Inference on Trending Panel Data

Peter M Robinson* and Carlos Velasco
London School of Economics and Universidad Carlos III de Madrid

November 30, 2015

Abstract

Semiparametric panel data modelling and statistical inference with fractional stochastic trends, nonparametrically time-trending individual effects, and general cross-sectional correlation and heteroscedasticity in innovations is developed. The fractional stochastic trends allow for a wide range of nonstationarity, indexed by a memory parameter, nesting the familiar $I(1)$ case and allowing for parametric short-memory. The individual effects can nonparametrically vary simultaneously across time and across units. The cross-sectional covariance matrix is also nonparametric. The main focus is on estimation of the time series parameters. Two methods are considered, both of which entail an only approximate differencing out of the individual effects, leaving an error which has to be taken account of in our theory. In both cases we obtain standard asymptotics, with a central limit theorem, over a wide range of possible parameter values, and unlike the nonstandard asymptotics for autoregressive parameter estimates at a unit root. For statistical inference, consistent estimation of the limiting covariance matrix of the parameter estimates requires consistent estimation of a functional of the cross-sectional covariance matrix. We examine efficiency loss due to cross-sectional correlation in a spatial model example. A Monte Carlo study of finite-sample performance is included.

Keywords: Semiparametric panel data modelling, Nonparametrically time-trending individual effects, Nonparametric cross-sectional correlation and heteroscedasticity, Spatial model, Parametric fractional dependence, Consistency, Asymptotic normality.

JEL classifications: C12, C13, C23

....

* Corresponding author. Tel. 44 20 7955 7516. fax 44 20 7955 6592

E-mail address: p.m.robinson@lse.ac.uk

1. Introduction

A considerable literature has developed around the theme of nonstationary panel data with individual and temporal effects. The nonstationarity can have both stochastic and deterministic origins. Unit roots in an autoregressive setting have been the main representation of stochastic trends, while individual and temporal effects are often modelled in a separable, additive way, so that temporal effects are common over the cross-section and individual ones stay constant over time, and explanatory variables, including linear and other deterministic trends, can also feature. Autoregressive models are also commonly adopted in the nonstationary time series and cointegration literature. However, the latter has also developed fractional time series modelling, which can nest $I(1)$ behaviour in a continuum of nonstationary possibilities, with the degree of nonstationarity characterized by a memory parameter, which can be estimated from the data. This approach has the advantages of flexibility and of yielding standard asymptotics, and possible local efficiency of inference, unlike the nonstandard theory which usually emerges from an autoregressive setting. Likewise, the fixed-design nonparametric regression literature suggests a modelling of deterministic trends which is less prone to specification error than parametric functions. Robinson and Velasco (2015) employed fractional models with individual effects in a panel data setting, but with no provision for other time-trending features such as deterministic trends. Another aspect of their work was the assumption of cross-sectional independence and homoscedasticity, conditional on individual effects, which is increasingly seen as restrictive, and sometimes replaced by factor modelling, or spatial modelling.

The present paper considers semiparametric modelling of panel data, such that dynamic parametric fractional stochastic trends are complemented by stochastic or deterministic nonparametrically time trending individual effects and allowance for cross-sectional correlation and heteroscedasticity of a nonparametric form, entailing greater generality than factor or spatial models. The number of time series observations, T, is large, and the number of cross-sectional ones, N, can be large (increasing with T) or small. The latter setting is also considered by Robinson and Velasco (2015), but our introduction of nonparametric temporal variation of individual effects and relaxation of cross-sectional independence and homoscedasticity strengthens the need for an asymptotic theory based on increasing T. Large-T panel data are increasingly available, for example very long time series of prices of several stocks, monthly or quarterly macroeconometric series for several countries, and repeated micro-economic surveys.

We consider an observable array $\{y_{it}\}$, $i = 1, \ldots, N$, $t = 0, 1, \ldots, T$. The vectors of

N cross-sectional observations $y_t = (y_{1t}, ..., y_{Nt})'$, the prime denoting transposition, are assumed to be generated by the semiparametric model

$$\lambda_t (L; \theta_0) (y_t - \alpha_t) = \varepsilon_t, \quad t = 0, 1, \ldots, T. \tag{1}$$

The parametric aspect of (1) is due to the $(p + 1) \times 1$ parameter vector θ_0, which is known only to lie in a given compact subset Θ of R^{p+1}. In (1) L is the lag operator, and for any $\theta \in \Theta$ and each $t \geq 0$, $\lambda_t (L; \theta)$ is a scalar function given by

$$\lambda_t (L; \theta) = \sum_{j=0}^{t} \lambda_j (\theta) L^j, \tag{2}$$

where the $\lambda_j (\theta)$ are given functions to be defined subsequently. The unobservable vectors ε_t have elements with zero mean and are uncorrelated and homoscedastic across t. Thus, (1) is an autoregressive representation for the $y_t - \alpha_t$ with initial condition at $t = 0$, where the number of terms in (2) keeps increasing with t. The assumptions on $\lambda_t (L; \theta)$ that we will impose are aimed at covering fractionally integrated autoregressive moving average (FARIMA) sequences $y_t - \alpha_t$, with unknown memory parameter that can lie in either the stationary or nonstationary regions. Other aspects of (1) are nonparametric. We do not require uncorrelatedness or homoscedasticity across the elements of ε_t, allowing it to have covariance matrix that is unrestricted apart from remaining positive definite with increasing N, thereby to reflect possible cross-sectional correlation and heteroscedasticity of a nonparametric nature, though it can also be assumed to be diagonal, to reflect a lack of cross-sectional correlation but the possibility of nonparametric heteroscedasticity. The vectors α_t consist of stochastic or deterministic unobservable individual effects that can time-trend in a nonparametric way, and in a manner that can vary across elements of the vector, where with N increasing the familiar incidental parameters problem arises. Our allowance for temporally varying individual effects and cross-sectional correlation and heteroscedasticity of innovations, and our relaxation of temporal independence of innovations to martingale difference structure, extend the scope of the model of Robinson and Velasco (2015) to a practically significant degree, but as a consequence reduces the range of memory parameter values covered and limits the degree to which N can increase relative to T.

The main goal of the paper is to justify statistical inference on the unknown parameter vector θ_0. Detailed regularity conditions, later employed in establishing the properties of consistency and asymptotic normality, are described in the following section. In Section 3 an estimate of θ_0, based on initial first-differencing of (1), is shown to be consistent and asymptotically normal, and feasible statistical inference is justified. In Section 4 a more refined estimate, with some advantages, is similarly analysed. Section 5 employs a spatial model to illustrate relative efficiency. A Monte Carlo study of finite-sample performance is reported in Section 6. Section 7 contains some final comments. The proofs of theorems are included in an appendix.

2. Theoretical setting

The present section presents detailed regularity conditions on the model introduced in the previous section, which will be assumed to hold in our theoretical results.

The function $\lambda_t(L;\theta)$ defined in (2) is regarded as truncating the expansion

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

which has the structure

$$\lambda(L;\theta) = \Delta^{\delta}\psi(L;\xi),$$

where δ is a scalar, ξ is a $p \times 1$ vector, $\theta = (\delta, \xi')'$and the functions Δ^{δ} and $\psi(L;\xi)$ are described as follows. Defining the difference operator $\Delta = 1 - L$, 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 non-integer $\delta > 0$, 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 $0/0 = 1$ and $1(.)$ to be the indicator function; $\psi(L;\xi)$ is a known function of its arguments such that for complex-valued x, $|\psi(x;\xi)| \neq 0$, $|x| \leq 1$ and is continuously differentiable in ξ, and in the expansion

$$\psi(L;\xi) = \sum_{j=0}^{\infty} \psi_j(\xi) L^j,$$

the coefficients $\psi_j(\xi)$ satisfy

$$\psi_0(\xi) = 1, \ \ \left|\psi_j(\xi)\right| + \left\|\dot{\psi}_j(\xi)\right\| = O\left(\exp\left(-c(\xi) j\right)\right), \tag{3}$$

where $\dot{\psi}_j(\xi) = (\partial/\partial\xi)\psi(\xi)$ and $c(\xi)$ is a positive-valued function of ψ. Note that

$$\lambda_j(\theta) = \sum_{k=0}^{j} \pi_{j-k}(\delta)\psi_k(\xi), \quad j \geq 0. \tag{4}$$

In general it is assumed that (3) holds for all $\xi \in \Xi$ with $c(\psi)$ satisfying

$$\inf_{\Xi} c(\xi) = c^* > 0. \tag{5}$$

We impose the identifiability condition that, for all $\xi \neq \xi_0$, $|\psi(x;\xi)| \neq |\psi(x;\xi_0)|$ on a subset of $\{x : |x| = 1\}$ of positive Lebesgue measure. Define

$$\phi(L;\xi) = \psi^{-1}(L;\xi) = \sum_{j=0}^{\infty} \phi_j(\xi) L^j$$

and

$$\chi(L;\xi) = \frac{\partial}{\partial\theta} \log\lambda(L;\theta) = (\log\Delta, (\partial/\partial\xi')\log\psi(L;\xi))' = \sum_{j=0}^{\infty} \chi_j(\xi) L^j,$$

where the prime denotes transposition and

$$\chi_j(\xi) = \left(\chi_{1j}(\xi), \chi_{2j}'(\xi)\right)', \ \chi_{1j}(\xi) = -j^{-1}, \ \ \chi_{2j}(\xi) = \sum_{k=1}^{j} \phi_k(\xi) \dot{\psi}_{j-k}(\xi),$$

with the $\phi_k(\xi)$ given by $\sum_{k=0}^{\infty} \phi_k(\xi) L^j = \psi(L;\xi)^{-1}$. Then define the $(p+1)\times(p+1)$ matrix

$$B(\xi) = \sum_{j=1}^{\infty} \chi_j(\xi) \chi_j'(\xi) = \begin{bmatrix} \pi^2/6 & -\sum_{j=1}^{\infty} \chi_{2j}'(\xi)/j \\ -\sum_{j=1}^{\infty} \chi_{2j}(\xi)/j & \sum_{j=1}^{\infty} \chi_{2j}(\xi)\chi_{2j}'(\xi) \end{bmatrix},$$

and assume $B(\xi_0)$ is non-singular. The conditions on $\psi(L;\xi)$ are satisfied by the coefficients in stationary and invertible autoregressive moving average sequences, and the conditions on $\lambda(L;\theta)$ are satisfied by the coefficients in FARIMA sequences. The above setting is identical to that of Robinson and Velasco (2015), but we extend their model in the following three respects.

First, we relax their assumption of independence and identity of distribution of the unobservable elements ε_{it} of the vectors $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ across t to martingale difference structure for levels and squares/products, in particular, almost surely,

$$E(\varepsilon_{it} | \varepsilon_{j,t-s}, s \geq 1, j \geq 1) = 0, \ i \geq 1, \ t \geq 0, \tag{6}$$

$$E(\varepsilon_{it}\varepsilon_{ij} - \sigma_{ij0} | \varepsilon_{k,t-s}, \ \varepsilon_{l,t-s}, s \geq 1, \ k,l \geq 1) = 0, \ \ i,j \geq 1, \ t \geq 0, \tag{7}$$

where (6) implies $E(\varepsilon_{it}) = 0$. We impose also the moment condition

$$\sup_{i,t} E\varepsilon_{it}^4 < \infty \tag{8}$$

and the fourth cumulant condition

$$\sup_t \sum_{i,j,k,l=1}^{N} |cum(\varepsilon_{it}, \varepsilon_{jt}, \varepsilon_{kt}, \varepsilon_{lt})| = O(N), \tag{9}$$

both of which are automatically satisfied if ε_{it} is Gaussian.

Secondly, we relax their assumption of homoscedasticity and lack of correlation across elements of the vectors ε_t, in allowing $E\varepsilon_t\varepsilon_t' = \Sigma_{0N}, \;\; t = 0, 1, \ldots, T,$ for an $N \times N$ matrix $\Sigma_{0N} = (\sigma_{ij0})$ which is assumed to stay positive definite and have bounded elements with increasing N but is otherwise unknown, to reflect possible cross-sectional correlation and heteroscedasticity, or else is restricted to be diagonal, to reflect an assumed lack of cross-sectional correlation, but the possibility of heteroscedasticity. Since N is allowed to increase, Σ_{0N} can in either case be regarded as nonparametric. Specifically, we assume that

$$\overline{\lim_{N\to\infty}} \left(\|\Sigma_{0N}\| + \left\|\Sigma_{0N}^{-1}\right\| \right) < \infty, \tag{10}$$

(where $\|.\|$ denotes spectral norm), and existence of

$$\sigma_0^2 = \lim_{N\to\infty} \frac{1}{N} tr\left(\Sigma_{0N}\right), \tag{11}$$

and of

$$\kappa_0 = \lim_{N\to\infty} \frac{1}{N} tr\left(\Sigma_{0N}^2\right), \tag{12}$$

Of course $N^{-1}tr\left(\Sigma_{0N}\right) \leq \|\Sigma_{0N}\|$ while also, from the inequality $tr\left(ABB'A'\right) \leq \|A\|^2 tr\left(BB'\right)$ (see e.g. Horn and Johnson (1988, p. 313)),
$N^{-1}tr\left(\Sigma_{0N}^2\right) \leq \|\Sigma_{0N}\| N^{-1}tr\left(\Sigma_{0N}\right)$, so the first component of (10) implies $\sigma_0^2 + \kappa_0 < \infty$, while the second implies $\sigma_0^2 > 0$. Condition (10) effectively upper- and lower-bounds variances and mildly limits the extent of cross-sectional dependence. Under cross-sectional uncorrelatedness assumption (10) reduces to

$$\sup_i \left(\sigma_{ii0} + \sigma_{ii0}^{-1}\right) < \infty.$$

Our final extension allows the $\alpha_t = (\alpha_{1t}, ..., \alpha_{Nt})'$ in (1) to vary with t, being vectors of unobserved nonparametric (possibly stochastic, and cross-sectionally dependent) trending individual effects, such that $\alpha_{it} = \alpha_i\left(t/T\right),$ for functions $\alpha_i\left(u\right)$ which satisfy a (possibly stochastic) Lipschitz condition

$$\sup_{i,t} \left|\alpha_i\left(t\right) - \alpha_i\left(t+u\right)\right| = O_p\left(|u|\right) \;\; as \;\; u \to 0. \tag{13}$$

