Peter M. Robinson, Carlos Velasco

Efficient inference on fractionally integrated panel data models with fixed effects

Article (Published version) (Refereed)

Journal of Econometrics 185 (2015) 435–452

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Efficient inference on fractionally integrated panel data models with fixed effects

Peter M. Robinson [a,*], Carlos Velasco [b]

[a] London School of Economics, United Kingdom

[b] Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Received 21 March 2013
Received in revised form
27 June 2014
Accepted 12 December 2014
Available online 31 December 2014

JEL classification:
C12
C13
C23

Keywords:
Panel data
Fractional time series
Estimation
Testing
Bias correction

ABSTRACT

A dynamic panel data model is considered that contains possibly stochastic individual components and a common stochastic time trend that allows for stationary and nonstationary long memory and general parametric short memory. We propose four different ways of coping with the individual effects so as to estimate the parameters. Like models with autoregressive dynamics, ours nests $I(1)$ behaviour, but unlike the nonstandard asymptotics in the autoregressive case, estimates of the fractional parameter can be asymptotically normal. For three of the estimates, establishing this property is made difficult due to bias caused by the individual effects, or by the consequences of eliminating them, which appears in the central limit theorem except under stringent conditions on the growth of the cross-sectional size N relative to the time series length T, though in case of two estimates these can be relaxed by bias correction, where the biases depend only on the parameters describing autocorrelation. For the fourth estimate, there is no bias problem, and no restrictions on N. Implications for hypothesis testing and interval estimation are discussed, with central limit theorems for feasibly bias-corrected estimates included. A Monte Carlo study of finite-sample performance is included.

1. Introduction

Important features of many econometric models for panel data are unobserved individual fixed effects and temporal dynamics that possibly allow for nonstationarity. When the cross-sectional dimension is large the individual effects cause an incidental parameters problem that heavily determines methodology, which has been predominately developed in the context of autoregressive, including possibly unit root, dynamics. A recent textbook treatment is Hsiao (2014). The present paper focuses on the incidental parameters problem in the context of fractional dynamics, which offer some advantages over autoregressions. A simple model for an observable array $\{y_{it}\}$ is

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

for $i=1,\ldots,N$, $t=0,1,\ldots,T$. The unobserved individual effects $\{\alpha_i, i\geq 1\}$ are subject to little, if any, more detailed specification in the sequel; the unobserved innovations $\{\varepsilon_{it}, i\geq 1, t\geq 0\}$ are throughout assumed to be independent and identically distributed (iid) and to satisfy $E\varepsilon_{it}=0$, $E\varepsilon_{it}^4<\infty$; θ_0 is a $(p+1)\times 1$ parameter vector, known only to lie in a given compact subset Θ of R^{p+1}; L is the lag operator; for any $\theta\in\Theta$ and each $t\geq 0$,

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

truncates the expansion

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

where the $\lambda_j(\theta)$ are given functions. We are concerned with $\lambda(L;\theta)$ having the particular structure

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

where δ is a scalar, ξ is a $p\times 1$ vector, $\theta=\left(\delta,\xi'\right)'$, the prime denoting transposition, and the functions Δ^{δ} and $\psi(L;\xi)$ are described as follows. With $\Delta=1-L$, Δ^{δ} has the expansion

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

for non-integer $\delta>0$, while for integer $\delta=0,1,\ldots,\pi_j(\delta)=1(j=0,1,\ldots,\delta)\left(-1\right)^j\delta\left(\delta-1\right)\cdots\left(\delta-j+1\right)/j!$, taking $0/0=$

* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
E-mail address: p.m.robinson@lse.ac.uk (P.M. Robinson).

http://dx.doi.org/10.1016/j.jeconom.2014.12.003

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 in the expansion

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

the coefficients $\psi_j(\xi)$ satisfy

$$\psi_0(\xi) = 1, \qquad \psi_j(\xi) = O(\exp(-c(\xi)j)), \tag{3}$$

where $c(\xi)$ is a positive-valued function of ψ. Note that

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

The fractional operator Δ^δ bestows possible stationary (when $0 < \delta < 1/2$) or nonstationary (when $\delta \geq 1/2$) long memory on $y_{it} - \alpha_i$, while $\psi(L;\xi)$ adds possible short memory structure, for example representing the autoregressive operator of a stationary and invertible autoregressive moving average process with combined order p, or of an exponential spectrum model (Bloomfield (1973)). The truncation in (2) is motivated mainly by a desire to allow for $\delta \geq 1/2$, when $\Delta^{-\delta}$, and thus $\lambda^{-1}(L;\theta)$, do not converge. On the other hand we can write (1) as

$$y_{it} = \alpha_i + \lambda_t^{-1}(L;\theta_0)\varepsilon_{it} = \alpha_i + \lambda^{-1}(L;\theta_0)\{\varepsilon_{it} 1(t \geq 0)\}.$$

It is possible that ξ_0 is empty, i.e. $p = 0$ and $\psi(x;\xi) \equiv 1$ *a priori*, in which case for each i, $y_{it} - \alpha_i$ has pure fractional dynamics. Our interest is in statistical inference on $\theta_0 = \left(\delta_0, \xi_0'\right)'$, and especially on δ_0 with ξ_0 regarded as a nuisance parameter.

For each i we can call $y_{it} - \alpha_i$ an $I(\delta_0)$ process. Temporarily taking $\psi(x;\xi) \equiv 1$ for simplicity, we can write

$$y_{it} = \alpha_i + \sum_{j=0}^{t} \pi_j(-\delta_0)\varepsilon_{i,t-j}, \tag{5}$$

whence when $\delta_0 = 1$,

$$y_{it} = \alpha_i + \sum_{j=0}^{t} \varepsilon_{i,t-j}. \tag{6}$$

The latter results also on taking $\rho = 1$ in the autoregressive scheme popular in the dynamic panel data literature:

$$y_{it} = \alpha_i + \sum_{j=0}^{t} \rho^j \varepsilon_{i,t-j}. \tag{7}$$

The typical alternatives to $\rho = 1$ covered by (7) are the stationary ones $\rho \in (-1, 1)$ or the explosive ones $\rho > 1$. Other versions of the autoregressive panel data model are

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \varepsilon_{it}, \quad t > 0, \tag{8}$$

and

$$y_{it} = \alpha_i + u_{it}, \quad u_{it} = \rho u_{i,t-1} + \varepsilon_{it},\ t > 0, \tag{9}$$

with $\rho \in (-1, 1]$; note that (9) implies that

$$y_{it} = (1-\rho)\alpha_i + \rho y_{i,t-1} + \varepsilon_{it}, \quad t > 0,$$

so that α_i is eliminated when $\rho = 1$. The usual aim in (7), (8) or (9) is estimating ρ or unit root testing. As one recent reference, Han and Phillips (2010) develop inference based on generalized method-of-moment estimates. Note that in the fractional model (5), the weights $\pi_j(-\delta_0)$ have decay or growth that is, unlike in (7), not exponential but algebraic, since, for any δ,

$$\pi_j(\delta) = \frac{1}{\Gamma(-\delta)} j^{-\delta-1}(1 + O(j^{-1})) \quad \text{as } j \to \infty. \tag{10}$$

The moving average weights in the more general model (1) have the same rate, in particular, by (4) and summation-by-parts,

$$\begin{aligned}
\lambda_j(\theta) &= \sum_{k=0}^{j-1}\left(\pi_{j-k}(\delta) - \pi_{j-k-1}(\delta)\right)\sum_{l=0}^{k}\psi_l(\xi) + \sum_{l=0}^{j}\psi_l(\xi) \\
&= \psi(1;\xi)\sum_{k=0}^{j-1}\left(\pi_{j-k}(\delta) - \pi_{j-k-1}(\delta)\right) + \psi(1;\xi) \\
&\quad - \sum_{k=0}^{j-1}\left(\pi_{j-k}(\delta) - \pi_{j-k-1}(\delta)\right)\sum_{l=k+1}^{\infty}\psi_l(\xi) - \sum_{l=j+1}^{\infty}\psi_l(\xi) \\
&= \psi(1;\xi)\pi_j(\delta) \\
&\quad + O\left(\sum_{k=0}^{j-1}(j-k)^{-\delta-2}\exp(-c(\xi)k) + \exp(-c(\xi)j)\right) \\
&= \frac{\psi(1;\xi)}{\Gamma(-\delta)} j^{-\delta-1}(1 + O(j^{-1})) \quad \text{as } j \to \infty,
\end{aligned} \tag{11}$$

using (10) and (3), where we note that the exponential decay requirement in the latter ensures that (11) holds for all $\delta > 0$.

As is well known from the time series literature the fractional class described by $\lambda_t(L;\theta_0)$ has a smoothness at $\delta_0 = 1$ (and elsewhere) that the autoregressive class lacks. A consequence established in that literature is that large sample inference based on an approximate Gaussian pseudo likelihood can be expected to entail standard limit distribution theory; in particular, Lagrange multiplier tests on θ_0 (for example of the $I(1)$ hypothesis $\delta_0 = 1$) are asymptotically χ^2 distributed with classical local power properties, and estimates of θ_0 are asymptotically normally distributed with the usual parametric rate (see Robinson (1991, 1994), Beran (1995), Velasco and Robinson (2000), Hualde and Robinson (2011)). This is the case whether δ_0 lies in the stationary region $(0, 1/2)$ or the nonstationary one $[1/2, \infty)$ (or, also, the negative dependent region $(-\infty, 0)$).

If N is regarded as fixed while $T \to \infty$, (1) is just a multivariate fractional model, with a vector, possibly stochastic, location. But in many practical applications N is large, and even when smaller than T, is more reasonably treated as diverging in asymptotic theory if T is. In that case inference on θ_0 is considerably complicated by an incidental parameters problem. In this paper we present and justify several approaches that resolve this question. We throughout employ asymptotic theory with respect to T diverging, where either N increases with T or stays fixed, and both cases are covered by indexing with respect to T only. In (1) the interest is in estimating θ_0 (efficiently, perhaps with some *a priori* knowledge on the range of allowed values) and testing hypotheses such as $I(1)$, $\delta_0 = 1$, or of absence of short memory structure, which might entail $\psi_0 = 0$. Hassler et al. (2011) have recently developed tests in a panel with a more general temporal dependence 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 $T \to \infty$.

The following section introduces four rival estimates of θ_0. Three are versions of time series conditional-sum-of-squares (CSS) estimates, recently treated in a general fractionally integrated setting by Hualde and Robinson (2011), one of which ignores the fixed effects, while the other two correct for them by regression and first differencing, respectively. The fourth is a Gaussian pseudo-maximum likelihood estimate (PMLE) based on the differenced model, and is somewhat more onerous computationally. Section 3 contains consistency theorems. In Section 4 the estimates are shown to be asymptotically normal. For the 3 CSS estimates, unless the restriction on the growth of N relative to T is very stringent, asymptotic biases in the central limit theorem are present, though for two of them bias-correction is possible. The PMLE suffers no such bias. In Section 5 we describe the implications of our results for hypothesis testing and interval estimation, numerically compare biases, and justify feasible bias correction. Section 6 consists of a Monte Carlo study of finite-sample performance of our methods. Section 7 discusses possible extensions. Theorem proofs appear in Appendix A. These depend in part on two Propositions, stated in Sections 3 and 4 but proved in Appendix B. Our proofs also use technical lemmas, stated and proved in Appendix C; we draw attention here to Lemma 3, which is a technical tool that is central to the consistency proofs, and Lemma 4, which is of some independent interest.

