. We now show that ρ_1 is an inflection point. Suppose instead that ρ_1 were a maximum (note that ρ_1 cannot be a minimum because $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho_1-}>0$). Then ρ_2 must have been a minimum (because $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{+\infty}>0$) and there would be no inflection point in the interval $[\rho_{ML},\infty)$. This would contradict that $\widetilde{l}^c(r)$ has at least one inflection point larger than ρ_{ML}, namely at $r=1$. Thus ρ_1 is an inflection point. Remains to show that $\rho_1=\rho_2=1$. Since $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho_1-}>0$ and ρ_1 is inflection point and $\frac{\partial\widetilde{l}^c(r)}{\partial r}=0$ has at most two solutions, we also have $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho_1+}>0$. Because $\rho_1>\rho_{ML}$, because $(T-1)^{-1}(\sigma^2_{ML}/\sigma^2_y+(r-\rho_{ML})^2)$ is strictly positive, increasing and strictly convex when

$r > \rho_{ML}$, and because $(T-1)^{-1}(\sigma^2_{ML}/\sigma^2_y + (r-\rho_{ML})^2)\frac{\partial \widetilde{l}^c(r)}{\partial r} = \zeta(r) - (r-\rho_{ML})$, we have $(\zeta(r) - (r-\rho_{ML}))|_{\rho_1 -} > 0$, $(\zeta(r) - (r-\rho_{ML}))|_{\rho_1} = 0$ and $(\zeta(r) - (r-\rho_{ML}))|_{\rho_1 +} > 0$. It follows that $\zeta'(\rho_1) = 1$. Because $\rho_{ML} > -1/2$, we have both $\xi'''(r) > 0$ and $\zeta''(r) > 0$ for all $r > \rho_{ML}$. This implies that $\zeta'(r) > 1$ for $r > \rho_1$ and hence that there exists no $r > \rho_1$ such that $\zeta(r) - (r-\rho_{ML}) = 0$. It follows that $\rho_1 = \rho_2 = \rho = 1$. Similarly, we obtain that $\text{plim}_{N\to\infty}\widehat{\theta}(\rho_1) = \text{plim}_{N\to\infty}\widehat{\theta}(\rho_2) = \theta_0$ $\square$

Lemma 4 *Let $\widetilde{\Theta}^c_N$ be the set of roots of $\frac{\partial \widetilde{l}^c_N}{\partial r} = 0$ corresponding to local maxima of $\widetilde{l}^c_N$ on the interval $[-1,\infty)$. Let Assumption 1 hold and $\rho = 1$. Then $\lim_{N\to\infty}\Pr(\widetilde{\Theta}^c_N = \varnothing) > 0$.*

Proof: Let $\kappa(r) = \xi'(r)\{\widehat{\sigma}^2_{ML}/(s^2_y(1-r^2_{xy_{-1}})) + (r-\widehat{\rho}_{ML})^2\} - (r-\widehat{\rho}_{ML})$. To save space we only prove the lemma for the model without covariates so that $r^2_{xy_{-1}} = 0$.

Note that $\frac{\partial \widetilde{l}^c_N}{\partial r} = 0 \Leftrightarrow \kappa(r) = 0$. When $\rho = 1$,

$$
\begin{aligned}
s^2_y &= (T-1)^{-1}N^{-1}\textstyle\sum_{i=1}^N \varepsilon_i'\Phi' Q\Phi\varepsilon_i \equiv (T-1)^{-1}A_N, \\
\widehat{\rho}_{ML} - 1 &= (\textstyle\sum_{i=1}^N \varepsilon_i'\Phi' Q\Phi\varepsilon_i)^{-1}\textstyle\sum_{i=1}^N \varepsilon_i'\Phi' Q\varepsilon_i \equiv A_N^{-1}B_N, \quad \text{and} \\
\widehat{\sigma}^2_{ML} &= (T-1)^{-1}N^{-1}\textstyle\sum_{i=1}^N \varepsilon_i'(\Phi(1-\widehat{\rho}_{ML}) + I)'Q(\Phi(1-\widehat{\rho}_{ML}) + I)\varepsilon_i = \\
&\quad (T-1)^{-1}(C_N - A_N^{-1}B_N^2)
\end{aligned}
$$

where $C_N \equiv N^{-1}\sum_{i=1}^N \varepsilon_i' Q\varepsilon_i$.

When r is close to one, $\xi'(r) \approx \xi'(1) + \xi''(1)(r-1) + \frac{1}{2}\xi'''(1)(r-1)^2$. Note that $\xi'(1) = \frac{1}{2}$, $\xi''(1) = \frac{1}{6}(T-2)$ and $\xi'''(1) = \frac{1}{12}(T-2)(T-3)$.

Let $z = r-1$. When r is close to one, $\kappa(r) \approx (\frac{1}{2} + \frac{1}{6}(T-2)z + \frac{1}{24}(T-2)(T-3)z^2)(A_N^{-1}C_N - 2A_N^{-1}B_N z + z^2) - z + A_N^{-1}B_N \approx \frac{1}{2}A_N^{-1}C_N + A_N^{-1}B_N + (\frac{1}{6}(T-2)A_N^{-1}C_N - A_N^{-1}B_N - 1)z + (\frac{1}{2} + \frac{1}{24}(T-2)(T-3)A_N^{-1}C_N - \frac{1}{3}(T-2)A_N^{-1}B_N)z^2 \equiv \widetilde{\kappa}(z)$ where in the last step we have dropped the z^3-term and the z^4-term which are negligible when r is close enough to one. Note that when $T = 2$, the approximations can be replaced by exact equalities.

Note that $\widetilde{\kappa}(z) = 0 \Leftrightarrow 6A_N\widetilde{\kappa}(z) = 0$. Solving $6A_N\widetilde{\kappa}(z) = 0$ gives: $z_{1,2} = \{(6A_N + 6B_N - (T-2)C_N) \pm \sqrt{D_N}\}/(6A_N + \frac{1}{2}(T-2)(T-3)C_N - 4(T-2)B_N)$ where $D_N \equiv 36(A_N^2 + B_N^2 - A_NC_N) + (T-2)^2C_N^2 + 12(T-2)(B_NC_N - A_NC_N + 4B_N^2) - (3C_N + 6B_N)(T-2)(T-3)C_N$.

It is easily seen that $\text{plim}_{N\to\infty} A_N = \frac{1}{6}(T+1)\sigma^2$, $\text{plim}_{N\to\infty} B_N = -\frac{1}{2}\sigma^2$ and $\text{plim}_{N\to\infty} C_N = \sigma^2$. It follows that $\text{plim}_{N\to\infty}(6A_N + \frac{1}{2}(T-2)(T-3)C_N - 4(T-2)B_N) = 3(T-1)\sigma^2 + \frac{1}{2}(T-2)(T-3)\sigma^2 > 0$, $\text{plim}_{N\to\infty}(6A_N + 6B_N - (T-2)C_N) = 0$ and $\text{plim}_{N\to\infty} D_N = 0$ so that $\text{plim}_{N\to\infty} z_{1,2} = 0$, as predicted by lemma 2. However, in finite samples of any size we can have $D_N < 0$, so that $\widetilde{\kappa}(z) = 0$ does not have a real solution. $\Pr(D_N < 0)$ does not tend to zero when $N \to \infty$. We conclude that $\lim_{N\to\infty} \Pr(\widetilde{\Theta}_N^c = \varnothing) > 0$. $\square$

Proof of theorem 1:

Lemma 1 implies that $\widehat{\theta}_W$ and $\widehat{\theta}_C$ are uniquely defined when they exist. When $\widetilde{l}_N^c(r)$ and $\widetilde{l}_N(\theta)$ have local maxima on the interval $[-1, \infty)$ and the set $\widetilde{\Omega}$, respectively, this follows from lemma 1. We will now turn to the other claims of the theorem. To prove the consistency claims, we will verify the conditions of theorem 2.1 in Newey and McFadden (NMcF, 1994). To simplify matters and following NMcF, we will simply assume that the parameter space for θ is a very large compact subset of $\widetilde{\Omega}$, viz. $\overline{\Omega} = \overline{\Omega}_\rho \times \overline{\Omega}_{\sigma^2} \times \overline{\Omega}_\beta$ where $\overline{\Omega}_\rho = [-1, \rho_u]$ and $\overline{\Omega}_{\sigma^2} = [1/\sigma_u, \sigma_u]$ for some very large $\rho_u, \sigma_u \in \mathbb{R}^+$ and $\overline{\Omega}_\beta$ is a very large compact subset of $\mathbb{R}^K$.

We will first prove the claims for $\widehat{\theta}_C$. Using that $\frac{1}{(T-1)}\left|\frac{\partial \widetilde{l}_N^c(r)}{\partial r} - \frac{\partial \widetilde{l}^c(r)}{\partial r}\right| =$

$$\left|\frac{(r-\widehat{\rho}_{ML})s_y^2(1-r_{xy_{-1}}^2)\{\sigma_{ML}^2+(r-\rho_{ML})^2\sigma_y^2(1-\rho_{xy_{-1}}^2)\}-(r-\rho_{ML})\sigma_y^2(1-\rho_{xy_{-1}}^2)\{\widehat{\sigma}_{ML}^2+(r-\widehat{\rho}_{ML})^2 s_y^2(1-r_{xy_{-1}}^2)\}}{(\widehat{\sigma}_{ML}^2+(r-\widehat{\rho}_{ML})^2 s_y^2(1-r_{xy_{-1}}^2))(\sigma_{ML}^2+(r-\rho_{ML})^2\sigma_y^2(1-\rho_{xy_{-1}}^2))}\right| \leq$$

$$\left|\frac{(r-\widehat{\rho}_{ML})s_y^2(1-r_{xy_{-1}}^2)\{\sigma_{ML}^2+(r-\rho_{ML})^2\sigma_y^2(1-\rho_{xy_{-1}}^2)\}-(r-\rho_{ML})\sigma_y^2(1-\rho_{xy_{-1}}^2)\{\widehat{\sigma}_{ML}^2+(r-\widehat{\rho}_{ML})^2 s_y^2(1-r_{xy_{-1}}^2)\}}{\widehat{\sigma}_{ML}^2\sigma_{ML}^2}\right| \equiv$$

$|U(r)|$ and noting that the terms in the numerator of $U(r)$ are polynomials in r and that $\sigma_{ML}^2 = \operatorname{plim}_{N\to\infty}\widehat{\sigma}_{ML}^2 > 0$, it follows from $\operatorname{plim}_{N\to\infty} U(r) = 0\ \forall r \in \overline{\Omega}_\rho$ that $\operatorname{plim}_{N\to\infty}\sup_{r\in\overline{\Omega}_\rho}|U(r)| = 0$ and hence $\operatorname{plim}_{N\to\infty}\sup_{r\in\overline{\Omega}_\rho}\left|\frac{\partial \widetilde{l}_N^c(r)}{\partial r} - \frac{\partial \widetilde{l}^c(r)}{\partial r}\right| = 0$. In a similar way it can be shown that $\operatorname{plim}_{N\to\infty}\sup_{r\in\overline{\Omega}_\rho}\left|\left(\frac{\partial \widetilde{l}_N^c(r)}{\partial r}\right)^2 - \left(\frac{\partial \widetilde{l}^c(r)}{\partial r}\right)^2\right| = 0$ and $\operatorname{plim}_{N\to\infty}\sup_{r\in\overline{\Omega}_\rho}\left|\frac{\partial^2 \widetilde{l}_N^c(r)}{\partial r^2} - \frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}\right| = 0$. Also the polynomials $\left(\frac{\partial \widetilde{l}^c(r)}{\partial r}\right)^2$ and $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}$ are continuous on $\overline{\Omega}_\rho$. Next we need to distinguish between two cases, $\rho \neq 1$ and $\rho = 1$:

When $\rho \neq 1$, $\lim_{N\to\infty}\Pr(\widetilde{\Theta}_N^c = \varnothing) = 0$, where $\widetilde{\Theta}_N^c$ is the set of roots of $\frac{\partial \widetilde{l}_N^c}{\partial r} = 0$ corresponding to local maxima of $\widetilde{l}_N^c$ on the interval $[-1, \infty)$. Furthermore, it follows from lemma 2 above and theorem 2.1 in NMcF that $\widehat{\rho}_C$ converges in probability to ρ, which corresponds to a unique local maximum of $\widetilde{l}^c(r)$ on $\overline{\Omega}_\rho$. It then follows straightforwardly that $\widehat{\theta}_C$ $(= (\widehat{\rho}_C, \widehat{\sigma}^2(\widehat{\rho}_C, \widehat{\beta}(\widehat{\rho}_C)), \widehat{\beta}(\widehat{\rho}_C))'$) exists w.p.a.1. and is consistent.