Here there is no exogeneity requirement on $\alpha_i\left(t\right)$, with no restriction on possible dependence with the ε_{it}. However we find that if a strong exogeneity condition is imposed, and a related condition to (13) is added, we can slightly increase the range of values of δ_0 covered and relax the restrictions on increase of N with T. In particular:

$$\left\{\alpha_i\left(t\right) - \alpha_i\left(t+u\right), \;\; t, t+u \in (0,1], \;\; i \geq 1\right\} \text{ is independent of } \left\{\varepsilon_{it}, \;\; i \geq 1, t \geq 0\right\} \tag{14}$$

and

$$\sup_{i,t} E\left(\alpha_i(t) - \alpha_i(t+u)\right)^2 = O\left(u^2\right) \text{ as } u \to 0. \tag{15}$$

Note that neither (13) nor (15) implies the other, though they are closely related. To illustrate these various conditions, suppose $\alpha_i(t)$ has the formal separable infinite series representation

$$\alpha_i(t) = \beta_i + \sum_{j=1}^{\infty} \beta_{ij} g_{ij}(t), \ i \geq 1, \tag{16}$$

where β_i is completely unrestricted (as it can be fully differenced out), the functions $g_{ij}(t)$ are nonstochastic, and the β_{ij} can be random variables. Then since

$$|\alpha_i(t) - \alpha_i(t+u)| \leq \left(\sum_{j=1}^{\infty} \beta_{ij}^2 \sum_{j=1}^{\infty} \left(g_{ij}(t) - g_{ij}(t+u)\right)^2 \right)^{1/2}$$

a sufficient condition for (13) is

$$\sup_i \sum_{j=1}^{\infty} \beta_{ij}^2 = O_p(1), \ \sup_i \sum_{j=1}^{\infty} \left(g_{ij}(t) - g_{ij}(t+u)\right)^2 = O\left(u^2\right),$$

while the latter and

$$\sup_i E\left(\sum_{j=1}^{\infty} \beta_{ij}^2 \right) < \infty, \ \left\{\beta_{ij}, \ i, j \geq 1\right\} \text{ is independent of } \left\{\varepsilon_{it}, \ i \geq 1, t \geq 0\right\}$$

are collectively sufficient for (14) and (15). An example of (16) are polynomials in t, with degree and (possibly-stochastic) coefficients allowed to vary with i. In any case the trending individual effects are nonparametric, as in fixed-design nonparametric regression (see eg Cai (2007)), though here they can be stochastic, and also the fact that there are N of them, where N (like T) will be regarded as increasing in our asymptotic theory, lends a further nonparametric aspect. We can compare our approach with the familiar one (see e.g. Hsiao (2014)) in which our α_{it} is replaced by the addition of a separable individual effect and time effect, namely $\beta_i + \gamma_t$. Then first differencing, which is employed in the current paper, eliminates the individual effect β_i completely, but not the time effect γ_t. If, however we take $\gamma_t = \gamma(t/T)$, for a Lipschitz continuous function $\gamma(u)$, we obtain $\gamma_t - \gamma_{t-1} = O_p\left(T^{-1}\right)$, which becomes small as T diverges. We obtain the same error bound by differencing our α_{it} given condition (13), but clearly they afford more generality than the additive $\beta_i + \gamma(t/T)$, since our time trends can vary over individuals. It is thus possible that differenced estimation of θ_0 that ignores the α_{it} is more or less robust to their presence, and the extent of this will be examined below. The more specialised structure $\beta_i + \gamma(t/T)$ was also employed by Robinson (2012), but there the focus was on estimating the nonparametric function $\gamma(u)$. Note that the dependence on T of the α_t in (1) implies that the y_t form a triangular array but our notation suppresses reference to this fact.

3. First difference estimates

We consider in the current section one of the estimates of θ_0 proposed by Robinson and Velasco (2015) in their more specialised model, with our goal being to achieve the same property of consistency as if α_t were absent from (1) or constant over t and the elements of ε_t were cross-sectionally uncorrelated and homoscedastic, and asymptotic normality with the same convergence rate as before but with a different asymptotic variance matrix, and also to justify feasible large sample inference.

From (1) and (13) or (15) it follows that

$$\begin{aligned} \Delta y_t &= \Delta v_t + \Delta \alpha_{it} && (17) \\ &= \Delta v_t + O_p(T^{-1}), \text{ as } T \to \infty, && (18) \end{aligned}$$

where the final term on the right of (18) in fact represents a vector with elements that are $O_p(T^{-1})$ uniformly in i and t, and

$$v_t = \lambda_t^{-1}(L;\theta_0)\,\Delta\alpha_{it},\ t = 0,...,T. \quad (19)$$

For any $\theta \in \Theta$, we attempt to fully whiten the data by forming the $N \times T$ matrix

$$Z(\theta) = (z_1(\theta)\,....,z_T(\theta)),$$

where

$$z_t(\theta) = \lambda_t\left(L;\theta^{(-1)}\right)(\Delta y_t),\ t = 1,\dots,T, \quad (20)$$

with $\theta^{(-1)} = (\delta - 1, \xi')'$. From (27) of Robinson and Velasco (2015) and (17)

$$z_t(\theta) = \lambda_t(L;\theta)\,v_t - \tau_t(\theta)\,\varepsilon_0 + \lambda_t\left(L;\theta^{(-1)}\right)\Delta\alpha_{it} \quad (21)$$

where $\tau_t(\theta) = \lambda_t(1;\theta)$. We have $\lambda_t(L;\theta_0)\,v_t = \varepsilon_t$ but our estimates will be based on (20) as if the two error terms in (21) were absent, though our theoretical justification will take account of them.

Define the set $\Theta = D \times \Xi$, where Ξ is a compact subset of R^p and $D = [\underline{\delta}, \overline{\delta}]$, where $\underline{\delta} > \max\left(0, \delta_0 - \frac{1}{2}\right)$, $\overline{\delta} < \infty$, and we assume that $\delta_0 \in D$. The feasibly bias-corrected difference estimate of Robinson and Velasco (2015) is

$$\widehat{\theta}_T^D = \arg\min_{\theta\in\Theta} L_T^D(\theta) - T^{-1} b_T^D\left(\arg\min_{\theta\in\Theta} L_T^D(\theta)\right),$$

where

$$L_T^D(\theta) = \frac{1}{NT} tr\left(Z(\theta)\,Z(\theta)'\right), \quad (22)$$

$$b_T^D(\theta) = -B^{-1}(\xi)\left(S_{\tau\dot{\tau}T}(\theta) - S_{\tau\chi T}(\theta)\right),$$

$$S_{\tau\dot{\tau}T}(\theta)=\sum_{t=1}^{T}\tau_t(\theta)\,\dot{\tau}_t(\theta),\;\; S_{\tau\chi T}(\theta)=\sum_{t=1}^{T}\tau_t(\theta)\,\chi_t(\xi)\,,$$

where $\dot{\pi}_t(\delta)=(\partial/\partial\delta)\,\pi_t(\delta)$ and

$$\dot{\tau}_t(\theta)=\frac{\partial}{\partial\theta}\tau_t(\theta)=\left[\begin{array}{cc}\sum_{k=0}^{t}\dot{\pi}_k(\delta)\sum_{j=0}^{t-k}\psi_j(\xi) & \sum_{k=0}^{t}\pi_k(\delta)\sum_{j=0}^{t-k}\dot{\psi}_j'(\xi)\end{array}\right]', \tag{23}$$

The basic objective function (22) is of conditional sum of squares type, but corrupted by the second and third components on the right hand side of (21), whose presence will be accounted for in the theoretical development.

Finally define

$$\mu_0=\kappa_0/\sigma_0^4.$$

Theorem 3.1 *As $T\to\infty$*

$$\widehat{\theta}_T^D\to_p\;\theta_0. \tag{24}$$

If also (14) and (15) are not imposed but $\frac{1}{2}<\delta_0\leq 1$ and $NT^{1-2\delta_0}\log^2 T\to 0$ or $\delta_0>1$ and $NT^{-1}\to 0$, or if (14) and (15) are imposed with $\frac{1}{4}<\delta_0\leq\frac{1}{2}$ and $NT^{1-4\delta_0}\log^4 T\to 0$, as $T\to\infty$,

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^D-\theta_0\right)\to_d\mathcal{N}\left(0,\mu_0 B^{-1}(\xi_0)\right). \tag{25}$$

The consistency (24) requires no further restriction on δ_0, beyond $\delta_0\in D$, and no restriction on the rate of increase of N with T. However, comparing with Robinson and Velasco's (2015) Theorem 5.2 for their much more special model, we restrict δ_0 and N in order to achieve asymptotic normality (25), in particular requiring δ_0 to take nonstationary values when the α_t are not exogenous; this is due to bias produced by temporal variation in individual effects.

In order to base statistical inference on Theorem 3.1 we estimate $B(\xi_0)$ by $B\left(\widehat{\xi}_T^D\right)$, with $\widehat{\xi}_T^D$ denoting the final p elements of $\widehat{\theta}_T^D$, and estimate μ_0 by

$$\widehat{\mu}_T^D=Ntr\left(\widetilde{\Sigma}_N^2\left(\widehat{\theta}_T^D\right)\right)/tr^2\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)\right), \tag{26}$$

where

$$\widetilde{\Sigma}_N(\theta)=\frac{1}{T}Z(\theta)\,Z(\theta)'. \tag{27}$$

An analogous unrestricted estimate of the cross-sectional covariance matrix was considered by Robinson (2012) in the context of a panel model with no autocorrelation. If instead we maintain cross-sectional uncorrelatedness we take

$$\widetilde{\widetilde{\Sigma}}_N(\theta)=\frac{1}{T}diag\left(Z(\theta)\,Z(\theta)'\right). \tag{28}$$

Theorem 3.2 *Under the conditions of Theorem 3.1, as $T \to \infty$,*

$$B\left(\widehat{\xi}_T^D\right) \to_p B(\xi_0), \quad (29)$$

$$\widehat{\mu}_T^D \to_p \mu_0, \quad (30)$$

with $\widehat{\mu}_T^D$ given by (26) where $\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)$ is defined either by (27) or, if $\sigma_{ij0} = 0$ for all $i \neq j$, (28), $(NT)^{\frac{1}{2}}\left(B\left(\widehat{\xi}_T^D\right)/\widehat{\mu}_T^D\right)^{1/2}\left(\widehat{\theta}_T^D - \theta_0\right)$ converges in distribution to a vector of independent standard normal random variables.

4. Pseudo maximum likelihood estimate based on first differences

Our next estimate is the difference pseudo maximum likelihood estimate (PMLE) of Robinson and Velasco (2015). Define the $T \times T$ matrix, $\Omega_T(\theta) = (\omega_{st}(\theta))$, $\omega_{st}(\theta) = 1(s=t) + \tau_s(\theta)\tau_t(\theta)$, so that $\Omega_T(\theta_0)$ is proportional to the exact covariance matrix of the vector $(\varepsilon_{i1} - \tau_t(\theta_0)\varepsilon_{i0}, ..., \varepsilon_{iT} - \tau_T(\theta_0)\varepsilon_{i0})'$, cf (21). Unlike in the difference estimate of the previous section we thus allow for the initial value effect in our estimation, though as there we attempt to incorporate cross-sectional correlation or heteroscedasticity only in studentization, not in the point estimation of θ_0, and we have to contend in the theory with the $O_p(T^{-1})$ error of differencing the α_{it} in (21).

We estimate θ_0 by

$$\widehat{\theta}_T^P = \arg\min_{\theta \in \Theta} L_T^P(\theta),$$

where Θ is as defined in the previous section and

$$L_T^P(\theta) = |\Omega_T(\theta)|^{\frac{1}{T}} \widehat{\sigma}_T^2(\theta), \quad (31)$$

in which

$$\widehat{\sigma}_T^2(\theta) = \frac{1}{NT} tr\left(Z(\theta)\Omega_T^{-1}(\theta) Z(\theta)'\right).$$

.

Theorem 4.1 *As $T \to \infty$,*

$$\widehat{\theta}_T^P \to_p \theta_0. \quad (32)$$

If also (14) and (15) are not imposed but $\frac{1}{2} < \delta_0 \leq 1$ and $NT^{1-2\delta_0}\log^2 T \to 0$ or $\delta_0 > 1$ and $NT^{-1} \to 0$, or if (14) and (15) hold with $\frac{3}{8} < \delta_0 \leq \frac{1}{2}$ and $NT^{3-8\delta_0}\log^4 T \to 0 + NT^{-1}\log^2 T \to 0$, as $T \to \infty$,

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^P - \theta_0\right) \to_d \mathcal{N}\left(0, \mu_0 B^{-1}(\xi_0)\right). \quad (33)$$

The cost of allowing temporal variation in individual effects is greater than with the difference estimates, especially as the corresponding result (Theorem 4.4) of Robinson and Velasco (2015) imposed no restrictions on δ_0 and N.

Now define

$$\widehat{\mu}_T^P = Ntr\left(\widetilde{\Sigma}_N^2\left(\widehat{\theta}_T^P\right)\right)/tr^2\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^P\right)\right) \tag{34}$$

and denote by $\widehat{\xi}_T^P$ the final p elements of $\widehat{\theta}_T^P$.