2. Parameter estimation

We consider four different, but asymptotically equivalent and efficient, methods of estimating θ_0 in (1). All these estimates are implicitly-defined and entail optimization over $\Theta = D \times \Xi$, where Ξ is a compact subset of R^p and $D = [\underline{\delta}, \overline{\delta}]$, where

$$\underline{\delta} > \max\left(0, \delta_0 - \frac{1}{2}\right), \quad \delta_0 \in D, \tag{12}$$

which implies that $\delta_0 > 0$ and $\delta > \delta_0 - \frac{1}{2}$ for $\delta \in D$. The choice of D thus implies some prior belief about the whereabouts of δ_0, for example to cover the possibility $\delta_0 = 1$, D can only include nonstationary δ-values, $\delta > \frac{1}{2}$. On the other hand there is no upper limit on $\overline{\delta}$. In Hualde and Robinson's (2011) study of CSS estimates in the pure time series case, D is effectively unrestricted. There may accordingly be scope for relaxing our restrictions on D, though these restrictions appear to play a role in ensuring that the approximation errors stemming from the presence of the individual effects α_i, or from the measures we take to eliminate them, are small enough to enable our estimates to be consistent and asymptotically normally distributed. The choice of Ξ can naturally embody stationarity and invertibility restrictions on $\psi(L;\xi)$, for example, for $p = 1$ and in the first-order autoregressive case $\psi(L;\xi) = 1 - \xi L$, we might take $\Xi = [\eta - 1, 1 - \eta]$ for arbitrarily small positive η. In general it is assumed that (3) holds for all $\xi \in \Xi$ with $c(\psi)$ satisfying

$$\inf_{\Xi} c(\xi) = c^* > 0. \tag{13}$$

Also, it is assumed that for $\xi \in \Xi$, $\psi(x;\xi)$ is continuous in ξ and, 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. All our estimates optimize objective functions that cross-sectionally aggregate time series objective functions.

It is helpful to define

$$\tau_t(\theta) = \lambda_t(L;\theta)\,1 = \lambda_t(1;\theta),$$

so using (4),

$$\tau_t(\theta) = \sum_{j=0}^{t} \lambda_j(\theta) = \sum_{k=0}^{t} \pi_k(\delta) \sum_{j=0}^{t-k} \psi_j(\xi) = \sum_{k=0}^{t} \psi_k(\xi) \sum_{j=0}^{t-k} \pi_k(\delta). \tag{14}$$

In the pure fractional case $\psi(L;\xi) \equiv 1$ from summing coefficients of L^j on both sides of the identity $\Delta\Delta^{\delta-1} = \Delta^{\delta}$,

$$\tau_t(\theta) = \sum_{k=0}^{t} \pi_k(\delta) = \pi_t(\delta - 1). \tag{15}$$

2.1. Uncorrected CSS estimation

Our first approach is essentially CSS estimation which ignores the α_i. Define

$$L_T^U(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} (\lambda_t(L;\theta) y_{it})^2, \tag{16}$$

and

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

Notice that, writing

$$v_{it} = \lambda_t^{-1}(L;\theta_0)\,\varepsilon_{it}, \tag{17}$$

we have for all $\theta \in \Theta$,

$$\lambda_t(L;\theta) y_{it} = \lambda_t(L;\theta)(\alpha_i + v_{it}) = \lambda_t(L;\theta) v_{it} + \tau_t(\theta)\alpha_i. \tag{18}$$

The term $\tau_t(\theta)\alpha_i$ in (18) contributes a bias. From (11),

$$\tau_t(\theta) = \frac{\psi(1;\xi)}{\Gamma(1-\delta)} t^{-\delta} + O(t^{-\delta-1}). \tag{19}$$

Thus the bias decays to zero for $\delta > 0$, but more or less slowly, and its presence explains the need for asymptotic theory with $T \to \infty$, in order to achieve consistent estimation of θ_0.

2.2. Fixed effects CSS estimation

Instead of ignoring the α_i we now start from a CSS-type objective function based on fractionally differencing the $y_{it} - \alpha_i$, and then concentrate out the α_i. Define

$$L_T(\theta, \alpha_1, \ldots, \alpha_N) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} (\lambda_t(L;\theta)(y_{it} - \alpha_i))^2.$$

Differentiating gives

$$\frac{\partial}{\partial \alpha_i} L_T(\theta, \alpha_1, \ldots, \alpha_N) = \frac{-2}{NT} \sum_{t=0}^{T} (\lambda_t(L;\theta) y_{it} - \lambda_t(1;\theta)\alpha_i)\,\lambda_t(1;\theta), \quad i = 1, \ldots, N,$$

and thence

$$\hat{\alpha}_{iT}(\theta) = \frac{1}{S_{\lambda\lambda T}(\theta)} \sum_{t=0}^{T} (\lambda_t(L;\theta) y_{it})\,\tau_t(\theta), \quad i = 1, \ldots, N,$$

using (14) and defining

$$S_{\tau\tau T}(\theta) = 1 + \boldsymbol{\tau}_T'(\theta)\,\boldsymbol{\tau}_T(\theta),$$
$$\boldsymbol{\tau}_T(\theta) = (\tau_1(\theta), \ldots, \tau_T(\theta))'.$$

Thence introduce

$$L_T^F(\theta) = L_T\left(\theta, \hat{\alpha}_{1T}(\theta), \ldots, \hat{\alpha}_{NT}(\theta)\right) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} \left(\lambda_t(L;\theta)\left(y_{it} - \hat{\alpha}_{iT}(\theta)\right)\right)^2, \tag{20}$$

and

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

The summands in $L_T^F(\theta)$ are squared fractional residuals after regression on the final end effect $\tau_t(\theta)$.

Since

$$\hat{\alpha}_{iT}(\theta) = \frac{1}{S_{\tau\tau T}} \sum_{t=0}^{T} \tau_t(\theta)(\lambda_t(L;\theta) y_{it}) = \alpha_i + \frac{a_{iT}(\theta)}{S_{\tau\tau T}(\theta)},$$

where

$$a_{iT}(\theta) = \sum_{t=0}^{T} \tau_t(\theta)\,\lambda_t(L;\theta)\,v_{it}, \tag{21}$$

note that

$$\lambda_t(L;\theta)\left(y_{it}-\hat{\alpha}_i(\theta)\right)=\lambda_t(L;\theta)v_{it}-\frac{a_{iT}(\theta)\tau_t(\theta)}{S_{\tau\tau T}(\theta)}, \quad (22)$$

and by comparison with $\widehat{\theta}_T^U$ there is again a term contributing bias. We show that nevertheless $\widehat{\theta}_T^F$ is consistent though a bias correction may be desirable for statistical inference.

2.3. Differenced CSS estimation

Applying another standard approach to eliminating the α_i, first-differencing gives:

$$\Delta y_{it}=\Delta v_{it},\quad t=1,\ldots,T.$$

We might then attempt to fully whiten the data by forming the

$$z_{it}(\theta)=\lambda_{t-1}\left(L;\theta^{(-1)}\right)(\Delta y_{it}),\quad t=1,\ldots,T,$$

where $\theta^{(-1)}=\left(\delta-1,\xi'\right)'$. Define

$$L_T^D(\theta)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}^2(\theta) \quad (23)$$

and

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

Note that

$$\begin{aligned}z_{it}(\theta)&=\lambda_t\left(L;\theta^{(-1)}\right)(\Delta v_{it})+\left(\lambda_{t-1}\left(L;\theta^{(-1)}\right)\right.\\&\quad\left.-\lambda_t\left(L;\theta^{(-1)}\right)\right)(\Delta v_{it})\\&=\lambda_t(L;\theta)v_{it}-\lambda_t\left(\theta^{(-1)}\right)L^t\Delta\lambda_t^{-1}(L;\theta_0)\varepsilon_{it}\\&=\lambda_t(L;\theta)v_{it}-\lambda_t\left(\theta^{(-1)}\right)\varepsilon_{i0}.\end{aligned} \quad (24)$$

This results from

$$\lambda_{t-1}\left(L;\theta^{(-1)}\right)-\lambda_t\left(L;\theta^{(-1)}\right)=-\lambda_t\left(\theta^{(-1)}\right)L^t$$

and from

$$\begin{aligned}\lambda_t\left(L;\theta^{(-1)}\right)(\Delta v_{it})&=v_{it}+\sum_{j=1}^{t}\left(\lambda_j\left(\theta^{(-1)}\right)-\lambda_{j-1}\left(\theta^{(-1)}\right)\right)v_{i,t-j}\\&=\lambda_t(L;\theta)v_{it},\end{aligned}$$

because

$$\begin{aligned}\lambda_j\left(\theta^{(-1)}\right)-\lambda_{j-1}\left(\theta^{(-1)}\right)&=\sum_{k=0}^{j}\pi_k(\delta-1)\psi_{j-k}(\xi)\\&\quad-\sum_{k=0}^{j-1}\pi_k(\delta-1)\psi_{j-1-k}(\xi)\\&=\psi_j(\xi)+\sum_{k=1}^{j-1}\psi_{j-k}(\xi)\pi_k(\delta-1)\\&\quad-\sum_{k=1}^{j-1}\pi_{k-1}(\delta-1)\psi_{j-k}(\xi)\\&\quad+\pi_j(\delta-1)-\pi_{j-1}(\delta-1)\\&=\psi_j(\xi)+\sum_{k=1}^{j-1}\psi_{j-k}(\xi)(\pi_k(\delta-1)\\&\quad-\pi_{k-1}(\delta-1))-\pi_{j-1}(\delta-1)\\&\quad+\pi_j(\delta-1)\\&=\lambda_j(\theta),\end{aligned} \quad (25)$$

since for $t\geq 1$, from (15)

$$\pi_j(\delta-1)-\pi_{j-1}(\delta-1)=\pi_j(\delta). \quad (26)$$

From (25)

$$\lambda_t\left(\theta^{(-1)}\right)=\tau_t(\theta),$$

so

$$z_{it}(\theta)=\lambda_t(L;\theta)v_{it}-\tau_t(\theta)\varepsilon_{i0}. \quad (27)$$

In view of (18) and (27) there is a bias contribution of the same order as that for the uncorrected estimate $\widehat{\theta}_T^U$. But something has been gained because the α_i have been eliminated and, as with $\widehat{\theta}_T^F$, it will be possible to institute a bias-correction.