When $\rho = 1$, $\lim_{N\to\infty}\Pr(\widetilde{\Theta}_N^c = \varnothing) > 0$ by lemma 4, notwithstanding that $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_\rho = 0$. However, because $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_\rho = 0$, $\frac{\partial^2 \widetilde{l}^c(r)}{\partial r^2}|_\rho = 0$ and $\frac{\partial^3 \widetilde{l}^c(r)}{\partial r^3}|_\rho > 0$ ($\frac{\partial \widetilde{l}^c(r)}{\partial r}|_{\rho-} > 0$ and $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_{\rho+} > 0$, cf. proof of lemma 2), we have that $\widehat{\rho}_C$ exists w.p.a.1. and by lemma 2 above and theorem 2.1 in NMcF that $\widehat{\rho}_C$ converges in probability to ρ, which is a unique solution of $\frac{\partial \widetilde{l}^c(r)}{\partial r}|_\rho = 0$ on $\overline{\Omega}_\rho$. It follows that also when $\rho = 1$, $\widehat{\theta}_C$ exists w.p.a.1. and is consistent.

We now proceed to prove the claims for $\widehat{\theta}_W$. To prove consistency of $\widehat{\theta}_W$ we will make use of theorem 2.6 in NMcF to verify the conditions of their theorem 2.1.

Let $\widetilde{l}_{N,i}(\theta) = (T-1)\xi(r) - 0.5(T-1)\log s^2 - 0.5s^{-2}(y_i - ry_{i,-1} - X_i b)'Q(y_i - ry_{i,-1} - X_i b)$. In the notation of NMcF $\partial\widetilde{l}_{N,i}(\theta)/\partial\theta = g(z_i, \theta)$. We assume that $\theta_0 \in \overline{\Omega}$, which is compact.

It is easily checked that $\partial\widetilde{l}_{N,i}(\theta)/\partial\theta$ is continuous at each $\theta_0 \in \overline{\Omega}$ w.p.1. Furthermore, $E(\sup_{\theta\in\overline{\Omega}}(g(z_i,\theta)'g(z_i,\theta))) < \infty$. To complete the proof, we again need to distinguish between two cases, $\rho \neq 1$ and $\rho = 1$:

When $\rho \neq 1$, $\lim_{N\to\infty} \Pr(\widetilde{\Theta}_N = \varnothing) = 0$, where $\widetilde{\Theta}_N$ is the set of roots of $\frac{\partial\widetilde{l}_N}{\partial\theta} = 0$ corresponding to local maxima of $\widetilde{l}_N$ on the set $\widetilde{\Omega}$. Furthermore, it follows from lemma 2 above and theorem 2.6 in NMcF that $\widehat{\theta}_W$ converges in probability to θ_0, which corresponds to a unique local maximum of $\widetilde{l}(\theta)$ on $\overline{\Omega}$. Thus $\widehat{\theta}_W$ exists w.p.a.1. and is consistent.

When $\rho = 1$, $\lim_{N\to\infty} \Pr(\widetilde{\Theta}_N = \varnothing) > 0$ by lemma 4, notwithstanding that $\frac{\partial\widetilde{l}(r)}{\partial\theta}|_{\theta_0} = 0$. However, because $\frac{\partial\widetilde{l}(\theta)}{\partial\theta}|_{\theta_0} = 0$, $x'\left(\frac{\partial^2\widetilde{l}(\theta)}{\partial\theta\partial\theta'}|_{\theta_0}\right)x \leq 0\ \forall x \in \mathbb{R}^{2+K}$, $\det\left(\frac{\partial^2\widetilde{l}(\theta)}{\partial\theta\partial\theta'}|_{\theta_0}\right) = 0$, $\frac{\partial^2\widetilde{l}^c(r)}{\partial r^2}|_\rho = 0$ and $\frac{\partial^3\widetilde{l}^c(r)}{\partial r^3}|_\rho > 0$ ($\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho-} > 0$ and $\frac{\partial\widetilde{l}^c(r)}{\partial r}|_{\rho+} > 0$, cf. proof of lemma 2), we have that $\widehat{\theta}_W$ exists w.p.a.1. and by lemma 2 above and theorem 2.6 in NMcF that $\widehat{\theta}_W$ converges in probability to θ_0, which is a unique solution of $\frac{\partial\widetilde{l}^c(\theta)}{\partial\theta}|_{\theta_0} = 0$ on $\overline{\Omega}$. Thus also when $\rho = 1$, $\widehat{\theta}_W$ exists w.p.a.1. and is consistent.

The proofs of the claims for $\widehat{\theta}_F$ are similar. $\square$

Derivation of the minimum rate of convergence of $\widehat{\rho}_C$:

We first state some preliminary results. Let $\breve{\sigma}^2 = \widehat{\sigma}^2(1, \widehat{\beta}(1))$. Note that $\widehat{\beta}(1) = (\sum_{i=1}^N X_i'QX_i)^{-1}\sum_{i=1}^N X_i'Q\varepsilon_i$ and $\breve{\sigma}^2 = \frac{1}{(T-1)N}\sum_{i=1}^N(\varepsilon_i - X_i\widehat{\beta}(1))'Q(\varepsilon_i - X_i\widehat{\beta}(1))$. Let $\frac{\partial^j\widetilde{l}_N^c(1)}{\partial r^j} = \frac{\partial^j\widetilde{l}_N^c(r)}{\partial r^j}|_{r=1}$ and $\Phi = \Phi(1)$. Then

$$\frac{\partial\widetilde{l}_N^c(1)}{\partial r} = \breve{\sigma}^{-2}[(T-1)\xi'(1)(\breve{\sigma}^2 - \sigma^2) + (T-1)\xi'(1)\sigma^2 + N^{-1}\sum_{i=1}^N(\varepsilon_i - X_i\widehat{\beta}(1))'Q\Phi\varepsilon_i] \tag{29}$$

and

$$\frac{\partial^2\widetilde{l}_N^c(1)}{\partial r^2} = (T-1)\xi''(1) - \breve{\sigma}^{-2}N^{-1}\sum_{i=1}^N(\Phi\varepsilon_i + X_i\frac{\partial\widehat{\beta}(1)}{\partial r})'Q\Phi\varepsilon_i - \breve{\sigma}^{-4}\frac{\partial\widehat{\sigma}^2(1)}{\partial r}N^{-1}\sum_{i=1}^N(\varepsilon_i - X_i\widehat{\beta}(1))'Q\Phi\varepsilon_i,$$

where $\frac{\partial\widehat{\beta}(1)}{\partial r} = -(\sum_{i=1}^N X_i'QX_i)^{-1}\sum_{i=1}^N X_i'Q\Phi\varepsilon_i$ and $\frac{\partial\widehat{\sigma}^2(1)}{\partial r} = -2(T-1)^{-1}N^{-1}\sum_{i=1}^N(\varepsilon_i - X_i\widehat{\beta}(1))'Q(\Phi\varepsilon_i + X_i\frac{\partial\widehat{\beta}(1)}{\partial r})$.

Clearly $N^{1/2}\left(\frac{\partial\widetilde{l}_N^c(1)}{\partial r}\right) = O_p(1)$. Recall that $\xi''(1) - (T-1)^{-1}tr(\Phi'Q\Phi) + 2(\xi'(1))^2 = 0$. Therefore we also have $N^{1/2}\left(\frac{\partial^2\widetilde{l}_N^c(1)}{\partial r^2}\right) = O_p(1)$. Finally, we have

$$\begin{aligned}
\mathrm{p}\lim_{N\to\infty}\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3} &= \frac{\partial^3\widetilde{l}^c(1)}{\partial r^3} = \frac{T(T-1)(T+1)}{12} > 0,\\
\xi''''(1) &= \frac{1}{20}(T-2)(T-3)(T-4),\\
\mathrm{p}\lim_{N\to\infty}\frac{\partial^4\widetilde{l}_N^c(1)}{\partial r^4} &= \frac{\partial^4\widetilde{l}^c(1)}{\partial r^4} = (T-1)\,\xi''''(1) + \frac{1}{6}(T-1)(T^2-10T+7) \neq 0,\\
\xi'''''(1) &= \frac{1}{30}(T-2)(T-3)(T-4)(T-5) \quad \text{and}\\
\mathrm{p}\lim_{N\to\infty}\frac{\partial^5\widetilde{l}_N^c(1)}{\partial r^5} &= \frac{\partial^5\widetilde{l}^c(1)}{\partial r^5} = (T-1)\,\xi'''''(1) - \frac{1}{3}(T-1)(5T^2-20T+11) \neq 0.
\end{aligned}$$

We now derive the minimum rate of convergence of $\widehat{\rho}_C$. W.p.a.1 $\widehat{\rho}_C$ is a solution of the f.o.c. $\frac{\partial^2\widetilde{l}_N^c(r)}{\partial r^2}\frac{\partial\widetilde{l}_N^c(r)}{\partial r} = 0$. Let $G_N^c(r) = N^{3/4}\frac{\partial^2\widetilde{l}_N^c(r)}{\partial r^2}\frac{\partial\widetilde{l}_N^c(r)}{\partial r}$. Forming a Taylor expansion of $G_N^c(\widehat{\rho}_C)$ around $r=1$ gives that $\widehat{\rho}_C$ must solve

$$0 = G_N^c(1) + \sum_{j=1}^3\frac{1}{j!}\frac{\partial^j G_N^c(1)}{\partial r^j}(r-1)^j + P_{1,N}(N^{1/4}(r-1)),$$

where $P_{1,N}(N^{1/4}(r-1))$ is a polynomial in $N^{1/4}(r-1)$ with coefficients that are $o_p(1)$. That is, $\widehat{\rho}_C$ must solve

$$\begin{aligned}
0 \;=\;& N^{-1/4}N\frac{\partial^2\widetilde{l}_N^c(1)}{\partial r^2}\frac{\partial\widetilde{l}_N^c(1)}{\partial r} +\\
& N^{1/2}\left(\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3}\frac{\partial\widetilde{l}_N^c(1)}{\partial r} + \left(\frac{\partial^2\widetilde{l}_N^c(1)}{\partial r^2}\right)^2\right)N^{1/4}(r-1) +\\
& \frac{1}{2}N^{-1/4}N^{1/2}\left(\frac{\partial^4\widetilde{l}_N^c(1)}{\partial r^4}\frac{\partial\widetilde{l}_N^c(1)}{\partial r} + 3\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3}\frac{\partial^2\widetilde{l}_N^c(1)}{\partial^2 r}\right)N^{1/2}(r-1)^2 +\\
& \frac{1}{3!}\left(\frac{\partial^5\widetilde{l}_N^c(1)}{\partial r^5}\frac{\partial\widetilde{l}_N^c(1)}{\partial r} + 4\frac{\partial^4\widetilde{l}_N^c(1)}{\partial r^4}\frac{\partial^2\widetilde{l}_N^c(1)}{\partial r^2} + 3\left(\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3}\right)^2\right)N^{3/4}(r-1)^3 +\\
& P_{1,N}(N^{1/4}(r-1))
\end{aligned}$$

or equivalently

$$0 = N^{1/2}\left(\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3}\frac{\partial \widetilde{l}_N^c(1)}{\partial r}\right)N^{1/4}(r-1)+ \tag{30}$$
$$\frac{1}{2}\left(\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial^3 r}\right)^2 N^{3/4}(r-1)^3+P_{2,N}(N^{1/4}(r-1)),$$

where $P_{2,N}(N^{1/4}(r-1))$ is another polynomial in $N^{1/4}(r-1)$ with coefficients that are $o_p(1)$. It follows that $N^{1/4}(\widehat{\rho}_C-1)=O_p(1)$, i.e., the rate of convergence of $\widehat{\rho}_C$ is at least $N^{1/4}$.