Theorem 4.2 *Under the conditions of Theorem 4.1, as* $T\to\infty$*,*

$$\begin{aligned} B\left(\widehat{\xi}_T^P\right) &\to_p B\left(\xi_0\right), \\ \widehat{\mu}_T^P &\to_p \mu_0, \end{aligned}$$

with $\widehat{\mu}_T^D$ *given by (34) where* $\widetilde{\Sigma}_N\left(\widehat{\theta}_T^P\right)$ *is defined either using (27) or, if* $\sigma_{ij0}=0$ *for all* $i\neq j$, (28), $(NT)^{\frac{1}{2}}\left(B\left(\widehat{\xi}_T^P\right)/\widehat{\mu}_T^P\right)^{1/2}\left(\widehat{\theta}_T^P-\theta_0\right)$ *converges in distribution to a vector of independent standard normal random variables.*

5. Inefficiency of estimation

In the limiting covariance matrix $\mu_0 B^{-1}\left(\xi_0\right)$ in Theorems 3.1 and 4.1, the factor $\mu_0=1$ when the ε_{it} are homoscedastic and uncorrelated across i, but in general $\mu_0\geq 1$ so heteroscedasticity and/or cross-sectional correlation inflates the variance matrix in the limiting distribution by a scalar factor, relative to the outcome of Robinson and Velasco (2015). Note also that $\widehat{\mu}_T^D\geq 1$ whether it is based on either of the estimates (27) or (28) of Σ_{0N}.

The potential inefficiency of our estimates, or equivalently the degree of invalidity of the inference rules which assume homoscedasticity and lack of correlation across i, can be examined by considering a specific model for ε_t. The spatial moving average model is defined by

$$\varepsilon_t = \left(I_N+\rho W\right)\eta_t, \tag{35}$$

where I_N is the $N\times N$ identity matrix, W is an $N\times N$ user-chosen 'spatial weight' matrix with zero diagonal elements, taken here to be symmetric and normalised such that $\|W\|=1$, and correspondingly the scalar ρ satisfies $|\rho|<1$, while the elements of η_t are mutually uncorrelated with common variance ς^2 . Thus

$$tr\left(\Sigma_{0N}\right)=\varsigma^2 tr\left(\left(I+\rho W\right)^2\right)=\varsigma^2\left(N+\rho^2 tr\left(W^2\right)\right),$$

$$tr\left(\Sigma_{0N}^2\right)=\varsigma^4 tr\left(\left(I+\rho W\right)^4\right)=\varsigma^4\left(N+6\rho^2 tr\left(W^2\right)+4\rho^3 tr\left(W^3\right)+\rho^4 tr\left(W^4\right)\right),$$

and so

$$\begin{aligned}\frac{Ntr\left(\Sigma_{0N}^2\right)}{\left(tr\left(\Sigma_{0N}\right)\right)^2} &= \frac{N\left(N+6\rho^2 tr\left(W^2\right)+4\rho^3 tr\left(W^3\right)+\rho^4 tr\left(W^4\right)\right)}{\left(N+\rho^2 tr\left(W^2\right)\right)^2} \\ &= 1+\frac{4\rho^2 Ntr\left(W^2\right)+4\rho^3 Ntr\left(W^3\right)+\rho^4\left(Ntr\left(W^4\right)-tr\left(W^2\right)^2\right)}{\left(N+\rho^2 tr\left(W^2\right)\right)^2} \\ &\geq 1+\frac{4\rho^2 N\left(tr\left(W^2\right)+\rho tr\left(W^3\right)\right)}{\left(N+\rho^2 tr\left(W^2\right)\right)^2}. \qquad (36)\end{aligned}$$

This lower bound is 1 when there is no spatial correlation, $\rho = 0$, but in general (36) exceeds 1, noting that $tr\left(W^2\right)+\rho tr\left(W^3\right) > 0$ even when $-1 < \rho < 0$ since $\|W\| \leq 1$ implies $tr\left(W^3\right) \leq tr\left(W^2\right)$, and (36) increases in ρ^2. A simple W, proposed by Case (1991), is

$$W = I_r \otimes B_s,\ B_s = (s-1)^{-1}\left(1_s 1_s' - I_s\right), \qquad (37)$$

where $rs = N$ and 1_s is the $s \times 1$ vector of 1's, representing r districts each containing s farms, so farms are neighbours if and only if they lie in the same district and neighbours are equally weighted. Since

$$B_s^2 = (s-1)^{-2}\left((s-2)\,1_s 1_s' + I_s\right), B_s^3 = (s-1)^{-3}\left(\left(s^2-3s+3\right)1_s 1_s' - I_s\right),$$

we have

$$\begin{aligned}tr\left(W^2\right) &= r(s-1)^{-2}\left(s(s-2)+s\right) = N(s-1)^{-1}, \\ tr\left(W^3\right) &= r(s-1)^{-3}\left(s\left(s^2-3s+3\right)-s\right) = N(s-2)(s-1)^{-2},\end{aligned}$$

and the lower bound (36) becomes

$$1+\frac{4\rho^2\left((s-1)^{-1}+\rho(s-2)(s-1)^{-2}\right)}{\left(1+\rho^2(s-1)^{-1}\right)^2} = 1+\frac{4\rho^2\left(s-1+\rho(s-2)\right)}{\left(s-1+\rho^2\right)^2} \qquad (38)$$

A lower bound for μ_0 is obtained by letting $N \to \infty$, and if $s \to \infty$ this tends to 1, but if s stays fixed the bound exceeds 1, and again increases with ρ^2.

6. Monte Carlo simulations

In this section we report the results of a simulation study of the properties of our estimates in finite samples in the presence of cross-sectional dependence and trends. We extend a similar set-up of Robinson and Velasco (2015), where only constant fixed effects and independent and identically distributed ε_{it}, across both i and t, were employed.

We generate the ε_t as the spatial moving average (35), where η_t is $\mathcal{N}(0, I_N)$ and $\rho = 0.5$ or 0.9, while W is generated as in (37), setting $r = 4$ (when $N = 8$) or $r = 5$ ($N > 8$). We consider both a pure fractional model, so $p = 1$, $\theta = \delta$ and $\psi(L; \xi) \equiv 1$, and a model with FARIMA(1,δ,0) dynamics, so $p = 2$ and $\theta = (\delta, \xi)'$ with $\psi(L; \xi) = 1 - \xi L$.

We consider different choices of N, T and θ_0. In particular we employ three basic values of NT, namely $100, 200, 400$ with two combinations of N and T for each, to account for relatively short ($T = 10, 22$) and moderate time series ($T = 12, 25, 50$), and also $NT = 96$ for $T = 12$. The range of values of N thus varies from 8 through 20 from the smallest to the largest sample size. The values of δ_0 include a stationary one ($\delta_0 = 0.3$), which our theorems predict will be the most problematic from the point of view of bias, a moderately non-stationary one ($\delta_0 = 0.6$), a value close to the unit root ($\delta_0 = 0.9$), and a more nonstationary one ($\delta_0 = 1.2$). For FARIMA models we consider two autoregressive parameter values, $\xi = 0.5, 0.8$. Optimizations were carried out using the Matlab function fmincon with $D = [0.1, 1.5]$ and $\Xi = [-0.95, 0.95]$, and the results are based on 10,000 independent replications.

For all combinations of sample sizes and parameter values we report (scaled) empirical bias of both the feasibly bias-corrected difference estimate $\widehat{\theta}_T^D$ and the PMLE $\widehat{\theta}_T^P$, root-mean square error (MSE) and empirical size of the corresponding t or Wald test based on estimates of μ and B for the asymptotic variance as studied in Theorems 3.2 and 4.2 ("corrected" tests) to account for cross sectional dependence, while we also report the empirical size for tests based on a pooled estimate of the variance of innovations, which would be only valid in case of uncorrelated homoscedastic innovations ("uncorrected" tests).

The results in Tables 1-8 concern the pure fractional case, $\theta = \delta$. We first consider the case without trends, where constant fixed effects are exactly removed by first differencing. Table 1 provides a bias comparison of the estimates of δ_0. In general bias is not affected by cross-sectional correlation compared with the results in Robinson and Velasco (2015), who did not allow for such, and typically reduces with increase of δ_0 and T as expected. Bias of the difference estimate $\widehat{\delta}_T^D$ for $\delta_0 = 0.3$ can be of an order of magnitude larger than in the other cases for the smallest T for a given NT, despite bias-correction having a large beneficial effect (results without bias correction are not reported here). The PMLE $\widehat{\delta}_T^P$ does much better in this difficult setup, but the difference for larger δ_0 is much smaller. MSE results in Table 2 confirm the consistency of estimates, no clear superiority of any of the two estimates apart from the bias effect, and increase in the variance of estimates with cross-sectional correlation (that is, with increasing ρ). The empirical size of the properly studentized t-test is sensitive to the parameter and sample size values, though they converge with increasing NT to the nominal value. In general, performance tends to deteriorate with the larger ρ and smaller δ_0, and PMLE-based tests do better than difference ones. Tests not using consistent estimates of μ_0 under cross-sectional correlation are systematically very oversized in all cases, cf. Table 4.

We repeat the experiment in Tables 5-8 for $\rho = 0.8$ but with a linear trend specified as $\alpha_i(u) = \beta_i u$ (cf (16)), where the β_i are generated independently from the $\mathcal{N}(0, \gamma^2)$ distribution with $\gamma = 1, 3$ (and our estimates are of course invariant to fixing a temporally constant component of the individual effects at zero). The larger value of γ can generate relatively large trends that can dominate the behaviour of the time series as shown by the bias and MSE results in Tables 5 and 6, respectively. However, for the smaller γ, first differencing seems to account properly for the heterogenous deterministic component, though bias is substantially increased, as is MSE, for the small values of NT and T. In this case PMLE-based t-tests perform in a similar way as in the absence of trend for not too small T and NT and $\delta_0 \geq 0.6$.

In Tables 9-14 we consider the results for the FARIMA model, again considering for only the value of ρ, 0.9. The estimation of δ_0 is substantially affected by the autoregressive short run dynamic component, and the bias in Table 9 can change sign with the value of δ_0 and of T, for $\xi_0 = 0.5$, the results improving in most cases with increasing δ_0, while bias is always positive for the largest ξ_0, 0.8, so persistence is incorporated in the estimation of δ_0 in finite samples. In general, the PMLE dominates bias-corrected difference estimates again. There is an overall increase in variability in Table 10 compared to Table 2, since both parameter estimates are highly correlated. Estimation results for ξ_0 in Tables 11-12 are parallel to the ones for δ_0: large bias for small NT and δ_0, negative bias in all cases for $\xi_0 = 0.8$, but no clear pattern for $\xi_0 = 0.5$, and MSE decreasing with NT, T and δ_0. The feasible asymptotic inference results reported in Tables 12 and 14 confirm previous ideas on the need for consistent estimation of μ_0 to account for cross-sectional correlation, though now oversizing is more severe for the larger values of T across the whole range of values of δ_0 and ξ_0.

7. Final Comments

In a semiparametric panel data model with fractional dynamics we have established desirable and useful asymptotic properties of estimates of time series parameters that are robust to nonparametric, time-varying individual effects and to cross-sectional correlation and heteroscedasticity, at the cost of restrictions on the range of possible values of the memory parameter and the rate of increase of N with T. Some further issues that might be considered are as follows.

1. As an alternative to our nonparametric estimation of the factor μ_0 that inflates the limiting covariance matrix we could invest in a parametric model for the covariance matrix of ε_t, such as a factor model (see eg Ergemen and Velasco (2014)) or a spatial model (cf. the discussion in Section 5). With a correct specification improved finite-sample properties are likely to result, though a misspecified parameterization would lead to inconsistent estimation of μ_0 and thus invalidate inferences based on Theorems 3.2 and 4.2.

2. Point estimates of θ_0 that use either our nonparametric estimates of Σ_{0N}, or parametric ones such as just described, to correct for cross-sectional correlation/heteroscedasticity to the extent of being asymptotically efficient in a Gaussian context, can be constructed. Their investigation would be worthwhile but likely entail a considerable amount of further work and possibly some further restrictions on δ_0 and T.

3. Whereas we impose the same dynamics over the cross-section, Hassler, Demetrescu and Tarcolea (2011) developed tests in a panel with a fractional structure which is allowed to vary across units, and with allowance for cross-sectional dependence, but without allowing for individual effects and keeping N fixed as $T \to \infty$. In our context where N can increase it would be possible to keep the number of time series parameters fixed by assuming they are constant within finitely-many known cross-sectional subsets, over which the parameters can vary.

4. Though the nonstochastic conditional covariance matrix of innovations assumed in (7) is common in the time series literature, the implications of allowing for conditional heteroscedasticity could be explored.

5. Observable explanatory variables might be allowed to enter, in either a parametric or nonparametric way. In the former case, if they are linearly involved our differencing will leave only their first differences and initial value, but with nonlinear or nonparametric modelling we will get a difference of the functions, which structure, in the nonparametric case, needs to be exploited via additive nonparametric regression methodology in order to minimize a curse of dimensionality.

