Econometric Theory, **36**, 2020, 185–222.
doi:10.1017/S0266466619000069

ESTIMATION FOR DYNAMIC PANEL DATA WITH INDIVIDUAL EFFECTS

PETER M. ROBINSON
London School of Economics

CARLOS VELASCO
Universidad Carlos III de Madrid

The article discusses statistical inference in parametric models for panel data. The models feature dynamics of a general nature, individual effects, and possible explanatory variables. The focus is on large-cross-section inference on Gaussian pseudo maximum likelihood estimates with temporal dimension kept fixed, partially complementing and extending recent work of the authors. We focus on a particular kind of initial condition but go on to discuss implications of alternative initial conditions. Some possible further developments are briefly reviewed.

1. INTRODUCTION

The proliferation of econometric panel data sets has prompted considerable development in the modeling of such data and consequent methods of point estimation and statistical inference, with associated theoretical justification. The literature goes back a long way and includes also much work by statisticians under the heading of longitudinal data. A fundamental monograph is Hsiao (2014).

In general, we describe a panel data set as a rectangular array of scalars y_{it}, $i = 1, \ldots, N$, $t = 0, 1, \ldots, T$, so we have observations on N cross-sectional units at $T + 1$ consecutive, equally spaced points of time, along with perhaps observable explanatory variables. Except for the very simplest of models, such as linear regression under classical conditions, only asymptotic justification of statistical models is feasible. Mirroring the longitudinal character of many data sets, where T can be very small, even $T = 2$, it is often most reasonable to develop asymptotic theory with $N \to \infty$ but T kept fixed. However, theory that requires N to diverge is problematic when, for example, the model for y_{it} incorporates an additive, unobserved, individual effect parameter or random variable ζ_i. Thus, the number of unknowns increases with N, indeed there are precisely N such ζ_i. This is the 'incidental parameters' problem pointed out by Neyman and Scott (1948), and several approaches have been suggested for dealing with it, and with more general versions of the problem. Typically, these involve some procedure for eliminating, or approximately eliminating, the ζ_i (which are commonly regarded as nuisance

We are grateful to Peter Phillips and three referees for constructive comments that have improved the article. Address correspondence to Peter M. Robinson, London School of Economics, London, UK; e-mail: p.m.robinson@lse.ac.uk.

parameters), leaving us to study features of interest in a transformed or modified model.

These features typically include one or more of the following: explanatory variables (modeled parametrically or nonparametrically), instantaneous temporal effects (such as an unknown additive quantity varying over time), cross-sectional dependence, and modeling of temporal dependence. Many econometric panel data models have described the latter in terms of autoregressive (AR) models or autoregressive moving average (ARMA) models, to include the possibility of an $I(1)$ unit root, reflecting the preoccupation with unit roots in much of the macroeconometric time series literature. The literature on AR and ARMA panel data models is now very well developed, including much work on unit root testing, modifying econometric time series methods (for a recent review see Moon, Perron, and Phillips, 2015). Some of it concerns asymptotics with $N \to \infty$ but T kept fixed, but also there is work, motivated by some data sets, which entails $T \to \infty$ and N fixed, or with both $T \to \infty$ and $N \to \infty$, including with N increasing at some rate as a function of T or vice versa; the precise form of the model often dictates what type of asymptotics is possible or desirable.

Despite the popularity of AR models, there is in principle any number of dynamic models, even any number which nest $I(1)$ behaviour. One such class which has been studied a good deal in the time series literature is that of fractional models. Whereas AR models cover certain $I(0)$ processes (when the AR coefficient lies in the stationary region) as well as certain $I(\delta)$ processes for any integer δ, and also explosive processes, fractional processes describe certain $I(\delta)$ processes for real values of δ. These can include negative δ, to describe antipersistence or noninvertibility, but the main interest has been in moderately positive values of δ, such as $\delta \in (0,2]$, where $\delta \in (0,1/2)$ implies stationarity and $\delta \geq 1/2$ implies nonstationarity. The most striking difference between AR models and fractional models with respect to statistical inference is as follows: whereas in an AR setting the limit distribution as $T \to \infty$ of statistics such as LS estimates of the AR coefficient α are asymptotically normally distributed with $T^{1/2}$ norming rate when $|\alpha| < 1$, and have a nonstandard limit distribution with rate T when $\alpha = 1$ (and different behaviours again when $\alpha = -1$ and $|\alpha| > 1$), in fractional models, on the other hand, there exist estimates (which have an approximate Gaussian maximum likelihood interpretation and corresponding efficiency properties) of the memory parameter δ, which are asymptotically normal with norming rate $T^{1/2}$ whatever the value of δ (see Hualde and Robinson, 2011a). This latter property is due to an essential ‘smoothness’ of the fractional model. Thus, whereas testing $\alpha = 1$ against $\alpha \neq 1$ typically involves a nonstandard approximate distribution, testing $\delta = 1$ against $\delta \neq 1$ involves a standard approximate distribution.

Motivated by this time series experience, Robinson and Velasco (2015, 2017) have developed asymptotic statistical inference on certain panel data models with fractional dynamics. All their asymptotics is based on T diverging, with N allowed to be either fixed or increasing with T. The requirement that $T \to \infty$ in the simple model of Robinson and Velasco (2015) (hereafter RV) is due to the use

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

of an approximation to the Gaussian pseudo likelihood and the need for the effect of an initial condition to be asymptotically negligible. The work of Robinson and Velasco (2017) in a more general model needs $T \to \infty$ not only for that same reason but also because of the nonparametric modeling of (possibly time-varying) individual effects and the nonparametric estimation of the cross-sectional covariance matrix. Fractional modeling of panel data has also been studied by Hassler, Demetrescu, and Tarcolea (2011). Fractional models might also be used in place of AR and the ARMA ones used by, e.g., Ejrnaes and Browning (2014), for income dynamics.

However, as with AR-based modeling, it is possible to develop theory that is based on N diverging with T fixed, as in the classical longitudinal data setting. This seems particularly feasible when there is no cross-sectional dependence, whence we can appeal essentially to a central limit theorem for a weighted sum of N independent random variables, rather than (in the $T \to \infty$ theory) to a central limit theorem for a sum of dependent or approximately whitened time series observations. It might be argued that in the time series ‘long memory’ literature, in which fractional models have been extensively used, it is often considered natural to expect that T be large. However, parametric fractional models generate formulae for point estimates and other statistics for any T, just as AR models do, and it seems equally legitimate to study them in the fixed T case. Using a similar estimation approach to ours for a model with AR dynamics, Han and Phillips (2013) found peculiarities in the objective function, with implications for asymptotic theory with T diverging, that are removed in asymptotic theory with N diverging.

The distinctive properties of AR and fractional time series models alluded to above all arise in the asymptotic $T \to \infty$ regime, and when T is kept fixed in a panel data setting they are no longer relevant. Thus, in the $N \to \infty$, fixed T case essentially similar asymptotic properties would arise from any number of parameterizations of time series dynamics. We consider a modeling strategy which is general in this sense, extending work for particular parameterizations in the fixed T case, and complementing some of the work which depends on a diverging T. In particular, the article partly complements RV, like them incorporating individual effects but also including explanatory variables as well as generalizing the dynamics and relying on diverging N rather than T; the explanatory variables could include period-fixed effects accounting for any potential trend or level change. As usual, on the one hand, parsimony in modeling is desirable, on the other, misspecification is to be avoided, and our asymptotic theory can be applied in model testing (e.g., by Wald, likelihood-ratio, and Lagrange multiplier type tests) as well as in interval estimation.

One additional issue which we explore is the generalization of initial conditions. A stationary time series process, such as a linear process, is typically modeled over all time points in $Z = \{0, \pm 1, \ldots\}$. But allowance for nonstationary dynamics of a process requires initial conditions (on the process prior to the observation period) to ensure the variance remains finite at any time point t, even if it diverges as $t \to \infty$. For an AR(1) process (with a possible unit root in mind),

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

it suffices to impose an initial condition on the process at a single t, e.g., that it is zero at $t = -1$. But for fractional processes, possible nonstationarity requires in general a condition on an infinite past, e.g., that the process takes zero values for $t < -1$. (The 'e.g.' is important here, because the condition might equally be imposed only for $t < -m$ for any integer $m \geq 1$.) But the choice of m is part of the model specification in that an incorrect choice can lead to inconsistent estimation of parameters of interest, particularly when T is kept fixed in the asymptotics.

The following section describes our dynamic model with fixed effects and possible regressors (under the simplest, and most usual, kind of initial conditions that cover fractional models, for example). Section 3 employs one of the standard approaches to eliminating individual effects, first differencing. Section 4 develops Gaussian pseudo maximum likelihood estimation of unknown parameters, motivated by its asymptotic efficiency when Gaussianity holds and its retention of consistency and asymptotic normality, with the same norming, under more general conditions. The conditions, and strong consistency and asymptotic normality properties, are described in Section 5 (with proofs left to the final Sections 9 and 10, the latter focussing on the evaluation of the asymptotic variance matrix under Gaussianity). Section 6 contains a Monte Carlo study of finite sample performance, while Section 7 discusses modifications to the methodology under alternative initial conditions. Section 8 summarizes and briefly lists possible extensions to modeling and inference.

2. DYNAMIC PANEL MODEL WITH INDIVIDUAL EFFECTS AND REGRESSORS

The observable scalar array $\{y_{it}\}$ and $q \times 1$ vector array $\{x_{it}\}$ are supposed to be related by the model

$$\lambda_t(L;\theta_0)\left(y_{it} - \zeta_i - x_{it}'\beta_0\right) = \varepsilon_{it}, \tag{1}$$

$$\varepsilon_{it} = 0,\ t < 0, \tag{2}$$

$$x_{it} = 0,\ t < 0, \tag{3}$$

for $i = 1, \ldots, N$, $t = 0, 1, \ldots, T \geq 2$, with the prime denoting transposition.

The ingredients of (1) are described as follows. For each t, the ε_{it} are independent and identically distributed (iid) across $i = 1, \ldots, N$. For each i, the ε_{it} are uncorrelated across $t = 0, 1, \ldots, T$, with mean zero and unknown, finite, positive variance σ_0^2. The ζ_i are unobserved fixed effects. The x_{it} consists of explanatory variables. The $p \times 1$ vector θ_0 and $q \times 1$ vector β_0 have unknown elements whose estimation is of interest, though in an important special case $\beta_0 = 0$ *a priori*, so the model contains no explanatory variables. Denoting by L the lag operator, and θ any admissible value of θ_0, we define the operator

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

$$\lambda_t(L;\theta)=\sum_{j=0}^{t}\lambda_j(\theta)\,L^j,$$

where the $\lambda_j(\theta)$ are known functions of θ with $\lambda_0(\theta)=1$ for all θ. The truncation means that the y_{it} need not be defined for $t<0$. For fixed T, the $\lambda_j(\theta)$ can be chosen quite arbitrarily, but leading choices are associated with regarding $\lambda_t(L;\theta)$ as truncating the expansion

$$\lambda(L;\theta)=\sum_{j=0}^{\infty}\lambda_j(\theta)\,L^j, \tag{4}$$

where $\lambda(L;\theta)$ is one of the AR operators originally arising in the stationary time series literature, though here no stationarity assumptions are imposed on parameters, especially as T remains fixed in our theory.

For example, taking θ_j to be the jth element of θ:

(i) The autoregressive moving average or autoregressive integrated moving average operator

$$\lambda(L;\theta)=\left(1-\sum_{j=1}^{p_1}\theta_j L^j\right)\left(1+\sum_{j=p_1+1}^{p}\theta_j L^{j-p_1}\right)^{-1}, \tag{5}$$

where $0\le p_1\le p$, with the understanding that the first and second sums are, respectively, void when $p_1=0$ (the pure MA case) and $p_1=p$ (the pure AR case). The dynamic panel data literature has heavily stressed the AR(1) case $p_1=p=1$, in which there has been great interest in testing the unit root hypothesis $\theta_{01}=1$, taking θ_{0j} to be the jth element of θ_0.

(ii) The fractional operator (see Adenstedt, 1974)

$$\lambda(L;\theta)=\Delta^{\theta}, \tag{6}$$

with $\Delta=1-L$, so $p=1$. The operator Δ^{δ} has the expansion

$$\Delta^{\delta}=\sum_{j=0}^{\infty}\pi_j(\delta)\,L^j,\ \ \pi_j(\delta)=\frac{\Gamma(j-\delta)}{\Gamma(-\delta)\Gamma(j+1)},$$

for noninteger $\delta>0$, while for integer $\delta=0,1,\ldots,$
$\pi_j(\delta)=1(j=0,1,\ldots,\delta)\,(-1)^j\,\delta(\delta-1)\cdots(\delta-j+1)/j!$, taking $0/0=1$. This case has been studied by RV, as has the hybrid model

(iii)

$$\lambda(L;\theta)=\Delta^{\theta_1}\left(1-\sum_{j=2}^{p_1+1}\theta_j L^{j-1}\right)\left(1+\sum_{j=p_1+2}^{p}\theta_j L^{j-1-p_1}\right)^{-1}, \tag{7}$$

where the first and second sums are, respectively, void when $p_1 = 0$ and $p_1 = p - 1$. This is known as a fractional ARIMA model (FARIMA$(p_1, \theta_1, p - p_1 - 1)$) so (6) is FARIMA$(0, \theta, 0)$.

Condition (2) is an initial condition, which ensures that $y_{it} - \zeta_i - x_{it}'\beta_0$ has bounded variance even if $\lambda_t(L;\theta_0)^{-1}\varepsilon_{it}$ has infinite variance when (2) does not hold but instead the conditions on ε_{it} in the second paragraph of the current section hold for all $t = 0, \pm 1, \pm 2, \ldots,$ as is the case for 'nonstationary' filters $\lambda(L;\theta)$, for example in (5) when at least one zero of $1 - \sum_{j=1}^{p_1} \theta_j z^j$ falls on or in the unit circle on the complex plane or in (6) or (7) when $\theta \geq 1/2$ or $\theta_1 \geq 1/2$, respectively (for the latter model even when all zeroes of $1 - \sum_{j=2}^{p_1+1} \theta_j z^{j-1}$ fall outside the unit circle). An initial condition such as (2) would not be needed were we to assume $\lambda(L;\theta_0)$ is a 'stationary' filter, but the dynamic panel literature has paid a good deal of attention to the possibility of nonstationarity, in particular an AR unit root. Condition (3) is a similar initial condition on x_{it}.

We single out two important restrictions that are implied by our assumptions. One is that each cross-sectional unit has the same dynamics. The other is that conditional on the ζ_i and x_{it}, y_{it} is cross-sectionally independent. In this connection, our asymptotic theory entails N diverging while T is kept fixed, so the ζ_i cannot be consistently estimated and their presence is an obstacle to consistent estimation of θ_0 and β_0, indicating an incidental parameters problem which the following section commences by eliminating.