Proof of theorem 2 and corollary 1:

We will show in the proof below that $Z_{1,N}\xrightarrow{d} Z_1 \sim N(0,48T^{-2}((T-1)(T+1))^{-1})$ and that there exists a sequence $\{U_N\}$ with $U_N=O_p(N^{-1/2})$ such that if $Z_{1,N}+U_N>0$, then $M_N^c(r)$, which is given in (13), has two local minima attained at values $\widetilde{\rho}$ such that $N^{1/2}(\widetilde{\rho}-1)^2=Z_{1,N}+o_p(1)$, whereas if $Z_{1,N}+U_N<0$, then $M_N^c(r)$ has one local minimum attained at $r=\widehat{\rho}$ with $N^{1/2}(\widehat{\rho}-1)^2=o_p(1)$. Furthermore, if $Z_{1,N}+U_N>0$, then the sign of $N^{1/4}(\widehat{\rho}-1)$ is determined by the remainder $R_{1,N}^c(N^{1/4}(\widehat{\rho}-1))$ in (13). We first examine this remainder:

$$N^{1/4}R_{1,N}^c(N^{1/4}(\widehat{\rho}-1))=N^{1/4}(\widehat{\rho}-1)R_{2,N}^c(N^{1/2}(\widehat{\rho}-1)^2)+R_{3,N}^c(N^{1/4}(\widehat{\rho}-1))$$

where

$$R_{2,N}^c(N^{1/2}(\widehat{\rho}-1)^2)=2N\frac{\partial^2 \widetilde{l}_N^c(1)}{\partial r^2}\frac{\partial \widetilde{l}_N^c(1)}{\partial r}+N^{1/2}\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3}\frac{\partial^2 \widetilde{l}_N^c(1)}{\partial r^2}N^{1/2}(\widehat{\rho}-1)^2+$$
$$\frac{1}{3}N^{1/2}\frac{\partial^4 \widetilde{l}_N^c(1)}{\partial r^4}\frac{\partial \widetilde{l}_N^c(1)}{\partial r}N^{1/2}(\widehat{\rho}-1)^2+\frac{1}{6}\frac{\partial^4 \widetilde{l}_N^c(1)}{\partial r^4}\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial^3 r}N(\widehat{\rho}-1)^4 \quad \text{and}$$

$$R_{3,N}^c(N^{1/4}(\widehat{\rho}-1))=o_p(1).$$

However, if $N^{1/2}(\widehat{\rho}-1)^2=Z_{1,N}+o_p(1)$, then $R_{2,N}^c(N^{1/2}(\widehat{\rho}-1)^2)=o_p(1)$. Therefore we need to consider $R_{3,N}^c(N^{1/4}(\widehat{\rho}-1))$. We have

$$N^{1/2}R_{3,N}^c(N^{1/4}(\widehat{\rho}-1))=N^{1/4}R_{4,N}^c(N^{1/2}(\widehat{\rho}-1)^2)+$$
$$N^{5/4}(\widehat{\rho}-1)^5R_{5,N}^c(N^{1/2}(\widehat{\rho}-1)^2)+R_{6,N}^c(N^{1/4}(\widehat{\rho}-1))$$

where

$$R^c_{4,N}(N^{1/2}(\widehat{\rho}-1)^2) = (N^{1/2}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2})^2 N^{1/2}(\widehat{\rho}-1)^2 + (\frac{8}{4!}N^{1/2}\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2} + \frac{2}{4!}N^{1/2}\frac{\partial^5\widetilde{l}^c_N(1)}{\partial r^5}\frac{\partial\widetilde{l}^c_N(1)}{\partial r})N(\widehat{\rho}-1)^4 + (\frac{30}{6!}\frac{\partial^5\widetilde{l}^c_N(1)}{\partial r^5}\frac{\partial^3\widetilde{l}^c_N(1)}{\partial r^3} + \frac{20}{6!}(\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4})^2)N^{3/2}(\widehat{\rho}-1)^6 + o_p(1),$$

$$R^c_{5,N}(N^{1/2}(\widehat{\rho}-1)^2) = \frac{2}{5!}N^{1/2}\frac{\partial^6\widetilde{l}^c_N(1)}{\partial r^6}\frac{\partial\widetilde{l}^c_N(1)}{\partial r} + \frac{10}{5!}N^{1/2}\frac{\partial^5\widetilde{l}^c_N(1)}{\partial r^5}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2} + \frac{1}{5!}\frac{\partial^6\widetilde{l}^c_N(1)}{\partial r^6}\frac{\partial^3\widetilde{l}^c_N(1)}{\partial r^3}N^{1/2}(\widehat{\rho}-1)^2 + \frac{10}{6!}\frac{\partial^5\widetilde{l}^c_N(1)}{\partial r^5}\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4}N^{1/2}(\widehat{\rho}-1)^2 \quad \text{and}$$

$$R^c_{6,N}(N^{1/4}(\widehat{\rho}-1)) = o_p(1).$$

If $N^{1/2}(\widehat{\rho}-1)^2 = -2(\frac{\partial^3\widetilde{l}^c_N(1)}{\partial r^3})^{-1}(N^{1/2}\frac{\partial\widetilde{l}^c_N(1)}{\partial r}) + o_p(1)$, then

$$R^c_{4,N}(N^{1/2}(\widehat{\rho}-1)^2) = (N^{1/2}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2})^2 N^{1/2}(\widehat{\rho}-1)^2 + \frac{1}{3}N^{1/2}\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2}N(\widehat{\rho}-1)^4 + \frac{1}{36}(\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4})^2 N^{3/2}(\widehat{\rho}-1)^6 + o_p(1).$$

It follows that if $Z_{1,N} + U_N > 0$, then the value of $M^c_N(r)$ is in fact minimized at

$$N^{1/2}(\widehat{\rho}-1)^2 = Z_{1,N} + N^{-1/2}R^c_{7,N} + o_p(N^{-1/2})$$

where

$$R^c_{7,N} \equiv (\frac{\partial^3\widetilde{l}^c_N(1)}{\partial r^3})^{-2}(-2(N^{1/2}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2})^2 - \frac{4}{3}(\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4})(N^{1/2}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2})Z_{1,N} - \frac{1}{6}(\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4})^2 Z^2_{1,N}),$$

and

$$R^c_{2,N}(N^{1/2}(\widehat{\rho}-1)^2) = (N^{1/2}\frac{\partial^3\widetilde{l}^c_N(1)}{\partial r^3}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2} + \frac{1}{6}\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4}\frac{\partial^3\widetilde{l}^c_N(1)}{\partial r^3}Z_{1,N}) \times N^{-1/2}R^c_{7,N} + o_p(N^{-1/2}).$$

We also have $R^c_{4,N}(N^{1/2}(\widehat{\rho}-1)^2) = O_p(1)$ and

$$R^c_{5,N}(N^{1/2}(\widehat{\rho}-1)^2) = \frac{10}{5!}N^{1/2}\frac{\partial^5\widetilde{l}^c_N(1)}{\partial r^5}\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2} + \frac{10}{6!}\frac{\partial^5\widetilde{l}^c_N(1)}{\partial r^5}\frac{\partial^4\widetilde{l}^c_N(1)}{\partial r^4}Z_{1,N} + o_p(1).$$

We conclude that if $Z_{1,N} + U_N > 0$, then $N^{1/2}(\widehat{\rho}-1)^2 = Z_{1,N} + O_p(N^{-1/2})$,

$$\begin{aligned} &R^c_{1,N}(N^{1/4}(\widehat{\rho}-1)) = \\ &\quad N^{-1/2}R^c_{4,N}(N^{1/2}(\widehat{\rho}-1)^2) + N^{1/4}(\widehat{\rho}-1)N^{-3/4}R^c_N(N^{1/2}(\widehat{\rho}-1)^2) + o_p(N^{-3/4}) \end{aligned}$$

where

$$R^c_N(N^{1/2}(\widehat{\rho}-1)^2) \equiv N^{1/2}R^c_{2,N}(N^{1/2}(\widehat{\rho}-1)^2) + Z^2_{1,N}R^c_{5,N}(N^{1/2}(\widehat{\rho}-1)^2) + o_p(1),$$

and

$$sgn(N^{1/4}(\widehat{\rho}-1)) = sgn(-R^c_N(N^{1/2}(\widehat{\rho}-1)^2)).$$

It also follows that $U_N = O_p(N^{-1/2})$.

If $Z_{1,N} + U_N < 0$, then $M^c_N(r)$ has one local minimum and its value is minimized at $N^{1/2}(\widehat{\rho}-1)^2 = o_p(1)$.

Next we derive the limiting distributions of $N^{1/2}\left(\frac{\partial\widetilde{l}^c_N(1)}{\partial r}\right)$ and $N^{1/2}\left(\frac{\partial^2\widetilde{l}^c_N(1)}{\partial r^2}\right)$. Using that under normality of ε_i for any constant $T\times T$ matrices M_1 and M_2, $E(\varepsilon_i'M_1\varepsilon_i\varepsilon_i'M_2\varepsilon_i) = \sigma^4(tr(M_1)tr(M_2) + tr(M_1M_2 + M_1'M_2))$, we find that

$$\begin{aligned} N^{1/2}(\breve{\sigma}^2-\sigma^2)/\sigma^2 &\xrightarrow{d} V_1 \sim N(0, 2/(T-1)), \\ N^{1/2}(N^{-1}\sum_{i=1}^N \varepsilon_i'\Phi'Q\Phi\varepsilon_i - \sigma^2 tr(\Phi'Q\Phi))/\sigma^2 &\xrightarrow{d} V_2 \sim N(0, 2tr(\Phi'Q\Phi\Phi'Q\Phi)), \\ N^{1/2}(N^{-1}\sum_{i=1}^N \varepsilon_i'Q\Phi\varepsilon_i - \sigma^2 tr(Q\Phi))/\sigma^2 &\xrightarrow{d} V_3 \sim N(0, (tr(Q\Phi Q\Phi) + tr(\Phi'Q\Phi))), \end{aligned}$$

and

$$\begin{aligned} E(V_1V_2) &= 2(T-1)^{-1}tr(Q\Phi'Q\Phi) = \frac{1}{6}(T+1), \\ E(V_1V_3) &= 2(T-1)^{-1}tr(Q\Phi) = -1, \quad \text{and} \\ E(V_2V_3) &= 2tr(\Phi'Q\Phi Q\Phi) = -\frac{1}{6}(T-1)(T+1). \end{aligned}$$