Appendix

Proof of Theorem 3.1

To prove (24) we first prove consistency of $\arg\min_{\theta\in\Theta} L_T^D(\theta)$. The proof of this extends that of Theorem 3.3 of Robinson and Velasco (2015) (hereafter RV) which uses their Lemma 2, which clearly still holds in our setting, and their Proposition 1, which needs extending because we have relaxed their iid assumption on ε_{it}. First consider RV's $A_T(\theta)$, and their $U(\theta)$, which is defined as there but with σ_0^2 defined in (11). $A_T(\theta) - A_T(\theta_0) - U(\theta)$ is

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\nu_j^2(\theta)\sum_{t=0}^{T-j}\left(\varepsilon_{it}^2-\sigma_{ii0}\right)+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{k=0}^{j-1}\nu_j(\theta)\nu_k(\theta)\varepsilon_{i,t-j}\varepsilon_{i,t-k}\\
&+\left(\frac{1}{N}tr\left(\Sigma_{0N}\right)-\sigma_0^2\right)\frac{1}{T}\sum_{j=1}^{T}(T-j+1)\nu_j^2(\theta)\\
&-\frac{\sigma_0^2}{T}\sum_{j=1}^{T}\left(j-1\right)\nu_j^2(\theta)-\sigma_0^2\sum_{j=T+1}^{\infty}\nu_j^2(\theta) \qquad (39)
\end{aligned}
$$

(there is a typo in the last term on p.449 of RV). As in RV

$$
\sup_{\Theta}\left|\sum_{i=1}^{N}\sum_{j=1}^{T}\nu_j^2(\theta)\sum_{t=0}^{T-j}\left(\varepsilon_{it}^2-\sigma_{ii0}\right)\right|\le\sum_{i=1}^{N}\sup_{\Theta}\left|\sum_{j=1}^{T}\nu_j^2(\theta)\sum_{t=0}^{T-j}\left(\varepsilon_{it}^2-\sigma_{ii0}\right)\right|,
$$

which is uniformly $o_p(NT)$ much as in RV using also the proof of Theorem 1 of Hualde and Robinson (2011) (6) and (8). The second term in (39) is uniformly $o_p(1)$ in much the same way, and the remaining terms are uniformly $o(1)$ from (11) and (52) of RV. Since from (21) $L_T^D(\theta) = \sum_{t=1}^{T}\left\|\lambda_t(L;\theta)v_t - \tau_t(\theta)\varepsilon_0 + \lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_t\right\|^2$ it remains to show that the term in $\Delta\alpha_t$ contributes negligibly. First,

$$
\begin{aligned}
\sup_{\Theta}\sum_{t=1}^{T}\left\|\lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_t\right\|^2 &\le \sup_t\left\|\Delta\alpha_t\right\|^2\sum_{t=1}^{T}\left(\sup_{\Theta}\sum_{j=0}^{t-1}\left|\lambda_j\left(\theta^{(-1)}\right)\right|\right)^2\\
&= O_p\left(NT^{-2}\sum_{t=1}^{T}\left(t^{1-\underline{\delta}}+1\right)^2\right)\\
&= O_p\left(N\left(T^{1-2\underline{\delta}}+T^{-1}\right)\right)=o_p\left(NT\right) \qquad (40)
\end{aligned}
$$

as $T\to\infty$ for $\underline{\delta}>0$. Next using the Cauchy inequality,

$$
\begin{aligned}
&\sup_{\Theta}\left|\sum_{t=1}^{T}\left(\lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_t\right)'\lambda_t\left(L;\theta\right)v_t\right|\\
\le&\left(\sup_{\Theta}\sum_{t=1}^{T}\left\|\lambda_t\left(L;\theta\right)v_t\right\|^2\right)^{1/2}\left(\sup_{\Theta}\sum_{t=1}^{T}\left\|\lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_t\right\|^2\right)^{1/2}\\
=&\ O_p\left(\left(NT\right)^{1/2}\right)o_p\left(\left(NT\right)^{1/2}\right)=o_p\left(NT\right), \qquad (41)
\end{aligned}
$$

since the first term converges uniformly to a bounded function after standardization by $(NT)^{-1}$, while proceeding in a similar way,

$$\begin{aligned}
&\sup_{\Theta}\left|\sum_{t=1}^{T}\left(\lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_t\right)'\tau_t(\theta)\varepsilon_0\right| \\
\leq &\left(\sup_{\Theta}\sum_{t=1}^{T}\|\tau_t(\theta)\varepsilon_{i0}\|^2\right)^{1/2}\left(\sup_{\Theta}\sum_{t=1}^{T}\left\|\lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_t\right\|^2\right)^{1/2} \\
= &\; o_p\left((NT)^{1/2}\right)o_p\left((NT)^{1/2}\right) = o_p(NT). \qquad (42)
\end{aligned}$$

Thus consistency of $\arg\min_{\theta\in\Theta} L_T^D(\theta)$ is established, and thence straightforwardly (24), using smoothness properties of b_T^D.

The latter are also used, along with (24), in proving (25), as in RV's Theorem 5.2, which incidentally required weaker conditions on δ_0 and N than ours, so we do not repeat the details. But we have to extend RV's Theorem 4.3 on $\arg\min_{\theta\in\Theta} L_T^D(\theta) - b_T^D(\theta_0)$ to our setting, and this requires first extending the CLT for scores in Proposition 2 of RV. We can write their w_T as

$$w_T = \frac{1}{2(NT)^{\frac{1}{2}}}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}A_{iT}(\theta_0) = \frac{1}{(NT)^{\frac{1}{2}}}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\chi_{t-j}(\theta_0)\varepsilon_j'\varepsilon_t.$$

Thus

$$\begin{aligned}
Ew_Tw_T' &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\sum_{s=1}^{T}\sum_{k=0}^{s-1}\chi_{t-j}(\theta_0)\varepsilon_j'\varepsilon_t\varepsilon_s'\varepsilon_k\chi_{s-k}'(\theta_0) \\
&= \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\sum_{k=0}^{s-1}\chi_{t-j}(\theta_0)\varepsilon_j'\Sigma_{0N}\varepsilon_k\chi_{t-k}'(\theta_0) \\
&= \frac{tr\left(\Sigma_{0N}^2\right)}{NT}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\chi_{t-j}(\theta_0)\chi_{t-j}'(\theta_0).
\end{aligned}$$

Then given (12) and much as in RV,

$$w_T \to_d \mathcal{N}\left(0, \kappa_0 B(\xi_0)\right).$$

Apart from the extra terms discussed below the score and Hessian are handled much as in RV, where the latter has probability limit $2\sigma_0^2 B(\xi_0)$, with σ_0^2 is as in (11).

We consider first the extra terms in the score when (14) and (15) are not imposed. We have

$$(NT)^{\frac{1}{2}}\frac{\partial}{\partial\theta}L_T^D(\theta_0) = \frac{2}{(NT)^{\frac{1}{2}}}\sum_{t=1}^{T}\left(\left(f_t - \varepsilon_0\dot{\tau}_t^0\right)' + \dot{\lambda}_{t-1}^0(L)\Delta\alpha_t'\right)\left(\left(\varepsilon_t - \tau_t^0\varepsilon_0\right) + \Delta\alpha_t\lambda_{t-1}^0(L)\right),$$

where

$$\begin{aligned}
\tau_t^0 &= \tau_t^0(\theta_0),\ \dot{\tau}_t^0 = \dot{\tau}_t(\theta_0),\ \dot{\tau}_t(\theta) = \frac{\partial \tau_t(\theta)}{\partial \theta'}, \\
\lambda_{t-1}^0(L) &= \lambda_{t-1}\left(L;\theta_0^{(-1)}\right),\ \dot{\lambda}_{t-1}^0(L) = \dot{\lambda}_{t-1}\left(L;\theta_0^{(-1)}\right), \\
\dot{\lambda}_{t-1}\left(L;\theta^{(-1)}\right) &= \frac{\partial \lambda_{t-1}\left(L;\theta^{(-1)}\right)}{\partial \theta'},\quad f_t = \sum_{j=0}^{t-1} \varepsilon_t \chi_{t-j}'(\theta_0).
\end{aligned}$$

The additional terms contributing to the asymptotic bias of $(NT)^{\frac{1}{2}} \frac{\partial}{\partial \theta} L_T^D(\theta_0)$ that are not covered in RV are thus

$$\begin{aligned}
&\frac{2}{(NT)^{\frac{1}{2}}} \sum_{t=1}^T \left(\Delta\alpha_t \dot{\lambda}_{t-1}^0(L)\right)' \left(\varepsilon_t - \tau_t^0 \varepsilon_0\right)' + \frac{2}{(NT)^{\frac{1}{2}}} \sum_{t=1}^T \left(f_t - \varepsilon_0 \dot{\tau}_t^0\right)' \left(\Delta\alpha_t \lambda_{t-1}^0(L)\right) \\
&+ \frac{2}{(NT)^{\frac{1}{2}}} \sum_{t=1}^T \left(\Delta\alpha_t \dot{\lambda}_{t-1}^0(L)\right)' \left(\Delta\alpha_t \lambda_{t-1}^0(L)\right). \qquad (43)
\end{aligned}$$

Denoting $\lambda_j^0 = \lambda_j\left(\theta_0^{(-1)}\right)$, we have $\dot{\lambda}_j^0 = \frac{\partial}{\partial \theta'} \lambda_j\left(\theta_0^{(-1)}\right) = O\left(\left|\lambda_j^0\right| \log j\right) = O\left(j^{-\delta_0} \log t\right)$, and thus

$$\begin{aligned}
&\left\| \frac{2}{(NT)^{\frac{1}{2}}} \sum_{t=1}^T \left(\Delta\alpha_t \dot{\lambda}_{t-1}^0(L)\right)' \left(\varepsilon_t - \tau_t^0 \varepsilon_0\right) \right\| \\
\leq\ & \sup_{i,t} |\Delta\alpha_{it}| \frac{2N^{\frac{1}{2}}}{(NT)^{\frac{1}{2}}} \sum_{t=1}^T \left(\sum_{j=0}^{t-1} \left\| \dot{\lambda}_j^0 \right\| \right) \left\| \varepsilon_t - \tau_t^0 \varepsilon_0 \right\| \\
=\ & O_p\left(NT^{-1} (NT)^{-\frac{1}{2}} \sum_{t=1}^T \left(t^{1-\delta_0} + 1\right) \log t \right) \\
=\ & O_p\left(T^{-\frac{3}{2}} N^{\frac{1}{2}} \left(T^{2-\delta_0} + T\right) \log T \right) = O_p\left(N^{\frac{1}{2}} \left(T^{\frac{1}{2}-\delta_0} + T^{-\frac{1}{2}} \right) \log T \right)
\end{aligned}$$

which is $o_p(1)$ as $NT^{1-2\delta_0} \log^2 T + NT^{-1} \log^2 T \to 0$, with $\delta_0 > \frac{1}{2}$. The second term in (43) can be bounded similarly, given $E\left\| f_t - \varepsilon_0 \dot{\tau}_t^0 \right\| = O\left(N^{1/2}\right)$. The third is

$$\begin{aligned}
& O_p\left(NT^{-2} (NT)^{-\frac{1}{2}} \sum_{t=1}^T \left(t^{1-\delta_0} + 1\right)^2 \log t \right) \\
=\ & O_p\left(N^{\frac{1}{2}} T^{-\frac{5}{2}} \left(T^{3-2\delta_0} + T\right) \log T \right) = O_p\left(N^{\frac{1}{2}} \left(T^{\frac{1}{2}-2\delta_0} + T^{-3/2} \right) \log T \right),
\end{aligned}$$

which is $o_p(1)$ since $NT^{1-4\delta_0} \log^2 T + NT^{-1} \to 0.$ with $\delta_0 > \frac{1}{4}$.

Now impose (14) and (15). Let K denote a generic finite constant. From (14) the first term in (43) has zero mean (because $E[\varepsilon_{it}] = 0$) and, using (10), variance bounded by

$$\begin{aligned}
& \frac{K}{T^3} \|\Sigma_{0N}\| \sum_{t=1}^{T} \left(t^{1-\delta_0} + \log t\right)^2 \log^2 t \\
& + \frac{K}{T^3} \|\Sigma_{0N}\| \left(\sum_{t=1}^{T} \left(t^{1-\delta_0} + \log t\right) t^{-\delta_0} \log t \right)^2 \\
= & \; O\left(T^{-3}\left(T^{3-2\delta_0} + T \log T\right) \log^2 T\right) + O\left(T^{-3}\left(T^{2(1-\delta_0)} + \log T\right)^2 \log^2 T\right) \\
= & \; O\left(\left(T^{-2\delta_0} + T^{1-4\delta_0}\right) \log^3 T\right) = o(1)
\end{aligned}$$

since $\delta_0 > \frac{1}{4}$, so is negligible, because

$$\begin{aligned}
& \left\| E\left[\left(\dot{\lambda}^0_{t-1}(L)\, \Delta\alpha_{it}\right)\left(\dot{\lambda}^0_{t'-1}(L)\, \Delta\alpha_{it'}\right)\right]\right\| \\
\leq & \; E\left\|\dot{\lambda}^0_{t-1}(L)\, \Delta\alpha_{it}\right\|^2 \\
= & \; O\left(E\left|\Delta\alpha_{it}\right|^2 \left(\sum_{j=0}^{t-1} j^{-\delta_0} \log j\right)^2\right) \\
= & \; O\left(T^{-2}\left(t^{1-\delta_0} + 1\right)^2 \log^2 t\right).
\end{aligned}$$

The second term in (43) can be bounded similarly, and the last term is

$$\begin{aligned}
O_p\left(T^{-2}(NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(t^{1-\delta_0} + \log t\right)^2 \log t\right) &= O_p\left(N^{\frac{1}{2}} T^{-\frac{5}{2}} \left(T^{3-2\delta_0} + T \log^2 T\right) \log T\right) \\
&= O_p\left(N^{\frac{1}{2}}\left(T^{\frac{1}{2}-2\delta_0} + T^{-3/2} \log^2 T\right) \log T\right),
\end{aligned}$$

which is $o_p(1)$ since $NT^{1-4\delta_0} \log^2 T + NT^{-1} \to 0$ with $\delta_0 > \frac{1}{4}$ □