2.4. Pseudo maximum likelihood estimation

The previous estimates all employ versions of the CSS principal, where the Gaussian pseudo-likelihood is approximated by ignoring potential dependence and heteroscedasticity in the approximately whitened data. Here we develop a PMLE based on the fractionally adjusted first differences $z_{it}(\theta)$ (as distinct from, for example, the PMLE based on pure first differences Δy_{it} in an autoregressive setting of Hsiao et al. (2002)). From (27), for $t\geq 1$

$$z_{it}(\theta_0)=\varepsilon_{it}-\tau_t(\theta_0)\varepsilon_{i0},$$

whence, denoting $\sigma_0^2=E\varepsilon_{it}^2$,

$$Cov\left(z_{is}(\theta_0),z_{it}(\theta_0)\right)=\sigma_0^2\omega_{st}(\theta_0),$$

where

$$\omega_{st}(\theta)=1(s=t)+\tau_s(\theta)\tau_t(\theta).$$

Introduce the $T\times T$ matrix $\Omega_T(\theta)=(\omega_{st}(\theta))$ and the $T\times 1$ vectors $\mathbf{z}_{iT}(\theta)=(z_{i1}(\theta),\ldots,z_{iT}(\theta))'$, $i=1,\ldots,N$. Define the approximate Gaussian pseudo log-likelihood

$$\begin{aligned}Q_T\left(\theta,\sigma^2\right)&=-\sum_{i=1}^{N}\left\{\frac{T}{2}\log(2\pi)+\frac{T}{2}\log\sigma^2+\frac{1}{2}\log|\Omega_T(\theta)|\right.\\&\quad\left.+\frac{1}{2\sigma^2}\mathbf{z}_{iT}'(\theta)\Omega_T^{-1}(\theta)\mathbf{z}_{iT}(\theta)\right\}.\end{aligned}$$

Differentiating,

$$\frac{\partial}{\partial\sigma^2}Q_T\left(\theta,\sigma^2\right)=-\sum_{i=1}^{N}\left\{\frac{T}{2\sigma^2}-\frac{1}{2\sigma^4}\mathbf{z}_{iT}'(\theta)\Omega_T^{-1}(\theta)\mathbf{z}_{iT}(\theta)\right\},$$

leads to

$$\widehat{\sigma}_T^2(\theta)=\frac{1}{NT}\sum_{i=1}^{N}\mathbf{z}_{iT}'(\theta)\Omega_T^{-1}(\theta)\mathbf{z}_{iT}(\theta),$$

and the concentrated function

$$\begin{aligned}Q_T\left(\theta,\widehat{\sigma}_T^2(\theta)\right)&=-\left\{\frac{NT}{2}\log(2\pi)+\frac{NT}{2}\log\widehat{\sigma}_T^2(\theta)\right.\\&\quad\left.+\frac{N}{2}\log|\Omega_T(\theta)|+\frac{NT}{2}\right\}\\&=-\frac{NT}{2}(1+\log(2\pi))-\frac{NT}{2}\log\widehat{\sigma}_T^2(\theta)\\&\quad-\frac{N}{2}\log|\Omega_T(\theta)|.\end{aligned}$$

Thus define

$$\begin{aligned}L_T^P(\theta)&=\exp\left\{-\frac{2}{NT}Q_T^P\left(\theta,\widehat{\sigma}^2(\theta)\right)-(1+\log(2\pi))\right\}\\&=|\Omega_T(\theta)|^{\frac{1}{T}}\widehat{\sigma}_T^2(\theta),\end{aligned}$$

and the PMLE

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

For computations, note the formulae

$$\Omega_T^{-1}(\theta)=I_T-\frac{\tau_T(\theta)\tau_T'(\theta)}{S_{\tau\tau T}(\theta)},\qquad|\Omega_T(\theta)|=S_{\tau\tau T}(\theta). \quad (28)$$

3. Consistency

Consistency proofs are facilitated by noting that all four of the objective functions introduced in the previous section are approximately equal, and are of the form

$$L_T(\theta) = A_T(\theta) + B_T(\theta),$$

where

$$A_T(\theta) = \frac{1}{N}\sum_{i=1}^{N} A_{iT}(\theta), \qquad A_{iT}(\theta) = \frac{1}{T}\sum_{t=0}^{T} \left(\lambda_t(L;\theta)\, v_{it}\right)^2$$

and $B_T(\theta)$ is a measurable function of ε_{it}, $1 \le i \le N$, $t \le T$, of smaller order of magnitude. Hualde and Robinson (2011) showed under conditions on ε_{1t}, $t = 1, 2, \ldots$, that are implied by ours, that the statistic

$$\widetilde{\theta}_T^1 = \arg\min_{\theta\in\Theta} A_{1T}(\theta)$$

is consistent for θ_0. They were thus concerned with the single time series case, but due to the identity of distribution across i, and model constancy across i, their results easily extend to establish consistency of

$$\widetilde{\theta}_T = \arg\min_{\theta\in\Theta} A_T(\theta).$$

We state first the following Proposition which is used to prove consistency of each of our estimates, along with Theorem 1 of Hualde and Robinson (2011). Define

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta} L_T(\theta).$$

Proposition 1. *Let*

$$\sup_{\theta\in\Theta} |B_T(\theta)| \to_p 0, \quad \text{as } T\to\infty. \tag{29}$$

Then as $T\to\infty$

$$\widehat{\theta}_T \to_p \theta_0.$$

Theorem 3.1. *If* $\sum_{i=1}^{N}\alpha_i^2 = O_p(N)$ *as* $N\to\infty$, *as* $T\to\infty$,

$$\widehat{\theta}_T^U \to_p \theta_0.$$

Theorem 3.2. *As* $T\to\infty$,

$$\widehat{\theta}_T^F \to_p \theta_0.$$

Theorem 3.3. *As* $T\to\infty$,

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

Theorem 3.4. *As* $T\to\infty$,

$$\widehat{\theta}_T^P \to_p \theta_0.$$

Note that these results hold without assumptions on N, which might be fixed or increase with T at any rate.

4. Asymptotic normality

The following Proposition is not new when $N = 1$ (cf. Robinson (1991)), but we include it to demonstrate that N may increase with T. Define

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

and, for $\psi(L;\xi)$ differentiable in ψ,

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

so

$$\chi_j(\xi) = \left(\chi_{1j}(\xi), \chi_{2j}'(\xi)\right)',$$

where

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

with $\dot{\psi}_j(\xi) = (\partial/\partial\xi)\,\psi(\xi)$, and we assume (cf (3)) that $\dot{\psi}_j(\xi) = O(\exp(-c(\xi)\, j))$. Define also

$$w_T = \frac{1}{2(NT)^{\frac{1}{2}}}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}A_{iT}(\theta_0) = \frac{1}{(NT)^{\frac{1}{2}}}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it} f_{it},$$

where

$$f_{it} = \sum_{j=0}^{t-1}\chi_{t-j}(\theta_0)\,\varepsilon_{ij}, \tag{30}$$

noting that

$$\frac{\partial}{\partial\theta}A_{iT}(\theta) = \frac{2}{T}\sum_{t=0}^{T}\left(\lambda_t(L;\theta)\, v_{it}\right)\frac{\partial}{\partial\theta}\lambda_t(L;\theta)\, v_{it}$$

and $\lambda_t(L;\theta_0)\, v_{it} = \varepsilon_{it}$ while, for $t\ge 1$,

$$\begin{aligned}\frac{\partial}{\partial\theta}\lambda_t(L;\theta_0)\, v_{it} &= \frac{\partial}{\partial\theta}\lambda(L;\theta_0)\,\lambda_t^{-1}(L;\theta_0)\,\varepsilon_{it}\\ &= \chi(L;\xi)\,\lambda(L;\theta_0)\,\lambda_t^{-1}(L;\theta_0)\,\varepsilon_{it}\\ &= \chi(L;\xi)\left\{\varepsilon_{it}1(t\ge 0)\right\}\\ &= f_{it}.\end{aligned} \tag{31}$$

Introduce 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.

Proposition 2. *As* $T\to\infty$,

$$w_T \to_d \mathcal{N}\left(0, \sigma_0^4 B(\xi_0)\right). \tag{32}$$

Now assume that $\theta_0 \in Int(\Theta)$, and for all x on the complex unit circle $\psi(x;\xi)$ is twice continuously differentiable in ψ in a neighbourhood of ξ_0; note that (3) implies that all derivatives in x on $|x| = 1$ of $\psi(x;\xi)$, and thus of $\phi(x;\xi)$, exist and are bounded.

Theorem 4.1. *Let* $\sum_{i=1}^{N}\alpha_i^2 = O_p(N)$ *as* $N\to\infty$. *When* $\delta_0 > \frac{1}{4}$, *as* $T\to\infty$,

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

if, as $T\to\infty$, $NT^{1-4\delta_0}\log^2 T \to 0$ *when* $\delta_0 \in \left(\frac{1}{4}, \frac{1}{2}\right)$, $NT^{-1}\log^4 T \to 0$ *when* $\delta_0 = \frac{1}{2}$, *and* $NT^{-1}\to 0$ *when* $\delta_0 > \frac{1}{2}$.

Note that Theorem 4.1, like Theorems 3.1–3.4, allows N to grow, but a slower rate than T, and arbitrarily slowly for δ_0 close enough to $\frac{1}{4}$ from above, and no central limit theorem is available when $\delta_0 < \frac{1}{4}$.

Define $\dot{\pi}_t(\delta) = (\partial/\partial\delta)\,\pi_t(\delta)$, so from (14),

$$\dot{\tau}_t(\theta) = \frac{\partial}{\partial\theta}\tau_t(\theta) = \left[\sum_{k=0}^{t}\dot{\pi}_k(\delta)\sum_{j=0}^{t-k}\psi_j(\xi) \quad \sum_{k=0}^{t}\pi_k(\delta)\sum_{j=0}^{t-k}\dot{\psi}_j'(\xi)\right]'. \tag{34}$$

Let

$$S_{\tau\dot{\tau}T}(\theta) = \sum_{t=1}^{T}\tau_t(\theta)\,\dot{\tau}_t(\theta).$$

Theorem 4.2. *As $T \to \infty$,*

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^F - \theta_0 - T^{-1}b_T^F(\theta_0)\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right), \tag{35}$$

where

$$b_T^F(\theta) = B^{-1}(\xi)\frac{S_{\tau\dot{\tau}T}(\theta)}{S_{\tau\tau T}(\theta)},$$

with $b_T^F(\theta) = O\left(\log T\,1(\delta \le \frac{1}{2}) + 1(\delta > \frac{1}{2})\right)$. Thus

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^F - \theta_0\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right) \tag{36}$$

if, as $T \to \infty$, $NT^{-1}\log^2 T \to 0$ when $\delta_0 \le \frac{1}{4}$, and if $NT^{-1} \to 0$ otherwise.

When $\delta_0 > \frac{1}{2}$, the restrictions on N for (36) are the same as those for (33) for $\widehat{\theta}_T^U$ but when $\delta_0 \le \frac{1}{2}$ they are weaker, and do not strengthen with decreasing δ_0, indeed (36), unlike (33), holds for $\delta_0 \in \left(0, \frac{1}{4}\right]$. Moreover, whereas Theorem 4.1, like Theorem 3.1, imposes some restriction on the α_i, this is avoided in Theorem 4.2. The recentering in (35) avoids any restrictions on N. Note that $b_T^F(\theta)$ is a known function of θ. When $\delta < 1$, in the pure fractional case $\psi(L;\xi) \equiv 1$, from (15) $\tau_t(\theta) > 0$, whence Lemma 3 in Appendix C implies that $S_{\tau\dot{\tau}T}(\theta) < 0$, and thus $b_T^F(\theta) < 0$.

Define

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

Theorem 4.3. *When $\delta_0 > \frac{1}{4}$, as $T \to \infty$,*

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^D - \theta_0 - T^{-1}b_T^D(\theta_0)\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right), \tag{37}$$

where

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

and $b_T^D(\theta) = O(T^{1-2\delta}\log T\,1(\delta < \frac{1}{2}) + \log^2 T\,1(\delta = \frac{1}{2}) + 1(\delta > \frac{1}{2}))$. Thus, when $\delta_0 > \frac{1}{4}$,

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^D - \theta_0\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right) \tag{38}$$

if, as $T \to \infty$, $NT^{1-4\delta_0}\log^2 T \to 0$ when $\delta_0 \in \left(\frac{1}{4}, \frac{1}{2}\right)$, $NT^{-1}\log^4 T \to 0$ when $\delta_0 = \frac{1}{2}$, and $NT^{-1} \to 0$ when $\delta_0 > \frac{1}{2}$.