It is now easily seen that

$$\begin{aligned}
N^{1/2}\left(\frac{\partial \widetilde{l}_N^c(1)}{\partial r}\right) \xrightarrow{d} V_4 &\equiv (T-1)\xi'(1)V_1 + V_3 = \frac{1}{2}(T-1)V_1 + V_3, \\
N^{1/2}\left(\frac{\partial^2 \widetilde{l}_N^c(1)}{\partial r^2}\right) \xrightarrow{d} V_5 &\equiv (2(T-1)\xi''(1) - tr(\Phi' Q \Phi))V_1 - V_2 - 4\xi'(1)V_3 = \\
&\quad \frac{1}{6}(T-1)(T-5)V_1 - V_2 - 2V_3 \quad \text{and} \\
Z_{1,N} \xrightarrow{d} Z_1 &\equiv \left(-\frac{1}{2}\frac{\partial^3 \widetilde{l}^c(1)}{\partial r^3}\right)^{-1} V_4 = -24\left(T(T-1)(T+1)\right)^{-1} V_4,
\end{aligned}$$

where we have used that $\xi'(1) = \frac{1}{2}$, $\xi''(1) = \frac{1}{6}(T-2)$, $tr(\Phi' Q \Phi) = \frac{1}{6}(T-1)(T+1)$ and $\frac{\partial^3 \widetilde{l}^c(1)}{\partial r^3} = \frac{T(T-1)(T+1)}{12}$. Clearly

$$Z_{2,N} = N^{1/2}(\breve{\sigma}^2 - \sigma^2) \xrightarrow{d} Z_2 \equiv \sigma^2 V_1.$$

By using the delta method, we obtain

$$N^{1/2}(\widehat{\sigma}_n^2 - \sigma^2) = N^{1/2}(\widehat{\sigma}^2/\widehat{\rho} - \sigma^2) = N^{1/2}((\widehat{\sigma}^2 - \sigma^2) - \sigma^2(\widehat{\rho} - 1)) + o_p(1).$$

Noting that

$$\begin{aligned}
N^{1/2}(\widehat{\sigma}^2 - \sigma^2) &= Z_{2,N} + 2N^{-1/2}(1-\widehat{\rho})(T-1)^{-1}\sum_{i=1}^N \varepsilon_i' Q \Phi \varepsilon_i + \\
&\quad N^{-1/2}(\widehat{\rho}-1)^2(T-1)^{-1} \times \sum_{i=1}^N \varepsilon_i' \Phi' Q \Phi \varepsilon_i + o_p(1) \quad \text{and} \\
K_+ &= \sigma^2 tr(\Phi' Q \Phi)/(T-1),
\end{aligned}$$

we find that

$$N^{1/2}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{d} (Z_2 + K_+ Z_1) \times \mathbf{1}\{Z_1 > 0\} + Z_2 \times \mathbf{1}\{Z_1 \leq 0\}.$$

It also follows that

$$N^{1/4}(\widehat{\sigma}_C^2 - \sigma^2) - N^{1/4}(\widehat{\rho}_C - 1)\sigma^2 = o_p(1).$$

Finally, it is easily seen that

$$Z_{3,N} \xrightarrow{d} Z_3 \sim N(0, \sigma^2(\Sigma_{xqx})^{-1}),$$

that $E(Z_1Z_2) = E(Z_1Z_3) = E(Z_2Z_3) = 0$, and that $N^{1/2}R^c_{2,N}(Z_{1,N} + U_N) \xrightarrow{d} R^c_2$, $R^c_{5,N}(Z_{1,N} + U_N) \xrightarrow{d} R^c_5$ and $R^c_N(Z_{1,N} + U_N) \xrightarrow{d} R^c$ for some R^c_2, R^c_5 and R^c.

Next let

$$\tilde{R}^c_2 \equiv \left(\frac{\partial^3 \widetilde{l}^c_N(1)}{\partial r^3}\right)^{-1} \left(-2V_5^3 - \frac{5}{3}\frac{\partial^4 \widetilde{l}^c(1)}{\partial r^4}V_5^2 Z_1 - \frac{7}{18}(\frac{\partial^4 \widetilde{l}^c(1)}{\partial r^4})^2 V_5 Z_1^2 - \frac{1}{36}(\frac{\partial^4 \widetilde{l}^c(1)}{\partial r^4})^3 Z_1^3\right)$$

and

$$\tilde{R}^c_5 \equiv \frac{1}{12}\frac{\partial^5 \widetilde{l}^c(1)}{\partial r^5}(V_5 + \frac{1}{6}\frac{\partial^4 \widetilde{l}^c(1)}{\partial r^4}Z_1).$$

When $Z_1 > 0$, $R^c_2 = \tilde{R}^c_2$, $R^c_5 = \tilde{R}^c_5$ and

$$R^c = \tilde{R}^c_2 + Z_1^2 \tilde{R}^c_5. \tag{31}$$

Noting that $tr(Q\Phi Q\Phi) = -\frac{1}{12}(T-1)(T-5)$ and $tr(\Phi' Q\Phi\Phi' Q\Phi) = \frac{1}{180}(2T^4 + 5T^2 - 7)$, we have

$$\begin{aligned} V_4 &\sim N(0, \frac{1}{12}(T-1)(T+1)), \\ V_5 &\sim N(0, \frac{1}{90}(T-1)(T+1)(2T^2+7)) \quad \text{and} \\ & E(V_4 V_5) = -\frac{1}{12}(T-1)(T+1). \end{aligned}$$

We can decompose V_5 as $V_5 = -V_4 + V_0$ so that $E(V_0V_4) = 0$ and

$$V_0 \sim N(0, \frac{1}{180}(T-1)(T+1)(4T^2-1)).$$

Let

$$\begin{aligned} \kappa &= -2\left(\frac{\partial^3 \widetilde{l}^c_N(1)}{\partial r^3}\right)^{-1} \frac{\partial^4 \widetilde{l}^c(1)}{\partial r^4}, \\ \tilde{\kappa}_0(T) &= 2 - \frac{5}{3}\kappa + \frac{7}{18}\kappa^2 - \frac{1}{36}\kappa^3 + \frac{1}{18}\left(\frac{\partial^3 \widetilde{l}^c_N(1)}{\partial r^3}\right)^{-1} \frac{\partial^5 \widetilde{l}^c(1)}{\partial r^5}(\kappa - 6) \quad \text{and} \\ \tilde{\kappa}_2(T) &= 6 - \frac{5}{3}\kappa. \end{aligned}$$

Then we have

$$\tilde{R}_2^c + Z_1^2 \tilde{R}_5^c = (\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3})^{-1}(\tilde{\kappa}_0(T)V_4^3 + \tilde{\kappa}_1(T)V_4^2 V_0 + \tilde{\kappa}_2(T)V_4 V_0^2 + \tilde{\kappa}_3(T)V_0^3)$$

for some $\tilde{\kappa}_1(T)$ and $\tilde{\kappa}_3(T)$. It is easily verified that $\tilde{\kappa}_0(T) > 0$ and $\tilde{\kappa}_2(T) > 0$ for $T \geq 4$, $\tilde{\kappa}_0(2) = 0$, $\tilde{\kappa}_0(3) < 0$, $\tilde{\kappa}_2(2) < 0$ and $\tilde{\kappa}_2(3) < 0$. Furthermore, because V_0 is a Gaussian r.v. with mean zero, the conditional p.d.f. of $\tilde{\kappa}_1(T)V_4^2 V_0 + \tilde{\kappa}_3(T)V_0^3$ given V_4 (or equivalently, given Z_1) is symmetric around zero. Also, $V_0^2 \geq 0$. Noting that $B^c = \mathbf{1}(R^c > 0)$, it is now easily seen that $E((-1)^{B^c} Z_1^{1/2} | Z_1, Z_1 > 0) > 0$ when $T \geq 4$, while $E((-1)^{B^c} Z_1^{1/2} | Z_1, Z_1 > 0) < 0$ when $T = 2$ or $T = 3$. We can conclude that $E((-1)^{B^c} Z_1^{1/2} | Z_1 > 0) > 0$ when $T \geq 4$, while $E((-1)^{B^c} Z_1^{1/2} | Z_1 > 0) < 0$ when $T = 2$ or $T = 3$. $\square$

Derivation of the rate of convergence of $\widehat{\rho}_F$ and the limiting distribution of $\widehat{\theta}_F$ when $\rho = 1$:

Let

$$\begin{aligned}\Psi_{N,n}(\theta_n) &\equiv (\Psi_{\rho,N,n}(r), \Psi_{\sigma_n^2,N,n}(\theta_n), \Psi'_{\beta,N,n}(\theta_n))' = \\ &(\frac{\partial \widetilde{l}_N^c(r)}{\partial r}, s_n^2 r \frac{\partial \widetilde{l}_{N,n}(\theta_n)}{\partial s_n^2}, s_n^2 r \frac{\partial \widetilde{l}_{N,n}(\theta_n)}{\partial b'})' \quad \text{and}\end{aligned}$$

$$M_N(\theta_n) = N\left(\Psi'_{N,n}(\theta_n)\right) W_N\left(\Psi_{N,n}(\theta_n)\right).$$

Let G_N be an $(2+K) \times (2+K)$ matrix with

$$G_{N,1,1} = \frac{1}{2}\frac{\partial^3 \widetilde{l}_N^c(1)}{\partial r^3}, \quad G_{N,2,2} = \frac{\partial \Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial s_n^2}, \quad G_{N,2,3} = \frac{\partial \Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial b},$$
$$G_{N,3,3} = \frac{\partial \Psi_{\beta,N,n}(\theta_*)}{\partial b'}, \quad G_{N,2,1} = \frac{1}{2}\frac{\partial^2 \Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r^2}$$

and the other elements of G_N equal to zero. Note that $\operatorname{plim}_{N\to\infty} G_N = G$ has full rank. Similarly to the analysis for $M_N^c(r)$, we consider a Taylor expansion of $M_N(\theta_n)$ around $\theta_n = \theta_*$. Let $\widehat{\theta} = \widehat{\theta}_F$, $\widehat{\theta}_n = \widehat{\theta}_{n,F}$ and $\widehat{\omega}_n = ((\widehat{\rho}-1)^2, (\widehat{\sigma}_n^2 - \sigma^2), \widehat{\beta}')'$. Substituting $\widehat{\theta}_n$ for θ_n we obtain

$$\begin{aligned}M_N(\widehat{\theta}_n) &= N\left(\Psi'_{N,n}(\theta_*)\right) W_N\left(\Psi_{N,n}(\theta_*)\right) + 2N^{1/2}\Psi'_{N,n}(\theta_*)W_N G_N N^{1/2}\widehat{\omega}_n + \\ &\quad N^{1/2}\widehat{\omega}_n' G_N' W_N G_N N^{1/2}\widehat{\omega}_n + R_{1,N}(N^{1/4}(\widehat{\rho}-1)),\end{aligned} \tag{32}$$

where $R_{1,N}(N^{1/4}(\widehat{\rho}-1)) = o_p(1)$.