Proof of Theorem 3.2

Given Theorem 3.1 and continuity of $B(\xi)$ it suffices to prove (30). We give the proof only for (27) because that for (28) is simpler. From (11), $\widehat{\mu}^D_T - \mu_0$ differs by $o(1)$ from

$$Ntr\left(\widetilde{\Sigma}^2_N\left(\widehat{\theta}^D_T\right)\right) / tr^2\left(\widetilde{\Sigma}_N\left(\widehat{\theta}^D_T\right)\right) - Ntr\left(\Sigma^2_{0N}\right) / tr^2\left(\Sigma_{0N}\right),$$

which is

$$\begin{aligned}
& Ntr\left(\widetilde{\Sigma}^2_N\left(\widehat{\theta}^D_T\right) - \Sigma^2_{0N}\right) / tr^2\left(\Sigma_{0N}\right) \\
& - Ntr\left(\widetilde{\Sigma}^2_N\left(\widehat{\theta}^D_T\right)\right)\left(tr^2\left(\widetilde{\Sigma}_N\left(\widehat{\theta}^D_T\right)\right) - tr^2\left(\Sigma_{0N}\right)\right) / \left(tr^2\left(\widetilde{\Sigma}_N\left(\widehat{\theta}^D_T\right)\right) tr^2\left(\Sigma_{0N}\right)\right)
\end{aligned}$$

so the result follows on showing that

$$tr\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)-\Sigma_{0N}\right) = o_p\left(N\right), \qquad (44)$$

$$tr\left(\widetilde{\Sigma}_N^2\left(\widehat{\theta}_T^D\right)-\Sigma_{0N}^2\right) = o_p\left(N\right), \qquad (45)$$

noting that (44) implies that $tr\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)\right)/N$ has a positive, finite probability limit. We have

$$\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)-\Sigma_{0N}=\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)-\widetilde{\Sigma}_N\left(\theta_0\right)\right)+\left(\widetilde{\Sigma}_N\left(\theta_0\right)-\Sigma_{0N}\right).$$

Now

$$\widetilde{\Sigma}_N\left(\theta_0\right)=\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t+s_t\right)\left(\varepsilon_t+s_t\right)'=A_T+R_T,$$

where

$$A_T=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t',\;\; R_T=\frac{1}{T}\sum_{t=1}^{T}\left(s_t\varepsilon_t'+\varepsilon_t s_t'+s_t s_t'\right),$$

with $s_t=\lambda_{t-1}\left(L;\theta_0^{(-1)}\right)\Delta\alpha_t-\tau_t\left(\theta_0\right)\varepsilon_0$. Thus

$$tr\left(\widetilde{\Sigma}_N\left(\theta_0\right)-\Sigma_{0N}\right)=tr\left(A_T-\Sigma_{0N}+R_T\right).$$

Since

$$E\left(\varepsilon_{it}^2-\sigma_{0ii}\right)\left(\varepsilon_{jt}^2-\sigma_{0jj}\right)=2\sigma_{0ij}^2+cum\left(\varepsilon_{it},\varepsilon_{it},\varepsilon_{jt},\varepsilon_{jt}\right),$$

we have

$$\begin{aligned} Etr^2\left(A_T-\Sigma_{0N}\right) &= E\left(\frac{1}{T}\sum_{t=1}^{T}\left(\left\|\varepsilon_t\right\|^2-tr\left(\Sigma_{0N}\right)\right)\right)^2 \\ &= \frac{1}{T^2}E\left(\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\left(\varepsilon_{it}^2-\sigma_{0ii}\right)\right)\left(\sum_{j=1}^{N}\left(\varepsilon_{jt}^2-\sigma_{0jj}\right)\right)\right) \\ &= \frac{1}{T}\sum_{i,j=1}^{N}\left(2\sigma_{0ij}^2+cum\left(\varepsilon_{it},\varepsilon_{it},\varepsilon_{jt},\varepsilon_{jt}\right)\right) \\ &= \frac{2}{T}tr\left(\Sigma_{0N}^2\right)+\frac{1}{T}\sum_{i,j=1}^{N}cum\left(\varepsilon_{it},\varepsilon_{it},\varepsilon_{jt},\varepsilon_{jt}\right) \\ &= O\left(\frac{N}{T}\right)=o\left(N^2\right) \end{aligned}$$

using (12) and (9), since $NT\rightarrow\infty$. Also

$$tr\left(R_T\right)=tr\left(\frac{1}{T}\sum_{t=1}^{T}\left(s_t\varepsilon_t'+\varepsilon_t s_t'+s_t s_t'\right)\right),$$

where

$$\frac{1}{T}\sum_{t=1}^{T}\|s_t\|^2 \leq \frac{2}{T}\sum_{t=1}^{T}\left\|\lambda_t\left(\theta_0^{(-1)}\right)\varepsilon_0\right\|^2+\frac{2}{T}\sum_{t=1}^{T}\left\|\lambda_{t-1}\left(L;\theta_0^{(-1)}\right)\Delta\alpha_t\right\|^2$$
$$= o_p(N),\ \text{as } T\to\infty,$$

as in (40)-(42). Also

$$tr\left(\frac{1}{T}\sum_{t=1}^{T}s_t\varepsilon_t'\right)\leq\left(\frac{1}{T}\sum_{t=1}^{T}\|s_t\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|\varepsilon_t\|^2\right)^{1/2}=o_p(N)$$

from the above and the fact that the second factor in brackets is $O_p(N)$. Thus we have shown that $tr\left(\widetilde{\Sigma}_N(\theta_0)-\Sigma_{0N}\right)=o_p(N)$. Next, from Theorem 3.1 and for $\overline{\theta}$ such that $\left\|\overline{\theta}-\theta_0\right\|\leq\left\|\widehat{\theta}_T^D-\theta_0\right\|$,

$$\begin{aligned} tr\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right)-\widetilde{\Sigma}_N(\theta_0)\right) &= \frac{1}{T}\sum_{t=1}^{T}\left(\left\|z_t\left(\widehat{\theta}_T^D\right)\right\|^2-\|z_t(\theta_0)\|\right)\\ &= \frac{2}{T}\sum_{t=1}^{T}z_t'(\theta_0)\frac{\partial z_t\left(\overline{\theta}\right)}{\partial\theta'}\left(\widehat{\theta}_T^D-\theta_0\right)\\ &= O_p\left(\left(\frac{1}{NT^3}\sum_{t=1}^{T}\|z_t(\theta_0)\|^2\sum_{t=1}^{T}\left\|\frac{\partial z_t\left(\overline{\theta}\right)}{\partial\theta'}\right\|^2\right)^{1/2}\right). \\ &. \end{aligned} \tag{46}$$

From (21),

$$z_t(\theta_0)=\varepsilon_t-\tau_t(\theta_0)\varepsilon_0+\lambda_{t-1}\left(L;\theta_0^{(-1)}\right)\Delta\alpha_t=\varepsilon_t+O_p(1),$$

and thus

$$\sum_{t=1}^{T}\|z_t(\theta_0)\|^2=O_p(NT).$$

Next

$$\frac{\partial z_t'(\theta)}{\partial\theta}=\frac{\partial\lambda_t(L;\theta)}{\partial\theta}v_t'+\dot{\tau}_t(\theta)\varepsilon_0'+\dot{\lambda}_{t-1}\left(L;\theta^{(-1)}\right)'\Delta\alpha_t'.$$

As in RV the derivative with respect to δ dominates, and we have

$$\begin{aligned}\left\|\frac{\partial\lambda_t(L;\theta)}{\partial\theta}v_t'\right\| &\leq K\sum_{j=1}^{t}(\log j)\,j^{\delta_0-\delta-1}\left(\sum_{i=1}^{N}\varepsilon_{i,t-j}^2\right)^{1/2}\\ &= O_p\left(N^{1/2}\left(t^{\delta_0-\delta}+1\right)\log^2 t\right),\end{aligned}$$

while

$$\dot{\tau}_t(\theta)=O\left((\log t)\,t^{-\delta}\right),\quad \left\|\dot{\lambda}_{t-1}\left(L;\theta^{(-1)}\right)'\Delta\alpha_t'\right\|=O_p\left(\frac{N^{1/2}}{T}\sum_{j=1}^{t}(\log j)\,j^{-\delta}\right)=O_p\left(\frac{N^{1/2}t}{T}\right).$$

Thus from Theorem 3.1 and with T large enough and any $\epsilon > 0$

$$\left\| \frac{\partial z_t\left(\overline{\theta}\right)}{\partial \theta'} \right\| = O_p\left(N^{1/2}\left(t^{\epsilon} + t^{\epsilon-\delta_0}\right)\right) = O_p\left(N^{1/2}t^{\epsilon}\right),$$

uniformly in t, and so

$$\sum_{t=1}^{T} \left\| \frac{\partial z_t\left(\overline{\theta}\right)}{\partial \theta'} \right\|^2 = O_p\left(NT^{1+2\epsilon}\right).$$

It follows that (46) is

$$O_p\left(\frac{NT^{1+\epsilon}}{N^{1/2}T^{3/2}}\right) = O_p\left(N^{1/2}T^{\epsilon-1/2}\right) = o_p\left(N\right),$$

so $tr\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) - \widetilde{\Sigma}_N\left(\theta_0\right)\right) = o_p\left(N\right)$. The proof of (44) is completed. Next, to establish (45) we have

$$\begin{aligned}
\left|tr\left(\widetilde{\Sigma}_N^2\left(\widehat{\theta}_T^D\right) - \Sigma_{0N}^2\right)\right| &= \left|tr\left(\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) - \Sigma_{0N}\right)\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) + \Sigma_{0N}\right)\right)\right| \\
&\leq \left(tr\left(\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) - \Sigma_{0N}\right)^2\right)\right)^{1/2} tr\left(\left(\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) + \Sigma_{0N}\right)^2\right)\right)^{1/2},
\end{aligned}$$

where

$$\begin{aligned}
tr\left(\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) - \Sigma_{0N}\right)^2\right) &= tr\left(\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) - \widetilde{\Sigma}_N\left(\theta_0\right) + \widetilde{\Sigma}_N\left(\theta_0\right) - \Sigma_{0N}\right)^2\right) \\
&\leq 2tr\left(\left(\widetilde{\Sigma}_N\left(\widehat{\theta}_T^D\right) - \widetilde{\Sigma}_N\left(\theta_0\right)\right)^2\right) + 2tr\left(\left(\widetilde{\Sigma}_N\left(\theta_0\right) - \Sigma_{0N}\right)^2\right).
\end{aligned}$$

Now

$$\begin{aligned}
tr\left(\left(\widetilde{\Sigma}_N\left(\theta_0\right) - \Sigma_{0N}\right)^2\right) &= tr\left(\left(A_T - \Sigma_{0N} + R_T\right)^2\right) \\
&\leq 2tr\left(\left(A_T - \Sigma_{0N}\right)^2\right) + tr\left(R_T^2\right),
\end{aligned}$$

where

$$\begin{aligned}
Etr\left(\left(A_T - \Sigma_{0N}\right)^2\right) &= E\left(tr\left(\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t\varepsilon_t' - \Sigma_{0N}\right)\right)^2\right) \\
&= \frac{1}{T^2}E\left(\sum_{t=1}^{T}\left(\left\|\varepsilon_t\right\|^4 - tr\left(\Sigma_{0N}^2\right)\right) + tr\left(\sum_{s,t=1,s\neq t}^{T}\left(\varepsilon_s\varepsilon_s'\varepsilon_t\varepsilon_t' - \Sigma_{0N}^2\right)\right)\right) \\
&\leq \frac{1}{T^2}\sum_{s=1}^{T}E\left\|\varepsilon_t\right\|^4 \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\left(3tr\left(\Sigma_{0N}^2\right) + \sum_{i,j,k,l=1}^{N} cum\left(\varepsilon_{it},\varepsilon_{jt},\varepsilon_{kt},\varepsilon_{lt}\right)\right) \\
&= O\left(\frac{N}{T}\right),
\end{aligned}$$

from (12) and (9). Thus $tr\left((A_T-\Sigma_{0N})^2\right)=o_p(N)$. From the proof of (44) it is readily seen that $tr\left(R_T^2\right)=o_p(N)$, to complete the proof of (45). $\square$

Proof of Theorem 4.1

The proof of (32) extends that of RV's Theorem 4.4, inspection of which indicates that it suffices to consider the additional term in $L_T^D(\theta)-\widehat{\sigma}^2(\theta)$ beyond ones of the type considered in the proof of Theorem 3.1, namely

$$\frac{\left|\sum_{t=1}^{T}\tau_t(\theta)\,\lambda_{t-1}\left(L;\theta^{(-1)}\right)\Delta\alpha_{it}\right|^2}{S_{\tau\tau T}(\theta)},\tag{47}$$

which comes from $\left\{\mathbf{z}_{iT}'(\theta)\,\boldsymbol{\tau}_T(\theta)\right\}^2/S_{\tau\tau T}(\theta)$. We can bound (47) by

$$\begin{aligned}
&\sup_{i,t}|\Delta\alpha_{it}|^2\sup_{D,\Xi}\frac{\left|\sum_{t=1}^{T}|\tau_t(\theta)|\sum_{j=0}^{t-1}\left|\lambda_j\left(\theta^{(-1)}\right)\right|\right|^2}{S_{\tau\tau T}(\theta)}\\
&= O_p\left(T^{-2}\sup_{D}\frac{\left|\sum_{t=1}^{T}t^{-\delta}\left(t^{1-\delta}+1\right)\right|^2}{T^{1-2\delta}+1}\right)\\
&= O_p\left(T^{-2}\sup_{D}\frac{\left(T^{2(1-\delta)}+1\right)^2}{T^{1-2\delta}+1}\right)\\
&= O_p\left(T^{1-2\underline{\delta}}+1\right)
\end{aligned}$$

so its contribution to $L_T^D(\theta)-\widehat{\sigma}^2(\theta)$ is $O_p\left(T^{-2\underline{\delta}}+T^{-1}\right)=o_p(1)$ since $\underline{\delta}>0$. The proof of (32) then follows straightforwardly from those of RV's Theorem 3.4 and our Theorem 3.1.