The result (38) is the same as (33) for $\widehat{\theta}_T^U$, except that it imposes no restrictions on the α_i. As with (35) for $\widehat{\theta}_T^F$, (37) avoids any restrictions on N, but it requires $\delta_0 > \frac{1}{4}$, as (33) and (38), due to the slow convergence of the term causing the bias in the CSS which are of similar nature in both the uncorrected and difference CSS estimation. The bias term $b_T^D(\theta)$ lacks the deflating factor $S_{\tau\tau T}^{-1}(\theta) \le 1$ of $\widehat{\theta}_T^F$, making it of larger order of magnitude than $b_T^F(\theta)$ when $\delta_0 \le \frac{1}{2}$, and it also involves the additional term $S_{\tau\chi T}(\theta)$. This is $O(1)$ for all δ (see Lemma 1 in Appendix C) and is thus dominated asymptotically by $S_{\tau\dot{\tau}T}(\theta)$ when $\delta_0 \le \frac{1}{2}$. In the pure fractional case $\psi(L;\xi) \equiv 1$, when $\delta < 1$ from (15) $\tau_t(\theta) > 0$, and thus $S_{\tau\chi T}(\theta) < 0$, and since $S_{\tau\dot{\tau}T}(\theta) < 0$ as previously observed, there is some cancellation in the bias, while when $\delta > 1$, $\sum_{t=1}^{\infty}\pi_t(\delta-1) = -1$, and it is readily seen that $S_{\tau\chi T}(\theta) > 0$ for all large enough T.

Theorem 4.4. *As $T \to \infty$,*

$$(NT)^{\frac{1}{2}}\left(\widehat{\theta}_T^P - \theta_0\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right). \tag{39}$$

Theorem 4.4 demonstrates superiority of $\widehat{\theta}_T^P$ in that it imposes no restrictions on N or δ, to compensate for its somewhat greater computational complexity relative to our other estimates.

5. Statistical inference

In the present section we develop the results of the previous section for statistical inference on θ_0. Our results allow for example testing of a short memory composite null hypothesis $\delta_0 = 0$ against long memory alternatives, testing an $I(1)$ composite hypothesis $\delta_0 = 1$, and testing the short memory component, for example the pure fractional null $\psi(L;\xi) = 1$ with δ_0 unspecified, or the composite null $\psi(L;\xi) = 1 - \sum_{j=1}^{p}\psi_{j0}L^j$, with unspecified δ_0 and $\psi_{j0}, j = 1, \ldots, p$. Theorems 4.1–4.4 suggest that $\widehat{\theta}_T^F$, $\widehat{\theta}_T^D$ and $\widehat{\theta}_T^P$ are more useful than $\widehat{\theta}_T^U$ with potential for bias correction of $\widehat{\theta}_T^F$, $\widehat{\theta}_T^D$, thereby relaxing the restrictions on the rate of increase of N relative to T, and since inference based on is $\widehat{\theta}_T^P$ straightforward, requiring no bias correction, we do not discuss this further in the present section.

We first consider Wald hypothesis testing on δ_0, focusing on the pure fractional case $\psi(L;\xi) \equiv 1$. The leading case, mentioned in the Introduction, of testing the $I(1)$ null $\delta_0 = 1$, turns out to be the most favourable. Since $\tau_t(1) = 0$, $1 \le t \le T$, it follows that $b_T^F(1) = b_T^D(1) = 0$. Thus the results (36) and (38) are respectively identical to (35) and (37), and so $(NT)^{1/2}\left(\widehat{\delta}_T^F - 1\right)$ and $(NT)^{1/2}\left(\widehat{\delta}_T^D - 1\right)$ are asymptotically $\mathcal{N}\left(0, 6/\pi^2\right)$ with no restrictions on N. Another case that is sometimes of interest is the $I(2)$ hypothesis $\delta_0 = 2$. It is easy to see that $S_{\tau\tau T}(2) = S_{\tau\dot{\tau}T}(2) = S_{\tau\chi T}(2) = 1$, so $b_T^F(2) = -6/\pi^2$, $b_T^D(2) = 0$, and $\widehat{\delta}_T^F$ is simply bias-corrected, while no correction of $\widehat{\delta}_T^D$ is needed. In general, for other null hypotheses, for example $\delta_0 = \frac{1}{2}$ (the boundary between the stationary and nonstationary regions), we can carry out the bias correction by evaluating $b_T^F(\delta_0)$ and $b_T^D(\delta_0)$ at the null, which is straightforward given Lemma 3 in Appendix C, and applying (35) and (37).

Some numerical comparisons of the biases are of interest. Tables 1 and 2 present the scaled biases of $\widehat{\delta}_T^F$ and $\widehat{\delta}_T^D$ for selected values of T and δ. We find that $b_T^F(\delta)$ decreases monotonically in δ and in T, sharing the sign of δ, whereas $b_T^D(\delta)$ is positive and increasing in $|\delta - 1|$ (though not symmetrically) and is mostly decreasing in T (note that scaling with respect to T has already been carried out).

For interval estimation $b_T^F(\theta_0)$ and $b_T^D(\theta_0)$ need to be estimated, while when a short memory parameter vector ψ_0 is present this must be estimated even for hypothesis testing on δ_0 only. We introduce the feasibly bias-corrected estimates

$$\widetilde{\theta}_T^F = \widehat{\theta}_T^F - T^{-1}b_T^F\left(\widehat{\theta}_T^F\right),$$
$$\widetilde{\theta}_T^D = \widehat{\theta}_T^D - T^{-1}b_T^D\left(\widehat{\theta}_T^D\right).$$

The following theorems indicate that these estimates entail stronger restrictions on N (and in some cases on δ_0) than the infeasible bias-corrected ones featured in (35) and (37), but milder restrictions than the uncorrected ones $\widehat{\theta}_T^F$ and $\widehat{\theta}_T^D$. In particular, $\widetilde{\theta}_T^D$

Table 1
Scaled asymptotic bias $b_T^F(\delta) \times 100/T$ of fixed effect estimate, $\psi(L;\xi) \equiv 1$.

T	δ: 0.3	0.6	0.9	1.0	1.1	1.4
5	−17.77	−11.04	−2.25	0	1.76	4.77
10	−11.54	−6.64	−1.17	0	0.85	2.24
100	−2.25	−1.04	−0.13	0	0.08	0.21

Table 2
Scaled asymptotic bias $b_T^D(\delta) \times 100/T$ of differenced estimate, $\psi(L;\xi) \equiv 1$.

T	δ: 0.3	0.6	0.9	1.0	1.1	1.4
5	27.05	5.43	0.20	0	0.14	1.17
10	28.94	4.51	0.14	0	0.08	0.63
100	18.90	1.18	0.02	0	0.01	0.06

still requires $\delta_0 > \frac{1}{4}$ as $\widehat{\theta}_T^D$ and both estimates require N to increase slower than T^3, though as with $\widetilde{\theta}_T^U$, the rate for $\widetilde{\theta}_T^D$ is heavily δ_0-dependent, such that N cannot increase much faster than T when δ_0 approaches $\frac{1}{4}$ from above.

Theorem 5.1. *As* $T \to \infty$,

$$(NT)^{\frac{1}{2}}\left(\widetilde{\theta}_T^F - \theta_0\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right),$$

if, as $T \to \infty$, $NT^{-3}\log^6 T \to 0$ *when* $\delta_0 \le \frac{1}{2}$ *or* $NT^{-3} \to 0$ *when* $\delta_0 > \frac{1}{2}$.

Theorem 5.2. *When* $\delta_0 > \frac{1}{4}$, *as* $T \to \infty$,

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

if $NT^{1-8\delta_0}\log^6 T \to 0$ *when* $\delta_0 \in \left(\frac{1}{4}, \frac{1}{2}\right)$, *or if* $NT^{-3}\log^{10} T \to 0$ *when* $\delta_0 = \frac{1}{2}$, *or if* $NT^{-3} \to 0$ *when* $\delta_0 > \frac{1}{2}$.

Simplified corrections are possible that improve on our original F and D estimates, but by less than our feasible bias-corrected ones. In the case $\psi(L;\xi) \equiv 1$,

$$\begin{aligned} b_T^F(\delta) = &-\left\{\frac{6\log T}{\pi^2} + O(1)\right\} 1\left(\delta < \frac{1}{2}\right) \\ &-\left\{\frac{3\log T}{\pi^2} + O(1)\right\} 1\left(\delta = \frac{1}{2}\right) \\ &+\left\{\frac{6\omega(\delta)}{\pi^2} + O(T^{1-2\delta}\log T)\right\} 1\left(\delta > \frac{1}{2}\right), \end{aligned} \tag{40}$$

$$\begin{aligned} b_T^D(\delta) = &\left\{\frac{6T^{1-2\delta}\log T}{\pi^2(1-2\delta)\Gamma(1-\delta)^2} + O(T^{1-2\delta})\right\} 1\left(\delta < \frac{1}{2}\right) \\ &+\left\{\frac{3\log^2 T}{2\pi^3} + O(\log T)\right\} 1\left(\delta = \frac{1}{2}\right), \\ &-\left\{\frac{6\omega(\delta)\pi^{-2}}{(2\delta-1)B(\delta,\delta)} + \int_0^1 \left(\frac{(1-x)^{\delta-1}-1}{x}\right) dx \right. \\ &\left. + O\left(\frac{\log T}{T^{2\delta-1}} + T^{-\delta}\right)\right\} 1\left(\delta > \frac{1}{2}\right), \end{aligned} \tag{41}$$

where $\omega(\delta) = \psi(2\delta) - \psi(\delta) - (2\delta-1)^{-1}$, defining the digamma function $\psi(x) = (\partial/\partial x)\log\Gamma(x)$; this follows from Lemma 1 since with $\tau_j(\theta) = \pi_j(\delta-1)$, we have $\sum_{j=0}^{\infty}\tau_j^2(\theta) = ((2\delta-1)B(\delta,\delta))^{-1}$, $\sum_{j=0}^{\infty}\tau_j(\theta)\dot{\tau}_t(\theta) = \frac{1}{2}(\partial/\partial\theta)\sum_{j=0}^{\infty}\tau_j^2(\theta)$, $(\partial/\partial\delta)\log B(\delta,\delta) = 2(\psi(2\delta)-\psi(\delta))$ and $\sum_{j=1}^{\infty}\tau_j(\delta)/j = \int_0^1\left((1-x)^{\delta-1}-1\right)/x dx$. The leading terms in (40) and (41) could be used in simpler bias corrections. For example a simple bias-corrected Fixed effects estimate is $\widehat{\delta}_T^F - 6\log T/\left(\pi^2 T\right)$ for $\delta_0 < \frac{1}{2}$, where the correction is free of $\widehat{\delta}_T^F$. But which correction to use requires knowledge of whether or not we are in the stationary region, and the theoretical improvements over the original bias-uncorrected estimates are small, noting the approximation errors above and bearing in and that the effect of inserting estimates of δ_0 in most of the corrections needs to be taken into account. Tables 3 and 4 illustrate the approximations, and are directly comparable with those of Tables 1 and 2, respectively. The approximations work reasonably well when δ_0 is close to 1, but otherwise are less precise.

Table 3
Approximation (40) to asymptotic bias of fixed effect estimate $b_T^F(\delta) \times 100/T$.