3. DIFFERENCED MODEL

Of possible approaches to eliminate the ζ_i, we employ the popular one of first temporal differencing. Given (1) and (2), and defining

$$v_{it} = \lambda_t^{-1}(L;\theta_0)\varepsilon_{it}, \quad t = 0, \ldots, T, \ i = 1, \ldots, N, \tag{8}$$

we solve and then take first differences,

$$\Delta y_{it} - \Delta x_{it}'\beta_0 = \Delta v_{it}, \quad t = 1, \ldots, T, \ i = 1, \ldots, N. \tag{9}$$

In general, the Δv_{it} are not white noise, so (as in RV in the fractional case) we attempt a full whitening in order to estimate the parameters. For any θ, β, define

$$z_{it}(\theta,\beta) = \tau_{t-1}(L;\theta)\left(\Delta y_{it} - \Delta x_{it}'\beta\right), \ t = 1, \ldots, T, \ i = 1, \ldots, N, \tag{10}$$

where

$$\tau_t(L;\theta) = 1 + \sum_{j=1}^{t} \tau_j(\theta) L^j,$$

for

$$\tau_j(\theta) = \lambda_j(1;\theta) = \sum_{i=0}^{j} \lambda_i(\theta),$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

so that $\tau_t(L;\theta)$ truncates the expansion

$$\tau(L;\theta) = \sum_{j=0}^{\infty} \tau_j(\theta) L^j$$

of

$$\tau(L;\theta) = \lambda(L;\theta)/\Delta.$$

Then (cf. RV), for $t \geq 1$,

$$\begin{aligned} z_{it}(\theta,\beta) &= \tau_{t-1}(L;\theta)\Delta x_{it}'(\beta_0-\beta) + \tau_{t-1}(L;\theta)\Delta v_{it} \\ &= \tau_t(L;\theta)\Delta x_{it}'(\beta_0-\beta) + \left(\tau_{t-1}(L;\theta) - \tau_t(L;\theta)\right)\Delta x_{it}'(\beta_0-\beta) \\ &\quad + \tau_t(L;\theta)\Delta v_{it} + \left(\tau_{t-1}(L;\theta) - \tau_t(L;\theta)\right)\Delta v_{it} \\ &= \lambda_t(L;\theta) x_{it}'(\beta_0-\beta) - \tau_t(\theta)\Delta x_{i0}'(\beta_0-\beta) + \lambda_t(L;\theta) v_{it} - \tau_t(\theta)\Delta\varepsilon_{i0} \\ &= \lambda_t(L;\theta)\left\{v_{it} - x_{it}'(\beta-\beta_0)\right\} - \tau_t(\theta)\left\{\varepsilon_{i0} - x_{i0}'(\beta-\beta_0)\right\}, \end{aligned} \qquad \textbf{(11)}$$

with the last step imposing (2) and (3). Thus,

$$z_{it}(\theta,\beta_0) = \lambda_t(L;\theta) v_{it} - \tau_t(\theta)\varepsilon_{i0} \qquad \textbf{(12)}$$

and

$$z_{it}(\theta_0,\beta_0) = \varepsilon_{it} - \tau_t(\theta_0)\varepsilon_{i0}. \qquad \textbf{(13)}$$

As (13) indicates, the $z_{it}(\theta_0,\beta_0)$ are not white noise across t. As noted in RV, $\tau_t(\theta) = O\left(t^{-\theta_1}\right)$ as $t \to \infty$, in the case of models (ii) and (iii) of the preceding section (where, for example, $\tau(L;\theta) = \Delta^{\theta_1 - 1}$ in (ii)). Thus in these cases this corrupting term is negligible under stationary and nonstationary long memory, $\theta_1 > 0$, as $t \to \infty$, with faster the decay greater the memory, though on the other hand for negative dependence, $\theta_1 < 0$ (not discussed by RV), the corrupting term dominates as $t \to \infty$, while for $\theta_1 = 0$, it is generally of exact order $O(1)$. The latter is also the case for stationary versions of model (i). In general, with T fixed there is a nonnegligible source of bias incurred by employing methods that assume the $z_{it}(\theta_0,\beta_0)$ are uncorrelated across t.

If explanatory variables x_{it} are present, then in view of (10) we adopt the convention that in the original model (1) no element of x_{it} is constant across t, so that any intercept is incorporated in ζ_i. It is also important to acknowledge from (9) that in the presence of x_{it}, under the conditions we will impose to justify the more elaborate estimates proposed in the following section, β_0 can unsurprisingly instead be consistently and asymptotically normally estimated by, for example, ordinary least squares regression of the Δy_{it} on the Δx_{it}. Moreover, the norming factor in the central limit theorem under the same $N \to \infty$, fixed T regime is $N^{1/2}$, like our estimates. In view of the autocorrelation in the errors v_{it} which the general dynamics in (1) anticipate, ordinary least squares will generally be inefficient, but generalised least squares, employing an estimated $T \times T$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

error covariance matrix estimated simply from sums of squares and products of N least squares residuals without recourse to the parametric dynamics imposed in (1), will be asymptotically more efficient. Moreover, under conditions which imply that the limiting covariance matrix of the estimates proposed in the following section is block diagonal with respect to θ_0 and β_0, these estimates of β_0 will be asymptotically no more efficient than the computationally far simpler generalized least squares. However, the estimates of the following section seem worthwhile because the model (like RV's) may not include any x_{it}; because estimation of parameters θ_0 can help in inference on dynamics, such as possible nonstationarity and might be used in forecasting; because with correct specification of dynamics in (1) with p small relative to T our method may have better finite sample properties than the generalized least squares approach described above; and because (1) allows a comparison between our asymptotics and those of RV.

4. GAUSSIAN PSEUDO MAXIMUM LIKELIHOOD ESTIMATION

Gaussian pseudo maximum likelihood estimation is widely used in statistics and econometrics due to its asymptotic efficiency under Gaussianity and its consistency robustness under much broader conditions. It has been used in panel data models by, e.g., Hsiao, Pesaran, and Tahmiscioglu (2002). Given (10) define for $i = 1, \ldots, N$ the $T \times 1$ vectors

$$\begin{aligned} z_i(\theta,\beta) &= \left(z_{i1}(\theta,\beta), \ldots, z_{iT}(\theta,\beta)\right)' \\ &= \Upsilon(L;\theta)\left(\Delta y_i - \Delta x_i \beta\right), \end{aligned}$$

with the $T \times 1$ vector, $T \times q$ matrix and $T \times T$ diagonal matrix operator

$$\begin{aligned} \Delta y_i &= \left(\Delta y_{i1}, \ldots, \Delta y_{iT}\right)', \\ \Delta x_i &= \left(\Delta x_{i1}, \ldots, \Delta x_{iT}\right)', \\ \Upsilon(L;\theta) &= diag\left(\tau_0(L;\theta), \ldots, \tau_{T-1}(L;\theta)\right). \end{aligned}$$

From (13), $z_i(\theta_0, \beta_0)$ has zero mean vector and covariance matrix $\sigma_0^2 \Omega(\theta_0)$, where

$$\Omega(\theta) = I_T + \tau(\theta)\tau'(\theta), \tag{14}$$

introducing the $T \times 1$ vector

$$\tau(\theta) = \left(\tau_1(\theta), \ldots, \tau_T(\theta)\right)'$$

and with I_T the $T \times T$ identity matrix. Thus, define (cf. RV)

$$\widehat{\sigma}^2(\theta,\beta) = \frac{1}{NT}\sum_{i=1}^{N} z_i'(\theta,\beta)\,\Omega(\theta)^{-1} z_i(\theta,\beta) \tag{15}$$

and

$$L(\theta,\beta) = |\Omega(\theta)|^{1/T}\widehat{\sigma}^2(\theta,\beta). \tag{16}$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

The Gaussian pseudo maximum likelihood estimate (PMLE) is

$$\left(\widehat{\theta},\widehat{\beta}\right)=\arg\min_{\theta\in\Theta,\beta} L(\theta,\beta), \tag{17}$$

where Θ is a compact subset of R^p.

For computations note from RV that

$$\Omega(\theta)^{-1}=I_T-\frac{\tau(\theta)\tau'(\theta)}{|\Omega(\theta)|}, \tag{18}$$

$$|\Omega(\theta)|=1+\tau'(\theta)\tau(\theta), \tag{19}$$

while of course we can concentrate out β, defining

$$\beta(\theta)=\left(\sum_{i=1}^{N}(\Upsilon(L;\theta)\Delta x_i)'\Omega(\theta)^{-1}\Upsilon(L;\theta)\Delta x_i\right)^{-1}\sum_{i=1}^{N}(\Upsilon(L;\theta)\Delta x_i)'\Omega(\theta)^{-1}\Upsilon(L;\theta)\Delta y_i,$$

so that

$$\widehat{\theta}=\arg\min_{\theta\in\Theta} L\left(\theta,\widehat{\beta}(\theta)\right),\;\; \widehat{\beta}=\widehat{\beta}\left(\widehat{\theta}\right).$$

5. ASYMPTOTIC STATISTICAL PROPERTIES

To establish strong consistency of $\left(\widehat{\theta},\widehat{\beta}\right)$, we introduce the following assumptions.

Assumptions A. (*i*) The $\{\varepsilon_{it},x_{it},1\le i\le N,\ 1\le t\le T\}$ are iid across i.

(*ii*) $E(\varepsilon_{1t}|\,x_{1s},1\le s\le T)=0$, a.s., $0\le t\le T$.

(*iii*) $E(\varepsilon_{1s}\varepsilon_{1t}|\,x_{1r},1\le r\le T)=\sigma_0^2\delta_{st}$, a.s., $0\le s,t\le T$, for $\sigma_0^2<\infty$ and δ_{st} the Kronecker delta.

(*iv*) x_{it} does not contain an intercept and $E\|x_{1t}\|^2<\infty$, $0\le t\le T$.

(*v*) $\theta_0\in\Theta$, which is compact.

(*vi*) For $1\le t\le T$, the $\lambda_t(\theta)$ are continuous in θ.

(*vii*) For at least one $t\in[1,T]$, $\lambda_t(\theta)\neq\lambda_t(\theta_0)$ for all $\theta\in\Theta-\{\theta_0\}$.

(*viii*) The matrix

$$E\left((\Upsilon(L;\theta_0)\Delta x_1)'\Omega(\theta_0)^{-1}\Upsilon(L;\theta_0)\Delta x_1\right)$$

is positive definite.

Assumption (i) ensures that a strong law of large numbers for iid random variables can be used. Assumptions (ii) and (iii) require strong exogeneity of x_{it}. One or more elements of x_{it} might be deterministic and entail trending, e.g., linearly in t (because T is fixed), but x_{it} cannot include incidental tends. The first part of assumption (iv) merely means that any intercept is regarded as incorporated in the individual effects ζ_i, which are differenced out. Assumptions (v) and (vi) permit uniform convergence arguments and are readily checked, and, for example, Θ

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

can be chosen to cover the possibility of an AR unit root in model (5). Assumptions (vii) and (viii) ensure identifiability, with (viii) ruling out multicollinearity in regressors and (vii) being automatically satisfied in case of model (6), and satisfied in case of (5) and (7) if $p \le T$ and the autoregressive and moving average operators have no zeros in common.

THEOREM 1. *Let* (*1*), (*2*) *and Assumptions A hold. Then, as* $N \to \infty$, *almost surely* (*a.s.*)

$$\widehat{\theta} \to \theta_0, \ \ \widehat{\beta} \to \beta_0.$$

To develop a useful asymptotic normality result, as is standard we use Theorem 1 and the mean value theorem (see (A.10) in the proofs in Section 9 below) and consider the vector of first partial derivatives of $L(\theta, \beta)$ evaluated at θ_0, β_0. Now (cf. (A.12) of Section 9 below) define, for $j = 1, \ldots, p$,

$$\begin{aligned} r_{1ji}(\theta,\beta) = \frac{1}{T}\left|\Omega_T(\theta)\right|^{\frac{1}{T}} & \left(\frac{1}{T} tr\left(\Omega^{-1}(\theta)\,\Omega^j(\theta)\right) z_i'(\theta,\beta)\,\Omega(\theta)^{-1} z_i(\theta,\beta)\right. \\ & - z_i'(\theta,\beta)\,\Omega^{-1}(\theta)\,\Omega^j(\theta)\,\Omega^{-1}(\theta)\,z_i(\theta,\beta) \\ & \left. + 2\,\dot{z}_i^{j\prime}(\theta,\beta)\,\Omega^{-1}(\theta)\,z_i(\theta,\beta)\right). \end{aligned} \qquad \textbf{(20)}$$

Here,

$$\Omega^j(\theta) = \left(\partial/\partial\theta_j\right)\Omega(\theta) = \dot{\tau}^j(\theta)\,\tau(\theta)' + \tau(\theta)\,\dot{\tau}^j(\theta)' \qquad \textbf{(21)}$$

with

$$\dot{\tau}^j(\theta) = \left(\partial/\partial\theta_j\right)\tau(\theta), \qquad \textbf{(22)}$$

and $\dot{z}_i^{j\prime}(\theta,\beta)$ is the jth row of

$$\dot{z}_i'(\theta,\beta) = (\partial/\partial\theta)\,z_i'(\theta,\beta),$$

namely, the transpose of

$$\left(\partial/\partial\theta_j\right)\left(\Upsilon(L;\theta)\left(\Delta y_i - \Delta x_i\beta\right)\right) = \Xi^j(L;\theta)\left(\Delta y_i - \Delta x_i\beta\right),$$

where

$$\begin{aligned} \Xi^j(L;\theta) &= \left(\partial/\partial\theta_j\right)\Upsilon(L;\theta) \\ &= diag\left(\dot{\tau}_0^j(L;\theta), \ldots, \dot{\tau}_{T-1}^j(L;\theta)\right) \end{aligned}$$

with $\dot{\tau}_0^j(L;\theta) \equiv 0$ and for $t \ge 1$

$$\dot{\tau}_t^j(L;\theta) = \left(\partial/\partial\theta_j\right)\tau_t(L;\theta) = \sum_{k=1}^{t} \dot{\tau}_k^j(\theta)\,L^k,$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

in which the $\dot{\tau}_k^j(\theta)$ are given by $\dot{\tau}^j(\theta) = \left(\dot{\tau}_1^j(\theta), \ldots, \dot{\tau}_T^j(\theta)\right)'$. Next define (cf. (A.13) of Section 9 below)

$$r_{2i}(\theta,\beta) = -\frac{2}{T}\left|\Omega(\theta)\right|^{\frac{1}{T}}\left(\Upsilon(L;\theta)\Delta x_i\right)'\Omega(\theta)^{-1}z_i(\theta,\beta),$$

and then,

$$r_i(\theta,\beta) = \left(r_{1i}(\theta,\beta)', r_{21i}(\theta,\beta)'\right)',$$

and

$$C(\theta,\beta) = \frac{1}{N}\sum_{i=1}^{N} r_i(\theta,\beta)r_i(\theta,\beta)'.$$

Now define, for $j,k = 1,\ldots,p$,

$$\begin{aligned} b_{1jki}(\theta,\beta) = \frac{1}{T}\left|\Omega(\theta)\right|^{\frac{1}{T}}\Big(&\widehat{\sigma}^2(\theta,\beta)\,tr\left(\Omega^{-1}(\theta)\,\Omega^k(\theta)\,\Omega^{-1}(\theta)\,\Omega^j(\theta)\right) \\ &-\frac{1}{T}\widehat{\sigma}^2(\theta,\beta)\,tr\left(\Omega^{-1}(\theta)\,\Omega^j(\theta)\right)tr\left(\Omega^{-1}(\theta)\,\Omega^k(\theta)\right) \\ &-2\dot{z}_i^{k\prime}(\theta,\beta)\,\Omega^{-1}(\theta)\,\Omega^j(\theta)\,\Omega^{-1}(\theta)\,z_i(\theta,\beta) \\ &-2\dot{z}_i^{j\prime}(\theta,\beta)\,\Omega^{-1}(\theta)\,\Omega^k(\theta)\,\Omega^{-1}(\theta)\,z_i(\theta,\beta) \\ &+2\dot{z}_i^{j\prime}(\theta,\beta)\,\Omega^{-1}(\theta)\,\dot{z}_i^k(\theta,\beta)\Big) \end{aligned}$$

and the $q \times q$ matrix

$$B_{2i}(\theta) = 2\left|\Omega(\theta)\right|^{\frac{1}{T}}\left(\Upsilon(L;\theta)\Delta x_i\right)'\Omega(\theta)^{-1}\Upsilon(L;\theta)\Delta x_i/T.$$

Let $B_{1i}(\theta,\beta)$ be the $p \times p$ matrix with jkth element $b_{1jki}(\theta,\beta)$ and define the block diagonal matrix

$$B(\theta,\beta) = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} B_{1i}(\theta,\beta) & 0 \\ 0 & B_{2i}(\theta) \end{pmatrix}.$$

Alternative formulae might be used in place of $B_{1i}(\theta,\beta)$, but ours is a relatively simple one. For computations note again (19), (18) and, for example,

$$\begin{aligned} \Omega^{-1}(\theta)\,\Omega^j(\theta)\,\Omega^{-1}(\theta) &= \Omega^j(\theta) - \frac{\tau(\theta)\tau(\theta)'}{|\Omega(\theta)|}\Omega^j(\theta_0) - \Omega^j(\theta_0)\frac{\tau(\theta)\tau(\theta)'}{|\Omega(\theta)|} \\ &\quad + \frac{\tau(\theta)\tau(\theta)'}{|\Omega(\theta)|}\left(\dot{\tau}^j(\theta)\tau(\theta)' + \tau(\theta)\dot{\tau}^j(\theta)'\right)\frac{\tau(\theta)\tau(\theta)'}{|\Omega(\theta)|} \\ &= \frac{1}{|\Omega(\theta)|}\left(\dot{\tau}^j(\theta)\tau(\theta)' + \tau(\theta)\dot{\tau}^j(\theta)'\right) - 2\frac{\dot{\tau}^j(\theta)'\tau(\theta)}{|\Omega(\theta)|^2}\tau(\theta)\tau(\theta)', \end{aligned}$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

$$
\begin{aligned}
tr\left(\Omega^{-1}(\theta_0)\,\Omega^{j}(\theta_0)\right) &= tr\left(\left(I_T - \frac{\tau(\theta)\,\tau(\theta)'}{|\Omega(\theta)|}\right)\left(\dot{\tau}^{j}(\theta)\,\tau(\theta)' + \tau(\theta)\,\dot{\tau}^{j}(\theta)'\right)\right) \\
&= 2\dot{\tau}^{j}(\theta)'\,\tau(\theta)\left(1 - \frac{\tau(\theta)'\,\tau(\theta)}{|\Omega(\theta)|}\right) \\
&= 2\frac{\dot{\tau}^{j}(\theta)'\,\tau(\theta)}{|\Omega(\theta)|}.
\end{aligned}
$$

The proof of asymptotic normality is based on the following additional conditions.