To prove (33) we consider first the case where (14) and (15) are not imposed. We have first to bound the extra terms in $\widehat{\sigma}_T^2(\theta_0)$ depending on $\Delta\alpha_{it}$ in the normalized score based on L_T^D, namely

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right)^2+2\left(\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right)\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right\}\\
&-\frac{1}{S_{\tau\tau T}^0}\sum_{i=1}^{N}\left\{\left(\sum_{t=1}^{T}\tau_t^0\left(\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right)\right)^2+2\tau_t^0\left(\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right)\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right\}
\end{aligned}\tag{48}$$

We have

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right)^2 &= O_p\left(T^{-2}N\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)^2\right)\\
&= O_p\left(T^{-2}N\left(T^{3-2\delta_0}+T\right)\right)=O_p\left(N\left(T^{1-2\delta_0}+T^{-1}\right)\right),
\end{aligned}$$

so its contribution to $\widehat{\sigma}_T^2(\theta_0)$ is $O_p\left(T^{-2\delta_0}+T^{-2}\right)$. Next

$$\begin{aligned}\sum_{i=1}^N\sum_{t=1}^T 2\left(\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right)\lambda_{t-1}^0(L)\,\Delta\alpha_{it} &= O_p\left(T^{-1}N\sum_{t=1}^T\left(t^{1-\delta_0}+1\right)\right)\\ &= O_p\left(T^{-1}N\left(T^{2-\delta_0}+T\right)\right),\end{aligned}$$

which is $O_p\left(N\left(T^{1-\delta_0}+1\right)\right)$, so its contribution to $\widehat{\sigma}_T^2(\theta_0)$ is $O_p\left(T^{-\delta_0}+T^{-1}\right)$. Finally

$$\begin{aligned}\frac{1}{S_{\tau\tau T}^0}\sum_{i=1}^N\left(\sum_{t=1}^T\tau_t^0\left(\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right)\right)^2 &= O_p\left(S_{\tau\tau T}^{0-1}T^{-2}N\left(\sum_{t=1}^T t^{-\delta_0}\left(t^{1-\delta_0}+1\right)\right)^2\right)\\ &= O_p\left(S_{\tau\tau T}^{0-1}T^{-2}N\left(\left(T^{2(1-\delta_0)}+\log T\right)\right)^2\right)\\ &= O_p\left(S_{\tau\tau T}^{0-1}N\left(T^{2(1-2\delta_0)}+T^{-2}\log^2 T\right)\right)\\ &= O_p\left(NT^{1-2\delta_0}+NT^{2(1-2\delta_0)}+NT^{-2}\log^2 T\right),\end{aligned}$$

so its contribution to $\widehat{\sigma}_T^2(\theta_0)$ is $O_p\left(T^{-2\delta_0}+T^{1-4\delta_0}+T^{-3}\log^2 T\right)$, as is that of

$$\frac{2}{S_{\tau\tau T}^0}\sum_{i=1}^N\tau_t^0\left(\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right)\lambda_{t-1}^0(L)\,\Delta\alpha_{it}.$$

Overall the contribution of the extra terms in $\widehat{\sigma}_T^2(\theta_0)$ to $(NT)^{1/2}(\partial/\partial\theta)L_T^P(\theta_0)$ is

$$\begin{aligned}&O_p\left((NT)^{1/2}\left(T^{-\delta_0}+T^{1-4\delta_0}+T^{-1}\right)\right)\\ =\;&O_p\left(\left(N\left(T^{1-2\delta_0}+T^{3-8\delta_0}+T^{-1}\right)\right)^{1/2}\right)=o_p(1),\end{aligned}$$

since $NT^{1-2\delta_0}+NT^{-1}\to 0$ and $\delta_0>\frac{1}{2}$, and they do not contribute to the asymptotic bias. With the notation $[a_t]_1^T$ meaning the $T\times 1$ column vector with tth element a_t, with respect to the contribution of $\frac{\partial}{\partial\theta_j}\widehat{\sigma}_T^2(\theta)$ to $(\partial/\partial\theta)L_T^P(\theta_0)$ we have to consider the extra terms in $\mathbf{z}_{iT}(\theta_0)=[z_{it}(\theta_0)]_1^T$ depending on $\lambda_{t-1}^0(L)\,\Delta\alpha_{it}$,

$$\begin{aligned}&\frac{2}{NT}\sum_{i=1}^N\left[\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right]_1^{T\prime}\Omega_T^{-1}(\theta_0)\,\Omega_T^j(\theta_0)\,\Omega_T^{-1}(\theta_0)\,\mathbf{z}_{iT}(\theta_0)\\ =\;&\frac{2}{NT}\sum_{i=1}^N\left[\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right]_1^{T\prime}\Omega_T^{-1}(\theta_0)\,\Omega_T^j(\theta_0)\,\Omega_T^{-1}(\theta_0)\left[\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right]_1^T\\ &+\frac{2}{NT}\sum_{i=1}^N\left[\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right]_1^{T\prime}\Omega_T^{-1}(\theta_0)\,\Omega_T^j(\theta_0)\,\Omega_T^{-1}(\theta_0)\left[\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right]_1^T,\end{aligned}\tag{49}$$

and those in $\dot{\mathbf{z}}_{iT}(\theta_0)=[\dot{z}_{it}(\theta_0)]_1^T$ depending on $\dot{\lambda}_{t-1}^0(L)\,\Delta\alpha_{it}$,

$$\frac{2}{NT}\sum_{i=1}^N\left[\lambda_{t-1}^0(L)\,\Delta\alpha_{it}\right]_1^{T\prime}\Omega_T^{-1}(\theta_0)\,\dot{\mathbf{z}}_{iT}(\theta_0)+\frac{2}{NT}\sum_{i=1}^N\left[\dot{\lambda}_{t-1}^0(L)\,\Delta\alpha_{it}\right]_1^{T\prime}\Omega_T^{-1}(\theta_0)\,\mathbf{z}_{iT}(\theta_0).\tag{50}$$

Since $\Omega_T^{-1}(\theta_0) = I_T - \frac{\boldsymbol{\tau}_T^0 \boldsymbol{\tau}_T^{0\prime}}{S_{\tau\tau T}^0}$, $\Omega_T^j(\theta_0) = \dot{\boldsymbol{\tau}}_{j,T}^0 \boldsymbol{\tau}_T^{0\prime} + \boldsymbol{\tau}_T^0 \dot{\boldsymbol{\tau}}_{j,T}^{0\prime}$,where $\dot{\boldsymbol{\tau}}_{j,T}^{0\prime} = \left(\dot{\tau}_{j,1}^0, \ldots, \dot{\tau}_{j,T}^0\right)$ and $\dot{\tau}_{j,k}^0 = O\left(k^{-\delta_0} \log k\right)$ as $k \to \infty$, the first extra term is

$$
\begin{aligned}
&\frac{2}{NT} \sum_{i=1}^{N} \left[\lambda_{t-1}^0(L)\Delta\alpha_{it}\right]_1^{T\prime} \Omega_T^{-1}(\theta_0)\, \Omega_T^j(\theta_0)\, \Omega_T^{-1}(\theta_0) \left[\lambda_{t-1}^0(L)\Delta\alpha_{it}\right]_1^T \\
=\;& O_p\left(T^{-3} \sum_t \left(t^{1-\delta_0}+1\right) t^{-\delta_0} \sum_r \left(t^{1-\delta_0}+1\right) t^{-\delta_0} \log t\right) \\
&+O_p\left(T^{-3} S_{\tau\tau T}^{0-1} S_{\dot{\tau}\tau T}^0 \left(\sum_t \left(t^{1-\delta_0}+1\right) t^{-\delta_0}\right)^2\right) \\
=\;& O_p\left(T^{-3}\left(T^{2(1-\delta_0)} + \log T\right)^2 \log T\right) \\
=\;& O_p\left(T^{1-4\delta_0} \log T + T^{-3} \log^3 T\right),
\end{aligned}
$$

while using

$$
E\left|\sum_{t=1}^{T} \tau_t^0 \varepsilon_{it}\right| \leq E\left[\left|\sum_{t=1}^{T} \tau_t^0 \varepsilon_{it}\right|^2\right]^{1/2} = \left[\sigma_0^2\left(S_{\tau\tau T}^0 - 1\right)\right]^{1/2} = O\left(T^{\frac{1}{2}-\delta_0} + \log^{\frac{1}{2}} T\right),
$$

and similarly

$$
\begin{aligned}
E\left\|\sum_{t=1}^{T} \dot{\tau}_t^0 \varepsilon_{it}\right\| &= O\left(T^{\frac{1}{2}-\delta_0} \log T + \log T\right), \\
E\left\|\sum_{t=1}^{T} \dot{\tau}_t^0 \tau_t^0 \varepsilon_{i0}\right\| &= O\left(S_{\tau\dot{\tau}T}^0\right) = O\left(T^{1-2\delta_0} \log T + \log^2 T\right),
\end{aligned}
$$

the second one is

$$
\begin{aligned}
&\frac{2}{NT} \sum_{i=1}^{N} \left[\lambda_{t-1}^0(L)\Delta\alpha_{it}\right]_1^{T} \Omega_T^{-1}(\theta_0)\, \Omega_T^j(\theta_0)\, \Omega_T^{-1}(\theta_0) \left[\varepsilon_{it} - \tau_t^0 \varepsilon_{i0}\right]_1^T \\
=\;& O_p\left(T^{-2} \sum_t \left(t^{1-\delta_0}+1\right) t^{-\delta_0} \left(T^{1-2\delta_0} \log T + \log^2 T\right)\right) \\
&+O_p\left(T^{-2} S_{\tau\tau T}^{0-1} S_{\dot{\tau}\tau T}^0 \left(\sum_t \left(t^{1-\delta_0}+1\right) t^{-\delta_0}\right)\left(T^{1-2\delta_0} \log T + \log^2 T\right)\right) \\
=\;& O_p\left(T^{-2}\left(T^{2(1-\delta_0)} + \log T\right)\left(T^{1-2\delta_0} \log T + \log^2 T\right)\right) \\
=\;& O_p\left(T^{1-4\delta_0} \log T + T^{-2} \log^3 T\right).
\end{aligned}
$$

Their joint contribution to $(NT)^{1/2}(\partial/\partial\theta) L_T^P(\theta_0)$ is thus

$$
\begin{aligned}
&O_p\left((NT)^{1/2}\left(T^{1-4\delta_0} \log T + T^{-2} \log^3 T\right)\right) \\
=\;& O_p\left(\left(N\left(T^{3-8\delta_0} \log^2 T + T^{-3} \log^4 T\right)\right)^{1/2}\right),
\end{aligned}
$$

which is $o_p(1)$ under $NT^{1-2\delta_0}\log^2 T + NT^{-1} \to 0$ and $\delta_0 > \frac{1}{2}$. The extra terms due to the new element in $\dot{z}_{it}(\theta_0) = f_{it} - \dot{\tau}_t^0 \varepsilon_{i0} + \dot{\lambda}_{t-1}^0(L)\Delta\alpha_{it}$, where f_{it} is the ith row of f_t, are

$$\begin{aligned}
&\frac{2}{NT}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(\dot{\lambda}_{t-1}^0(L)\Delta\alpha_{it}\right)z_{it}(\theta_0) + \dot{z}_{it}(\theta_0)\left(\lambda_{t-1}^0(L)\Delta\alpha_{it}\right)\right) \\
&-\frac{2}{NTS_{\tau\tau T}^0}\sum_{i=1}^{N}\left(\left(\sum_{t=1}^{T}\left(\dot{\lambda}_{t-1}^0(L)\Delta\alpha_{it}\right)\tau_t^0\right)\left(\sum_{t=1}^{T}z_{it}(\theta_0)\tau_t^0\right)\right. \\
&\left.+\left(\sum_{t=1}^{T}\dot{z}_{it}(\theta_0)\tau_t^0\right)\left(\sum_{t=1}^{T}\left(\lambda_{t-1}^0(L)\Delta\alpha_{it}\right)\tau_t^0\right)\right). \qquad (51)
\end{aligned}$$

Using similar methods as before, the first term is

$$O_p\left(T^{-2}\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)\log t\right) = O_p\left(T^{-2}\left(T^{2-\delta_0}+1\right)\log T\right).$$

Next, since

$$\begin{aligned}
E\left|\sum_{t=1}^{T}z_{it}(\theta_0)\tau_t^0\right| &= O\left(S_{\tau\tau T}^0\right) = O\left(T^{1-2\delta_0}+\log T\right), \\
E\left\|\sum_{t=1}^{T}\dot{z}_{it}(\theta_0)\tau_t^0\right\| &= O\left(S_{\tau\dot{\tau}T}^0\right) = O\left(S_{\tau\tau T}^0\log T\right),
\end{aligned}$$

the other terms are

$$\begin{aligned}
&O_p\left(\frac{1}{T^2 S_{\tau\tau T}^0}\left(\left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)t^{-\delta_0}\log t\right)S_{\tau\tau T}^0 + S_{\tau\dot{\tau}T}^0\left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)t^{-\delta_0}\right)\right)\right) \\
= \;& O_p\left(\frac{1}{T^2}\left(\left(\left(T^{2(1-\delta_0)}+\log T\right)\log T\right)+\left(T^{2(1-\delta_0)}+\log T\right)\right)\log T\right) \\
= \;& O_p\left(\left(T^{-2\delta_0}+T^{-2}\log T\right)\log T\right).
\end{aligned}$$

Thus the contribution to the standardized score of these two terms is

$$\begin{aligned}
&O_p\left((NT)^{\frac{1}{2}}\left(T^{-2\delta_0}+T^{-2}\log T\right)\log T\right) \\
= \;& O_p\left(\left(N\left(T^{1-4\delta_0}+T^{-3}\log^2 T\right)\log^2 T\right)^{\frac{1}{2}}\right),
\end{aligned}$$

which is $o_p(1)$ since $NT^{1-2\delta_0}\log^2 T + NT^{-1} \to 0$ for $\delta_0 > \frac{1}{2}$. This completes the proof when (14) and (15) are not imposed.