T	δ: 0.3	0.6	0.9	1.0	1.1	1.4
5	−19.57	−45.57	−2.55	0	1.64	4.21
10	−14.00	−22.79	−1.28	0	0.82	2.11
100	−2.80	−2.28	−0.13	0	0.08	0.21

Table 4
Approximation (41) to asymptotic bias of difference estimate $b_T^D(\delta) \times 100/T$.

T	δ: 0.3	0.6	0.9	1.0	1.1	1.4
5	55.27	82.64	0.44	0	0.20	1.28
10	52.17	41.32	0.22	0	0.10	0.64
100	26.21	4.13	0.02	0	0.01	0.06

6. Simulations

In this section we conduct a simulation study of the finite sample properties of our estimates of θ_0. in the pure fractional case, $\theta = \delta$. We concentrate on the Fixed Effects and Difference estimates and the PML estimates, in both original and feasible bias-corrected forms, and the PML estimates, but not for Uncorrected estimates, which heavily depend on the magnitude of the fixed effects α_i relative to the idiosyncratic errors ε_{it}, whereas the others are invariant to the specification of α_i.

We focus first on the pure fractional case, $\theta = \delta$. We generate the ε_{it} as standard normal, noting that the estimates are invariant to the variance of ε_{it}. We consider different choices of N, T and δ_0. In particular we set $T = 5, 10$ and 100 as in Tables 1–4, and to consider the effect of increasing the overall sample size, we used when $T = 5, 10$ three combinations of NT (100, 200 and 400) so the range of values of N oscillates from $N = 20$ to 80 for $T = 5$ and from $N = 10$ to 40 for $T = 10$, while when $T = 100$ we took only $NT = 200$ and 400, i.e. $N = 2$ and 4 (thus omitting the case $NT = T = 100$ since we cannot remove fixed effects with a single time series). The values of δ_0 include a stationary one ($\delta_0 = 0.3$), which is the most problematic from the point of view of bias, a moderately non-stationary one ($\delta_0 = 0.6$), values around the unit root ($\delta_0 = 0.9, 1.0, 1.1$), and a more nonstationary one ($\delta_0 = 1.4$). Optimizations were carried out using the Matlab function fminbnd with $D = [0.1, 1.5]$, and the results are based on 10,000 independent replications.

We first explore the accuracy of the asymptotic approximations for the biases in Theorems 4.2–4.3, and whether feasible bias correction produces better centering properties. In Table 5 we observe that the uncorrected Fixed Effects estimate $\widehat{\delta}_T^F$ has a bias in line with that predicted in Table 1 when $\delta_0 = 0.3$ and $T = 5$, but in general it has larger bias (in absolute value) than predicted by the magnitude of $b_T^F(\delta_0)/T$ for large T and small δ_0. For $\delta_0 \ge 1.0$ the bias is small, as predicted, and the accuracy of the approximation improves with increasing N. The right panel of Table 5 shows that feasible bias correction removes a large fraction of the bias of $\widehat{\delta}_T^F$ when $\delta_0 = 0.3$, but for all the smallish δ_0 the biases, while reduced, are still substantial. In some cases the biases of $\widehat{\delta}_T^F$ and $\widetilde{\delta}_T^F$ do not change monotonically with δ_0 and T. For the Difference estimate $\widehat{\delta}_T^D$ we observe that Table 6 shows more monotonic properties of $b_T^D(\delta_0)/T$ found in Table 2, even for the smaller NT, and that bias correction works in $\widetilde{\delta}_T^D$ quite well when $\delta_0 \ge 0.6$. Table 7 illustrates the far

Table 5
100× Empirical bias of fixed effect estimates $\widehat{\delta}_T^F, \widetilde{\delta}_T^F$.

	Uncorrected estimates $\widehat{\delta}_T^F$						Bias-corrected estimates $\widetilde{\delta}_T^F = \widehat{\delta}_T^F - b_T^F(\widehat{\delta}_T^F)/T$					
δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$						$NT = 100$					
5	−19.95	−45.42	−19.43	−6.51	−0.33	2.37	−0.42	−26.69	−12.57	−4.69	−1.62	−2.48
10	−17.80	−21.13	−4.27	−1.81	−0.42	0.50	−5.11	−11.34	−2.22	−1.41	−1.11	−1.73
T	$NT = 200$						$NT = 200$					
5	−20.00	−48.28	−14.06	−2.49	1.25	3.69	−0.47	−29.01	−8.34	−1.70	−0.52	−1.26
10	−18.92	−20.23	−2.88	−0.91	0.24	1.51	−6.15	−10.32	−1.17	−0.72	−0.57	−0.76
100	−12.95	−7.99	−1.54	−0.71	−0.19	0.43	−5.05	−3.07	−0.76	−0.63	−0.57	−0.66
T	$NT = 400$						$NT = 400$					
5	−20.00	−49.62	−9.11	−0.88	1.75	4.49	−0.47	−30.13	−4.48	−0.57	−0.18	−0.53
10	−19.58	−19.32	−2.20	−0.45	0.59	2.01	−6.77	−9.37	−0.65	−0.36	−0.28	−0.29
100	−13.38	−7.31	−1.11	−0.35	0.13	0.83	−5.45	−2.44	−0.40	−0.31	−0.28	−0.26

Table 6
100× Empirical bias of difference estimates $\widehat{\delta}_T^D, \widetilde{\delta}_T^D$.

	Uncorrected estimates $\widehat{\delta}_T^D$						Bias-corrected estimates $\widetilde{\delta}_T^D = \widehat{\delta}_T^D - b_T^D(\widehat{\delta}_T^D)/T$					
δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$						$NT = 100$					
5	21.80	5.64	−0.76	−1.05	−0.85	−0.46	11.58	1.17	−1.36	−1.28	−1.09	−1.58
10	17.91	3.63	−0.93	−1.10	−1.00	−0.93	6.04	−0.56	−1.31	−1.22	−1.13	−1.53
T	$NT = 200$						$NT = 200$					
5	22.34	6.13	−0.28	−0.56	−0.36	0.66	13.01	2.22	−0.66	−0.67	−0.54	−0.51
10	18.63	4.14	−0.44	−0.60	−0.50	−0.00	8.07	0.50	−0.69	−0.66	−0.60	−0.62
100	15.12	2.59	−0.50	−0.59	−0.54	−0.33	4.04	−0.37	−0.65	−0.62	−0.59	−0.64
T	$NT = 400$						$NT = 400$					
5	22.65	6.42	−0.00	−0.29	−0.09	1.23	13.77	2.79	−0.29	−0.34	−0.25	0.03
10	19.06	4.47	−0.15	−0.31	−0.21	0.43	9.13	1.10	−0.34	−0.34	−0.30	−0.20
100	15.71	2.94	−0.20	−0.30	−0.24	0.06	5.56	0.22	−0.32	−0.31	−0.29	−0.26

Table 7
100× Empirical bias of PML estimate $\widehat{\delta}_T^P$.

δ_0:	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$					
5	0.18	−1.86	−1.17	−0.98	−0.84	−1.22
10	−0.58	−1.43	−1.09	−1.00	−0.94	−1.27
T	$NT = 200$					
5	−0.31	−0.98	−0.58	−0.49	−0.42	−0.47
10	−0.54	−0.76	−0.56	−0.51	−0.47	−0.52
100	−0.55	−0.65	−0.56	−0.53	−0.52	−0.57
T	$NT = 400$					
5	−0.31	−0.46	−0.28	−0.23	−0.20	−0.17
10	−0.33	−0.38	−0.28	−0.26	−0.24	−0.23
100	−0.28	−0.31	−0.27	−0.26	−0.25	−0.24

superior bias properties of the PML estimate $\widehat{\delta}_T^P$, which are much better than those of the previous bias-corrected estimates.

Tables 8–10 report (scaled) Monte Carlo square error across simulations for the three estimates, in both uncorrected and feasible bias-corrected versions in case of $\widehat{\delta}_T^F$ and $\widehat{\delta}_T^D$. For all estimates, performance improves with increasing δ_0, T and NT, predominately monotonically, and with bias correction when implemented. The asymptotic standard error, $\left(6/\pi^2\right)/(NT)$, which gives 0.61, 0.30 and 0.15 for $NT = 100$, 200 and 400, respectively, are poorly approximated for low δ_0, but in a number of cases quite well approximated for larger δ_0.

Tables 11–13 report empirical coverage of 95% confidence intervals for δ_0 based on our central limit theorems. The $\widehat{\delta}_T^P$ estimate achieves the most accurate coverage, although the results leave something to be desired when $\delta_0 = 0.3$ and 0.6, but the bias-corrected $\widetilde{\delta}_T^F$ and $\widetilde{\delta}_T^D$ also generally perform reasonably, at least for the larger δ_0, especially by comparison with intervals based on uncorrected estimates.

We next incorporated autoregressive short memory, FAR(1), taking $p = 1$ and $\psi(L; \xi) = 1 - \xi L$, generating data for the same 6 δ_0 values as before, and for $\xi_0 = -0.5, -0.3, 0.3, 0.5$. In view of the large number of potential cases, Tables 14 and 15 report Monte Carlo bias and mean square error respectively for only the, superior, PMLE. As might be expected the biases of and mean square errors of $\widehat{\delta}_T^P$ increase markedly relative to the pure fractional case, though interestingly by no means always increasing with $|\xi_0|$, the figures for $\widehat{\xi}_T^P$ being broadly comparable. Empirical coverages of $\widehat{\delta}_T^P$ likewise deteriorated relative to Table 13, and are not reported in order to conserve on space.

7. Final comments

We have established asymptotic properties of four estimates of the time series parameters in the fractional panel model (1), finding that the simplest one is the least useful practically, two others are useful at least after bias-correction which may limit the magnitude of N relative to T, and the fourth requires no bias correction and is valid for all sequences N as T increases. We have focused on a relatively simple model in order to get ideas across, as even here some details are complicated, but a number of modifications and extensions are possible.

1. All our procedures are justified under large-T asymptotics. It seems possible to consider methods that are likely to be valid under $N \to \infty$ and/or $T \to \infty$, in particular a *PMLE* based not on the fractionally adjusted first differences $z_{it}(\theta)$ but on the pure first differences Δy_{it}. However, though theory with $N \to \infty$ only seems relatively straightforward, the covariance matrix of the vector $(\Delta y_{i2}, \ldots, \Delta y_{iT})$ does not have the simple identity-plus-rank-one-matrix structure of $\Omega_T(\theta)$, and theory with $T \to \infty$ seems harder than for our methods, and may also require focusing on $\delta_0 \in (1/2, 3/2)$.

2. There are alternative ways of introducing short memory parameterizations. If $\delta_0 < 1/2$ is assumed we can employ the untruncated model

$$\lambda(L; \theta_0)(y_{it} - \alpha_i) = \varepsilon_{it}, \tag{42}$$

Table 8
Empirical MSE×100 of fixed effect estimates $\widehat{\delta}_T^F$, $\widetilde{\delta}_T^F$.