Assumptions B. (i) $E\varepsilon_{1t}^{4} < \infty$.

(ii) For $1 \le t \le T$ the $\lambda_t(\theta)$ are twice continuously differentiable in θ.

(iii) θ_0 is an interior point of Θ.

(iv) In a neighbourhood of θ_0, β_0, the matrix $EB(\theta, \beta)$ is nonsingular.

(v) The matrix $EC(\theta_0, \beta_0)$ is nonsingular.

Condition (i) seems unavoidable. Condition (ii) can be verified by inspection. Condition (iii) is standard. Our other conditions ensure existence of the matrices in (iv) and (v), with (v) being a local identifiability condition, which, along with (iv), may be checkable for specific choices of the $\lambda_t(\theta)$.

THEOREM 2. *Let* (*1*), (*2*) *and Assumptions A and B hold. Then, as* $N \to \infty$,

$$
\left(NB\left(\widehat{\theta}, \widehat{\beta}\right) C\left(\widehat{\theta}, \widehat{\beta}\right)^{-1} B\left(\widehat{\theta}, \widehat{\beta}\right)\right)^{1/2} \begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{\beta} - \beta_0 \end{pmatrix} \to_d N(0, I_{p+q}),
$$

where $D^{1/2}$ denotes the unique nonnegative definite square root of a positive definite matrix D.

Under normality of ε_{1t}, $\widehat{\theta}, \widehat{\beta}$ are asymptotically efficient and we may replace the studentizing factor in the theorem by $N^{1/2}(T/2)\widehat{\sigma}^{-1}\left(\widehat{\theta}\right)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{T}} C\left(\widehat{\theta}, \widehat{\beta}\right)^{1/2}$ or by $N^{1/2}(T/2)^{1/2}\widehat{\sigma}^{-1}\left(\widehat{\theta}\right)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{2T}} B\left(\widehat{\theta}, \widehat{\beta}\right)^{1/2}$, the additional normalization for C and B correcting for L being proportional to a likelihood, instead of being a log-likelihood (see Section 10 below). Furthermore, it would be possible to use restricted versions replacing the elements not in the $p \times p$ and $q \times q$ matrices spanning their main diagonals by zeroes, reflecting the asymptotic independence of $\widehat{\theta}$ and $\widehat{\beta}$. Comparison with Theorem 2 of RV, which covers models (ii) and (iii) in Section 1 with $T \to \infty$ (but without explanatory variables), illustrates the much greater complexity of the asymptotic variance matrix based on $N \to \infty$ and fixed T asymptotics compared to $T \to \infty$ asymptotics. This can be better understood by inspecting the formulae for $\partial L(\theta_0, \beta_0)/\partial\theta$ in the proof in Section 8, in particular the term (A.15), which is the ith summand

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

in its jth element. It contains terms in ε_{it}^2, $t = 1, \ldots, T$, which can be seen to make an asymptotically negligible contribution, by a law of large numbers, only as $T \to \infty$. There are also the terms involving ε_{i0}^2 for which this argument does not apply, but collecting them together reveals that they make a negligible contribution. The dominating term in (A.15) as $T \to \infty$ is the penultimate one, and this is the basis for the central limit theorem of RV. But though their formula for asymptotic variance estimation is far simpler than ours, ours is still valid for large T (if N also is large) and might be preferred in case T is feared too small for RV's formula to provide a good approximation.

6. FINITE-SAMPLE PERFORMANCE

In this section, we explore by Monte Carlo simulations finite sample performance especially for estimation of θ_0, including our asymptotic variance estimates developed for finite T in comparison with the asymptotic version of RV obtained for increasing T, though we also compare our PMLE of β_0 with GLS and OLS estimation.

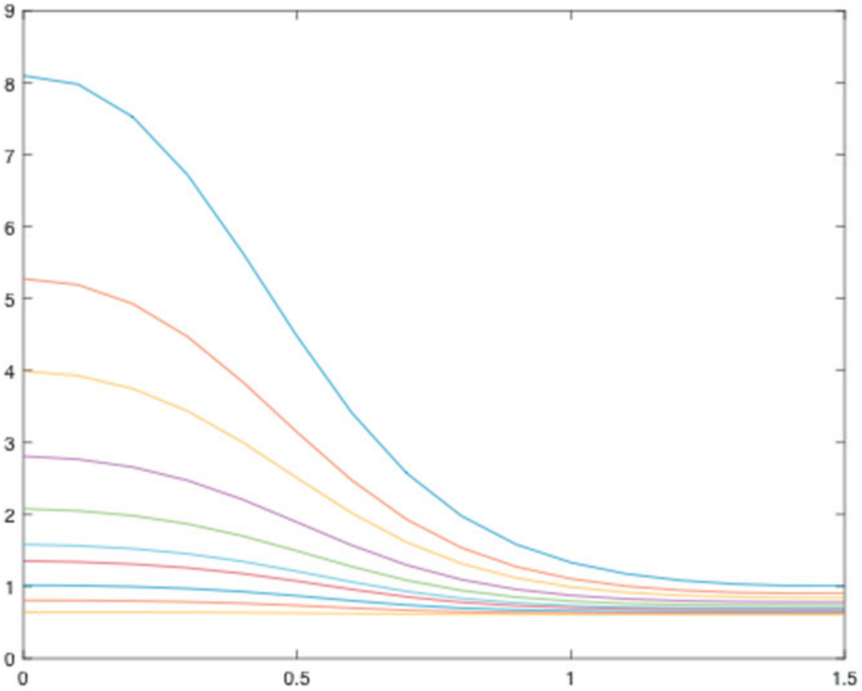

FIGURE 1. Asymptotic variance of PMLE estimates for the FARIMA$(0, \delta_0, 0)$ model (6), $\delta_0 \in [0, 1.5]$ in horizontal axis, each line corresponds to $T \in \{3, 4, 5, 7, 10, 15, 20, 40, 100, 1{,}000\}$, from top to bottom.

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

TABLE 1. Simulated size t-test $N = 100$, model (6), FARIMA$(0,\theta_0,0)$. Gaussian innovations. Nominal size 5%

θ_0	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	3.68	2.32	2.51	2.47	9.07
	10	4.38	2.78	2.95	2.72	14.53
	5	5.48	2.96	3.41	2.56	21.89
	4	5.81	3.46	3.65	3.01	24.85
	3	7.13	4.58	4.26	3.44	29.07
0.6	20	5.65	5.09	5.07	5.07	11.92
	10	5.91	4.93	5.06	5.06	17.91
	5	6.51	4.50	5.08	4.72	28.23
	4	7.34	4.81	5.77	5.64	33.94
	3	6.94	4.69	5.23	5.13	39.75
1.0	20	5.47	4.75	4.82	4.86	6.91
	10	6.05	4.79	5.20	5.19	9.05
	5	5.54	4.37	4.61	4.65	12.56
	4	6.19	4.76	5.18	5.05	15.35
	3	5.50	4.35	4.65	4.67	18.70
1.4	20	5.68	4.63	4.92	4.94	6.32
	10	6.07	4.94	5.10	5.18	7.89
	5	5.54	4.65	4.73	4.83	9.13
	4	6.39	4.77	5.32	5.26	11.27
	3	6.23	4.83	5.15	5.17	12.91

We initially consider models without regressors, thus estimating θ_0 only and implying a focus on the first diagonal block in B and C. We consider the following (alternative) studentization factors for $\widehat{\theta}$ in Theorem 2:

1. $\left(NB\left(\widehat{\theta}\right)C\left(\widehat{\theta}\right)^{-1}B\left(\widehat{\theta}\right)\right)^{1/2}$ as in Theorem 2, denoted as BCB.
2. $N^{1/2}(T/2)\widehat{\sigma}^{-1}\left(\widehat{\theta}\right)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{T}}C\left(\widehat{\theta}\right)^{1/2}$, following the discussion after Theorem 2 for Gaussian series (setting $\sigma_0^2 = 1$ $wlog$), the factor $(T/2)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{T}}$ correcting for L being only proportional to the likelihood, denoted as C.
3. $N^{1/2}(T/2)^{1/2}\widehat{\sigma}^{-1}\left(\widehat{\theta}\right)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{2T}}B\left(\widehat{\theta}\right)^{1/2}$, exploiting the proportionality of $B_0(\theta_0)$ and $C_0(\theta_0)$ under Gaussianity, denoted as B.
4. $\left(NB_0\left(\widehat{\theta}\right)C_0\left(\widehat{\theta}\right)^{-1}B_0\left(\widehat{\theta}\right)\right)^{1/2} = N^{1/2}(T/2)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{T}}C_0\left(\widehat{\theta}\right)^{1/2}$ $= N^{1/2}(T/2)^{1/2}\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{2T}}B_0\left(\widehat{\theta}\right)^{1/2}$ where $C_0(\theta_0)$ and $B_0(\theta_0)$ are the expectations of $C(\theta_0)$ and $B(\theta_0)$ under Gaussianity as evaluated in Appendix B using the same normalizations as in 2. and 3., denoted as $(BCB)_0$.
5. Asymptotic variance for $T \to \infty$ from RV, denoted as $Asymp$.

TABLE 2. Simulated size t-test $N = 200$, model (6), FARIMA$(0, \theta_0, 0)$. Gaussian innovations. Nominal size 5%

θ_0	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	3.02	2.70	2.54	2.67	9.35
	10	3.59	2.68	2.69	2.63	14.17
	5	4.63	2.91	2.80	2.73	22.29
	4	4.34	2.89	2.76	2.61	24.03
	3	5.56	3.68	3.60	2.85	29.39
0.6	20	5.57	5.01	5.17	5.12	11.81
	10	5.62	4.78	4.94	4.87	17.41
	5	5.43	4.76	4.89	4.96	28.35
	4	5.98	5.01	5.29	5.09	33.51
	3	5.47	4.74	4.73	4.90	40.92
1.0	20	5.44	4.98	5.13	5.13	7.12
	10	5.31	4.74	4.92	4.92	8.61
	5	5.32	4.57	4.71	4.69	12.61
	4	5.44	4.99	4.87	4.85	15.09
	3	5.29	4.81	5.00	5.04	18.43
1.4	20	5.40	4.88	5.10	5.09	6.70
	10	5.42	4.83	4.91	4.88	7.68
	5	5.33	4.76	4.93	4.83	9.59
	4	5.85	5.13	5.05	5.06	11.26
	3	5.65	4.91	5.16	5.17	13.06

We first calculate in the simple fractional model (6), i.e., FARIMA$(0, \theta_0, 0)$, the asymptotic variance $(BCB)_0$ of the PMLE of the memory parameter θ_0 computed for a grid of values of $\theta_0 \in [0, 1.5]$ and $T \in \{3, 4, 5, 7, 10, 15, 20, 40, 100, 1{,}000\}$. In Figure 1, we plot $T B_0(\theta_0)^{-1} C_0(\theta_0) B_0(\theta_0)^{-1}$ to make possible comparisons with the asymptotic variance of $\sqrt{NT}\left(\widehat{\theta} - \theta_0\right)$ as $T \to \infty$. The (scaled) asymptotic variance decreases very fast with T until $T = 20$, so we focus in our Monte Carlo simulations on these values, where the asymptotic variance based on $T \to \infty$ asymptotics severely underestimates the actual variability. There is also an important (monotone) dependence of the asymptotic variance on the persistence of the model for finite T since our estimation method is based on predifferenced data. Thus, capturing with precision values of θ_0 lower that 0.5 seems very challenging, while from about $\theta_0 = 1$ upwards the asymptotic variance shows an almost flat pattern, at least for $T \geq 5$. For $\theta_0 = 1.5$ and $T = 1{,}000$, the finite T variance $(BCB)_0$ equals 0.6107, which is very close to the asymptotic one (for T increasing), namely $6/\pi^2 = 0.6079$, but for $T = 3$ it equals 1.0059, 65% larger. However, for $\theta_0 = 0$, the finite $T = 3$ variance is about 13 times larger than the asymptotic $T \to \infty$ one.

We next explore the performance of the various studentizations in inference based on Theorem 2 by means of the simulated size of a nominal 5% Wald test under different model designs. We consider two basic cross-section

TABLE 3. Simulated size t-test $N = 100$, model (6), FARIMA$(0,\theta_0,0)$. Exponential innovations. Nominal size 5%.

θ_0	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	4.36	2.81	2.84	2.37	9.04
	10	4.65	3.74	2.88	2.84	14.70
	5	6.09	5.24	3.63	3.36	22.45
	4	7.14	6.92	4.49	4.26	27.06
	3	7.62	10.04	6.30	6.28	32.13
0.6	20	5.77	5.69	5.16	5.20	12.20
	10	6.85	7.54	6.59	6.64	20.45
	5	6.36	8.22	6.05	6.70	31.71
	4	7.03	9.26	6.92	7.49	37.52
	3	6.85	9.92	6.50	7.50	45.10
1.0	20	6.14	5.07	5.04	5.10	7.04
	10	6.46	5.05	5.17	5.20	9.11
	5	7.11	4.44	4.82	4.80	12.62
	4	6.88	4.25	4.44	4.47	14.48
	3	7.13	4.45	4.76	4.77	18.39
1.4	20	5.84	4.98	4.88	4.91	6.45
	10	6.59	5.30	5.44	5.55	8.03
	5	6.88	6.13	5.80	5.94	11.03
	4	6.69	5.83	5.41	5.55	11.99
	3	7.08	6.21	5.98	6.02	14.46

sizes, $N = 100, 200$, and five different values for the time series dimension $T \in \{3,4,5,10,20\}$. First, we consider the FARIMA$(0,\theta_{10},0)$ model (6) with $p = 1$ and $\theta_{10} \in \{0.0, 0.6, 1.0, 1.4\}$ for Gaussian ε_{it} and also for highly asymmetric exponential innovations for which the Gaussian standardization $(BCB)_0$ should be inappropriate. Then, we consider the FARIMA$(1,\theta_0,0)$ model (7) with $p = 2$ for Gaussian ε_{it} and the same values of θ_0 with $\theta_{20} \in \{-0.5, 0.5\}$.

In Tables 1 and 2, we report the simulated size for the t-test based on our five possible standardizations for the Gaussian FARIMA$(0,\theta_0,0)$ model and $N = 100$ and 200, respectively. Except for the asymptotic $T \to \infty$ standardization (which is very oversized except for the largest value $T = 20$ and the most persistent models), all provide reasonable approximations to the nominal 5% size for both values of N. The $T \to \infty$ standardization is particularly problematic in the zero-persistence case $\theta_0 = 0$, likely due to bias in estimation based on initially differenced data and the higher variance as reported in Figure 1.