Let

$$W_N = \begin{bmatrix} W_{N,1,1} & \underline{W}_{N,1,2} \\ \underline{W}_{N,2,1} & \underline{W}_{N,2,2} \end{bmatrix} \text{ and } G_N = \begin{bmatrix} G_{N,1,1} & \underline{G}_{N,1,2} \\ \underline{G}_{N,2,1} & \underline{G}_{N,2,2} \end{bmatrix},$$

where $\underline{W}_{N,2,1} = \underline{W}'_{N,1,2}$, $\underline{G}_{N,2,1} = (G_{N,2,1}, 0')'$ and $\underline{G}'_{N,1,2} = 0$ are $(K+1)-$vectors, and let

$$\Psi_{N,n}(\theta_*) = (\Psi_{\rho,N,n}(1), \underline{\Psi}'_{N,n}(\theta_*))' \quad \text{and} \quad \widehat{\omega}_n = ((\widehat{\rho}-1)^2, \widehat{\underline{\omega}}'_n)'.$$

Then we have the following result:

Lemma 5 *There exists a sequence $\{\widetilde{U}_N\}$ with $\widetilde{U}_N = o_p(1)$ such that if $Z_{1,N}+\widetilde{U}_N > 0$, then the value of $M_N(\widehat{\theta}_n)$ in (32) is minimized at $N^{1/2}(\widehat{\rho}-1)^2 = Z_{1,N} + o_p(1)$ and $N^{1/2}\widehat{\underline{\omega}}_n = N^{1/2}\widehat{\underline{\omega}}_+$ where $N^{1/2}\widehat{\underline{\omega}}_+ = \underline{G}^{-1}_{N,2,2}\underline{G}_{N,2,1}G^{-1}_{N,1,1}N^{1/2}\Psi_{\rho,N,n}(1) - \underline{G}^{-1}_{N,2,2}N^{1/2}\underline{\Psi}_{N,n}(\theta_*) + o_p(1)$ (i.e., at $N^{1/2}\widehat{\omega}_n = -G_N^{-1}N^{1/2}\Psi_{N,n}(\theta_*) + o_p(1)$), whereas if $Z_{1,N} + \widetilde{U}_N < 0$, the value of $M_N(\widehat{\theta}_n)$ is minimized at $N^{1/2}(\widehat{\rho}-1)^2 = o_p(1)$ and $N^{1/2}\widehat{\underline{\omega}}_n = N^{1/2}\widehat{\underline{\omega}}_-$ where $N^{1/2}\widehat{\underline{\omega}}_- \equiv N^{1/2}\widehat{\underline{\omega}}_+ + K_{-,N}Z_{1,N}$ with $K_{-,N} \equiv \underline{G}^{-1}_{N,2,2}\underline{W}^{-1}_{N,2,2}\underline{W}_{N,2,1}G_{N,1,1} + \underline{G}^{-1}_{N,2,2}\underline{G}_{N,2,1}$ (i.e., at $N^{1/2}\widehat{\omega}_n = (o_p(1), N^{1/2}\widehat{\underline{\omega}}'_-)'$).*

Proof of lemma 5: Minimizing $M_N(\widehat{\theta}_n)$ given in (32) w.r.t. $N^{1/2}\widehat{\omega}_n$ is equivalent to minimizing

$$\widetilde{M}_N(\widehat{\theta}_n) \equiv 2(G_N^{-1}N^{1/2}\Psi_{N,n}(\theta_*))'\widetilde{W}_N N^{1/2}\widehat{\omega}_n + N^{1/2}\widehat{\omega}'_n\widetilde{W}_N N^{1/2}\widehat{\omega}_n + R_{1,N}(N^{1/4}(\widehat{\rho}-1))$$

w.r.t. $N^{1/2}\widehat{\omega}_n$, where $\widetilde{W}_N = G'_N W_N G_N$. Since W_N is PD and G_N has full rank, $\widetilde{W}_N$ is also PD. Partition $\widetilde{W}_N$ as $\begin{bmatrix} \widetilde{W}_{N,1,1} & \widetilde{W}_{N,1,2} \\ \widetilde{W}_{N,2,1} & \widetilde{W}_{N,2,2} \end{bmatrix}$ where $\widetilde{W}_{N,1,1}$ is a scalar and $\widetilde{W}_{N,2,1} = \widetilde{W}'_{N,1,2}$ is a $(K+1)-$vector. Given the value of $N^{1/2}(\widehat{\rho}-1)^2$, $\widetilde{M}_N(\widehat{\theta}_n)$ is minimized at $N^{1/2}\widehat{\underline{\omega}}_n = N^{1/2}\widehat{\underline{\omega}}_+ - \widetilde{W}^{-1}_{N,2,2}\widetilde{W}_{N,2,1}(N^{1/2}(\widehat{\rho}-1)^2 - Z_{1,N})$. Substituting this expression for $N^{1/2}\widehat{\underline{\omega}}_n$ in $\widetilde{M}_N(\widehat{\theta}_n)$ and noting that $G^{-1}_{N,1,1}N^{1/2}\Psi_{\rho,N,n}(1) = -Z_{1,N}$ gives

$$\widetilde{M}_N(\widehat{\theta}_n) = (-2Z_{1,N}N^{1/2}(\widehat{\rho}-1)^2 + N(\widehat{\rho}-1)^4)(\widetilde{W}_{N,1,1} - \widetilde{W}'_{N,2,1}\widetilde{W}^{-1}_{N,2,2}\widetilde{W}_{N,2,1}) + \widetilde{R}_N + o_p(1), \tag{33}$$

where $\widetilde{R}_N$ does not depend on $\widehat{\theta}_n$. Noting that $\mathrm{plim}_{N\to\infty}(\widetilde{W}_{N,1,1} - \widetilde{W}'_{N,2,1}\widetilde{W}^{-1}_{N,2,2}\widetilde{W}_{N,2,1}) > 0$ (because $\mathrm{plim}_{N\to\infty}\widetilde{W}_N$ is PD) and $\widetilde{W}^{-1}_{N,2,2}\widetilde{W}_{N,2,1} = \underline{G}^{-1}_{N,2,2}\underline{W}^{-1}_{N,2,2}\underline{W}_{N,2,1}G_{N,1,1} + \underline{G}^{-1}_{N,2,2}\underline{G}_{N,2,1}$ $= K_{-,N}$, the claims in the lemma follow straightforwardly. $\square$

Proof of theorem 3:

According to lemma 5, if $Z_{1,N}+\widetilde{U}_N>0$, then the value of $M_N(\widehat{\theta}_n)$ in (32) is minimized at $N^{1/2}(\widehat{\rho}-1)^2 = Z_{1,N}+o_p(1)$ and $N^{1/2}\underline{\widehat{\omega}}_n = N^{1/2}\underline{\widehat{\omega}}_+$. The sign of $N^{1/4}(\widehat{\rho}-1)$ is such that it minimizes the value of $R_{1,N}(N^{1/4}(\widehat{\rho}-1))$ in (32) where

$$\begin{aligned}
&N^{1/4}R_{1,N}(N^{1/4}(\widehat{\rho}-1)) = 2N^{1/2}(\Psi_{N,n}(\theta_*)+\frac{\partial\Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n)'W_N \times\\
&N^{1/2}(\frac{\partial\Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r\partial w_n'}\underline{\widehat{\omega}}_n)N^{1/4}(\widehat{\rho}-1)+\\
&(\frac{2}{3!}N^{1/2}(\Psi_{N,n}(\theta_*)+\frac{\partial\Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^3\Psi_{N,n}(\theta_*)}{\partial r^3}+\\
&\quad N^{1/2}(\frac{\partial\Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r\partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r^2})N^{3/4}(\widehat{\rho}-1)^3+\\
&\frac{1}{3!}\frac{\partial^3\Psi_{N,n}'(\theta_*)}{\partial r^3}W_N\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r^2}N^{5/4}(\widehat{\rho}-1)^5+o_p(1).
\end{aligned}\tag{34}$$

We can write (34) as $N^{1/4}R_{1,N}(N^{1/4}(\widehat{\rho}-1)) = N^{1/4}(\widehat{\rho}-1)R_{2,N}(N^{1/4}(\widehat{\rho}-1))+R_{3,N}(N^{1/4}(\widehat{\rho}-1))$ where $R_{3,N}(N^{1/4}(\widehat{\rho}-1)) = o_p(1)$. Noting that

$$\begin{aligned}
&N^{1/2}\Psi_{\sigma_n^2,N,n}(\theta_*)\xrightarrow{d}\frac{(T-1)}{2}V_1,\quad N^{1/2}\Psi_{\beta,N,n}(\theta_*)\xrightarrow{d}V_6\sim N(0,\sigma^2\Sigma_{xqx}),\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial\Psi_{N,n}(\theta_*)}{\partial w_n'} = \mathrm{p}\lim_{N\to\infty}(0,\frac{\partial\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial w_n},\frac{\partial\Psi_{\beta,N,n}'(\theta_*)}{\partial w_n})',\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial s_n^2} = -\frac{(T-1)}{2\sigma^2},\quad \mathrm{p}\lim_{N\to\infty}\frac{\partial\Psi_{\beta,N,n}'(\theta_*)}{\partial b} = -\Sigma_{xqx},\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial b} = \mathrm{p}\lim_{N\to\infty}\frac{\partial\Psi_{\beta,N,n}'(\theta_*)}{\partial s_n^2} = 0,\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r\partial w_n'} = \mathrm{p}\lim_{N\to\infty}(0,\frac{\partial^2\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r\partial w_n},0)',\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial^2\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r\partial s_n^2} = -\frac{(T-1)}{2\sigma^2},\quad \mathrm{p}\lim_{N\to\infty}\frac{\partial^2\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r\partial\beta} = 0,\\
&\mathrm{p}\lim_{N\to\infty}N^{-1}\sum\nolimits_{i=1}^{N}(X_i'Q\Phi\Phi'QX_i) = \Sigma_{xq\phi\phi'qx},\\
&N^{1/2}\frac{\partial\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r}\xrightarrow{d}-V_3,\quad N^{1/2}\frac{\partial\Psi_{\beta,N,n}(\theta_*)}{\partial r}\xrightarrow{d}V_7\sim N(0,\sigma^2\Sigma_{xq\phi\phi'qx}),\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial^4\widetilde{l}_N^c(1)}{\partial r^4} = \frac{\partial^4\widetilde{l}^c(1)}{\partial r^4} = \frac{1}{60}(T-1)(T+1)(3T^2-20T-2),\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial^3\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r^3} = 0,\quad \mathrm{p}\lim_{N\to\infty}\frac{\partial^3\Psi_{\beta,N,n}(\theta_*)}{\partial r^3} = 0,\\
&\mathrm{p}\lim_{N\to\infty}\frac{\partial^2\Psi_{\sigma_n^2,N,n}(\theta_*)}{\partial r^2} = tr(\Phi'Q\Phi)\quad\text{and}\quad \mathrm{p}\lim_{N\to\infty}\frac{\partial^2\Psi_{\beta,N,n}(\theta_*)}{\partial r^2} = 0,
\end{aligned}$$

and recalling that $N^{1/2}\frac{\partial\widetilde{l_N}^c(1)}{\partial r}\xrightarrow{d}V_4$, $N^{1/2}\frac{\partial^2\widetilde{l_N}^c(1)}{\partial r^2}\xrightarrow{d}V_5$ and $\text{plim}_{N\to\infty}\frac{\partial^3\widetilde{l}_N^c(1)}{\partial r^3}=\frac{\partial^3\widetilde{l}^c(1)}{\partial r^3}=\frac{1}{12}T(T-1)(T+1)$, in other words, noting that

$$
\begin{aligned}
&N^{1/2}\Psi_{N,n}(\theta_*)\xrightarrow{d}\Psi_n(\theta_*),\quad N^{1/2}\frac{\partial\Psi_{N,n}(\theta_*)}{\partial r}\xrightarrow{d}\Psi_{n,\rho}(\theta_*),\\
&p\lim_{N\to\infty}\frac{\partial\Psi_{N,n}(\theta_*)}{\partial w_n}=\Psi_{n,\omega}(\theta_*),\\
&p\lim_{N\to\infty}\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r\partial w_n'}=\Psi_{n,\rho\omega}(\theta_*),\\
&p\lim_{N\to\infty}\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r^2}=\Psi_{n,\rho\rho}(\theta_*)\quad\text{and}\\
&p\lim_{N\to\infty}\frac{\partial^3\Psi_{N,n}(\theta_*)}{\partial r^3}=\Psi_{n,\rho\rho\rho}(\theta_*),
\end{aligned}
$$

it follows that $\text{plim}_{N\to\infty}G_N=G$ and that if $Z_{1,N}+\widetilde{U}_N>0$, then $N^{1/2}(\widehat{\rho}-1)^2\xrightarrow{d}Z_1$, $N^{1/2}\widehat{\underline{\omega}}_n=N^{1/2}\widehat{\underline{\omega}}_+\xrightarrow{d}\underline{\omega}_+=-\underline{G}_{2,2}^{-1}\underline{G}_{2,1}Z_1-\underline{G}_{2,2}^{-1}\underline{\Psi}_n(\theta_*)=-(\Psi_{n,\omega}'(\theta_*)W\Psi_{n,\omega}(\theta_*))^{-1}\times\Psi_{n,\omega}'(\theta_*)W(\Psi_n(\theta_*)+\frac{1}{2}\Psi_{n,\rho\rho}(\theta_*)Z_1)$ and $R_{2,N}(N^{1/4}(\widehat{\rho}-1))\xrightarrow{d}R_2$, which after lengthy but simple calculations can be shown to obey:

$$
R_2=(Z_1\frac{\partial^3\widetilde{l}^c(1)}{\partial r^3}+2V_4)(W_{1,1}-\underline{W}_{1,2}\underline{W}_{2,2}^{-1}\underline{W}_{2,1})(\frac{1}{6}\frac{\partial^4\widetilde{l}^c(1)}{\partial r^4}Z_1+V_5).\tag{35}
$$

Let $\Psi_{n,\omega}=\Psi_{n,\omega}(\theta_*)$ and $\mathcal{M}=-(\Psi_{n,\omega}'W\Psi_{n,\omega})^{-1}\Psi_{n,\omega}'W$. To derive (35) we have used that $W(I+\Psi_{n,\omega}\mathcal{M})=diag(W_{1,1}-\underline{W}_{1,2}\underline{W}_{2,2}^{-1}\underline{W}_{2,1},0,...,0)$ and $W(I+\Psi_{n,\omega}\mathcal{M})\Psi_{n,\rho\omega}(\theta_*)=0$.