	Uncorrected estimates $\widehat{\delta}_T^F$						Bias-corrected estimates $\widetilde{\delta}_T^F = \widehat{\delta}_T^F - b_T^F(\widehat{\delta}_T^F)/T$					
δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$						$NT = 100$					
5	4.00	22.95	15.76	6.69	2.64	0.75	0.02	8.64	8.90	4.08	1.79	0.66
10	3.51	8.09	2.05	1.37	1.10	0.68	0.55	3.98	1.38	1.06	0.94	0.67
T	$NT = 200$						$NT = 200$					
5	4.00	24.04	9.49	2.15	0.87	0.51	0.00	8.91	5.23	1.31	0.61	0.35
10	3.69	6.29	0.88	0.61	0.52	0.40	0.48	2.69	0.59	0.48	0.45	0.36
100	2.17	1.56	0.52	0.45	0.42	0.35	0.71	0.81	0.43	0.40	0.39	0.35
T	$NT = 400$						$NT = 400$					
5	4.00	24.73	4.32	0.62	0.42	0.42	0.00	9.16	2.26	0.39	0.29	0.19
10	3.86	4.94	0.2	0.29	0.26	0.25	0.49	1.76	0.27	0.23	0.22	0.20
100	2.09	0.99	0.25	0.22	0.20	0.19	0.58	0.41	0.21	0.19	0.19	0.18

Table 9
100× Empirical MSE of difference estimates $\widehat{\delta}_T^D$, $\widetilde{\delta}_T^D$.

	Uncorrected estimates $\widehat{\delta}_T^D$						Bias-corrected estimates $\widetilde{\delta}_T^D = \widehat{\delta}_T^D - b_T^D(\widehat{\delta}_T^D)/T$					
δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$						$NT = 100$					
5	5.79	1.43	1.19	1.20	1.19	0.82	3.96	1.95	1.37	1.25	1.17	0.80
10	4.09	1.01	0.92	0.93	0.93	0.73	3.17	1.56	1.03	0.96	0.92	0.72
T	$NT = 200$						$NT = 200$					
5	5.51	0.93	0.59	0.60	0.59	0.47	2.93	0.97	0.67	0.61	0.57	0.44
10	3.90	0.60	0.45	0.46	0.45	0.40	1.88	0.73	0.49	0.46	0.45	0.39
100	2.72	0.45	0.39	0.39	0.39	0.36	1.59	0.60	0.41	0.40	0.39	0.36
T	$NT = 400$						$NT = 400$					
5	5.40	0.70	0.30	0.30	0.30	0.27	2.51	0.54	0.33	0.30	0.29	0.24
10	3.85	0.42	0.22	0.23	0.22	0.21	1.42	0.36	0.24	0.23	0.22	0.21
100	2.68	0.27	0.19	0.19	0.19	0.18	0.95	0.29	0.20	0.19	0.19	0.18

Table 10
100× Empirical MSE of PML estimate $\widehat{\delta}_T^P$.

δ_0:	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$					
5	2.37	2.22	1.26	1.09	1.00	0.75
10	1.56	1.40	0.94	0.87	0.83	0.68
T	$NT = 200$					
5	1.46	1.08	0.60	0.52	0.48	0.40
10	0.90	0.67	0.45	0.42	0.40	0.36
100	0.63	0.50	0.39	0.37	0.36	0.34
T	$NT = 400$					
5	0.82	0.52	0.29	0.25	0.23	0.21
10	0.47	0.33	0.22	0.20	0.19	0.18
100	0.32	0.25	0.19	0.18	0.18	0.17

as in the stationary long memory literature. Without such a restriction on δ_0, we can consider, as in Hualde and Robinson (2011), for example,

$$\Delta_t^{\delta_0} (y_{it} - \alpha_i) = \psi (L; \xi_0) \varepsilon_{it}, \quad (43)$$

where $\Delta_t^{\delta} = \sum_{j=0}^{t} \pi_j (\theta) L^j$ (cf (2)). However, under both (42) and (43), $\widehat{\theta}_T^F$ and $\widehat{\theta}_T^D$ have additional bias components, due to $\{\varepsilon_{it}, t < 0\}$, that are again given functions of θ_0, and are of similar orders of magnitude to the biases so far encountered, but involve infinite series in general, and complicate matters considerably. Furthermore a PMLE of θ_0 under both (42) and (43) would involve an objective function far harder to handle theoretically and computationally than $L_T^P (\theta)$.

3. The iid requirement over t of the ε_{it} could be weakened to martingale difference and mild homogeneity assumptions as in Hualde and Robinson (2011), but for aesthetic reasons we keep the conditions simple by matching the iid assumption across i.

4. Variation in parameters across given subsets of the cross section can be accommodated relatively straightforwardly since it is only required that T increases in the asymptotics.

5. It would be possible to incorporate exogenous variables that vary with t, or with i and t, perhaps in a linear regression framework; this raises an additional initial values issue, see Hsiao et al. (2002).

6. Time trends can be introduced, perhaps with coefficients that vary over the cross section in the same way as the α_i, for example α_i can be replaced in (1) by $\alpha_i + \beta_i t$. The additional coefficients can be eliminated by extending our approaches for dealing with (1), for example taking second differences, but the details are more involved. Nonparametric trends can also be considered, see e.g. Robinson (2012).

7. We allow for cross-sectional dependence and heteroscedasticity in the y_{it} via the α_i. However, conditional on the α_i the y_{it} are cross-sectionally iid. It would be straightforward to relax this requirement in case of fixed N, such as by allowing $(\varepsilon_{1t}, \ldots, \varepsilon_{Nt})$ to have an unrestricted covariance matrix. For increasing N the covariance structure can thus be thought of as nonparametric, and more challenging to deal with, and in a different model with such structure Robinson (2012) found it necessary to heavily restrict the rate of increase of N with T. Alternatively a parametric form can be employed, such as a factor model (see Ergemen and Velasco, 2014) or, when there is knowledge of spatial locations or differences, a spatial model. Generally, cross-sectional dependence raises questions of robust inference and efficient estimation, but the bias issues encountered would remain much the same under cross-sectional dependence.

Acknowledgements

The authors are grateful for the comments of two referees, which have prompted several improvements. The first author's research was supported by a Cátedra de Excelencia at Universidad Carlos III de Madrid, Spanish Plan Nacional de I+d+I Grant SEJ2007-62908/ECON, and ESRC Grant ES/J007242/1. The second author's research was supported by Spanish Ministerio de Economía y Competitividad Grant ECO2012-31748.

Table 11
Empirical coverage of 95% CI based on $\widehat{\delta}_T^F, \widetilde{\delta}_T^F$.

	Uncorrected estimates $\widehat{\delta}_T^F$						Bias-corrected estimates $\widetilde{\delta}_T^F = \widehat{\delta}_T^F - b_T^F(\widehat{\delta}_T^F)/T$					
δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$						$NT = 100$					
5	0.17	4.18	56.17	70.08	76.82	96.08	99.90	3.94	65.28	79.29	84.15	92.36
10	14.34	36.83	78.00	83.57	86.42	95.51	98.30	53.69	84.70	87.72	89.10	93.66
T	$NT = 200$						$NT = 200$					
5	0.01	1.36	55.98	71.08	76.79	97.63	99.99	1.22	68.82	80.81	84.65	92.75
10	4.25	26.89	78.22	84.27	86.88	96.71	99.37	48.75	85.69	88.36	89.50	94.12
100	34.47	61.87	87.23	89.64	90.79	96.23	66.33	78.74	90.46	91.34	91.77	95.04
T	$NT = 400$						$NT = 400$					
5	0	0.18	54.20	71.41	75.85	67.41	100.00	0.19	70.20	81.60	85.16	93.17
10	0.39	14.63	77.75	84.57	86.91	85.98	99.88	41.74	86.03	88.66	89.75	94.59
100	16.27	52.86	87.56	90.04	91.09	91.50	59.08	77.94	90.87	91.72	92.10	92.35

Table 12
Empirical coverage of 95% CI based on $\widehat{\delta}_T^D, \widetilde{\delta}_T^D$.

	Uncorrected estimates $\widehat{\delta}_T^D$						Bias-corrected estimates $\widetilde{\delta}_T^D = \widehat{\delta}_T^D - b_T^D(\widehat{\delta}_T^D)/T$					
δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$						$NT = 100$					
5	25.38	79.33	84.26	84.05	84.08	93.01	49.49	73.94	81.94	83.65	84.83	92.20
10	37.15	87.32	89.09	88.98	89.07	93.90	59.10	80.50	87.57	88.64	89.38	93.45
T	$NT = 200$						$NT = 200$					
5	5.84	71.93	83.90	83.75	83.94	94.13	36.60	72.45	81.91	83.56	84.79	92.79
10	11.91	83.00	89.31	89.16	89.14	94.75	50.94	80.10	87.96	89.01	89.54	94.01
100	24.68	89.49	91.70	91.61	91.64	95.27	59.11	84.83	90.95	91.53	91.81	94.97
T	$NT = 400$						$NT = 400$					
5	0.26	58.03	83.87	83.77	84.02	83.80	20.17	69.13	81.95	83.68	84.9	86.33
10	0.93	74.37	89.32	89.18	89.29	89.67	37.27	79.32	88.03	89.10	89.76	90.34
100	4.56	85.27	91.93	91.83	91.91	92.17	51.33	84.91	91.28	91.79	92.11	92.28

Table 13
Empirical coverage of 95% CI based on $\widehat{\delta}_T^P$.

δ_0:	0.3	0.6	0.9	1.0	1.1	1.4
T	$NT = 100$					
5	59.04	71.41	83.97	86.23	87.81	93.60
10	73.30	81.69	89.13	90.25	90.92	94.35
T	$NT = 200$					
5	58.88	71.45	84.41	86.74	88.21	94.10
10	73.34	81.97	89.59	90.67	91.30	94.94
100	82.29	87.63	91.91	92.42	92.69	95.41
T	$NT = 400$					
5	59.08	71.63	84.98	87.36	88.85	90.36
10	73.47	82.22	89.94	91.11	91.72	92.32
100	82.56	87.94	92.34	92.96	93.28	93.45

Appendix A. Proofs of theorems

Proof of Theorem 3.1. From (16) and (18), $L_T^U(\theta) = L_T(\theta) + (NT)^{-1}\sum_{i=1}^N \alpha_i a_{iT}(\theta) + (NT)^{-1}\sum_{t=0}^T \tau_t^2(\theta)\sum_{i=1}^N \alpha_i^2$. We check Proposition 1. From (12), $\zeta = \frac{1}{2} - \left(\delta_0 - \underline{\delta}\right) > 0$. Thus, using Lemma 2 and $\sum_{i=1}^N \alpha_i^2 = O_p(N)$,

$$\sup_{D,\Xi}\left|\frac{2}{NT}\sum_{i=1}^N \alpha_i a_{iT}(\theta)\right| \le \frac{2}{TN}\left(\sum_{i=1}^N \alpha_i^2 \sum_{i=1}^N \sup_{D,\Xi} a_{iT}^2(\theta)\right)^{1/2}$$
$$= O_p\left(T^{-1}\sup_i E\left[\sup_{D,\Xi} a_{iT}^2(\theta)\right]^{1/2}\right)$$
$$= O_p\left(T^{\delta_0-\underline{\delta}+\max(\frac{1}{2}-\underline{\delta},0)-1}\log^2 T\right)$$
$$= O_p\left(T^{-\zeta+\max(-\underline{\delta},-\frac{1}{2})}\log^2 T\right), \quad (44)$$

which is $o_p(1)$, on choosing $\underline{\delta} > 0$. From Lemma 1, $S_{\tau\tau T}(\theta) \le KT^{2\max(\frac{1}{2}-\delta,0)}\log T$. Thus verification of (29), and thence the proof, is completed by the estimate

$$\sup_{D,\Xi}\frac{S_{\tau\tau T}(\theta)}{NT}\sum_{i=1}^N \alpha_i^2 = O_p\left(T^{-2\underline{\delta}}\log T\right) = o_p(1). \quad \square \quad (45)$$