Tables 3 and 4 deal with the same design but using (centred) exponential innovations for which the t-tests based on the first three finite T standardizations work similarly as for Gaussian innovations. However, the $(BCB)_0$ standardization provides a marginally worse simulated size for both values of N, while the asymptotic one remains unreliable.

TABLE 4. Simulated size t-test $N = 200$, model (6), FARIMA$(0, \theta_0, 0)$. Exponential innovations. Nominal size 5%.

θ_0	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	3.06	2.86	2.24	2.51	9.45
	10	3.61	3.71	2.72	2.77	14.44
	5	4.23	4.79	3.12	3.26	22.36
	4	4.84	7.18	4.06	4.45	27.08
	3	5.86	9.97	5.43	6.01	32.36
0.6	20	5.41	5.93	5.30	5.44	12.33
	10	5.55	6.63	5.89	5.88	19.49
	5	5.85	8.77	6.66	7.07	32.26
	4	5.60	9.50	6.59	7.16	37.68
	3	5.62	11.16	6.94	7.98	46.23
1.0	20	5.59	5.09	4.97	5.03	7.05
	10	5.79	4.77	4.79	4.86	8.57
	5	5.99	4.72	4.71	4.74	12.18
	4	6.43	4.82	4.75	4.79	14.67
	3	6.48	4.39	4.59	4.84	18.01
1.4	20	5.80	5.13	5.12	5.12	6.68
	10	5.24	5.26	4.98	4.89	7.64
	5	6.31	6.26	5.83	5.84	10.77
	4	6.39	6.89	6.13	6.25	12.97
	3	6.22	6.90	6.02	6.10	15.19

Tables 5 and 6 provide the equivalent results for the Wald test corresponding to testing the Gaussian FARIMA$(1, \theta_{10}, 0)$ model, with tables named "a" for $\theta_{20} = 0.5$, and tables "b" for $\theta_{20} = -0.5$. Here, the asymptotic standardization is again unable to approximate the actual variability of the estimates, with the exception of the (simultaneously) largest N, T, and θ_{10}. The standardizations based on finite T asymptotic theory have problems for the smallest N, the performance of BCB being inferior to that of C or B, which in turn is similar to that of $(BCB)_0$.

For the PMLE of β_0, we consider the same experimental design and specifications as for Tables 1–2, while for each i, x_{it} is generated as a univariate Gaussian FARIMA$(0, \theta_x, 0)$ time series with $\theta_x = 0.6$ (Tables 7a and 7b) and $\theta_x = 1.0$ (Tables 8a and 8b) with innovations having variance equal to that of ε_{it}, the results not being very sensitive to this choice. We do not report parallel results for PMLE of θ_0, because they are very similar to those of Tables 1 and 2, but consider two alternative estimates of β_0. These are OLS and GLS estimates based on first differences Δy_i and Δx_i, with the GLS estimate using the OLS residuals to estimate the covariance matrix of $\Delta y_i - \Delta x_i \beta$. We report simulated bias for $\beta_0 = 1$ for the three estimates, relative efficiency with respect to the PMLE based on empirical mean square error, and simulated size of a Wald test for β_0 with 5% nominal size using the results of Theorem 2 for the PMLE and direct adaptations for the GLS and OLS estimates.

TABLE 5a. Simulated size Wald-test $N = 100$, model (7), FARIMA$(1,\theta_{10},0)$, $\theta_{20} = 0.5$. Gaussian innovations. Nominal size 5%

θ_{10}	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.5	20	14.55	8.18	8.89	8.26	31.89
	10	12.60	7.93	7.21	7.90	34.81
	5	14.16	7.69	6.82	7.42	41.45
	4	17.46	8.46	8.13	8.18	44.60
	3	21.10	9.90	9.95	9.97	49.57
0.6	20	11.68	10.12	10.62	10.56	23.54
	10	14.94	12.31	13.25	13.10	34.56
	5	14.07	11.63	12.35	12.29	39.08
	4	15.07	12.56	13.40	13.49	42.54
	3	15.59	12.75	14.27	14.43	43.54
1.0	20	10.80	9.48	9.56	9.63	18.64
	10	14.71	12.39	12.94	13.04	29.15
	5	13.37	11.02	11.55	11.75	33.68
	4	14.23	11.75	12.48	12.43	36.98
	3	15.11	12.34	13.45	13.21	38.83
1.4	20	10.54	9.27	9.55	9.59	18.49
	10	15.13	12.66	13.23	13.20	28.95
	5	13.19	10.90	11.44	11.56	33.22
	4	13.54	11.51	12.10	12.13	35.91
	3	14.06	11.67	12.45	12.41	37.66

In terms of bias, the three estimates perform similarly, both in terms of magnitude and sign of the bias. The sign can shift with the values of T and θ_0 and the magnitude mainly falls with increasing T and N. The relative efficiency of GLS is typically higher than 97% for the smallest values of T, but deteriorates for $T = 20$ and $N = 100$, possibly due to the lack of precision in the estimation of a moderately large covariance matrix of residuals, the results improving substantially for $N = 200$. OLS estimates can be very inefficient for $\theta_0 = 0.0$ due to first differencing introducing strong (negative) correlation in the regression error terms with efficiency deteriorating as T increases from 0.68 (resp. 0.61) for $T = 3$ to 0.36 (resp. 0.13) for $T = 20$ and $\theta_x = 0.6$ (resp. 1.0). However, for other values of θ_0, the relative performance of OLS is more stable across T. For $\theta_0 = 1$, OLS estimation performs similarly to PMLE for both $\theta_x = 0.6$ and 1.0, as in this case first differencing is prewhitening exactly the errors. For the other values of θ_0, the relative performance of OLS estimation worsens for the larger value of θ_x.

Simulated size is very good for the Wald tests based on the PMLE and OLS estimate, improving with N, but not being monotone with T. The GLS-based test is seriously oversized for the largest values of T (and small N), related with the efficiency problems associated with the dimension of the residual sample covariance matrix.

TABLE 5b. Simulated size Wald-test $N = 100$, model (7), FARIMA$(1, \theta_{10}, 0)$, $\theta_{20} = -0.5$. Gaussian innovations. Nominal size 5%

θ_{10}	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	6.39	3.16	3.84	3.21	10.60
	10	8.49	3.79	5.05	3.79	17.71
	5	12.27	4.20	5.98	3.86	28.34
	4	14.05	5.12	6.29	4.38	32.02
	3	16.73	6.82	7.84	5.61	37.52
0.6	20	6.62	5.10	5.24	5.19	13.34
	10	7.38	5.23	5.60	5.44	22.89
	5	11.00	4.81	6.46	5.09	42.54
	4	14.24	5.39	7.62	5.38	50.57
	3	21.55	6.41	10.50	6.37	64.83
1.0	20	6.32	5.02	5.12	5.12	7.97
	10	6.70	5.20	5.26	5.34	11.33
	5	6.08	4.44	4.74	4.73	20.17
	4	5.88	4.08	4.23	4.20	27.10
	3	7.14	4.39	5.02	4.68	41.06
1.4	20	6.32	5.03	5.16	5.17	7.09
	10	6.52	5.14	5.15	5.16	9.39
	5	6.44	4.55	4.89	5.06	14.44
	4	6.54	4.62	4.76	4.87	18.59
	3	6.51	4.52	5.01	4.89	29.18

7. GENERALIZATION OF INITIAL CONDITIONS

Condition (2) is quite drastic, requiring $\varepsilon_{it} = 0$ for all $t < 0$, motivated by nonstationary versions of fractional models (6) and (7) whereas for the nonstationary AR(1) covered by (5) it is only necessary to impose a condition on ε_{it} for a single t (for an extensive discussion of initial values in the AR(1) case see Anderson and Hsiao, 1981). The time series version of (2) (see e.g., Hualde and Robinson, 2011a) has sometimes been imposed in the fractional literature with little discussion, whereas in fact it plays a crucial role in model specification, and consequent asymptotic statistical properties. Also, in a time series context, Johansen and Nielsen (2010) instead treat initial conditions as bounded constants.

Here, we consider the impact on our panel model by replacing (2) by the condition

$$\varepsilon_{it} = 0, \text{ a.s. } t < -m, \tag{23}$$

for a specified positive integer m. This condition was considered in a time series context by Hualde and Robinson (2011b) and Johansen and Nielsen (2016). The larger m, the closer one appears to get to the initial-condition-free setup usual in the stationary time series literature. However, (23) for $m \geq 1$ is not really a milder assumption than (2), rather, it replaces $\varepsilon_{it} = 0$, $-1 \leq t \leq -m$ by the assumption

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

TABLE 6a. Simulated size Wald-test $N = 200$, model (7), FARIMA$(1,\theta_{10},0)$, $\theta_{20} = 0.5$. Gaussian innovations. Nominal size 5%

θ_{10}	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	12.82	8.46	8.98	8.41	34.01
	10	10.37	7.83	6.84	7.68	37.47
	5	10.33	7.27	6.35	7.11	43.07
	4	11.67	7.36	6.55	6.86	46.65
	3	14.81	7.52	7.42	7.41	52.36
0.6	20	9.19	8.32	8.32	8.33	23.02
	10	13.82	12.50	12.92	12.96	37.82
	5	13.49	12.27	12.60	12.72	44.75
	4	14.19	12.88	13.16	13.15	46.27
	3	15.52	13.73	14.39	14.30	45.30
1.0	20	8.01	7.03	7.28	7.32	17.32
	10	13.73	12.23	12.82	12.79	31.59
	5	13.10	11.60	12.25	12.36	39.93
	4	13.86	12.44	12.75	12.79	41.73
	3	14.21	12.66	13.17	13.18	41.85
1.4	20	7.89	6.94	7.18	7.18	16.97
	10	13.69	12.54	12.97	12.90	31.07
	5	13.24	11.70	12.18	12.21	39.34
	4	13.66	12.22	12.69	12.69	40.79
	3	13.61	12.10	12.60	12.51	40.75

that these ε_{it} are iid across i with the same distribution as the iid nondegenerate ε_{it} for $t \geq 0$. We retain the initial condition (3) on x_{it}. We now require that $T \geq m+2$.

Corresponding to (23), we write in place of (1)

$$\lambda_{t+m}(L;\theta_0)\left(y_{it} - \zeta_i - x'_{it}\beta_0\right) = \varepsilon_{it}, \tag{24}$$

for $i = 1,\ldots,N$, $t = 0,1,\ldots,T$, where $\lambda_{t+m}(L;\theta) = \sum_{j=0}^{t+m}\lambda_j(\theta)L^j$ truncates $\lambda(L;\theta)$. Now redefine

$$\upsilon_{it} = \lambda_{t+m}^{-1}(L;\theta_0)\,\varepsilon_{it}, \tag{25}$$

and (cf. (11))

$$\begin{aligned}
z_{it}(\theta,\beta) &= \tau_{t-1}(L;\theta)\,\Delta x'_{it}(\beta_0-\beta) + \tau_{t-1}(L;\theta)\,\Delta\upsilon_{it}\\
&= \tau_{t+m}(L;\theta)\,\Delta x'_{it}(\beta_0-\beta) + \left(\tau_{t-1}(L;\theta) - \tau_{t+m}(L;\theta)\right)\Delta x'_{it}(\beta_0-\beta)\\
&\quad + \tau_{t+m}(L;\theta)\,\Delta\upsilon_{it} + \left(\tau_{t-1}(L;\theta) - \tau_{t+m}(L;\theta)\right)\Delta\upsilon_{it}\\
&= \lambda_{t+m}(L;\theta)\,x'_{it}(\beta_0-\beta) - \sum_{j=t}^{t+m}\tau_j(\theta)\,\Delta x'_{i,t-j}(\beta_0-\beta)\\
&\quad + \lambda_{t+m}(L;\theta)\,\upsilon_{it} - \sum_{j=t}^{t+m}\tau_j(\theta)\,\Delta\upsilon_{it-j}\\
&= \lambda_{t+m}(L;\theta)\left\{\upsilon_{it} - x'_{it}(\beta-\beta_0)\right\} - \tau_t^{t+m}(\theta)'\,\Delta\upsilon_i^m + \tau_t(\theta)\,x'_{i0}(\beta-\beta_0),
\end{aligned}$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

TABLE 6b. Simulated size Wald-test $N = 200$, model (7), FARIMA$(1,\theta_{10},0)$, $\theta_{20} = -0.5$. Gaussian innovations. Nominal size 5%

θ_{10}	T	BCB	C	B	$(BCB)_0$	$Asymp$
0.0	20	5.06	3.63	3.74	3.58	11.42
	10	6.74	3.53	4.08	3.58	17.83
	5	9.25	4.22	5.10	3.78	28.96
	4	10.53	4.12	5.04	3.67	32.00
	3	12.35	5.51	6.04	4.27	37.92
0.6	20	6.01	5.19	5.16	5.25	13.48
	10	6.22	5.09	5.08	5.10	22.92
	5	8.29	5.24	6.00	5.22	41.33
	4	9.57	4.79	6.03	4.88	49.91
	3	15.32	5.71	8.42	5.69	64.41
1.0	20	5.86	5.24	5.08	5.13	8.11
	10	5.68	4.77	4.93	4.96	10.99
	5	5.48	4.79	4.77	4.78	19.55
	4	5.48	4.36	4.51	4.51	25.35
	3	5.90	4.59	4.80	4.69	40.57
1.4	20	5.77	5.23	5.28	5.27	7.24
	10	5.76	4.85	4.91	4.95	8.71
	5	5.51	4.73	4.90	4.83	14.17
	4	5.66	4.82	4.95	4.93	17.52
	3	5.97	4.80	5.00	4.94	28.66

where

$$\tau_t^{t+m}(\theta) = (\tau_t(\theta), \ldots, \tau_{t+m}(\theta))'$$

and

$$\Delta v_i^m = \left(\Delta v_{i0}, \ldots, \Delta v_{i,-m}\right)'.$$

Thus,

$$\begin{aligned} z_{it}(\theta,\beta_0) &= \lambda_{t+m}(L;\theta)\, v_{it} - \tau_t^{t+m}(\theta)'\,\Delta v_i^m, \\ z_{it}(\theta_0,\beta_0) &= \varepsilon_{it} - \tau_t^{t+m}(\theta_0)'\,\Delta v_i^m. \end{aligned} \qquad \textbf{(26)}$$

Now for $-m \le t \le 0$,

$$\begin{aligned} \Delta v_{it} &= \lambda_{t+m}^{-1}(L;\theta_0)\,\varepsilon_{it} - \lambda_{t+m}^{-1}(L;\theta_0)\,\varepsilon_{i,t-1} \\ &= \sum_{0}^{m+t} \phi_j(\theta_0)\,\varepsilon_{i,t-j} - \sum_{0}^{m+t-1} \phi_j(\theta_0)\,\varepsilon_{i,t-j-1} \\ &= \varepsilon_{it} + \sum_{1}^{m+t} \left(\phi_j(\theta_0) - \phi_{j-1}(\theta_0)\right)\varepsilon_{i,t-j}, \end{aligned} \qquad \textbf{(27)}$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

TABLE 7a. Estimation of slope coefficient β_0 : $N = 100$, $x_{it} \sim \mathrm{FARIMA}(0, 0.6, 0)$, $\lambda(L;\theta) \sim \mathrm{FARIMA}(0, \theta_0, 0)$. Gaussian innovations. Nominal size 5%