Recall that $Z_1=-(\frac{1}{2}\frac{\partial^3\widetilde{l}^c(1)}{\partial r^3})^{-1}V_4$. Hence $R_{2,N}=o_p(1)$. This result also follows directly from (34) and $N^{1/2}(\Psi_{N,n}(\theta_*)+\frac{\partial\Psi_{N,n}(\theta_*)}{\partial w_n'}\widehat{\underline{\omega}}_n+\frac{1}{2}\frac{\partial^2\Psi_{N,n}(\theta_*)}{\partial r^2}(\widehat{\rho}-1)^2)=o_p(1)$ when $Z_{1,N}+\widetilde{U}_N>0$. The latter limit result holds because in the just identified case $\Psi_{N,n}(\widehat{\theta}_n)=0$. Just as in the case of $N^{1/4}R_{1,N}^c(N^{1/4}(\widehat{\rho}-1))$ in the proof of theorem 2, we need to consider a higher order expansion of the remainder term $N^{1/4}R_{1,N}(N^{1/4}(\widehat{\rho}-1))$ in order to find the limiting distribution of the sign of $N^{1/4}(\widehat{\rho}-1)$ when $Z_1>0$, i.e., the distribution of B given $Z_1>0$. This means we need to consider $R_{3,N}(N^{1/4}(\widehat{\rho}-1))$. We have

$$
\begin{aligned}
N^{1/2}R_{3,N}(N^{1/4}(\widehat{\rho}-1))&=N^{1/4}R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2)+\\
&\quad N^{1/4}(\widehat{\rho}-1)R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)+R_{6,N}(N^{1/4}(\widehat{\rho}-1))
\end{aligned}
$$

where

$$
R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2)=
$$

$$
\begin{aligned}
&(N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\underline{\widehat{\omega}}_n)'W_N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\underline{\widehat{\omega}}_n)+\\
&\quad \frac{1}{2!}N(\kappa_{1,N},\ \kappa_{2,N},\ \kappa_{3,N})'W_N\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r^2}+\\
&\quad N(\Psi_{N,n}(\theta_*)+\frac{\partial \Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^2 \partial w_n'}\underline{\widehat{\omega}}_n)N^{1/2}(\widehat{\rho}-1)^2+\\
&(\frac{2}{3!}N^{1/2}(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3}+\\
&\quad \frac{1}{2!}N^{1/2}\frac{\partial^2 \Psi_{N,n}'(\theta_*)}{\partial r^2}W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^2 \partial w_n'}\underline{\widehat{\omega}}_n+\\
&\quad \frac{2}{4!}N^{1/2}(\Psi_{N,n}(\theta_*)+\frac{\partial \Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4})N(\widehat{\rho}-1)^4+\\
&(\frac{1}{4!}\frac{\partial^2 \Psi_{N,n}'(\theta_*)}{\partial r^2}W_N\frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4}+\frac{1}{3!3!}\frac{\partial^3 \Psi_{N,n}'(\theta_*)}{\partial r^3}W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3})N^{3/2}(\widehat{\rho}-1)^6,
\end{aligned}
$$

$$
\begin{aligned}
&R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)=N^{3/2}(\kappa_{1,N},\ \kappa_{2,N},\ \kappa_{3,N})'W_N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\underline{\widehat{\omega}}_n)+\\
&(\frac{1}{3!}N(\kappa_{1,N},\ \kappa_{2,N},\ \kappa_{3,N})'W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3}+\\
&\quad \frac{2}{3!}N(\Psi_{N,n}(\theta_*)+\frac{\partial \Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^3 \partial w_n'}\underline{\widehat{\omega}}_n+\\
&\quad N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^2 \partial w_n'}\underline{\widehat{\omega}}_n)N^{1/2}(\widehat{\rho}-1)^2+\\
&(\frac{2}{5!}N^{1/2}(\Psi_{N,n}(\theta_*)+\frac{\partial \Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^5 \Psi_{N,n}(\theta_*)}{\partial r^5}+\frac{1}{3!}N^{1/2}\frac{\partial^2 \Psi_{N,n}'(\theta_*)}{\partial r^2}W_N\frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^3 \partial w_n'}\underline{\widehat{\omega}}_n+\\
&\quad \frac{2}{4!}N^{1/2}(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r}+\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\underline{\widehat{\omega}}_n)'W_N\frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4}+\frac{1}{3!}N^{1/2}\frac{\partial^3 \Psi_{N,n}'(\theta_*)}{\partial r^3}W_N\frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^2 \partial w_n'}\underline{\widehat{\omega}}_n)\times\\
&N(\widehat{\rho}-1)^4+(\frac{1}{5!}\frac{\partial^2 \Psi_{N,n}'(\theta_*)}{\partial r^2}W_N\frac{\partial^5 \Psi_{N,n}(\theta_*)}{\partial r^5}+\frac{2}{3!4!}\frac{\partial^3 \Psi_{N,n}'(\theta_*)}{\partial r^3}W_N\frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4})N^{3/2}(\widehat{\rho}-1)^6 \quad \text{and}
\end{aligned}
$$

$$
R_{6,N}(N^{1/4}(\widehat{\rho}-1))=o_p(1).
$$

with $\kappa_{k,N}=\underline{\widehat{\omega}}_n'K_{k,N}\underline{\widehat{\omega}}_n$ and $K_{k,N}=\frac{\partial^2 \Psi_{k,N,n}(\theta_*)}{\partial w_n \partial w_n'}$ for $k=1,2,...,K+2$, where $\Psi_{k,N,n}(\theta)$ is the kth element of $\Psi_{N,n}(\theta)$.

Note that $R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2)=O_p(1)$ and $R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)=O_p(1)$. Recalling that in the just identified case $\Psi_{N,n}(\widehat{\theta}_n)=0$ and hence $N^{1/2}(\Psi_{N,n}(\theta_*)+\frac{\partial \Psi_{N,n}(\theta_*)}{\partial w_n'}\underline{\widehat{\omega}}_n+\frac{1}{2}\frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r^2}(\widehat{\rho}-1)^2)=o_p(1)$, we can simplify the expressions for $R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2)$ and

$R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)$:

$$
\begin{aligned}
&R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2) = \\
&N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r} + \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\widehat{\underline{\omega}}_n)' W_N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r} + \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\widehat{\underline{\omega}}_n \\
&\quad \frac{1}{2!}(\kappa_{1,N},\ \kappa_{2,N},\ \kappa_{3,N})' W_N \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r^2}) N^{1/2}(\widehat{\rho}-1)^2 + \\
&\frac{2}{3!}N^{1/2}(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r} + \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\widehat{\underline{\omega}}_n)' W_N \frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3} N(\widehat{\rho}-1)^4 + \\
&\frac{1}{3!3!}\frac{\partial^3 \Psi_{N,n}'(\theta_*)}{\partial r^3} W_N \frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3} N^{3/2}(\widehat{\rho}-1)^6 + o_p(1), \quad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
&R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2) = N^{3/2}(\kappa_{1,N},\ \kappa_{2,N},\ \kappa_{3,N})' W_N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r} + \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\widehat{\underline{\omega}}_n) + \\
&(\frac{1}{3!}N(\kappa_{1,N},\ \kappa_{2,N},\ \kappa_{3,N})' W_N \frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^3} + \\
&\quad N(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r} + \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\widehat{\underline{\omega}}_n)' W_N \frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^2 \partial w_n'}\widehat{\underline{\omega}}_n) N^{1/2}(\widehat{\rho}-1)^2 + \\
&(\frac{2}{4!}N^{1/2}(\frac{\partial \Psi_{N,n}(\theta_*)}{\partial r} + \frac{\partial^2 \Psi_{N,n}(\theta_*)}{\partial r \partial w_n'}\widehat{\underline{\omega}}_n)' W_N \frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4} + \\
&\quad \frac{1}{3!}N^{1/2}\frac{\partial^3 \Psi_{N,n}'(\theta_*)}{\partial r^3} W_N \frac{\partial^3 \Psi_{N,n}(\theta_*)}{\partial r^2 \partial w_n'}\widehat{\underline{\omega}}_n) N(\widehat{\rho}-1)^4 + \\
&(\frac{2}{3!4!}\frac{\partial^3 \Psi_{N,n}'(\theta_*)}{\partial r^3} W_N \frac{\partial^4 \Psi_{N,n}(\theta_*)}{\partial r^4}) N^{3/2}(\widehat{\rho}-1)^6 + o_p(1). \qquad (36)
\end{aligned}
$$

We can easily show that if $Z_{1,N} + \widetilde{U}_N > 0$, then $R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2) \neq o_p(1)$ and $R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2) \neq o_p(1)$. In particular, the terms involving $N^{3/2}(\widehat{\rho}-1)^6$ in the expressions for $R_{4,N}(N^{1/2}(\widehat{\rho}-1)^2)$ and $R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)$ in (36) do not vanish or cancel out when $N \to \infty$.

We conclude that if $Z_{1,N} + \widetilde{U}_N > 0$, then $N^{1/2}(\widehat{\rho}-1)^2 = Z_{1,N} + O_p(N^{-1/2})$ and

$$sgn(N^{1/4}(\widehat{\rho}-1)) = sgn(-R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)).$$

If $Z_{1,N} + \widetilde{U}_N < 0$, then $M_N(\widehat{\theta}_n)$ has one local minimum and its value is minimized at $N^{1/4}(\widehat{\rho}-1) = o_p(1)$. It also follows that $\widetilde{U}_N = O_p(N^{-1/2})$.

It is easily seen that $R_{5,N}(Z_{1,N}+\widetilde{U}_N) \xrightarrow{d} R_5$ for some R_5. Likewise, when $Z_{1,N}+\widetilde{U}_N > 0$, $R_{5,N}(Z_{1,N}+\widetilde{U}_N) \xrightarrow{d} \tilde{R}_5$ for some $\tilde{R}_5$. We obtain a formula for $\tilde{R}_5$ from the expression for $R_{5,N}(N^{1/2}(\widehat{\rho}-1)^2)$ in (36) by replacing appropriately scaled versions of the derivatives of $\Psi_{N,n}(\theta)$ at θ_* by their stochastic limits, $N^{1/2}\widehat{\underline{\omega}}_n$ by $\underline{\omega}_+$, and $N^{1/2}(\widehat{\rho}-1)^2$ by Z_1.