Proof of Theorem 3.2. From (20) and (22), $L_T^F(\theta) = L_T(\theta) - \left(S_{\tau\tau T}(\theta)NT\right)^{-1}\sum_{i=1}^N a_{iT}^2(\theta)$. We again check Proposition 1. From Lemma 1, $S_{\tau\tau T}(\theta) \ge \eta T^{2\max(\frac{1}{2}-\delta,0)}$ for some $\eta > 0$. Thus adapting the methods of Lemma 2,

$$\sup_{D,\Xi}\frac{1}{S_{\tau\tau T}(\theta)NT}\sum_{i=1}^N a_{iT}^2(\theta) = O_p\left(\frac{1}{T}\sup_i E\left[\sup_{D,\Xi}\frac{a_{iT}^2(\theta)}{S_{\tau\tau T}(\theta)}\right]\right)$$
$$= O_p\left(\frac{\log^4 T}{T^{1-2(\delta_0-\underline{\delta})}}\right) = O_p\left(\frac{\log^4 T}{T^{2\zeta}}\right)$$
$$= o_p(1). \quad \square$$

Proof of Theorem 3.3. From (23) and (27), $L_T^D(\theta)$ is

$$L_T(\theta) - \frac{1}{NT}\sum_{i=1}^N \varepsilon_{i0}^2 - \frac{2}{N}\sum_{i=1}^N \varepsilon_{i0}\frac{1}{T}\sum_{t=1}^T \tau_t(\theta)\lambda_t(L;\theta)v_{it}$$
$$+ \frac{S_{\tau\tau T}(\theta)-1}{NT}\sum_{i=1}^N \varepsilon_{i0}^2 = L_T(\theta) - \frac{2}{NT}\sum_{i=1}^N \varepsilon_{i0}a_{iT}(\theta)$$
$$+ \frac{S_{\tau\tau T}(\theta)-1}{NT}\sum_{i=1}^N \varepsilon_{i0}^2 + O_p\left(T^{-1}\right)$$

uniformly. Then as in (44) and (45), $\sup_{D,\Xi}\left|\sum_{i=1}^N \varepsilon_{i0}a_{iT}(\theta)\right| + \sup_{D,\Xi}\left(S_{\tau\tau T}(\theta)-1\right)\sum_{i=1}^N \varepsilon_{i0}^2 = o_p(NT)$, to check Proposition 1. $\square$

Table 14
100× Empirical bias of PML estimate FAR(1) model $(\widehat{\delta}_T^P, \widehat{\xi}_T^P)$.

ξ_0		$\widehat{\delta}_T^P$						$\widehat{\xi}_T^P$					
	δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
−0.5	T	$NT = 100$						$NT = 100$					
	5	−6.81	−19.15	−13.90	−9.28	−5.17	1.74	2.02	6.97	5.18	2.95	0.93	−1.56
	10	−4.56	−6.68	−4.53	−3.65	−2.83	−0.90	2.46	3.41	2.26	1.85	1.51	1.14
	T	$NT = 200$						$NT = 200$					
	5	−8.19	−18.67	−10.38	−6.15	−2.91	2.96	2.34	6.51	2.93	0.81	−0.76	−2.66
	10	−3.91	−4.61	−2.93	−2.21	−1.50	0.34	1.72	1.84	0.83	0.48	0.20	−0.14
	100	−1.92	−2.05	−1.85	−1.62	−1.35	−0.57	1.23	1.16	0.94	0.83	0.73	0.63
	T	$NT = 400$						$NT = 400$					
	5	−8.30	−16.99	−7.85	−4.40	−1.70	3.69	2.08	5.44	1.28	−0.44	−1.72	−3.40
	10	−2.91	−3.29	−2.05	−1.44	−0.79	0.96	1.07	0.98	0.14	−0.16	−0.41	−0.72
	100	−1.06	−1.09	−1.07	−0.89	−0.65	0.07	0.60	0.47	0.31	0.22	0.14	0.06
−0.3	T	$NT = 100$						$NT = 100$					
	5	−8.14	−21.19	−14.19	−9.27	−5.24	1.10	3.15	10.45	7.88	5.08	2.80	0.54
	10	−4.90	−8.21	−5.42	−4.35	−3.43	−1.54	2.88	5.12	3.55	2.96	2.54	2.20
	T	$NT = 200$						$NT = 200$					
	5	10.63	−22.07	−10.06	−5.33	−1.99	3.31	4.37	11.09	4.87	1.96	0.05	−1.50
	10	−4.54	−5.41	−2.91	−2.10	−1.34	0.35	2.47	2.97	1.32	0.90	0.57	0.37
	100	−2.22	−2.39	−2.03	−1.77	−1.48	−0.76	1.55	1.56	1.27	1.15	1.05	1.00
	T	$NT = 400$						$NT = 400$					
	5	11.49	−20.30	−6.48	−2.85	−0.25	4.39	4.63	9.89	2.07	−0.11	−1.50	−2.61
	10	−3.54	−3.43	−1.77	−1.12	−0.47	1.12	1.85	1.59	0.42	0.09	−0.17	−0.32
	100	−1.14	−1.10	−1.05	−0.85	−0.61	0.04	0.79	0.62	0.48	0.39	0.31	0.30
0.3	T	$NT = 100$						$NT = 100$					
	5	−3.65	−6.56	−2.56	−1.61	−1.16	−1.56	1.08	3.45	1.46	1.36	1.79	4.61
	10	0.61	−6.25	−6.04	−5.70	−5.35	−5.03	−1.60	3.67	3.87	3.86	3.89	4.61
	T	$NT = 200$						$NT = 200$					
	5	−8.90	−9.92	−4.80	−3.62	−2.82	−2.67	6.51	7.77	4.75	4.44	4.55	6.85
	10	−0.69	−6.64	−5.83	−5.33	−4.88	−4.31	0.33	5.12	4.82	4.68	4.61	5.11
	100	0.19	−6.13	−6.55	−6.24	−6.00	−5.59	−0.75	4.71	5.30	5.16	5.11	5.21
	T	$NT = 400$						$NT = 400$					
	5	−13.07	−12.52	−6.00	−4.71	−3.85	−3.69	10.94	11.13	6.69	6.25	6.28	8.55
	10	−0.56	−6.17	−5.45	−4.80	−4.22	−3.60	0.87	5.49	5.24	4.94	4.73	5.18
	100	1.54	−3.51	−4.14	−3.86	−3.62	−3.37	−1.39	2.98	3.73	3.61	3.54	3.77
0.5	T	$NT = 100$						$NT = 100$					
	5	−0.90	1.36	4.09	4.73	5.29	5.27	−0.94	−3.10	−4.01	−3.93	−3.80	−2.31
	10	3.74	0.47	0.80	1.10	1.36	1.17	−4.87	−3.27	−3.19	−3.19	−3.15	−2.24
	T	$NT = 200$						$NT = 200$					
	5	−3.56	−0.02	3.45	4.05	4.58	5.23	2.32	−0.92	−2.47	−2.32	−2.13	−1.34
	10	2.72	−0.80	−0.34	−0.09	0.20	0.41	−3.06	−0.94	−0.88	−0.81	−0.79	−0.29
	100	3.71	−2.76	−3.39	−3.22	−3.10	−2.88	−4.41	0.69	1.45	1.44	1.47	1.60
	T	$NT = 400$						$NT = 400$					
	5	−6.13	−1.85	2.32	3.17	3.82	4.80	5.46	1.58	−0.69	−0.80	−0.74	−0.34
	10	2.00	−2.06	−1.50	−1.08	−0.72	−0.23	−1.59	1.16	1.12	1.01	0.96	1.16
	100	4.62	−2.53	−3.10	−2.90	−2.67	−2.38	−4.48	1.44	2.14	2.08	2.00	2.04

Proof of Theorem 3.4. We have $L_T^P(\theta) = \widehat{\sigma}_T^2(\theta) + (|\Omega_T(\theta)|^{\frac{1}{T}} - 1)\widehat{\sigma}_T^2(\theta) = L_T^D(\theta) + \left\{\widehat{\sigma}_T^2(\theta) - L_T^D(\theta)\right\} + \left(|\Omega_T(\theta)|^{\frac{1}{T}} - 1\right)\widehat{\sigma}_T^2(\theta)$. In view of Theorem 3.3, checking Proposition 1 entails verifying that

$$\sup_{D,\Xi}\left|L_T^D(\theta) - \widehat{\sigma}_T^2(\theta)\right| = o_p(1),$$
$$\sup_{D,\Xi}\left||\Omega_T(\theta)|^{\frac{1}{T}} - 1\right| = o(1), \tag{46}$$

since this and (from Theorem 1 of Hualde and Robinson (2011)) $\sup L_T^D(\theta) = O_p(1)$ imply that $\sup \widehat{\sigma}_T^2(\theta) = O_p(1)$. From (28),

$$L_T^D(\theta) - \widehat{\sigma}^2(\theta) = \frac{1}{NT}\sum_{i=1}^{N} \mathbf{z}_{iT}'(\theta)\left\{I_T - \Omega_T^{-1}(\theta)\right\}\mathbf{z}_{iT}(\theta)$$
$$= \frac{1}{NTS_{\tau\tau T}(\theta)}\sum_{i=1}^{N}\left\{\mathbf{z}_{iT}'(\theta)\,\boldsymbol{\tau}_T(\theta)\right\}^2. \tag{47}$$

Now $\mathbf{z}_{iT}'(\theta)\,\boldsymbol{\tau}_T(\theta) = \sum_{t=1}^{T}\tau_t(\theta)\left(\lambda_t(L;\theta)v_{it} - \tau_t(\theta)\varepsilon_{i0}\right) = a_{iT}(\theta) - \varepsilon_{i0}\left(S_{\tau\tau T}(\theta) - 1\right)$. Then using Lemma 1 and proceeding as in Lemma 2,

$$\sup_i E\left[\sup_{D,\Xi}\left|\frac{\left\{\mathbf{z}_{iT}'(\theta)\,\boldsymbol{\tau}_T(\theta)\right\}^2}{S_{\tau\tau T}(\theta)}\right|\right] \le \sup_i 2E\left[\sup_{D,\Xi}\frac{a_{iT}^2(\theta)}{S_{\tau\tau T}(\theta)}\right] + 2\sigma_0^2\left(1 + \sup_{D,\Xi}S_{\tau\tau T}^{-2}(\theta)\right),$$

which is $O\left(T^{2(\delta_0-\underline{\delta})}\log^4 T + 1\right)$ so, noting that $\delta_0 - \underline{\delta} \ge 0$, (47) is $O_p\left(T^{2(\delta_0-\underline{\delta})-1}\log^4 T\right) = O_p\left(T^{-2\zeta}\log^4 T\right) = o_p(1)$ uniformly, to check the first part of (46). Finally, from (28), for K a generic finite constant,

$$|\Omega_T(\theta)|^{\frac{1}{T}} - 1 = S_{\tau\tau T}(\theta)^{\frac{1}{T}} - 1 = O\left(T^{-1}\left(S_{\tau\tau T}(\theta) - 1\right)\right)$$
$$\le KT^{2\max(\frac{1}{2}-\underline{\delta},0)-1}\log T \le KT^{-2\underline{\delta}}\log T \to 0,$$

uniformly, to check the last part of (46). □

Proof of Theorem 4.1. By the mean value theorem, $0 = (\partial/\partial\theta)\,L_T^U\left(\widehat{\theta}_T^U\right) = (\partial/\partial\theta)\,L_T^U(\theta_0) + M_T^U\left(\widehat{\theta}_T^U - \theta_0\right)$, where M_T^U is the matrix $\left(\partial^2/\partial\theta\partial\theta'\right)L_T^U(\theta)$ with each row evaluated at a mean value

Table 15
100× Empirical MSE of PML estimates FAR(1) model $(\widehat{\delta}_T^P, \widehat{\xi}_T^P)$.