		Bias			Rel. efficiency		Size Wald-test		
θ_0	T	PMLE	GLS	OLS	GLS	OLS	PMLE	GLS	OLS
0.0	20	0.0002	0.0001	0.0001	0.7957	0.3511	5.47	12.39	5.57
	10	–0.0003	–0.0003	–0.0003	0.9081	0.4514	5.24	8.36	5.52
	5	–0.0009	–0.0009	–0.0009	0.9574	0.5773	5.58	6.75	5.10
	4	0.0004	0.0004	0.0007	0.9633	0.6124	5.42	6.73	5.17
	3	–0.0002	–0.0002	–0.0018	0.9726	0.6743	5.69	6.28	5.87
0.6	20	0.0001	0.0000	0.0001	0.8551	0.8574	5.77	11.57	5.67
	10	–0.0002	–0.0002	–0.0002	0.9229	0.8629	5.40	7.82	5.39
	5	–0.0004	–0.0004	–0.0004	0.9533	0.8752	5.49	6.64	5.00
	4	0.0011	0.0010	0.0011	0.9667	0.8828	5.14	5.92	5.01
	3	–0.0004	–0.0003	–0.0010	0.9723	0.8969	6.03	6.66	5.85
1.0	20	0.0000	–0.0000	0.0000	0.8622	1.0005	5.69	11.23	5.71
	10	–0.0001	–0.0002	–0.0001	0.9260	1.0006	5.30	7.81	5.35
	5	–0.0001	–0.0001	–0.0001	0.9511	1.0032	5.40	6.55	5.35
	4	0.0013	0.0011	0.0013	0.9664	1.0036	5.15	6.15	5.05
	3	–0.0007	–0.0006	–0.0006	0.9718	1.0061	6.06	6.83	5.82
1.4	20	–0.0000	0.0000	–0.0000	0.8532	0.8255	5.44	11.32	5.62
	10	–0.0001	–0.0002	–0.0000	0.9233	0.8135	5.26	7.70	5.22
	5	0.0000	0.0000	0.0002	0.9500	0.8169	5.36	6.62	5.29
	4	0.0012	0.0011	0.0014	0.9654	0.8012	5.03	5.92	5.02
	3	–0.0006	–0.0006	–0.0004	0.9715	0.8114	5.94	7.06	5.77

TABLE 7b. Estimation of slope coefficient β_0 : $N = 200$, $x_{it} \sim \mathrm{FARIMA}(0, 0.6, 0)$, $\lambda(L;\theta) \sim \mathrm{FARIMA}(0, \theta_0, 0)$. Gaussian innovations. Nominal size 5%

		Bias			Rel. efficiency		Size Wald-test		
θ_0	T	PMLE	GLS	OLS	GLS	OLS	PMLE	GLS	OLS
0.0	20	–0.0003	–0.0004	0.0001	0.9081	0.3571	5.18	7.78	5.03
	10	0.0002	0.0002	0.0002	0.9467	0.4409	5.15	6.49	5.32
	5	–0.0003	–0.0004	–0.0006	0.9756	0.5654	5.20	5.71	5.32
	4	0.0000	0.0002	0.0002	0.9770	0.6092	5.44	5.76	5.41
	3	–0.0001	–0.0001	–0.0005	0.9880	0.6767	5.29	5.57	5.00
0.6	20	–0.0001	–0.0001	0.0000	0.9208	0.8541	5.03	7.43	4.97
	10	–0.0000	–0.0001	–0.0001	0.9503	0.8553	5.30	6.75	5.32
	5	–0.0003	–0.0004	–0.0004	0.9715	0.8657	5.18	5.94	5.33
	4	0.0004	0.0005	0.0004	0.9769	0.8786	5.42	5.86	5.46
	3	–0.0007	–0.0007	–0.0008	0.9855	0.8907	5.18	5.54	5.10
1.0	20	0.0000	0.0000	0.0000	0.9220	1.0004	5.02	7.40	4.99
	10	–0.0002	–0.0002	–0.0002	0.9532	1.0011	5.08	6.50	5.09
	5	–0.0002	–0.0003	–0.0002	0.9704	1.0001	5.65	6.20	5.62
	4	0.0005	0.0006	0.0005	0.9766	1.0002	5.19	5.68	5.14
	3	–0.0009	–0.0009	–0.0009	0.9850	1.0032	4.99	5.51	4.96
1.4	20	0.0000	0.0001	0.0001	0.9210	0.8289	4.91	7.43	5.09
	10	–0.0002	–0.0003	–0.0003	0.9522	0.8202	5.00	6.13	5.37
	5	–0.0001	–0.0002	–0.0001	0.9718	0.8110	5.45	6.23	5.29
	4	0.0005	0.0006	0.0005	0.9778	0.8237	5.15	5.39	4.92
	3	–0.0010	–0.0009	–0.0010	0.9866	0.8296	5.09	5.44	4.80

TABLE 8a. Estimation of slope coefficient β_0 : $N = 100$, $x_{it} \sim$ FARIMA$(0, 1.0, 0)$, $\lambda(L;\theta) \sim$ FARIMA$(0, \theta_0, 0)$. Gaussian innovations. Nominal size 5%

		Bias			Rel. efficiency		Size Wald-test		
θ_0	T	PMLE	GLS	OLS	GLS	OLS	PMLE	GLS	OLS
0.0	20	0.0001	0.0001	0.0001	0.6832	0.1338	5.52	15.46	5.54
	10	–0.0002	–0.0001	–0.0003	0.8884	0.2534	5.60	8.60	5.42
	5	–0.0006	–0.0006	–0.0010	0.9562	0.4423	6.04	7.13	5.19
	4	0.0001	0.0002	0.0002	0.9611	0.5005	5.59	6.73	5.40
	3	–0.0004	–0.0004	–0.0021	0.9765	0.6059	5.95	6.49	5.69
0.6	20	0.0001	0.0001	0.0001	0.8376	0.6514	5.79	11.93	5.78
	10	–0.0003	–0.0003	–0.0003	0.9190	0.7220	5.25	8.09	5.34
	5	–0.0005	–0.0004	–0.0007	0.9536	0.7984	5.60	6.76	5.20
	4	0.0008	0.0007	0.0008	0.9691	0.8313	5.34	6.27	5.04
	3	–0.0005	–0.0004	–0.0012	0.9759	0.8584	5.87	6.31	5.56
1.0	20	0.0001	–0.0000	0.0001	0.8605	1.0020	5.58	11.48	5.58
	10	–0.0003	–0.0004	–0.0003	0.9228	1.0017	5.38	7.62	5.36
	5	–0.0004	–0.0004	–0.0004	0.9519	1.0038	5.28	6.80	5.21
	4	0.0012	0.0010	0.0012	0.9671	1.0050	5.02	5.89	5.00
	3	–0.0008	–0.0006	–0.0008	0.9735	1.0081	6.16	6.64	6.03
1.4	20	0.0000	–0.0000	–0.0000	0.8397	0.6030	5.50	11.68	5.55
	10	–0.0002	–0.0003	–0.0004	0.9187	0.6422	5.32	7.82	5.19
	5	–0.0002	–0.0002	–0.0001	0.9495	0.7044	5.42	6.65	5.28
	4	0.0013	0.0012	0.0015	0.9646	0.7036	5.08	6.19	5.36
	3	–0.0008	–0.0006	–0.0005	0.9719	0.7589	6.15	7.11	6.05

TABLE 8b. Estimation of slope coefficient β_0 : $N = 200$, $x_{it} \sim$ FARIMA$(0, 1.0, 0)$, $\lambda(L;\theta) \sim$ FARIMA$(0, \theta_0, 0)$. Gaussian innovations. Nominal size 5%

		Bias			Rel. efficiency		Size Wald-test		
θ_0	T	PMLE	GLS	OLS	GLS	OLS	PMLE	GLS	OLS
0.0	20	–0.0002	–0.0002	0.0000	0.8588	0.1335	4.91	8.39	5.14
	10	0.0000	0.0000	0.0003	0.9473	0.2452	5.03	6.40	5.42
	5	–0.0001	–0.0001	–0.0006	0.9794	0.4288	5.51	5.95	5.18
	4	–0.0004	–0.0003	–0.0003	0.9796	0.5010	5.41	5.68	5.26
	3	0.0002	0.0002	–0.0002	0.9869	0.5986	5.39	5.54	5.24
0.6	20	–0.0002	–0.0002	–0.0001	0.9182	0.6546	5.19	7.71	5.22
	10	0.0001	0.0001	0.0001	0.9499	0.7128	5.23	6.51	5.42
	5	–0.0002	–0.0003	–0.0004	0.9763	0.7862	5.24	5.83	5.12
	4	–0.0000	0.0000	0.0000	0.9792	0.8215	5.61	6.01	5.35
	3	–0.0006	–0.0006	–0.0007	0.9851	0.8537	5.23	5.65	5.18
1.0	20	–0.0001	–0.0001	–0.0001	0.9229	1.0009	5.19	7.51	5.16
	10	–0.0000	–0.0001	–0.0000	0.9518	1.0025	5.19	6.61	5.12
	5	–0.0003	–0.0004	–0.0003	0.9733	1.0018	5.51	6.14	5.37
	4	0.0002	0.0003	0.0002	0.9770	1.0007	5.40	5.86	5.36
	3	–0.0011	–0.0010	–0.0010	0.9842	1.0035	5.26	5.68	5.25
1.4	20	0.0000	0.0000	0.0000	0.9173	0.6026	5.04	7.63	4.82
	10	–0.0002	–0.0002	–0.0002	0.9517	0.6505	5.07	6.43	5.30
	5	–0.0002	–0.0003	–0.0002	0.9710	0.6980	5.52	6.34	5.46
	4	0.0004	0.0005	0.0002	0.9769	0.7424	5.32	5.73	5.02
	3	–0.0012	–0.0011	–0.0014	0.9857	0.7703	4.98	5.34	4.89

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

where the moving average weights $\phi_j(\theta)$ are defined by

$$\lambda(L;\theta)^{-1} = \sum_0^\infty \phi_j(\theta) L^j \tag{28}$$

and the second term in (27) is absent for $t = -m$. Then, we can write

$$\Delta v_i^m = U(\theta_0)\varepsilon_i^m,$$

where we introduce

$$\varepsilon_i^m = \left(\varepsilon_{i0}, \varepsilon_{i,-1}, \ldots, \varepsilon_{i,-m}\right)'$$

and the $(m+1) \times (m+1)$ upper-triangular matrix

$$U(\theta) = \begin{pmatrix} 1 & \phi_1(\theta) - 1 & \phi_2(\theta) - \phi_1(\theta) & \ldots & \phi_m(\theta) - \phi_{m-1}(\theta) \\ 0 & 1 & \phi_1(\theta) - 1 & \ldots & \phi_{m-1}(\theta) - \phi_{m-2}(\theta) \\ 0 & 0 & 1 & \ldots & \phi_{m-2}(\theta) - \phi_{m-3}(\theta) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \phi_1(\theta) - 1 \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

It follows from (26) that

$$z_i(\theta_0, \beta_0) = \varepsilon_i - \tau^m(\theta_0) U(\theta_0) \varepsilon_i^m,$$

where

$$\tau^m(\theta) = \left(\tau_1^{1+m}(\theta), \ldots, \tau_T^{T+m}(\theta)\right)'.$$

Thus, $z_i(\theta_0, \beta_0)$ has covariance matrix $\sigma_0^2 \Omega(\theta_0)$, with (14) replaced by

$$\Omega(\theta) = I_T + \tau^m(\theta_0) U(\theta_0) U(\theta_0)' \tau^m(\theta_0)'. \tag{29}$$

We can thus employ the same definition of pseudo likelihood $L(\theta, \beta)$ (16) with the same formula (15) for $\widehat{\sigma}^2(\theta, \beta)$ in both cases redefining $\Omega(\theta)$ as (29), and thence the formula (17) for the estimates $\widehat{\theta}$, $\widehat{\beta}$ that are now based on the initial condition (23) in place of (2). Writing $W_m(\theta) = \tau^m(\theta) U(\theta)$, note that by Woodbury's identity

$$\Omega(\theta)^{-1} = I_T - W_m(\theta) \left(I_{m+1} + W_m(\theta)' W_m(\theta)\right)^{-1} W_m(\theta)'$$

and by Silvester's identity

$$|\Omega(\theta)| = \left|I_{m+1} + W_m(\theta)' W_m(\theta)\right|,$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

thereby reducing relevant computations to the inverse and determinant of a $(m+1)\times(m+1)$ matrix. Asymptotic properties analogous to those of Theorems 1 and 2 will follow under slightly modified assumptions.

If asymptotics with $T \to \infty$ are instead pursued it may be possible to allow $m \to \infty$ simultaneously. But in practice m must be chosen and the implications of increasing m will typically be an increase in the variance of y_{it} conditional on ζ_i and x_{it} – in particular there will be a monotonic increase with m if all moving average weights are nonnegative, as in the AR(1) case of (5) with positive θ_{01} (as under a unit root) or the long memory case $\theta_0 > 0$ of (6). The models for varying m, $m = 0, 1, \ldots$, are nonnested and, for given $\lambda(L;\theta)$, m might be determined by a suitable model-selection procedure.

There are other possible implications for the choice of initial conditions that might be studied in our model setting. Hahn (1999) compared semiparametric efficiency bounds under rival initial conditions. Moon, Perron, and Phillips (2007) compared initial conditions in unit root testing in panel models with incidental trends, where T diverges with N. Instead of the zero initial conditions assumed above, one might consider ones that are heterogeneous across i.

To explore the latter possibility, assume now

$$\varepsilon_{i,-1} = \xi_i, \ \varepsilon_{it} = 0, \text{ a.s. } t < -1, \tag{30}$$

where as we plan to eliminate it, no assumptions are required on ξ_i, as was the case with ζ_i. (An analogous argument to that below can apply to the alternative setting $\varepsilon_{i0} = \xi_i$, $\varepsilon_{it} = 0$, a.s. $t < 0$, and to ones with more than one heterogeneous initial condition.) With the notation in (28), $v_{it} = \lambda_t^{-1}(L;\theta_0)\varepsilon_{it} = \sum_{j=0}^{\infty}\phi_j(\theta_0)\varepsilon_{i,t-j}$, see (8), we have $y_{it} - \zeta_i - x_{it}'\beta_0 = \psi_{it}(\theta_0) + \phi_{t+1}(\theta_0)\xi_i$, where $\psi_{it}(\theta) = \sum_{j=0}^{t}\phi_j(\theta)\varepsilon_{i,t-j}$. Thence, from (9),

$$\Delta y_{it} - \Delta x_{it}'\beta_0 = \Delta\psi_{it}(\theta_0) + \Delta\phi_{t+1}(\theta_0)\xi_i, \quad t = 1, \ldots, T. \tag{31}$$

If $\Delta\phi_{t+1}(\theta_0) = 0$, ξ_i is eliminated, but assuming that is not the case, form $\left(\Delta y_{it} - \Delta x_{it}'\beta_0\right)/\Delta\phi_{t+1}(\theta_0) = \Delta\psi_{it}(\theta_0)/\Delta\phi_{t+1}(\theta_0) + \xi_i$, whence after further differencing

$$\Delta\left\{\left(\Delta y_{it} - \Delta x_{it}'\beta_0\right)/\Delta\phi_{t+1}(\theta_0)\right\} = \Delta\left\{\Delta\psi_{it}(\theta_0)/\Delta\phi_{t+1}(\theta_0)\right\}, \ t = 2, \ldots, T. \tag{32}$$

Noting that $\Delta\psi_{it}(\theta) = \varepsilon_{it} + \sum_{j=1}^{t}\Delta\phi_j(\theta)\varepsilon_{i,t-j}$, after rearrangement the right-hand side of (32) can be expressed as

$$\Delta\left\{\Delta\psi_{it}(\theta_0)/\Delta\phi_{t+1}(\theta_0)\right\} = \sum_{j=0}^{t}\chi_{jt}(\theta_0)\varepsilon_{i,t-j},$$

where

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

$$\chi_{0t}(\theta) = 1/\Delta\phi_{t+1}(\theta),$$
$$\chi_{1t}(\theta) = \left(\frac{\Delta\phi_1(\theta)}{\Delta\phi_{t+1}(\theta)} - \frac{1}{\Delta\phi_t(\theta)}\right),$$
$$\chi_{jt}(\theta) = \left(\frac{\Delta\phi_j(\theta)}{\Delta\phi_{t+1}(\theta)} - \frac{\Delta\phi_{j-1}(\theta)}{\Delta\phi_t(\theta)}\right), \ j = 2,\ldots,t.$$