Now imposing (14) and (15), we have first to reanalyze two terms in $\widehat{\sigma}_T^2(\theta_0)$ depending on $\Delta\alpha_{it}$ in the normalized score based on L_T^D, given in (48). Now

$$\sum_{i=1}^{N}\sum_{t=1}^{T}2\left(\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right)\lambda_{t-1}^0(L)\Delta\alpha_{it}$$

has zero mean and variance bounded by

$$
\begin{aligned}
& K\frac{N\left\|\Sigma_N\right\|}{T^2}\left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)^2+\left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)t^{-\delta_0}\right)^2\right) \\
= \; & O\left(NT^{-2}\left(T^{3-2\delta_0}+T\right)\right),
\end{aligned}
$$

so it is $O_p\left(N^{1/2}\left(T^{\frac{1}{2}-\delta_0}+T^{-\frac{1}{2}}\right)\right)$, and its contribution to $\widehat{\sigma}_T^2\left(\theta_0\right)$ is $O_p\left(N^{-1/2}\left(T^{-\frac{1}{2}-\delta_0}+T^{-\frac{3}{2}}\right)\right)$. Next,

$$
\frac{2}{S_{\tau\tau T}^{0}}\sum_{i=1}^{N}\tau_t^0\left(\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right)\lambda_{t-1}^0\left(L\right)\Delta\alpha_{it}
$$

has zero mean and variance bounded by

$$
\begin{aligned}
& \frac{KN}{T^2S_{\tau\tau T}^{02}}\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)^2+\frac{KN}{T^2S_{\tau\tau T}^{20}}\left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)t^{-\delta_0}\right)^2 \\
= \; & O\left(NT^{-2}\left(T^{3-2\delta_0}\log T+T^{1+2\delta_0}1_{\{\delta_0<1/2\}}+T^2\right)\log^2 T\right) \\
= \; & O\left(N\left(T^{1-2\delta_0}\log T+1\right)\log^2 T\right),
\end{aligned}
$$

so its contribution to $\widehat{\sigma}_T^2\left(\theta_0\right)$ is $O_p\left(\left(T^{-2\delta_0}\log T+T^{-1}\right)\log^2 T\right)$. Overall, the contribution of the extra terms in $\widehat{\sigma}_T^2\left(\theta_0\right)$ to $\left(NT\right)^{1/2}\left(\partial/\partial\theta\right)L_T^P\left(\theta_0\right)$ is

$$
\begin{aligned}
& O_p\left(\left(NT^{1-4\delta_0}\right)^{1/2}\right)+O_p\left(T^{-\delta_0}+T^{-1}\right) \\
& +O_p\left(\left(N\left(T^{3-8\delta_0}+T^{-2}\log^2 T+T^{1-4\delta_0}\log^4 T\right)\right)^{1/2}\right) \\
& +O_p\left(\left(N\left(T^{1-4\delta_0}\log^2 T+T^{-1}\right)\log^2 T\right)^{1/2}\right),
\end{aligned}
$$

which is $o_p\left(1\right)$, since $NT^{1-4\delta_0}\log^4 T+NT^{-1}\log^2 T\rightarrow 0$ and $\delta_0>\frac{1}{4}$, and they do not to contribute to the asymptotic bias. Regarding the contribution of $\frac{\partial}{\partial\theta_j}\widehat{\sigma}_T^2\left(\theta\right)$ to $\left(\partial/\partial\theta\right)L_T^P\left(\theta_0\right)$, we need to reconsider the term in (49)

$$
\frac{2}{NT}\sum_{i=1}^{N}\left[\lambda_{t-1}^0\left(L\right)\Delta\alpha_{it}\right]_1^{T\prime}\Omega_T^{-1}\left(\theta_0\right)\Omega_T^{j}\left(\theta_0\right)\Omega_T^{-1}\left(\theta_0\right)\left[\varepsilon_{it}-\tau_t^0\varepsilon_{i0}\right]_1^T,
$$

which has zero mean and variance

$$
\begin{aligned}
& O\left(N^{-1}T^{-4}S_{\dot{\tau}\tau T}^{02}\left(\sum_t\left(t^{1-\delta_0}+1\right)t^{-\delta_0}\right)^2\right) \\
= \; & O\left(N^{-1}T^{-4}\left(T^{1-2\delta_0}+\log T\right)^2\left(T^{2(1-\delta_0)}+\log T\right)^2\log^2 T\right) \\
= \; & O\left(N^{-1}\left(T^{2-8\delta_0}\log T+T^{-4\delta_0}\log^4 T+T^{-4}\log^6 T\right)\right),
\end{aligned}
$$

so this term is $O\left(_p\left(N^{-1}\left(T^{2-8\delta_0}\log T + T^{-4\delta_0}\log^4 T + T^{-4}\log^6 T\right)\right)^{1/2}\right)$. Then the joint contribution of all terms depending on the differenced trend to $(NT)^{1/2}(\partial/\partial\theta)L_T^P(\theta_0)$ is thus

$$
\begin{aligned}
&O_p\left(N\left(T^{3-8\delta_0}\log^4 T + T^{-5}\log^8 T\right)\right)^{1/2}\\
&+O_p\left(\left(T^{3-8\delta_0}\log T + T^{1-4\delta_0}\log^4 T + T^{-3}\log^6 T\right)\right)^{1/2},
\end{aligned}
$$

which is $o_p(1)$ since $NT^{3-8\delta_0}\log^4 T + NT^{-1}\log^2 T \to 0$ and $\delta_0 > \frac{3}{8}$. Now consider the term in (51) under (14) and (15):

$$
\frac{2}{NT}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(f_{it} - \dot{\tau}_t^0\varepsilon_{i0}\right)\left(\lambda_{t-1}^0(L)\Delta\alpha_{it}\right)\right)
$$

has zero mean and has variance

$$
\begin{aligned}
&O\left(N^{-1}T^{-4}\left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)^2\log T + \left(\sum_{t=1}^{T}\left(t^{1-\delta_0}+1\right)t^{-\delta_0}\log T\right)^2\right)\right)\\
= \;& O\left(N^{-1}T^{-4}\left(\left(T^{3-2\delta_0}+T\right)\log T + \left(\left(T^{2-2\delta_0}+\log T\right)\log T\right)^2\right)\right)\\
= \;& O\left(N^{-1}\left(\left(T^{-1-2\delta_0}+T^{-3}\right)\log T + \left(\left(T^{-2\delta_0}+T^{-2}\log T\right)\log T\right)^2\right)\right),
\end{aligned}
$$

and a similar result holds for the term depending on $\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\left(\dot{\lambda}_{t-1}^0(L)\Delta\alpha_{it}\right)$, so the contribution to the standardized score of this first term is

$$
\begin{aligned}
&O_p\left(\left(N\left(T^{1-4\delta_0}+T^{-2}\right)\log^2 T\right)^{1/2}\right)\\
&+O_p\left(\left(\left(T^{-2\delta_0}+T^{-2}\right)\log^2 T\right)^{1/2}\right) + O\left(\left(\left(T^{1-4\delta_0}+T^{-3}\log^2 T\right)\log^2 T\right)^{1/2}\right),
\end{aligned}
$$

which is $o_p(1)$ since $NT^{1-4\delta_0}\log^4 T + NT^{-1}\log^2 T \to 0$ for $\delta_0 > \frac{3}{8}$, while the second term is shown $o_p(1)$ as before. □

Proof of Theorem 4.2

This straightforwardly extends the proofs of Theorem 3.2 and Theorem 4.1, and is thus omitted. □

References

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163-188.

Case, A.C., 1991. Spatial patterns in household demand. Econometrica 59, 953-165.

Ergemen, Y.E. and C. Velasco, 2014. Estimation of fractionally integrated panels with fixed effects and cross-section dependence. Preprint UC3M.

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

Horn, R.A. and C.R. Johnson, 1988. Matrix analysis. Cambridge University Press.

Hsiao, C., 2014. Analysis of panel data, Third Edition. Cambridge University Press.

Hualde, J. and P. M. Robinson, 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. Annals of Statistics 39, 3152-3181.

Robinson, P. M., 2012. Nonparametric trending regression with cross-sectional dependence. Journal of Econometrics 169, 4-14.

Robinson, P. M. and C. Velasco, 2015. Efficient inference on fractionally integrated panel data models with fixed effects. Journal of Econometrics 185, 435-452.

Table 1. Empirical bias ×100. $I(\delta_0)$

		$\rho = 0.5$								$\rho = 0.9$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	3.73	-1.97	-2.14	-1.68	-0.33	-2.31	-1.87	-1.53	1.76	-3.12	-2.75	-2.13	-0.15	-2.86	-2.35	-1.93
12	**96**	1.98	-2.39	-2.20	-1.77	-0.51	-2.22	-1.96	-1.68	-0.46	-3.76	-2.92	-2.31	-0.34	-2.73	-2.50	-2.15
10	**200**	7.53	0.10	-1.00	-0.82	-0.52	-1.02	-0.89	-0.75	6.34	-0.58	-1.40	-1.14	-0.56	-1.56	-1.29	-1.06
25	**200**	1.11	-1.22	-1.02	-0.90	-0.75	-1.05	-0.93	-0.84	-0.55	-1.74	-1.28	-1.12	-0.83	-1.28	-1.15	-1.04
20	**400**	5.09	0.03	-0.42	-0.37	-0.33	-0.40	-0.37	-0.34	4.19	-0.31	-0.59	-0.52	-0.49	-0.62	-0.54	-0.49
50	**400**	0.57	-0.56	-0.48	-0.45	-0.39	-0.46	-0.45	-0.43	-0.46	-0.76	-0.59	-0.56	-0.48	-0.57	-0.55	-0.53

Table 2. Empirical Root-MSE ×100. $I(\delta_0)$

		$\rho = 0.5$								$\rho = 0.9$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	22.18	16.38	13.08	11.98	14.92	15.25	12.62	11.67	26.08	19.67	15.26	13.73	15.99	17.15	14.44	13.26
12	**96**	22.97	16.80	13.21	12.10	14.42	14.86	12.77	11.84	27.78	20.57	15.51	13.91	15.44	16.66	14.68	13.52
10	**200**	14.90	9.90	8.19	7.67	10.82	9.67	7.96	7.40	17.05	12.19	10.02	9.33	12.76	12.04	9.88	9.13
25	**200**	16.01	9.81	7.98	7.68	9.43	8.78	7.85	7.59	19.58	11.45	9.06	8.69	10.30	9.86	8.84	8.54
20	**400**	10.52	6.23	5.23	5.06	6.67	5.86	5.13	4.95	12.11	7.69	6.43	6.22	8.22	7.27	6.34	6.13
50	**400**	11.59	6.18	5.32	5.24	6.11	5.66	5.27	5.20	13.88	6.98	5.94	5.85	6.76	6.30	5.87	5.79

Table 3. Empirical Size (%) Corrected 5% t-test. $I(\delta_0)$

		$\rho = 0.5$								$\rho = 0.9$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	22.85	11.28	5.88	3.80	5.81	9.43	5.03	3.62	25.00	13.48	7.68	5.38	6.43	11.14	6.35	4.75
12	**96**	25.96	12.76	7.13	5.38	5.53	10.19	6.55	4.88	27.91	15.01	9.05	6.62	5.81	11.45	7.89	6.11
10	**200**	16.68	4.40	1.74	0.93	2.93	3.63	1.55	0.87	18.48	7.02	3.03	2.02	4.35	6.75	3.08	1.65
25	**200**	30.93	11.22	5.88	5.14	9.29	8.51	5.36	5.03	32.72	12.51	6.99	6.06	9.94	9.29	6.40	5.86
20	**400**	25.83	5.42	2.46	1.93	4.72	4.04	2.22	1.71	26.10	7.62	3.75	3.01	7.27	6.05	3.40	2.75
50	**400**	35.70	9.06	5.12	4.91	8.01	6.89	5.07	4.91	36.60	9.72	5.84	5.47	8.46	7.34	5.49	5.27

Table 4. Empirical Size (%) Uncorrected 5% t-test. $I(\delta_0)$

		$\rho = 0.5$								$\rho = 0.9$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	39.50	25.87	19.06	16.33	21.00	24.70	17.67	15.01	35.65	23.60	17.01	14.33	12.65	21.66	15.47	12.65
12	**96**	38.71	24.89	17.96	15.38	18.09	22.18	17.23	14.72	35.27	22.42	16.50	13.83	11.20	19.14	15.05	13.03
10	**200**	45.81	23.50	16.74	15.00	26.62	22.97	15.77	13.12	46.86	29.47	22.40	19.46	33.19	29.07	21.70	18.70
25	**200**	40.11	19.83	13.48	12.00	17.85	16.40	12.90	11.68	36.46	17.42	11.39	10.21	13.83	13.68	10.72	9.61
20	**400**	46.06	19.75	12.68	11.75	19.24	17.09	12.22	10.99	48.18	25.49	17.94	16.73	25.90	22.97	17.78	16.34
50	**400**	43.01	15.94	10.88	10.29	13.79	13.06	10.52	10.24	39.15	12.96	8.63	8.04	10.57	10.17	8.32	7.91

Table 5. Empirical bias ×100. $I(\delta_0)$, $\rho = 0.9$, linear trend $\beta_i\,(t/T)$, $\beta_i \sim IIN\,(0, \gamma^2)$.