Now, when $Z_1 > 0$, $R = \tilde{R}_5$. We also have $\text{plim}_{N\to\infty} K_{-,N} = K_-$. We conclude that

$$\begin{bmatrix} N^{1/4}(\widehat{\rho}_F - 1) \\ N^{1/2}\underline{\widehat{\omega}}_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (-1)^B Z_1^{1/2} \\ \underline{\omega}_+ \end{bmatrix} \mathbf{1}\{Z_1 > 0\} + \begin{bmatrix} 0 \\ \underline{\omega}_+ + K_- Z_1 \end{bmatrix} \mathbf{1}\{Z_1 \leq 0\}$$

where $B = \mathbf{1}(R > 0)$. Regarding Σ_ω we note that $E(V_6 V_4) = E(V_6 V_1) = 0$. $\square$

Derivation of generic formulae for Z_1, $\underline{\omega}_+$ and K_- given in a comment below theorem 3:

In this case we minimize (cf. the proof of lemma 5):

$$\begin{aligned} \widetilde{M}_N(\widehat{\theta}_n) \equiv\; & 2(N^{1/2}\Psi_{N,n}(\theta_*))' W N^{1/2}(\Psi_{n,\omega}\underline{\widehat{\omega}}_n + \frac{1}{2}\Psi_{n,\rho\rho}(\widehat{\rho} - 1)^2) + \\ & N(\Psi_{n,\omega}\underline{\widehat{\omega}}_n + \frac{1}{2}\Psi_{n,\rho\rho}(\widehat{\rho} - 1)^2)' W (\Psi_{n,\omega}\underline{\widehat{\omega}}_n + \frac{1}{2}\Psi_{n,\rho\rho}(\widehat{\rho} - 1)^2) + o_p(1) \end{aligned}$$

with respect to $N^{1/2}(\widehat{\rho} - 1)^2$ and $N^{1/2}\underline{\widehat{\omega}}_n$, where $\Psi_{n,\rho\rho} = \Psi_{n,\rho\rho}(\theta_*)$. The f.o.c.'s are:

$$\begin{aligned} \frac{1}{2}\Psi'_{n,\rho\rho} W N^{1/2}\Psi_{N,n}(\theta_*) + \frac{1}{2}\Psi'_{n,\rho\rho} W N^{1/2}(\Psi_{n,\omega}\underline{\widehat{\omega}}_n + \frac{1}{2}\Psi_{n,\rho\rho}(\widehat{\rho} - 1)^2) + o_p(1) &= 0, \\ \Psi'_{n,\omega} W N^{1/2}\Psi_{N,n}(\theta_*) + \Psi'_{n,\omega} W N^{1/2}(\Psi_{n,\omega}\underline{\widehat{\omega}}_n + \frac{1}{2}\Psi_{n,\rho\rho}(\widehat{\rho} - 1)^2) + o_p(1) &= 0 \end{aligned}$$

Solving and letting $N \to \infty$ yields $N^{1/2}(\widehat{\rho} - 1)^2 \xrightarrow{d} Z_1 = -2(\Psi'_{n,\rho\rho} W^{1/2} M_\omega W^{1/2} \Psi_{n,\rho\rho})^{-1} \times (\Psi'_{n,\rho\rho} W^{1/2} M_\omega W^{1/2} \Psi_n)$ and $N^{1/2}\underline{\widehat{\omega}}_n \xrightarrow{d} \underline{\omega}_+ = \mathcal{M}(\Psi_n + \frac{1}{2}\Psi_{n,\rho\rho} Z_1)$, where $\Psi_n = \Psi_n(\theta_*)$. When $Z_1 < 0$, these solutions are not allowed. In this case we solve the lower part of the system of f.o.c.'s for $N^{1/2}\underline{\widehat{\omega}}_n$ while $N^{1/2}(\widehat{\rho} - 1)^2 = o_p(1)$. Again letting $N \to \infty$ gives $N^{1/2}\underline{\widehat{\omega}}_n \xrightarrow{d} \underline{\omega}_+ + K_- Z_1 = \mathcal{M}\Psi_n$ so that $K_- = -\frac{1}{2}\mathcal{M}\Psi_{n,\rho\rho}$. $\square$

Proof of theorem 4:

We first prove that the restricted estimator $\widetilde{\theta}_n = \widetilde{\theta}_{N,n}$ that satisfies $A\widetilde{\theta}_{N,n} = a_N$, which includes a restriction on ρ_N, is root N consistent under the parameter sequence $\theta_{0,N,n}$ with $\theta_{0,N,n} \to \theta_*$ and $A\theta_{0,N,n} = a_N$ (so that $A\theta_* = a = \lim_{N\to\infty} a_N$). Consider the restricted reparametrized modified log-likelihood where $A\theta_{N,n} = a_N$ (e.g. $r = a_N$). Noting that this function converges uniformly in probability to a limiting function that is continuous on a compact parameter set Θ that contains θ_*, and is uniquely maximized at θ_*, the claim follows from Theorem 2.1 in NMcF. In a similar way we can prove that the restricted FE(Q)MLE for $\overline{\sigma}^2_{v,n}$ is consistent under the parameter sequence $\underline{\theta}_{0,N,n}$ with $\underline{\theta}_{0,N,n} \to \underline{\theta}_* \in \underline{\Theta}$ and $A\theta_{0,N,n} = a_N$, where $\underline{\Theta}$ is a compact parameter set.

Following Davidson and MacKinnon (1993, pp. 276-277), we obtain $N^{1/2}(\widetilde{\theta}_n - \theta_{0,N,n}) \overset{asy}{=} -\mathcal{H}^{-1}(I - A'(A\mathcal{H}^{-1}A')^{-1}A\mathcal{H}^{-1})N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\theta_{0,N,n})}{\partial \theta_n}$ with $\mathcal{H} = \mathcal{H}(\underline{\theta}_{0,N,n})$ and $\rho_N \neq 1$. We also have $N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n} \overset{asy}{=} N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\theta_{0,N,n})}{\partial \theta_n} + \mathcal{H}N^{1/2}(\widetilde{\theta}_n - \theta_{0,N,n})$. Hence $N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n} \overset{asy}{=} A'(A\mathcal{H}^{-1}A')^{-1}A\mathcal{H}^{-1}N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\theta_{0,N,n})}{\partial \theta_n}$ and $(A\mathcal{H}^{-1}(\widetilde{\underline{\theta}}_n)\mathcal{J}(\widetilde{\theta}_n)\mathcal{H}^{-1}(\widetilde{\underline{\theta}}_n)A')^{-1/2}A\mathcal{H}^{-1}(\widetilde{\underline{\theta}}_n)\times N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n} \overset{asy}{=} (A\mathcal{H}^{-1}(\underline{\theta}_{0,N,n})E_{\underline{\theta}_{0,N,n}}(\mathcal{J}(\theta_{0,N,n}))\mathcal{H}^{-1}(\underline{\theta}_{0,N,n})A')^{-1/2}A\mathcal{H}^{-1}(\underline{\theta}_{0,N,n}) \times N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\theta_{0,N,n})}{\partial \theta_n}$ under the parameter sequence $\underline{\theta}_{0,N,n}$ with $\underline{\theta}_{0,N,n} \to \underline{\theta}_*$, $A\theta_{0,N,n} = a_N$ and $\rho_N \neq 1$. Similarly we can show that $(A\widetilde{\mathcal{H}}^{-1}(\widetilde{\underline{\theta}}_n)\widetilde{\mathcal{J}}(\widetilde{\theta}_n)\widetilde{\mathcal{H}}^{-1}(\widetilde{\underline{\theta}}_n)A')^{-1/2}A\widetilde{\mathcal{H}}^{-1}(\widetilde{\underline{\theta}}_n)N^{1/2}\times \widetilde{S}(\widetilde{\theta}_n) \overset{asy}{=} (A\widetilde{\mathcal{H}}^{-1}(\underline{\theta}_{0,N,n})E_{\underline{\theta}_{0,N,n}}(\widetilde{\mathcal{J}}(\theta_{0,N,n}))\widetilde{\mathcal{H}}^{-1}(\underline{\theta}_{0,N,n})A')^{-1/2}A\widetilde{\mathcal{H}}^{-1}(\underline{\theta}_{0,N,n})N^{1/2}\widetilde{S}(\theta_{0,N,n})$ when $\underline{\theta}_{0,N,n} = \underline{\theta}_*$ with $A\theta_* = a$.

Let $S_{N,i}(\underline{\theta}_n) = A\mathcal{H}^{-1}(\underline{\theta}_n)\frac{\partial \widetilde{l}_{N,n,i}(\theta_n)}{\partial \theta_n}$, $S_{N,i} = S_{N,i}(\underline{\theta}_{0,N,n})$ and $\underline{\theta}_{0,N,n} \to \underline{\theta}_* \in \underline{\Theta}$ with $A\theta_{0,N,n} = a_N$ and $\rho_N \neq 1$. Under appropriate regularity conditions (cf. Bottai, 2003) $E_{\underline{\theta}_{0,N,n}}(S_{N,i}) = 0$, $Var_{\underline{\theta}_{0,N,n}}(S_{N,i}) = A\mathcal{H}^{-1}(\underline{\theta}_{0,N,n})E_{\underline{\theta}_{0,N,n}}(\mathcal{J}_i(\theta_{0,N,n}))\mathcal{H}^{-1}(\underline{\theta}_{0,N,n})A'$ and $\sup_i \sup_{\underline{\theta}_n \in \mathcal{N}} E_{\underline{\theta}_{0,N,n}}(|\lambda' S_{N,i}(\underline{\theta}_n)|^{\varsigma}) < \infty$ for some $\varsigma > 2$ and for all $\lambda \in \mathbb{R}^J$ where $\mathcal{N} \subset \underline{\Theta}$ is an open neighbourhood around $\underline{\theta}_*$. We also have $N^{-1}\sum_{i=1}^N Var_{\underline{\theta}_{0,N,n}}(\lambda' S_{N,i}) > 0$ uniformly in N for all $\lambda \in \mathbb{R}^J \backslash \{\mathbf{0}\}$. Thus the Lyapunov conditions are satisfied and by (a multivariate version of) Lindeberg's CLT for triangular arrays, $(\sum_{i=1}^N Var_{\underline{\theta}_{0,N,n}}(S_{N,i}))^{-1/2}\times \sum_{i=1}^N S_{N,i}$ converges under the parameter sequence $\underline{\theta}_{0,N,n}$ to $N(0, I_J)$. It follows that $(A\mathcal{H}^{-1}(\widetilde{\underline{\theta}}_n)\mathcal{J}(\widetilde{\theta}_n)\mathcal{H}^{-1}(\widetilde{\underline{\theta}}_n)A')^{-1/2}A\mathcal{H}^{-1}(\widetilde{\underline{\theta}}_n)N^{1/2}\frac{\partial \widetilde{l}_{N,n}(\widetilde{\theta}_n)}{\partial \theta_n} \xrightarrow{d} N(0, I_J)$ and $QLM(\theta_{0,n}) \xrightarrow{d} \chi^2(J)$ under the parameter sequence $\underline{\theta}_{0,N,n}$ with $\underline{\theta}_{0,N,n} \to \underline{\theta}_*$, $A\theta_{0,N,n} = a_N$ and $\rho_N \neq 1$.

Next let $\widetilde{S}_{N,i}(\underline{\theta}_n) = A\widetilde{\mathcal{H}}^{-1}(\underline{\theta}_n)\widetilde{S}(\theta_n)$, $\widetilde{S}_{N,i} = \widetilde{S}_{N,i}(\underline{\theta}_{0,N,n})$ and $\underline{\theta}_{0,N,n} = \underline{\theta}_* \in \underline{\Theta}$ with $A\theta_* = a$. Under appropriate regularity conditions (cf. Bottai, 2003) we can also show that $(\sum_{i=1}^N Var_{\underline{\theta}_{0,N,n}}(\widetilde{S}_{N,i}))^{-1/2}\sum_{i=1}^N \widetilde{S}_{N,i} \xrightarrow{d} N(0, I_J)$ and hence $QLM(\theta_{0,n}) \xrightarrow{d} \chi^2(J)$ when $\underline{\theta}_{0,N,n} = \underline{\theta}_*$ with $A\theta_* = a$.