Now form the $(T-1)\times 1$ vectors

$$w_i(\theta,\beta) = \left(\Delta\left\{\left(\Delta y_{i2} - \Delta x_{i2}'\beta\right)/\Delta\phi_3(\theta)\right\},\ldots,\Delta\left\{\left(\Delta y_{iT} - \Delta x_{iT}'\beta\right)/\Delta\phi_{T+1}(\theta)\right\}\right)',$$
$$\gamma_i(\theta) = \left(\sum_{j=0}^{2}\chi_{j2}(\theta)\varepsilon_{i,t-j},\ldots,\sum_{j=0}^{T}\chi_{jT}(\theta)\varepsilon_{i,t-j}\right)'.$$

Now $\gamma_i(\theta_0)$ has zero mean vector and covariance matrix $\sigma_0^2\Gamma(\theta_0)$, where $\Gamma(\theta)$ has (l,m)th element

$$\sum_{j=0}^{\min(l,m)+1}\chi_{j,l+1}(\theta)\,\chi_{j+|l-m|,m+1}(\theta).$$

A PMLE of θ_0,β_0, based on the transformed model (32) and under the heterogeneous initial condition (30), thus minimizes $|\Gamma(\theta)|^{1/(T-1)}\tilde{\sigma}^2(\theta,\beta)$, where $\tilde{\sigma}^2(\theta,\beta) = \frac{1}{N(T-1)}\sum_{i=1}^{N} w_i'(\theta,\beta)\,\Gamma(\theta)^{-1}w_i(\theta,\beta)$, cf. (15), (16). Its asymptotic properties can be derived from similar arguments to those used in proving Theorems 1 and 2 with the important exception of the identifiability argument. With respect to this, note that under the simple fractional model (6) for $\lambda_t(L;\theta)$ with a fractional unit root, $\theta_0 = 1$, in (1) we have $\Delta\phi_{t+1}(\theta_0) = 0$ for all t so that ξ_i is eliminated from (31). But the practitioner would not know that $\theta_0 = 1$ and so would be inclined to use the procedure based on (32). But this clearly breaks down at $\theta = 1$ so Θ must exclude $\theta = 1$ and so it must be assumed that $\theta_0 \neq 1$. However, we can cover the possibility of a fractional unit root under, say, the FARIMA$(1,\theta_1,0)$ structure (7), where Θ can be chosen to include $\theta_1 = 1$ so long as it also entails $0 < |\theta_2| < 1$.

8. FINAL COMMENTS

We have discussed inference in panel data models with general parametric dynamics, individual effects and possible linearly involved explanatory variables, with asymptotic theory based on cross-sectional dimension N diverging but temporal dimension T remaining fixed. For $T \to \infty$ similarly desirable asymptotic properties are available but typically with simpler formulae for the large sample variance matrix of estimates. Obviously, one might wish to consider a modified or more general panel data model, and given the literature and range of potential applications, the possibilities are too numerous to list in full. But we briefly mention some possible developments in connection with our model.

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

1. An alternative form of inference is prompted by the rather cumbersome covariance matrix of Theorem 2. A suitable bootstrap procedure can avoid this, and since it may achieve an Edgeworth correction is liable to have better finite sample properties than first-order asymptotic inference.
2. Though Theorem 2 does not assume normality of the ε_{it}, if normality does not hold one expects greater efficiency to be achievable by maximum likelihood estimates with respect to a correctly specified nonnormal distribution. However, not only may these be inconsistent if the density of ε_{it} is misspecified but also the joint density of the $\varepsilon_{it} - \tau_t(\theta_0)\varepsilon_{i0}$, $t = 1, \ldots, T$, is a convolution of the underlying density of ε_{it}, and is thus potentially complicated. In principle, at least, it would be possible to construct semiparametric estimates that achieve equal efficiency without parameterizing the density of ε_{it}, being adaptive in the sense of Stone (1975) and involving nonparametric estimation of the relevant density or score function.
3. Our iid (across i) assumption on ε_{it}, x_{it} allows strong consistency of estimates to be established under minimal moment conditions. But there is concern for robustness to departures from some of our assumptions. For example, it should be straightforward to extend our proof of consistency to allow for unconditional heteroscedasticity across i of the ε_{it} (heteroscedasticity across t can be incorporated in the model for fixed T). But our limiting covariance matrix estimate in Theorem 2 can be robustified with respect to unconditional heteroscedasticity across i via the nonparametric approach of Eicker (1963). On the other hand, unanticipated heteroscedasticity entails loss of efficiency, so more ambitiously one could develop asymptotically efficient estimates in a semiparametric extension of our model with conditionally (on x_{it}) heteroscedastic ε_{it}, with conditional variances estimated by nonparametric smoothing (as in, say, Robinson, 1987). Our estimates should also be consistency-robust to cross-sectional dependence, but valid large-N, fixed T inference would likely require specification and estimation of a parametric model for the $N \times N$ cross-sectional covariance matrix of the ε_{it}, based perhaps on a factor or spatial mode1. Likewise relaxation of, respectively, identity of distribution and independence of the x_{it} across i would be possible. Though we do not assume the ε_{it} are independent of the x_{it}, if our strong exogeneity assumption were relaxed to orthogonality generalized-method-of-moments estimates can be considered, albeit with some loss in efficiency.
4. The linearity of the regression component $x_{it}'\beta_0$ reflects popular practice, but in our setting it is easily extended to a general nonlinear parametric component, since our general dynamics require implicitly defined extremum estimation in any case so our asymptotic proofs can be straightforwardly modified. Nonparametric regression would entail a more challenging extension, but here we might consider a series approximation to the regression com-

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

ponent, where, in asymptotic theory, the number of terms would increase slowly with N.

5. Arellano and Bonhomme (2012) considered a static regression model, i.e., with $\lambda_t(L;\theta_0) \equiv 1$ in (1), but with possibly AR errors ε_{it}, in which our individual effect term ζ_i is generalized to $z_{it}'\varphi_i$, where z_{it} is a vector of observable explanatory variables and φ_i a vector of unknown individual-specific parameters. They discussed *inter alia,* nonparametric identification and estimation of the conditional distribution of φ_i. It may be of some interest to develop an extension to our dynamic model. Of course, many parametric, semiparametric, and nonparametric approaches can be applied to panel data, bearing in mind the need to reconcile issues of correct specification, parsimony, and curse-of-dimensionality, and reflecting absolute and relative sizes of N and T.
6. Fractional modeling could be applied to errors of static models, in place of the popular ARMA modeling. In particular, issues of identification and estimation might be studied when the shocks to the errors comprise both a permanent and transitory component, as in e.g., Ejrnaes and Browning (2014).
7. Our ζ_i could also be extended to incidental or heterogeneous trends (see Moon and Phillips, 1999; Moon, Perron, and Phillips, 2007). Here, for example, a polynomial trend with individual-specific slope might be eliminated by higher order differencing and then a modification of our PMLE developed.

REFERENCES

Adenstedt, R.K. (1974) On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 6, 1095–1107.

Anderson, T.W. & C. Hsiao (1981) Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598–606.

Arellano, M. & S. Bonhomme (2012) Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies* 79, 987–1020.

Eicker, F. (1963) Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics* 34, 447–456.

Ejrnaes, M. & M. Browning (2014) The persistent-transitory representation for earnings processes. *Quantitative Economics* 5, 555–581.

Hahn, J. (1999) How informative is the initial condition in the dynamic panel model with fixed effects? *Journal of Econometrics* 93, 309–326.

Han, C. & P.C.B. Phillips (2013) First difference maximum likelihood and dynamic panel estimation. *Journal of Econometrics* 175, 35–45.

Hassler, U., M. Demetrescu, & A.L. Tarcolea (2011) Asymptotic normal tests for integration in panels with cross-dependent units. *Advances in Statistical Analysis* 95, 187–204.

Hsiao, C. (2014) *Analysis of Panel Data*, 3rd ed. Cambridge University Press.

Hsiao, C., M.H. Pesaran, & A.K. Tahmiscioglu (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107–150.

Hualde, J. & P.M. Robinson (2011a) Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* 39, 3152–3181.

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

Hualde, J. & P.M. Robinson (2011b) Supplement to Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics*. Available at https://projecteuclid.org/download/suppdf_1/euclid.aos/1330958676.

Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 2, 633–643.

Johansen, S. & M. Nielsen (2010) Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* 158, 51–66.

Johansen, S. & M. Nielsen (2016) The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time-series models. *Econometric Theory* 32, 1095–1139.

Moon, H.R., B. Perron, & P.C. Phillips (2007) Incidental trends and the power of panel unit root tests. *Journal of Econometrics* 141, 416–459.

Moon, H.R., B. Perron, & P.C.B. Phillips (2015) Incidental parameters and dynamic panel models. In B.H. Baltagi (ed.), *The Oxford Handbook of Panel Data*, pp. 111–148. Oxford University Press.

Moon, H.R. & P.C.B. Phillips (1999) Maximum likelihood estimation in panels with incidental trends. *Oxford Bulletin of Economics and Statistics* 61, 711–747.

Nagar, A.L. (1959) The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575–595.

Neyman, J. & E. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–31.

Robinson, P.M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891

Robinson, P.M. & C. Velasco (2015) Efficient inference on fractionally integrated panel data models with fixed effects. *Journal of Econometrics* 185, 435–452.

Robinson, P.M. & C. Velasco (2017) Inference on trending panel data. *Journal of Econometrics*, forthcoming.

Stone, C.J. (1975) Adaptive maximum likelihood estimators of a location parameter. *Annals of Statistics* 3, 267–284.

APPENDIX A: Theorem Proofs

Proof of Theorem 1. We have

$$\widehat{\sigma}^2\left(\theta,\widehat{\beta}(\theta)\right)=\frac{1}{NT}\sum_{i=1}^{N}\left(\Upsilon(L;\theta)\left(\Delta y_i-\Delta x_i\widehat{\beta}(\theta)\right)\right)'\Omega(\theta)^{-1}\left(\Upsilon(L;\theta)\left(\Delta y_i-\Delta x_i\widehat{\beta}(\theta)\right)\right),$$

which straightforwardly equals

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\left(\Upsilon(L;\theta)\,\Delta y_i\right)'\Omega(\theta)^{-1}\left(\Upsilon(L;\theta)\,\Delta y_i\right)\\
&-\frac{1}{NT}\sum_{i=1}^{N}\left(\Upsilon(L;\theta)\,\Delta y_i\right)'\Omega(\theta)^{-1}\,\Upsilon(L;\theta)\,\Delta x_i\\
&\times\left(\frac{1}{NT}\sum_{i=1}^{N}\left(\Upsilon(L;\theta)\,\Delta x_i\right)'\Omega(\theta)^{-1}\,\Upsilon(L;\theta)\,\Delta x_i\right)^{-1}\\
&\times\frac{1}{NT}\sum_{i=1}^{N}\left(\Upsilon(L;\theta)\,\Delta x_i\right)'\Omega(\theta)^{-1}\,\Upsilon(L;\theta)\,\Delta y_i\\
&=A(y,y,\theta)-A(y,x,\theta)A(x,x,\theta)^{-1}A(x,y,\theta), \qquad \textbf{(A.1)}
\end{aligned}$$

where, for example,

$$A(x,y,\theta) = \frac{1}{NT}\sum_{i=1}^{N} \left(\Upsilon(L;\theta)\,\Delta x_i\right)' \Omega(\theta)^{-1}\, \Upsilon(L;\theta)\,\Delta y_i.$$

With the same kind of notation, (A.1) straightforwardly equals

$$A(\upsilon,\upsilon,\theta) - A(\upsilon,x,\theta)A(x,x,\theta)^{-1}A(x,\upsilon,\theta).$$

Assumption A(vi) and finiteness of T implies the $\tau_t(\theta)$ are continuous in θ, and thus uniformly continuous on the compact set Θ. Also, from (18), (19), $\Omega(\theta)^{-1}$ is continuous in θ, and thus uniformly continuous on the compact set Θ.

Now, from finiteness of T and conditions Assumptions A(i)–(vi), as $N \to \infty$, uniformly in $\theta \in \Theta$, a.s.

$$A(\upsilon,\upsilon,\theta) \to EA(\upsilon,\upsilon,\theta),\; A(\upsilon,x,\theta) \to EA(\upsilon,x,\theta) = 0, A(x,x,\theta) \to EA(x,x,\theta), \tag{A.2}$$

where the last expression is positive definite from condition (viii). Here, pointwise convergence follows from Assumption A(1), while equicontinuity, and thus uniform convergence, follow since Assumption A(vi) and finiteness of T implies the $\tau_t(\theta)$ are continuous in θ, and thus uniformly continuous on the compact set Θ, while from (18), (19), $\Omega(\theta)^{-1}$ is continuous in θ, and thus uniformly continuous on the compact set Θ. It follows that as $N \to \infty$, uniformly in $\theta \in \Theta$, a.s.

$$\widehat{\sigma}^2\left(\theta,\widehat{\beta}(\theta)\right) \to EA(\upsilon,\upsilon,\theta). \tag{A.3}$$

Now,

$$EA(\upsilon,\upsilon,\theta) = \frac{1}{T} tr\left(\Omega(\theta)^{-1} Ez_i(\theta,\beta_0)\, z_i'(\theta,\beta_0)\right),$$

where

$$\begin{aligned} z_i(\theta,\beta_0) &= \left(z_{i1}(\theta,\beta),\ldots,z_{iT}(\theta,\beta)\right)' \\ &= V(\theta)\,\varepsilon_i + \left(\nu(\theta) - \tau(\theta)\right)\varepsilon_{i0}, \end{aligned} \tag{A.4}$$

defining the $T \times T$ upper-triangular matrix

$$V(\theta) = \begin{pmatrix} 1 & \nu_1(\theta) & \nu_2(\theta) & \dots & \nu_{T-1}(\theta) \\ 0 & 1 & \nu_1(\theta) & \dots & \nu_{T-2}(\theta) \\ 0 & 0 & 1 & \dots & \nu_{T-3}(\theta) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 1 \end{pmatrix}, \tag{A.5}$$

where the $\nu_j(\theta)$ are given by

$$\nu(L;\theta,\theta_0) = \lambda(L;\theta)/\lambda(L;\theta_0) = \sum_{j=0}^{\infty} \nu_j(\theta)\,L^j, \tag{A.6}$$

and the $T \times 1$ vector

$$\nu(\theta) = (\nu_1(\theta), \nu_2(\theta), \ldots, \nu_T(\theta))'. \tag{A.7}$$

The evaluation (A.4) follows because

$$\begin{aligned} z_{it}(\theta,\beta_0) &= \tau_{t-1}(L;\theta)\Delta\upsilon_{it} \\ &= \lambda_t(L;\theta)\upsilon_{it} - \tau_t(\theta)\varepsilon_{i0}, \\ &= \lambda_t(L;\theta_0)^{-1}\lambda_t(L;\theta)\varepsilon_{it} - \tau_t(\theta)\varepsilon_{i0} \\ &= \sum_{j=0}^{t}\nu_j(\theta)\varepsilon_{i,t-j} - \tau_t(\theta)\varepsilon_{i0} \\ &= \sum_{j=0}^{t-1}\nu_j(\theta)\varepsilon_{i,t-j} + (\nu_t(\theta) - \tau_t(\theta))\varepsilon_{i0}. \end{aligned}$$

From (A.4)

$$Ez_i(\theta,\beta_0)z_i(\theta,\beta_0)'/\sigma_0^2 = V(\theta)'V(\theta) + (\nu(\theta) - \tau(\theta))(\nu(\theta) - \tau(\theta))',$$

and so

$$EA(\upsilon,\upsilon,\theta)/\sigma_0^2 = \frac{1}{T}tr\left(\Omega(\theta)^{-1}\left(V(\theta)'V(\theta) + (\nu(\theta) - \tau(\theta))(\nu(\theta) - \tau(\theta))'\right)\right). \tag{A.8}$$