		$\gamma = 1$								$\gamma = 3$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	10.29	1.70	-1.06	-1.71	5.54	1.05	-0.82	-1.53	36.90	19.83	7.70	1.45	32.95	18.71	7.58	1.54
12	**96**	7.90	0.45	-1.56	-1.90	4.66	0.25	-1.33	-1.74	33.28	16.59	5.49	0.36	29.26	15.52	5.41	0.45
10	**200**	15.21	4.58	0.52	-0.53	7.95	3.33	0.58	-0.45	43.83	25.12	10.98	3.32	40.85	24.26	10.91	3.39
25	**200**	5.20	0.16	-0.87	-1.03	2.71	0.14	-0.77	-0.98	23.75	8.96	1.57	-0.55	19.69	8.17	1.58	-0.50
20	**400**	10.67	2.43	0.10	-0.34	5.27	1.79	0.14	-0.30	32.32	14.80	4.32	0.64	28.36	13.83	4.28	0.67
50	**400**	3.40	0.00	-0.54	-0.60	2.02	0.06	-0.49	-0.56	17.64	4.85	0.29	-0.50	14.12	4.38	0.32	-0.46

Table 6. Empirical Root-MSE ×100. $I(\delta_0), \rho = 0.9$, linear trend $\beta_i\,(t/T)$, $\beta_i \sim IIN\,(0, \gamma^2)$.

		$\gamma = 1$								$\gamma = 3$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	23.42	16.87	14.11	13.33	17.32	15.73	13.50	12.90	39.40	23.42	14.22	12.28	36.23	22.65	13.91	11.98
12	**96**	23.77	17.29	14.45	13.55	16.60	15.74	13.92	13.19	36.38	21.08	13.51	12.57	33.10	20.41	13.27	12.29
10	**200**	19.99	11.73	9.40	9.06	14.99	11.51	9.37	8.92	44.75	26.49	13.69	9.05	42.06	25.73	13.56	8.94
25	**200**	16.60	10.52	8.95	8.75	10.53	9.57	8.71	8.58	26.01	12.47	8.43	8.55	22.14	11.85	8.31	8.40
20	**400**	14.22	7.46	6.24	6.13	9.30	7.08	6.17	6.05	33.00	15.91	7.15	5.99	29.26	15.06	7.10	5.92
50	**400**	12.14	6.83	6.00	5.93	6.89	6.24	5.92	5.86	19.43	7.77	5.81	5.89	15.77	7.35	5.75	5.83

Table 7. Empirical Size (%) Corrected 5% t-test. $I(\delta_0)$, $\rho = 0.9$, linear trend $\beta_i\,(t/T)$, $\beta_i \sim IIN\,(0, \gamma^2)$.

		$\gamma = 1$								$\gamma = 3$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	26.91	10.15	6.39	5.38	11.19	8.77	5.26	4.77	80.92	32.27	6.06	3.70	70.92	29.63	5.43	3.30
12	**96**	28.41	12.21	7.64	6.16	10.70	10.45	6.92	5.58	76.01	26.93	6.20	4.70	65.34	24.57	5.60	4.38
10	**200**	32.34	5.74	2.11	1.81	14.31	5.26	2.20	1.58	98.61	66.85	9.80	1.56	95.46	63.01	9.24	1.46
25	**200**	32.77	10.30	6.76	6.14	11.43	8.18	6.17	5.69	79.06	20.40	5.07	5.71	65.99	17.75	4.85	5.32
20	**400**	41.13	6.84	3.10	2.72	13.69	5.51	2.82	2.48	99.31	60.23	5.57	2.55	97.10	53.79	5.40	2.34
50	**400**	37.46	9.52	6.25	5.84	9.85	7.10	5.93	5.51	81.63	15.31	5.27	5.67	68.14	12.80	5.15	5.46

Table 8. Empirical Size (%) Uncorrected 5% t-test. $I(\delta_0)$, $\rho = 0.9$, linear trend $\beta_i\,(t/T)$, $\beta_i \sim IIN\,(0, \gamma^2)$.

		$\gamma = 1$								$\gamma = 3$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	37.04	19.42	14.24	13.21	17.84	18.30	12.98	11.90	83.26	44.92	14.67	10.55	73.67	42.46	13.96	9.58
12	**96**	34.93	19.47	14.48	13.32	15.25	18.07	13.77	12.16	76.47	35.04	12.63	10.90	65.64	32.93	12.15	10.30
10	**200**	61.65	30.31	19.79	18.28	39.19	29.42	19.56	17.63	99.75	90.91	43.13	17.74	98.78	89.03	42.91	17.19
25	**200**	36.18	15.08	10.94	10.28	14.70	12.46	10.17	9.61	78.77	26.09	9.08	9.60	65.35	23.35	8.51	8.93
20	**400**	61.28	24.91	17.23	16.24	32.95	22.25	16.76	15.90	99.80	83.99	23.58	15.18	98.96	79.55	23.52	14.95
50	**400**	39.53	12.57	8.92	8.67	11.96	10.08	8.47	8.32	81.53	19.22	7.95	8.57	67.40	16.27	7.72	8.21

Table 9. Empirical bias $\widehat{\delta}$ ×100. FARIMA(ξ_0, δ_0), $\rho = 0.9$.

		$\xi_0 = 0.5$								$\xi_0 = 0.8$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	5.00	-3.24	-4.18	-5.64	4.93	-3.15	-3.94	-5.46	6.70	4.47	3.75	0.91	5.84	2.39	2.23	0.35
12	**96**	1.42	-5.08	-5.68	-6.93	4.38	-4.40	-5.32	-6.75	6.48	4.39	3.72	1.22	6.07	2.53	2.45	0.67
10	**200**	8.78	-2.63	-3.99	-4.58	3.64	-3.40	-4.06	-4.78	5.88	4.08	3.52	1.50	5.07	3.01	2.97	1.52
25	**200**	-3.61	-7.81	-7.54	-7.63	0.57	-6.78	-7.09	-7.31	4.04	3.32	3.05	1.69	3.97	2.27	2.17	1.32
20	**400**	4.58	-4.96	-5.80	-5.64	-0.24	-5.63	-5.49	-5.53	3.91	3.55	3.41	2.28	3.86	3.06	3.05	2.33
50	**400**	-4.22	-6.48	-6.31	-6.25	-1.62	-6.18	-6.06	-6.01	2.93	2.54	2.44	1.85	2.88	2.15	2.08	1.67

Table 10. Empirical Root-MSE $\widehat{\delta}$ ×100. FARIMA(ξ_0, δ_0), $\rho = 0.9$.

		$\xi_0 = 0.5$								$\xi_0 = 0.8$							
		$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$				$\widehat{\delta}_T^D$				$\widehat{\delta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	35.05	30.66	29.43	26.75	26.61	29.05	28.73	26.62	24.83	26.69	26.22	23.14	23.66	25.80	25.64	22.54
12	**96**	37.05	31.58	29.86	27.35	25.91	29.14	29.09	27.22	24.52	26.14	25.87	22.87	23.56	25.60	25.51	22.54
10	**200**	27.61	24.82	24.56	23.27	23.80	25.25	24.46	23.41	22.45	23.80	23.65	21.08	21.78	23.18	23.12	20.62
25	**200**	33.73	26.29	24.97	24.33	20.11	24.53	24.50	24.08	19.83	21.13	21.01	19.20	19.62	20.66	20.56	18.96
20	**400**	22.21	20.74	20.76	20.41	18.29	21.18	20.43	20.23	18.31	18.87	18.87	17.29	18.20	18.53	18.51	17.07
50	**400**	29.78	21.02	20.18	19.97	16.05	20.20	19.79	19.65	16.84	16.92	16.88	15.92	16.56	16.69	16.65	15.79

Table 11. Empirical bias $\widehat{\xi}$ ×100. ARFI(ξ_0, δ_0), $\rho = 0.9$.

		$\xi_0 = 0.5$								$\xi_0 = 0.8$							
		$\widehat{\xi}_T^D$				$\widehat{\xi}\delta_T^P$				$\widehat{\xi}_T^D$				$\widehat{\xi}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	-8.90	-2.76	-1.46	0.31	-7.30	-1.74	-1.12	0.61	-9.91	-8.37	-7.60	-5.04	-8.26	-5.91	-5.73	-3.96
12	**96**	-6.00	-1.27	-0.22	1.36	-7.04	-0.71	-0.08	1.53	-9.97	-8.56	-7.80	-5.49	-8.67	-6.36	-6.27	-4.57
10	**200**	-9.81	-0.89	0.47	1.16	-5.06	0.22	0.84	1.65	-7.67	-6.51	-5.97	-4.14	-6.56	-5.27	-5.21	-3.78
25	**200**	0.85	3.98	4.03	4.22	-2.74	3.48	3.82	4.10	-6.27	-6.02	-5.75	-4.42	-5.72	-4.87	-4.78	-3.92
20	**400**	-4.94	2.90	3.80	3.66	-1.07	3.64	3.62	3.69	-5.14	-5.05	-4.93	-3.85	-4.88	-4.50	-4.51	-3.76
50	**400**	2.85	4.43	4.44	4.40	0.14	4.28	4.27	4.25	-4.43	-4.20	-4.10	-3.51	-4.08	-3.77	-3.72	-3.29

Table 12. Empirical Root-MSE $\widehat{\xi}$ ×100. FARIMA(ξ_0, δ_0), $\rho = 0.9$.

		$\xi_0 = 0.5$								$\xi_0 = 0.8$							
		$\widehat{\xi}_T^D$				$\widehat{\xi}_T^P$				$\widehat{\xi}_T^D$				$\widehat{\xi}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	31.22	29.39	29.58	26.96	28.27	29.33	29.31	26.82	24.49	24.22	23.32	19.49	22.28	21.26	20.97	18.33
12	**96**	31.96	29.52	29.61	27.33	27.95	29.21	29.31	27.21	24.43	24.14	23.20	19.60	22.60	21.33	21.08	18.48
10	**200**	26.25	25.00	25.23	24.10	24.07	25.34	25.21	24.17	21.25	20.86	20.17	17.12	19.71	19.42	19.28	16.80
25	**200**	29.66	24.83	24.70	24.25	21.24	24.2	24.42	24.01	18.42	18.76	18.45	16.14	17.73	17.53	17.40	15.68
20	**400**	21.23	20.77	21.11	20.89	18.38	21.13	20.84	20.70	16.51	16.52	16.35	14.37	16.10	15.92	15.88	14.28
50	**400**	27.08	20.44	20.27	20.15	16.81	20.09	19.95	19.85	15.10	14.87	14.72	13.46	14.51	14.36	14.27	13.25

Table 13. Empirical Size (%) Corrected 5% Wald-test. FARIMA(ξ_0, δ_0), $\rho = 0.9$.

		$\xi_0 = 0.5$								$\xi_0 = 0.8$							
		$\widehat{\theta}_T^D$				$\widehat{\theta}_T^P$				$\widehat{\theta}_T^D$				$\widehat{\theta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	14.95	10.36	7.90	3.71	11.25	9.03	6.93	3.25	6.58	5.10	4.47	3.35	5.13	3.61	3.38	2.96
12	**96**	15.26	12.10	9.91	5.88	12.11	10.51	8.85	5.20	11.32	12.68	12.35	11.15	9.44	11.41	10.94	10.12
10	**200**	10.72	5.61	4.30	2.58	9.03	5.60	4.15	2.37	5.76	7.71	7.53	6.61	5.37	7.19	6.85	6.25
25	**200**	23.65	11.95	13.12	11.4	11.72	12.14	12.14	10.47	12.41	14.53	14.40	12.66	11.34	14.06	13.94	12.57
20	**400**	14.61	8.73	10.20	9.68	9.77	9.54	9.55	9.28	10.45	11.13	11.08	8.64	9.70	10.74	10.65	8.35
50	**400**	25.70	14.93	14.15	13.84	10.27	14.36	13.41	13.16	14.38	13.80	13.75	10.86	12.94	13.50	13.58	10.71

Table 14. Empirical Size (%) Uncorrected 5% Wald-test. FARIMA(ξ_0, δ_0), $\rho = 0.9$.

		$\xi_0 = 0.5$								$\xi_0 = 0.8$							
		$\widehat{\theta}_T^D$				$\widehat{\theta}_T^P$				$\widehat{\theta}_T^D$				$\widehat{\theta}_T^P$			
	δ_0 :	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
T	NT																
10	**100**	28.50	22.78	19.24	12.26	23.75	21.15	17.41	11.14	19.58	23.25	21.50	18.29	17.35	21.13	19.90	17.19
12	**96**	25.83	21.98	19.25	13.34	21.68	20.18	17.67	12.33	20.08	23.85	22.76	20.33	17.83	22.66	21.38	19.03
10	**200**	40.61	31.25	28.56	25.70	36.92	31.52	27.32	25.17	29.18	32.17	30.98	25.44	27.86	30.98	29.51	23.74
25	**200**	30.70	18.38	19.32	17.78	17.30	18.17	18.19	16.95	18.73	20.14	19.53	16.50	17.03	19.53	18.92	16.35
20	**400**	41.77	30.40	29.00	28.31	33.46	31.49	28.10	27.41	31.66	30.84	30.46	24.70	31.17	30.32	29.68	23.79
50	**400**	30.06	19.08	17.86	17.44	14.36	18.30	17.13	16.66	18.47	17.97	17.62	13.92	16.70	17.28	17.13	13.72