We conclude that $\lim_{N\to\infty}\sup_{\underline{\theta}_0 \in \mathcal{N}} \left|\Pr_{\underline{\theta}_0}\{QLM(\theta_0) > \chi^2_{J,\alpha}\} - \alpha\right| = 0$. $\square$

References

[1] Ahn, S.C., and G.M. Thomas, 2004, Likelihood-based inference for dynamic panel data models, unpublished manuscript, Arizona State University, revised in 2006.

[2] Alvarez, J., and M. Arellano, 2004, Robust likelihood estimation of dynamic panel data models, unpublished manuscript, University of Alicante and CEMFI, Madrid.

[3] Anderson, T.W., and C. Hsiao, 1981, Estimation of dynamic models with error components, Journal of the American Statistical Association 76, 598-606.

[4] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of econometrics 18, 47-82.

[5] 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.

[6] Arellano, M., and S. Bonhomme, 2009, Robust priors in nonlinear panel data models, Econometrica 77, 489-536.

[7] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, Journal of econometrics 68, 29-51.

[8] Bester, C.A., and C. Hansen, 2007, Frequentist bias reduction via Bayesian priors, unpublished manuscript, Graduate School of Business, University of Chicago.

[9] Bottai, M., 2003, Confidence regions when the Fisher information is zero, Biometrika 90, 73-84.

[10] Bun, M.J.G., and M.A. Carree, 2005, Bias-Corrected Estimation in Dynamic Panel Data Models, Journal of business & economic statistics 23, 200-210; Correction: 363.

[11] Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of economic studies XLVII, 225-238.

[12] Dhaene, G., and K. Jochmans, 2015, Likelihood inference in an autoregression with fixed effects, Econometric Theory 32, 1178-1215.

[13] Dovonon, P., and A. Hall, 2018, The asymptotic properties of GMM and Indirect Inference under second-order identification, Journal of econometrics 205, 76-111.

[14] Dovonon, P., and E. Renault, 2013, Testing for conditionally heteroskedastic factors, Econometrica 81, 2561-2586.

[15] Hahn, J., and G. Kuersteiner, 2002, Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large, Econometrica 70, 1639-1657.

[16] Hallin, M., and C. Ley, 2014, Skew-symmetric distributions and Fisher information: The double sin of the skew-normal, Bernoulli 20, 1432-1453.

[17] Han, C., P.C.B. Phillips, and D. Sul, 2014, X-differencing and dynamic model estimation, Econometric Theory 30, 201-251.

[18] Harris, R.D.F., and E. Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed, Journal of econometrics 91, 201-226.

[19] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel models covering short time periods, Journal of econometrics 109, 107-150.

[20] Juodis, A., 2013, A note on bias-corrected estimation in dynamic panel data models, Economics Letters 118, 435-438.

[21] Kruiniger, H., 2001, On the estimation of panel regression models with fixed effects, appeared in 2002 as working paper no. 450, Queen Mary, University of London, revised in 2003.

[22] Kruiniger, H., 2008, Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model, Journal of Econometrics 144, 447-464.

[23] Kruiniger, H., 2009, GMM estimation and inference in dynamic panel data models with persistent data, Econometric theory 25, 1348-1391.

[24] Kruiniger, H., 2013, Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions, Journal of Econometrics 173, 175-188.

[25] Kruiniger, H., 2015, Uniform Quasi ML based inference for the panel AR(1) model with arbitrary initial conditions, mimeo.

[26] Lancaster, T., 2002, Orthogonal parameters and panel data, Review of economic studies LXIX, 647-667.

[27] Lee, L.F., and A. Chesher, 1986, Specification testing when score test statistics are identically zero, Journal of econometrics 31, 121-149.

[28] Madsen, E., 2009, GMM-based inference in the AR(1) panel data model for parameter values where local identification fails, unpublished manuscript, Centre for Applied Microeconometrics (CAM), University of Copenhagen.

[29] Moreira, M.J., 2009, A maximum likelihood method for the incidental parameter problem, Annals of Statistics 37, 3660–3696.

[30] Newey, W., and D. McFadden, 1994, Large sample estimation and hypothesis testing, Ch. 36 in Handbook of econometrics, Vol. 4, R.F. Engle and D. McFadden (eds.), North-Holland, Elsevier, Amsterdam.

[31] Neyman, J., and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, Econometrica 16, 1-32.

[32] Nickell, S., 1981, Biases in dynamic models with fixed effects, Econometrica 49, 1417-1426.

[33] Rotnitzky, A., D.R. Cox, M. Bottai, and J. Robins, 2000, Likelihood-based inference with singular information matrix, Bernoulli 6, 243-284.

[34] Sargan, J. D., 1983. Identification and lack of identification, Econometrica, 51, 1605-1633.

[35] Van der Vaart, A.W., 2000, Asymptotic statistics, Cambridge University Press.

Table 1: Estimators of ρ; Design S; 5000 replications.

N=100	T=4	NM	MMLC		FEML		REML	
σ_μ^2	ρ		bias	RMSE	bias	RMSE	bias	RMSE
1	0.50	.075	.019	.126	.023	.140	.017	.125
1	0.80	.396	-.010	.132	.010	.147	.038	.156
1	0.90	.468	-.040	.132	-.012	.135	.038	.149
1	0.95	.471	-.065	.139	-.009	.138	.042	.146
1	0.98	.485	-.076	.144	.009	.138	.038	.140
1	1.00	.481	-.084	.148	.026	.135	.035	.136
0	0.50	.079	.017	.125	.021	.138	.005	.098
0	0.80	.385	-.012	.131	.008	.146	.019	.139
0	0.90	.459	-.042	.132	-.013	.135	.030	.147
0	0.95	.474	-.064	.141	-.010	.139	.037	.145
0	1.00	.481	-.086	.150	.026	.135	.026	.135
25	0.50	.077	.016	.125	.019	.138	.022	.143
25	0.80	.400	-.010	.133	.011	.148	.045	.173
25	0.90	.461	-.043	.134	-.015	.136	.040	.159
25	0.95	.474	-.064	.140	-.010	.137	.041	.148
25	1.00	.479	-.089	.152	.025	.137	.035	.137

NM: relative frequency that $\widehat{\rho}_{LAN}$ does not exist (No Maximum).

Table 2: Estimators of ρ; Design NS; 5000 replications.

N=100	T=4	NM	MMLC		FEML		REML	
σ_μ^2	ρ		bias	RMSE	bias	RMSE	bias	RMSE
1	0.50	.326	.010	.143	.014	.141	.015	.142
1	0.80	.467	-.069	.148	.028	.161	.028	.163
1	0.90	.471	-.085	.153	.010	.146	.021	.149
1	0.95	.478	-.085	.150	-.001	.133	.016	.139
1	0.98	.483	-.088	.152	.005	.137	.024	.139
1	1.00	.481	-.084	.148	.026	.135	.035	.136

Table 3: Estimators of ρ; Design S; 5000 replications.

N=100	T=9	NM	MMLC		FEML		REML	
σ_μ^2	ρ		bias	RMSE	bias	RMSE	bias	RMSE
1	0.50	.000	.000	.042	.000	.042	.000	.041
1	0.80	.130	.006	.064	.008	.069	.005	.061
1	0.90	.375	-.004	.060	.007	.070	.011	.067
1	0.95	.455	-.020	.061	-.004	.064	.014	.067
1	0.98	.489	-.032	.063	.001	.062	.017	.063
1	1.00	.490	-.041	.068	.013	.057	.016	.058

Table 4: Estimators of ρ; Design S; 5000 replications.

N=500	T=4	NM	MMLC		FEML		REML	
σ_μ^2	ρ		bias	RMSE	bias	RMSE	bias	RMSE
1	0.50	.001	.003	.052	.002	.048	.002	.046
1	0.80	.306	.007	.084	.017	.099	.008	.077
1	0.90	.442	-.016	.082	-.003	.089	.015	.090
1	0.95	.482	-.035	.085	-.017	.087	.024	.095
1	0.98	.498	-.047	.090	-.014	.089	.029	.093
1	1.00	.512	-.054	.092	.018	.087	.024	.088
0	0.50	.001	.002	.053	.002	.050	.001	.042
0	0.80	.293	.007	.085	.019	.101	.004	.068
0	0.90	.438	-.018	.084	-.007	.090	.009	.085
0	0.95	.473	-.034	.085	-.016	.085	.020	.093
0	1.00	.493	-.056	.094	.018	.088	.018	.088
25	0.50	.002	.002	.054	.000	.049	.000	.049
25	0.80	.314	.009	.085	.017	.097	.058	.135
25	0.90	.443	-.016	.082	-.004	.088	.055	.118
25	0.95	.471	-.036	.084	-.016	.085	.043	.102
25	1.00	.489	-.058	.096	.018	.090	.023	.090

NM: relative frequency that $\widehat{\rho}_{LAN}$ does not exist (No Maximum).

Table 5: Estimators of ρ; Design NS; 5000 replications.

N=500	T=4	NM	MMLC		FEML		REML	
σ_μ^2	ρ		bias	RMSE	bias	RMSE	bias	RMSE
1	0.50	.180	.016	.091	.004	.064	.004	.064
1	0.80	.489	-.036	.088	.019	.104	.021	.105
1	0.90	.502	-.051	.094	.027	.101	.032	.102
1	0.95	.484	-.056	.094	.009	.091	.019	.092
1	0.98	.481	-.058	.096	-.004	.088	.011	.090
1	1.00	.512	-.054	.092	.018	.087	.024	.088

Table 6: Estimators of ρ; Design S; 5000 replications.

N=500	T=9	NM	MMLC		FEML		REML	
σ_μ^2	ρ		bias	RMSE	bias	RMSE	bias	RMSE
1	0.50	.000	.001	.020	.001	.020	.000	.017
1	0.80	.006	.003	.028	.002	.026	.001	.022
1	0.90	.272	.005	.039	.009	.046	.003	.033
1	0.95	.420	-.007	.036	.001	.041	.007	.039
1	0.98	.460	-.019	.039	-.006	.039	.009	.041
1	1.00	.488	-.027	.044	.010	.039	.012	.039

Table 7: Empirical size of Quasi LM test based on Modified Likelihood; Nominal size is 0.05; T=9; 10000 replications.

model	S-Normal		S-ChiSq		NS-Normal	
ρ	$N = 100$	$N = 500$	$N = 100$	$N = 500$	$N = 100$	$N = 500$
0.50	.0547	.0474	.0531	.0491	.0533	.0483
0.80	.0540	.0514	.0551	.0519	.0575	.0522
0.90	.0557	.0500	.0529	.0510	.0527	.0539
0.95	.0495	.0510	.0496	.0498	.0501	.0512
0.98	.0502	.0508	.0482	.0512	.0472	.0443
0.99	.0512	.0518	.0528	.0508	.0496	.0506

Table 8: Empirical power of Quasi LM test based on Modified Likelihood; $H_0 : \rho = 0.8$; Nominal size is 0.05; T=9; 5000 replications.

model	S-Normal		S-ChiSq		NS-Normal	
true ρ	$N = 100$	$N = 500$	$N = 100$	$N = 500$	$N = 100$	$N = 500$
0.50	.999	1.000	.966	.999	.993	1.000
0.60	.919	1.000	.828	.999	.781	1.000
0.70	.375	.926	.399	.878	.274	.783
0.90	.170	.743	.207	.775	.141	.667
0.95	.421	.989	.404	.956	.391	.982
0.99	.714	1.000	.529	.992	.726	1.000