By the inequality between arithmetic and geometric means, the last expression is no less than

$$\left|\Omega(\theta)^{-1}\left(V(\theta)'V(\theta) + (\nu(\theta) - \tau(\theta))(\nu(\theta) - \tau(\theta))'\right)\right|^{1/T}. \tag{A.9}$$

It is readily seen from (A.5) and (A.7) after evaluating the $\nu_j(\theta)$ from (A.6) (eg $\nu_1(\theta) = \lambda_1(\theta) - \lambda_1(\theta_0)$) that $V(\theta)'V(\theta) + (\nu(\theta) - \tau(\theta))(\nu(\theta) - \tau(\theta))'$ equals $\Omega(\theta)$ plus a symmetric matrix all of whose elements are functions of $\lambda_t(\theta) - \lambda_t(\theta_0)$, $t = 1, \ldots, T$ and θ_0 only, such that all eigenvalues of $\Omega(\theta)^{-1}\left(V(\theta)'V(\theta) + (\nu(\theta) - \tau(\theta))(\nu(\theta) - \tau(\theta))'\right)$ are equal, and thence the lower bound (A.9) is attained, when and only when $\lambda_t(\theta) = \lambda_t(\theta_0)$, $j = 1, \ldots, T$. But from condition (vii) the latter holds only when $\theta = \theta_0$. Since (A.9) equals $\left(1 + \tau'(\theta_0)\tau(\theta_0)\right)^{-1/T} = |\Omega(\theta_0)|^{-1/T}$at $\theta = \theta_0$ we have shown that $L(\theta,\beta(\theta))/\sigma_0^2$ converges uniformly a.s. to a function that is bounded below uniquely by the limit, 1, of $L(\theta_0,\beta(\theta_0))/\sigma_0^2$, since $\widehat{\sigma}^2\left(\theta_0,\widehat{\beta}(\theta_0)\right) \to \sigma_0^2$ a.s. from (A.3). Thus, by a standard argument (see e.g., Jennrich, 1969) it follows that $\widehat{\theta} \to \theta_0$ a.s. Using this property and some of the arguments above, finally, $\widehat{\beta} = \widehat{\beta}\left(\widehat{\theta}\right) = A(x,x,\widehat{\theta})^{-1}A(x,y,\widehat{\theta}) \to \beta_0$ a.s. ■

Proof of Theorem 2. By Assumptions B (ii) and the usual mean value theorem argument

$$\frac{\partial}{\partial(\theta',\beta')'}L\left(\widehat{\theta},\widehat{\beta}\right) = \frac{\partial}{\partial(\theta',\beta')'}L(\theta_0,\beta_0) + \widetilde{M}\begin{pmatrix}\widehat{\theta} - \theta_0 \\ \widehat{\beta} - \beta_0\end{pmatrix}, \tag{A.10}$$

where $\widetilde{M}$ is the matrix derived from

$$M(\theta,\beta) = \frac{\partial^2}{\partial(\theta',\beta')'\partial(\theta',\beta')}L(\theta,\beta) \tag{A.11}$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

by evaluating the jth row at $\theta = \widetilde{\theta}^{(j)}$, $\beta = \widetilde{\beta}^{(j)}$satisfying $\left\|\widetilde{\theta}^{(j)} - \theta_0\right\| \le \left\|\widehat{\theta} - \theta_0\right\|$, $\left\|\widetilde{\beta}^{(j)} - \beta_0\right\| \le \left\|\widehat{\beta} - \beta_0\right\|$. Now for $j = 1, \ldots, p$,

$$\frac{\partial}{\partial \theta_j} L(\theta_0, \beta_0) = \frac{1}{NT} \sum_{i=1}^{N} r_{1ji}(\theta_0, \beta_0), \tag{A.12}$$

and

$$\frac{\partial}{\partial \beta} L(\theta_0, \beta_0) = \frac{1}{N} \sum_{i=1}^{N} r_{2i}(\theta_0, \beta_0). \tag{A.13}$$

The expression (A.12) follows by noting that from the proof of Theorem 4.4 of RV the left hand side is

$$|\Omega(\theta_0)|^{\frac{1}{T}} \left(\frac{\widehat{\sigma}^2(\theta_0, \beta_0)}{T} tr \left(\Omega^{-1}(\theta_0) \Omega^j(\theta_0) \right) \right.$$
$$\left. - \frac{1}{NT} \sum_{i=1}^{N} z_i'(\theta_0, \beta_0) \Omega^{-1}(\theta_0) \Omega^j(\theta_0) \Omega^{-1}(\theta_0) z_i(\theta_0, \beta_0) \right)$$
$$+ \frac{2}{NT} \sum_{i=1}^{N} \dot{z}_i^{j\prime}(\theta_0, \beta_0) \Omega^{-1}(\theta_0) z_i(\theta_0, \beta_0),$$

and applying (15) and (20).

To analyze the right side of (A.12), we make heavy use of expressions in the proof of Theorem 4.4 of RV, reproducing them here because their central limit theorem is based on $T \to \infty$ whereas ours is based on $N \to \infty$; hence, the representation in (A.13) as an arithmetic mean over $i = 1, \ldots, N$, instead of one over t. First, we have

$$\widehat{\sigma}^2(\theta_0, \beta_0) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\varepsilon_{it} - \tau_t^0 \varepsilon_{i0} \right)^2 - \frac{1}{NT S_{\tau\tau}^0} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \tau_t^0 \left(\varepsilon_{it} - \tau_t^0 \varepsilon_{i0} \right) \right)^2$$

where

$$S_{\tau\tau}^0 = |\Omega(\theta_0)| = 1 + \tau^{0\prime} \tau^0,$$

and from RV

$$E\widehat{\sigma}^2(\theta_0, \beta_0) = \sigma_0^2.$$

Also, defining

$$S_{\tau\dot{\tau}j}^0 = \tau^{0\prime} \dot{\tau}^{0j}, \quad \dot{\tau}^{0j} = \dot{\tau}^j(\theta_0),$$

we have

$$z_i'(\theta_0, \beta_0) \Omega^{-1}(\theta_0) \Omega^j(\theta_0) \Omega^{-1}(\theta_0) z_i(\theta_0, \beta_0)$$
$$= \frac{2}{S_{\tau\tau}^0} \left(\sum_{t=1}^{T} \left(\varepsilon_{it} - \tau_t^0 \varepsilon_{i0} \right) \tau_t^0 \right) \left(\sum_{t=1}^{T} \left(\varepsilon_{it} - \tau_t^0 \varepsilon_{i0} \right) \dot{\tau}_t^{0j} \right)$$
$$-2 \frac{S_{\tau\dot{\tau}j}^0}{S_{\tau\tau}^{02}} \left(\sum_{t=1}^{T} \left(\varepsilon_{it} - \tau_t^0 \varepsilon_{i0} \right) \tau_t^0 \right)^2,$$

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

which has expectation $tr\left(\Omega^{-1}(\theta_0)\,\Omega^{j}(\theta_0)\right)$. Finally, $\dot{z}_i^{j\prime}(\theta_0,\beta_0)$ is the jth row of

$$\dot{z}_i'(\theta_0,\beta_0) = f_i - \dot{\tau}^{0}\varepsilon_{i0}, \tag{A.14}$$

in which we have the $p \times T$ matrices $\dot{\tau}^0 = \dot{\tau}(\theta_0) = \left(\dot{\tau}^{01\prime},\ldots,\dot{\tau}^{0p\prime}\right)'$ and $f_i = (f_{i1},\ldots,f_{iT})$, where

$$f_{it} = \sum_{j=0}^{t-1} \chi_{t-j}(\theta_0)\,\varepsilon_{ij},$$

with the vectors $\chi_j(\theta)$ defined by

$$\chi(L;\theta) = \frac{\partial}{\partial\theta}\log\lambda(L;\theta) = \sum_{j=0}^{\infty}\chi_j(\theta)\,L^{j}.$$

The representation (A.14) follows from

$$\begin{aligned}\frac{\partial}{\partial\theta} z_{it}(\theta,\beta) &= \frac{\partial}{\partial\theta}\lambda_t(L;\theta)\left\{v_{it} - x_{it}'(\beta-\beta_0)\right\} - \frac{\partial}{\partial\theta}\tau_t(\theta)\left\{\varepsilon_{i0} - x_{i0}'(\beta-\beta_0)\right\}\\ &= \chi(L;\theta)\,\lambda(L;\theta)\,\lambda_t^{-1}(L;\theta_0)\,\varepsilon_{it} - \chi(L;\theta)\,\lambda(L;\theta)\,x_{it}'(\beta-\beta_0)\\ &\quad -\dot{\tau}_t(\theta)\left\{\varepsilon_{i0} - x_{i0}'(\beta-\beta_0)\right\},\end{aligned}$$

and hence,

$$\begin{aligned}\frac{\partial}{\partial\theta} z_{it}(\theta,\beta_0) &= \chi(L;\theta)\,\lambda(L;\theta)\,\lambda_t^{-1}(L;\theta_0)\,\varepsilon_{it} - \dot{\tau}_t(\theta)\,\varepsilon_{i0},\\ \frac{\partial}{\partial\theta} z_{it}(\theta_0,\beta_0) &= \chi(L;\theta_0)\,\varepsilon_{it} - \dot{\tau}_t^{0}\varepsilon_{i0}\\ &= f_{it} - \dot{\tau}_t^{0}\varepsilon_{i0}.\end{aligned}$$

Then, with f_{it}^{j} the jth element of f_{it},

$$\begin{aligned}&\dot{z}_i^{j\prime}(\theta_0,\beta_0)\,\Omega^{-1}(\theta_0)\,z_i(\theta_0,\beta_0)\\ &= \sum_{t=1}^{T}\left(f_{it}^{j} - \dot{\tau}_t^{0j}\varepsilon_{i0}\right)\left(\varepsilon_{it} - \tau_t^{0}\varepsilon_{i0}\right) - \frac{1}{S_{\tau\tau}^{0}}\sum_{t=1}^{T}\left(f_{it}^{j} - \dot{\tau}_t^{0j}\varepsilon_{i0}\right)\tau_t^{0}\\ &\quad\times\left(\sum_{t=1}^{T}\left(\varepsilon_{it} - \tau_t^{0}\varepsilon_{i0}\right)\tau_t^{0}\right),\end{aligned}$$

which, as shown by RV, has expectation $\sigma_0^2 S_{\tau\dot{\tau}j}^{0}/S_{\tau\tau}^{0}$.

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

Altogether we have, writing $S^0_{\dot{\tau}\tau} = \left(S^0_{\tau\dot{\tau}1}, \ldots, S^0_{\tau\dot{\tau}p}\right)'$,

$$\begin{aligned} Tr_{1i}(\theta_0,\beta_0) = {} & \frac{S^0_{\dot{\tau}\tau}}{\left(S^0_{\tau\tau}\right)^{1-1/T}} \left\{ \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)^2 - \frac{1}{TS^0_{\tau\tau}}\left(\sum_{t=1}^{T}\tau_t^0\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\right)^2\right\} \\ & - \frac{2}{\left(S^0_{\tau\tau}\right)^{1-1/T}}\left(\sum_{t=1}^{T}\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\tau_t^0\right)\left(\sum_{t=1}^{T}\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\dot{\tau}_t^0\right) \\ & + 2\frac{\dot{\tau}^{0\prime}\tau^0}{\left(S^0_{\tau\tau}\right)^{2-1/T}}\left(\sum_{t=1}^{T}\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\tau_t^0\right)^2 \\ & + 2\left(S^0_{\tau\tau}\right)^{1/T}\sum_{t=1}^{T}\left(f_{it} - \dot{\tau}^0\varepsilon_{i0}\right)\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right) \\ & - \frac{2}{\left(S^0_{\tau\tau}\right)^{1-1/T}}\left(\sum_{t=1}^{T}\left(f_{it} - \dot{\tau}_t^0\varepsilon_{i0}\right)\tau_t^0\right)\left(\sum_{t=1}^{T}\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\tau_t^0\right), \end{aligned} \tag{A.15}$$

and $Er_{1ji}(\theta_0,\beta_0) = 0$. The $r_{2i}(\theta_0,\beta_0)$ can be similarly but more simply expressed and it is readily confirmed that $Er_{2i}(\theta_0,\beta_0) = 0$, and then from Assumptions A, conditions (i) and the central limit theorem for iid random vectors,

$$N^{1/2}\frac{\partial}{\partial\left(\theta',\beta'\right)'}L(\theta_0,\beta_0) \rightarrow_d N(0, EC(\theta_0,\beta_0)).$$

Next, using also Theorem 1 and Assumption B(ii)–(v), it follows much as in the proof of (A.2) that we have uniform convergence on a neighbourhood of (θ_0,β_0) of $B(\theta,\beta)$, $C(\theta,\beta)$ and $M(\theta,\beta)$, whence

$$\begin{aligned} C\left(\widehat{\theta},\widehat{\beta}\right) &\rightarrow_p EC(\theta_0,\beta_0), \\ B\left(\widehat{\theta},\widehat{\beta}\right) &\rightarrow_p EB(\theta_0,\beta_0), \\ \widetilde{M} - M(\theta_0,\beta_0) &\rightarrow_p 0 \end{aligned}$$

and

$$M(\theta_0,\beta_0) = B(\theta_0,\beta_0) + o_p(1) \rightarrow_p EB(\theta_0,\beta_0),$$

the $o_p(1)$ terms including negligible ones such as those in second derivatives of elements of $\Omega(\theta)$, $z_i(\theta,\beta)$. ■

APPENDIX B: Evaluation of the Asymptotic Variance of $\hat{\theta}$ under Gaussianity

Here, we pursue closed form expressions for the asymptotic variance of $\hat{\theta}$ that can be estimated by replacing θ_0 by $\widehat{\theta}$ when there are no regressors in the model. We suppress dependence of β in the notation and write $z_i^0 = z_i(\theta_0) = z_i(\theta_0,\beta_0) = \varepsilon_i - \tau(\theta_0)\varepsilon_{i0}$, with

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

a similar interpretation for $\dot{z}_i^{0j} = \dot{z}_i^{j}(\theta_0)$, its derivative with respect to θ_j. We also write $\Omega_0^{-1} = \Omega^{-1}(\theta_0)$ and $\Omega_0^{j} = \Omega^{j}(\theta_0)$.

We first evaluate the variance of the score of L for Gaussian data, setting $\sigma_0^2 = 1$ $w$log,

$$\begin{aligned} E\left[C(\theta_0)\right]_{j,k} &= E\left[r_{1i}(\theta_0) r_{1i}(\theta_0)'\right]_{j,k} \\ &= \frac{1}{T^2}\left|\Omega_T(\theta_0)\right|^{\frac{2}{T}} \sum_{a,b=1}^{3} E\left[\left(w_{aji}^0 - E\left[w_{aji}^0\right]\right)\left(w_{bki}^0 - E\left[w_{bki}^0\right]\right)\right] \\ &= \frac{1}{T^2}\left|\Omega_T(\theta_0)\right|^{\frac{2}{T}} \sum_{a,b=1}^{3} E\left[w_{aji}^0 w_{bki}^0\right] - E\left[w_{aji}^0\right] E\left[w_{bki}^0\right] \end{aligned}$$

for coordinates $j,k = 1,\ldots,p$, where

$$\begin{aligned} w_{1ji}^0 &= \frac{1}{T} tr\left(\Omega_0^{-1}\Omega_0^{j}\right) z_i' \Omega_0^{-1} z_i^0 \\ w_{2ji}^0 &= -z_i^{0\prime}\Omega_0^{-1}\Omega_0^{j}\Omega_0^{-1} z_i^0 \\ w_{3ji}^0 &= 2\dot{z}_i^{0j\prime}\Omega_0^{-1} z_i^0, \end{aligned}$$

and $E\left[w_{1ji}^0\right] = -E\left[w_{2ji}^0\right] = tr\left(\Omega_0^{-1}\Omega_0^{j}\right)$ because

$$\begin{aligned} E\left[z_i^{0\prime}\Omega_0^{-1} z_i^0\right] &= E\left[\left(\varepsilon_{it} - \tau_t(\theta_0)\varepsilon_{i0}\right)' \Omega(\theta_0)^{-1} \left(\varepsilon_{it} - \tau_t(\theta_0)\varepsilon_{i0}\right)\right] \\ &= tr\left\{\Omega_0^{-1} E\left[\left(\varepsilon_{it} - \tau_t(\theta_0)\varepsilon_{i0}\right)'\left(\varepsilon_{it} - \tau_t(\theta_0)\varepsilon_{i0}\right)\right]\right\} \\ &= tr\left\{I_T\right\} = T \end{aligned}$$

while

$$\begin{aligned} E\left[w_{2ji}^0\right] &= -E\left[z_i^{0\prime}\Omega_0^{-1}\Omega_0^{j}\Omega_0^{-1} z_i^0\right] \\ &= -tr\left(\Omega_0^{-1}\Omega_0^{j} E\left[z_i^{0\prime}\Omega_0^{-1} z_i^0\right]\right) \\ &= -tr\left(\Omega_0^{-1}\Omega_0^{j}\right) \end{aligned}$$

and $E\left[w_{3ji}^0\right] = tr\left(\Omega_0^{-1} E\left(z_i^0 \dot{z}_i^{0j\prime}\right)\right) = 0$ from the proof of Theorem 4.4 in RV.

Next,

$$\begin{aligned} E\left[w_{1ji}^0 w_{1ki}^0\right] &= \frac{1}{T^2} tr\left(\Omega_0^{-1}\Omega_0^{j}\right) tr\left(\Omega_0^{-1}\Omega_0^{k}\right)\left[\left(z_i^{0\prime}\Omega_0^{-1} z_i^0\right)^2\right] \\ &= \left(1 + \frac{2}{T}\right) tr\left(\Omega_0^{-1}\Omega_0^{j}\right) tr\left(\Omega_0^{-1}\Omega_0^{k}\right) \end{aligned}$$

because using Gaussianity, see Nagar (1959), with $E\left[\left(\varepsilon_i - \tau(\theta_0)\varepsilon_{i0}\right)\left(\varepsilon_i - \tau(\theta_0)\varepsilon_{i0}\right)'\right] = \Omega_0$,

$$\begin{aligned} E\left[\left(z_i^{0\prime}\Omega(\theta_0)^{-1} z_i^0\right)^2\right] &= E\left[\left(\left(\varepsilon_i - \tau(\theta_0)\varepsilon_{i0}\right)'\Omega(\theta_0)^{-1}\left(\varepsilon_i - \tau(\theta_0)\varepsilon_{i0}\right)\right)^2\right] \\ &= tr\left\{\Omega_0^{-1}\Omega_0\right\}^2 + 2tr\left\{\Omega_0^{-1}\Omega_0\Omega_0^{-1}\Omega_0\right\} \\ &= tr\left(I_T\right)^2 + 2tr\left(I_T\right) = T^2 + 2T. \end{aligned}$$

Next, using again Gaussianity,

$$E\left[w_{2ji}^{0} w_{2ki}^{0}\right] = E\left[\left(z_i^{0\prime}\Omega_0^{-1}\Omega_0^{j}\Omega_0^{-1}z_i^{0}\right)\left(z_i^{0\prime}\Omega_0^{-1}\Omega_0^{k}\Omega_0^{-1}z_i^{0}\right)\right]$$
$$= tr\left(\Omega_0^{-1}\Omega_0^{j}\right) tr\left(\Omega_0^{-1}\Omega_0^{k}\right) + 2tr\left(\Omega_0^{-1}\Omega_0^{j}\Omega_0^{-1}\Omega_0^{k}\right).$$

Write

$$z_i^0 = \Sigma^0\,\varepsilon_i, \quad \Sigma^0 = \Sigma\left(\theta_0\right), \;\; \Sigma = \Sigma\left(\theta\right) = \left(-\tau\left(\theta\right) \vdots\, I_T\right)$$

while from

$$\dot{z}_{it}^{0j} = \sum_{a=1}^{t} \chi_a^{j}\left(\theta_0\right)\varepsilon_{it-a} - \dot{\tau}_t\left(\theta_0\right)\varepsilon_{i0}$$

we have

$$\dot{z}_i^{0j} = A_j^0\,\varepsilon_i, \quad A_j^0 = A_j\left(\theta_0\right), \;\; A_j = A_j\left(\theta\right) = \left(\Xi^j\left(\theta\right) \vdots\, 0_{T\times 1}\right) - \left(\dot{\tau}^j\left(\theta\right) \vdots\, 0_{T\times T}\right)$$

where

$$\Xi^j\left(\theta\right) = \begin{pmatrix} \chi_1^j\left(\theta\right) & 0 & \cdots & 0 \\ \chi_2^j\left(\theta\right) & \chi_1^j\left(\theta\right) & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \chi_T^j\left(\theta\right) & \cdots & \chi_2^j\left(\theta\right) & \chi_1^j\left(\theta\right) \end{pmatrix}.$$

Then, writing $w_{3ji}^0 = 2\dot{z}_i^{0j\prime}\Omega_0^{-1}z_i^0 = \varepsilon_i^{\prime}B_j^0\varepsilon_i$ for symmetric $B_j^0 = \left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0 + \left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right)$,

$$E\left[w_{3ji}^{0} w_{3ki}^{0}\right] = 4E\left[\left(\dot{z}_i^{0j\prime}\Omega_0^{-1}z_i^{0}\right)\left(\dot{z}_i^{0k\prime}\Omega_0^{-1}z_i^{0}\right)\right]$$
$$= 2tr\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0 A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right) + 2tr\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right)$$
$$+2tr\left(\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right) + 2tr\left(\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right)$$
$$= 2tr\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right) + 2tr\left(\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)$$

because $tr\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\right) = tr\left(\Omega_0^{-1}\Sigma^0 A_j^{0\prime}\right) = tr\left(\Omega_0^{-1}E\left(z_i^0\dot{z}_i^{0j\prime}\right)\right) = 0$ and $tr\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0 A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right) = tr\left(\left(A_j^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right) = 0$ extending the arguments in the proof of Theorem 4.4 of RV.

Finally, the covariance terms are

$$E\left[w^0_{1ji}w^0_{2ki}\right] = -\frac{1}{T}tr\left(\Omega^{-1}(\theta_0)\,\Omega^j(\theta_0)\right)E\left[\left(z_i^{0\prime}\Omega_0^{-1}z_i^0\right)\left(z_i^{0\prime}\Omega_0^{-1}\Omega_0^k\Omega_0^{-1}z_i^0\right)\right]$$
$$= -\left(1+\frac{2}{T}\right)tr\left(\Omega_0^{-1}\Omega_0^j\right)tr\left(\Omega_0^{-1}\Omega_0^k\right)$$

because

$$E\left[\left(z_i^{0\prime}\Omega_0^{-1}z_i^0\right)\left(z_i^{0\prime}\Omega_0^{-1}\Omega_0^k\Omega_0^{-1}z_i^0\right)\right]$$
$$= tr\left(\Omega_0^{-1}\Omega_0\right)tr\left(\Omega_0^{-1}\Omega_0^k\Omega_0^{-1}\Omega_0\right) + 2tr\left(\Omega_0^{-1}\Omega_0\Omega_0^{-1}\Omega_0^k\Omega_0^{-1}\Omega_0\right)$$
$$= (T+2)\,tr\left(\Omega_0^{-1}\Omega_0^k\right)$$

with $E\left[w^0_{2ji}w^0_{1ki}\right] = E\left[w^0_{1ji}w^0_{2ki}\right]$, while

$$E\left[w^0_{1ji}w^0_{3ki}\right] = 2\frac{1}{T}tr\left(\Omega_0^{-1}\Omega_0^j\right)E\left[\left(z_i^{0\prime}\Omega_0^{-1}z_i^0\right)\left(\dot{z}_i^{0k\prime}\Omega_0^{-1}z_i^0\right)\right]$$
$$= 2\left(1+\frac{2}{T}\right)tr\left(\Omega_0^{-1}\Omega_0^j\right)tr\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right) = 0,$$

because

$$E\left[\left(z_i^{0\prime}\Omega_0^{-1}z_i^0\right)\left(\dot{z}_i^{0k\prime}\Omega_0^{-1}z_i^0\right)\right]$$
$$= tr\left(\Omega_0^{-1}\Omega_0\right)tr\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right) + tr\left(\Omega_0^{-1}\Omega_0A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right) + tr\left(\Omega_0^{-1}\Omega_0\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right)$$
$$= (T+2)\,tr\left(A_k^{\prime}\Omega_0^{-1}\Sigma^0\right) = 0.$$

Similarly $E\left[w^0_{3ji}w^0_{1ki}\right] = 0$, and

$$E\left[w^0_{2ji}w^0_{3ki}\right] = -2E\left[\left(z_i^{0\prime}\Omega_0^{-1}\Omega_0^j\Omega_0^{-1}z_i^0\right)\left(\dot{z}_i^{0k\prime}\Omega_0^{-1}z_i^0\right)\right]$$
$$= -2tr\left(\Sigma^{0\prime}\Omega_0^{-1}\Omega_0^j\Omega_0^{-1}\Sigma^0A_k^{\prime}\Omega_0^{-1}\Sigma^0\right) - 2tr\left(\Sigma^{0\prime}\Omega_0^{-1}\Omega_0^j\Omega_0^{-1}\Sigma^0\left(A_k^{0\prime}\Omega_0^{-1}\Sigma^0\right)^{\prime}\right)$$

and a similar expression follows for $E\left[w^0_{2ki}w^0_{3ji}\right]$ interchanging j and k.

Compiling all the terms we set

$$\left[C_0(\theta)\right]_{j,k} = \frac{1}{T^2}\left|\Omega(\theta)\right|^{\frac{2}{T}}\Big(-\frac{2}{T}tr\left(\Omega^{-1}\Omega^j\right)tr\left(\Omega^{-1}\Omega^k\right) + 2tr\left(\Omega^{-1}\Omega^j\Omega^{-1}\Omega^k\right)$$
$$+2tr\left(A_j^{\prime}\Omega^{-1}\Sigma\left(A_k^{\prime}\Omega^{-1}\Sigma\right)^{\prime}\right) + 2tr\left(\left(A_j^{\prime}\Omega^{-1}\Sigma\right)^{\prime}A_k^{\prime}\Omega^{-1}\Sigma\right)$$
$$-2tr\left(\Sigma^{\prime}\Omega^{-1}\Omega^j\Omega^{-1}\Sigma A_k^{\prime}\Omega^{-1}\Sigma\right) - 2tr\left(\Sigma^{\prime}\Omega^{-1}\Omega^j\Omega^{-1}\Sigma\left(A_k^{\prime}\Omega^{-1}\Sigma\right)^{\prime}\right)$$
$$-2tr\left(\Sigma^{\prime}\Omega^{-1}\Omega^k\Omega^{-1}\Sigma A_j^{\prime}\Omega^{-1}\Sigma\right) - 2tr\left(\Sigma^{\prime}\Omega^{-1}\Omega^k\Omega^{-1}\Sigma\left(A_j^{\prime}\Omega^{-1}\Sigma\right)^{\prime}\right)\Big)$$

where $\Omega^{-1} = \Omega^{-1}(\theta)$, $\Omega^j = \Omega^j(\theta)$, $A_j = A_j(\theta)$ and $\Sigma = \Sigma(\theta)$ are evaluated at θ instead of Ω_0^{-1}, etc., which are evaluated at θ_0.

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press

For the evaluation of the probability limit of the Hessian, $B_{1i}(\theta_0)$, we define $b^0_{1jki}(\theta)$ as $b_{1jki}(\theta)$ with $\widehat{\sigma}^2(\theta)$ replaced by σ_0^2 $(= 1\ w\text{log})$, $j,k = 1,\ldots,p$, so that

$$\begin{aligned}E\left[b^0_{1jki}(\theta_0)\right] = \frac{1}{T}\left|\Omega(\theta_0)\right|^{\frac{1}{T}}\Big(&tr\left(\Omega^{-1}(\theta_0)\,\Omega^k(\theta_0)\,\Omega^{-1}(\theta_0)\,\Omega^j(\theta_0)\right)\\ &-\frac{1}{T}tr\left(\Omega^{-1}(\theta_0)\,\Omega^j(\theta_0)\right)tr\left(\Omega^{-1}(\theta_0)\,\Omega^k(\theta_0)\right)\\ &-2tr\left(\Omega^{-1}(\theta_0)\,\Omega^j(\theta_0)\,\Omega^{-1}(\theta_0)\,E\left[z_i^0\dot{z}_i^{0k\prime}\right]\right)\\ &-2tr\left(\Omega^{-1}(\theta_0)\,\Omega^k(\theta_0)\,\Omega^{-1}(\theta_0)\,E\left[z_i^0\dot{z}_i^{0j\prime}\right]\right)\\ &+2tr\left(\Omega^{-1}(\theta_0)\,E\left[\dot{z}_i^{0k}\dot{z}_i^{0j\prime}\right]\right)\Big),\end{aligned}$$

and then set

$$\begin{aligned}\left[B_0(\theta)\right]_{j,k} = \frac{1}{T}\left|\Omega(\theta)\right|^{\frac{1}{T}}\Big(&tr\left(\Omega^{-1}(\theta)\,\Omega^k(\theta)\,\Omega^{-1}(\theta)\,\Omega^j(\theta)\right)\\ &-\frac{1}{T}tr\left(\Omega^{-1}(\theta)\,\Omega^j(\theta)\right)tr\left(\Omega^{-1}(\theta)\,\Omega^k(\theta)\right)\\ &-2tr\left(\Omega^{-1}(\theta)\,\Omega^j(\theta)\,\Omega^{-1}(\theta)\,E_{(\theta)}\left[z_i\dot{z}_i^{k\prime}\right]\right)\\ &-2tr\left(\Omega^{-1}(\theta)\,\Omega^k(\theta)\,\Omega^{-1}(\theta)\,E_{(\theta)}\left[z_i\dot{z}_i^{j\prime}\right]\right)\\ &+2tr\left(\Omega^{-1}(\theta)\,E_{(\theta)}\left[\dot{z}_i^k\dot{z}_i^{j\prime}\right]\right)\Big)\end{aligned}$$

where $E_{(\theta)}\left[z_i\dot{z}_i^{j\prime}\right] = \Sigma(\theta)\,A_j'(\theta) = \Sigma A_j'$ and $E_{(\theta)}\left[\dot{z}_i^k\dot{z}_i^{j\prime}\right] = A_k(\theta)\,A_j'(\theta) = A_kA_j'$.

Then, the equivalent normalizing matrices for $\left(\widehat{\theta}-\theta_0\right)$ based on these expected values under Gaussianity satisfy $\left(NB_0\left(\widehat{\theta}\right)C_0\left(\widehat{\theta}\right)^{-1}B_0\left(\widehat{\theta}\right)\right)^{1/2} = N^{1/2}(T/2)\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{T}}C_0\left(\widehat{\theta}\right)^{1/2} = N^{1/2}(T/2)^{1/2}\left|\Omega_T\left(\widehat{\theta}\right)\right|^{-\frac{1}{2T}}B_0\left(\widehat{\theta}\right)^{1/2}$ exploiting the proportionality of $B_0(\theta)$ and $C_0(\theta)$ under Gaussianity, because $\Sigma(\theta)\,\Sigma(\theta)' = \Omega(\theta)$, so that

$$tr\left(\left(A_j'\Omega^{-1}\Sigma\right)'A_k'\Omega^{-1}\Sigma\right) = tr\left(\Omega^{-1}A_jA_k'\Omega^{-1}\Sigma\Sigma'\right) = tr\left(\Omega^{-1}A_jA_k'\right),$$

$$tr\left(\Sigma'\Omega^{-1}\Omega^j\Omega^{-1}\Sigma A_k'\Omega^{-1}\Sigma\right) = tr\left(\Omega^{-1}\Omega^j\Omega^{-1}\Sigma A_k'\right),$$

the additional factors correcting for the fact of L being only proportional to the likelihood.

https://doi.org/10.1017/S0266466619000069 Published online by Cambridge University Press