ξ_0		$\widehat{\delta}_T^P$						$\widehat{\xi}_T^P$					
	δ_0:	0.3	0.6	0.9	1.0	1.1	1.4	0.3	0.6	0.9	1.0	1.1	1.4
−0.5	T	$NT = 100$						$NT = 100$					
	5	4.77	11.19	10.23	7.41	4.97	2.64	2.19	4.05	4.83	4.14	3.37	2.59
	10	3.27	3.83	2.40	2.07	1.83	1.55	1.73	2.14	1.84	1.74	1.67	1.60
	T	$NT = 200$						$NT = 200$					
	5	3.33	8.14	4.75	2.80	1.75	1.18	1.24	2.43	2.07	1.56	1.27	1.16
	10	2.00	1.76	1.03	0.89	0.80	0.70	0.94	0.97	0.80	0.77	0.75	0.73
	100	1.24	0.95	0.69	0.65	0.63	0.59	0.77	0.72	0.66	0.65	0.65	0.65
	T	$NT = 400$						$NT = 400$					
	5	2.31	5.51	2.06	1.16	0.77	0.67	0.70	1.36	0.84	0.67	0.62	0.65
	10	1.08	0.82	0.48	0.42	0.38	0.34	0.48	0.45	0.39	0.38	0.37	0.37
	100	0.60	0.43	0.32	0.30	0.29	0.27	0.37	0.34	0.32	0.32	0.31	0.31
−0.3	T	$NT = 100$						$NT = 100$					
	5	5.92	14.26	12.91	9.63	7.04	4.22	3.36	7.04	8.81	7.68	6.55	5.10
	10	4.04	5.56	3.85	3.33	2.98	2.51	2.67	3.97	3.64	3.43	3.26	3.05
	T	$NT = 200$						$NT = 200$					
	5	4.48	11.90	6.48	3.88	2.53	1.71	2.10	5.21	4.41	3.22	2.52	2.05
	10	2.67	2.66	1.39	1.19	1.05	0.92	1.58	1.87	1.38	1.30	1.22	1.18
	100	1.66	1.27	0.89	0.84	0.81	0.77	1.26	1.18	1.05	1.03	1.02	1.03
	T	$NT = 400$						$NT = 400$					
	5	3.51	8.80	2.50	1.38	0.98	0.93	1.34	3.45	1.66	1.20	1.04	1.01
	10	1.57	1.12	0.59	0.52	0.48	0.45	0.89	0.81	0.62	0.60	0.59	0.58
	100	0.82	0.55	0.40	0.38	0.36	0.35	0.63	0.55	0.50	0.49	0.49	0.49
0.3	T	$NT = 100$						$NT = 100$					
	5	8.99	10.18	8.07	7.65	7.36	5.91	7.60	8.91	8.44	8.24	8.09	6.76
	10	6.81	7.62	6.82	6.67	6.52	6.11	6.25	7.35	7.08	7.03	6.96	6.68
	T	$NT = 200$						$NT = 200$					
	5	7.23	8.05	6.03	5.61	5.28	4.37	5.87	7.14	6.37	6.14	5.92	5.24
	10	5.33	5.93	4.91	4.71	4.55	4.20	4.74	5.77	5.23	5.12	5.01	4.74
	100	3.96	4.88	4.52	4.36	4.28	4.06	3.79	4.95	4.79	4.69	4.63	4.47
	T	$NT = 400$						$NT = 400$					
	5	6.46	6.91	4.63	4.24	3.94	3.22	5.16	6.29	5.00	4.76	4.56	4.12
	10	4.03	4.43	3.60	3.35	3.15	2.83	3.54	4.42	3.90	3.71	3.55	3.31
	100	2.77	2.87	2.59	2.46	2.38	2.28	2.58	3.00	2.85	2.76	2.70	2.62
0.5	T	$NT = 100$						$NT = 100$					
	5	6.53	6.83	6.39	6.19	5.97	4.37	6.85	7.56	7.26	6.94	6.55	4.40
	10	6.74	6.05	5.62	5.53	5.42	4.76	6.98	6.40	6.06	5.96	5.84	5.04
	T	$NT = 200$						$NT = 200$					
	5	4.83	4.96	4.57	4.39	4.18	3.12	4.73	5.25	4.93	4.65	4.32	2.92
	10	5.42	4.47	4.01	3.92	3.83	3.45	5.33	4.55	4.18	4.09	3.98	3.53
	100	4.48	3.98	3.79	3.73	3.69	3.52	4.51	3.94	3.77	3.72	3.68	3.52
	T	$NT = 400$						$NT = 400$					
	5	3.57	3.57	3.22	3.04	2.83	2.04	3.37	3.63	3.29	3.05	2.76	1.77
	10	4.25	3.29	2.89	2.80	2.70	2.37	4.11	3.27	2.93	2.84	2.75	2.40
	100	3.40	2.80	2.63	2.57	2.51	2.36	3.28	2.77	2.61	2.57	2.52	2.37

between θ_0 and $\widehat{\theta}_T^U$. Now $(\partial/\partial\theta) L_T^U(\theta_0) = \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T(\frac{\partial}{\partial\theta}\lambda_t(L;\theta_0)\,y_{it})\lambda_t(L;\theta_0)\,y_{it}$, where, using (30) and (31), $\lambda_t(L;\theta_0)\,y_{it} = \varepsilon_{it}+\tau_t^0\alpha_i$, $\frac{\partial}{\partial\theta}\lambda_t(L;\theta_0)\,y_{it} = f_{it}+\dot{\tau}_t^0\alpha_i$, with $\tau_t^0 = \tau_t(\theta_0)$, $\dot{\tau}_t^0 = \dot{\tau}_t(\theta_0)$ for brevity, with similar abbreviating notation used subsequently. Thus

$$(NT)^{\frac{1}{2}}\frac{\partial}{\partial\theta}L_T^U(\theta_0) = 2w_T + \frac{2}{(NT)^{\frac{1}{2}}}\sum_{i=1}^N\alpha_i\sum_{t=1}^T\left(\dot{\tau}_t^0\varepsilon_{it}+\tau_t^0 f_{it}\right) + \frac{2}{(NT)^{\frac{1}{2}}}\sum_{i=1}^N\alpha_i^2\sum_{t=1}^T\dot{\tau}_t^0\tau_t^0. \qquad (48)$$

By Proposition 2, $w_T \to_d \mathcal{N}(0,\sigma_0^4 B(\xi_0))$. The second term on the right of (48) is $O_p\left(\left((N/T)\,E\left\|\sum_{t=1}^T\left(\dot{\tau}_t^0\varepsilon_{it}+\tau_t^0 f_{it}\right)\right\|^2\right)^{1/2}\right)$.

Now

$$E\left\|\sum_{t=1}^T\dot{\tau}_t^0\varepsilon_{it}\right\|^2 \le K\sum_{t=1}^T\left\|\dot{\tau}_t^0\right\|^2 \le K\sum_{t=1}^T\left(\sum_{j=1}^t\frac{(t-j)^{-\delta_0}}{j}\right)^2,$$

while, since $\sum_{t=1}^T\tau_t^0 f_{it} = \sum_{t=1}^T\tau_t^0\sum_{j=0}^{t-1}\chi_{t-j}^0\varepsilon_{ij} = \sum_{t=0}^{T-1}(\sum_{j=1}^{T-t}\tau_{j+t}^0\chi_j^0)\varepsilon_{it}$,

$$E\left\|\sum_{t=1}^T\tau_t^0 f_{it}\right\|^2 = \sigma_0^2\sum_{t=1}^{T-1}\left\|\sum_{j=1}^{T-t}\tau_{j+t}^0\chi_j^0\right\|^2 \le K\sum_{t=1}^{T-1}\left(\sum_{j=1}^{T-t}(j+t)^{-\delta_0}j^{-1}\right)^2.$$

Now for $\delta > 0$, $\sum_{j=1}^t j^{-1}(t-j)^{-\delta} = O\left(t^{-\delta}\log t + t^{-1}\right)$, $t > 0$, $\sum_{j=1}^{T-t} j^{-1}(t+j)^{-\delta} = O((t+1)^{-\delta}\log(t+1))$, $t \ge 0$, so the second term in (48) is

$$\left(\frac{N}{T}\right)^{\frac{1}{2}} O_p\left(\left\{\sum_{t=1}^T\left(t^{-2\delta_0}\log^2 t + t^{-2}\right)\right\}^{\frac{1}{2}} + 1\right) = \left(\frac{N}{T}\right)^{\frac{1}{2}} O_p\left(\frac{\log T}{T^{\delta_0-\frac{1}{2}}}1\left(\delta_0 < \frac{1}{2}\right) + \log^{\frac{3}{2}} T\,1\left(\delta_0 = \frac{1}{2}\right) + 1\right).$$

Likewise the final term in (48) is

$$O_p\left(\left(\frac{N}{T}\right)^{\frac{1}{2}}\sum_{t=1}^{T}\left(\frac{\log t}{t^{2\delta_0}}+\frac{1}{t^{\delta_0+1}}\right)\right)$$
$$=\left(\frac{N}{T}\right)^{\frac{1}{2}}O_p\left(T^{1-2\delta_0}\log T\,1(\delta_0<1/2)+\log^2 T\,1(\delta_0=1/2)\right.$$
$$\left.+1(\delta_0>1/2)\right).$$

Thus the last two terms in (48) are $o_p(1)$ under the stated conditions on N and T. The result follows if $\left(\partial^2/\partial\theta\partial\theta'\right)L_T^U(\theta_0)\to_p 2\sigma_0^2 B(\xi_0)$ and $M_T^U-\left(\partial^2/\partial\theta\partial\theta'\right)L_T^U(\theta_0)\to_p 0$. Because we do not have to contend with the $(NT)^{\frac{1}{2}}$ norming it is not hard to show that the α_i have negligible effect, and the second limit is established using Theorem 3.1 and the proof of Theorem 2 of Hualde and Robinson (2011), while to save space we justify the first only in the pure fractional case $\psi(L;\xi)\equiv 1$. Writing $\chi_t(L)=-\sum_{j=1}^{t}j^{-1}L^j$, $\left(\partial^2/\partial\delta^2\right)L_T^U(\delta)$ is

$$\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left\{\left(\chi_t^2(L)\,\Delta_t^{\delta}\alpha_i+\chi_t^2(L)\,\Delta_t^{\delta-\delta_0}\varepsilon_{it}\right)\left(\Delta_{t+1}^{\delta}\alpha_i+\Delta_t^{\delta-\delta_0}\varepsilon_{it}\right)\right.$$
$$+\left(\chi_t(L)\,\Delta_t^{\delta}\alpha_i+\chi_t(L)\,\Delta_t^{\delta-\delta_0}\varepsilon_{it}\right)$$
$$\left.\times\left(\chi_t(L)\,\Delta_t^{\delta}\alpha_i+\chi_t(L)\,\Delta_t^{\delta-\delta_0}\varepsilon_{it}\right)\right\},$$

and so $\left(\partial^2/\partial\delta^2\right)L_T^U(\delta_0)$ *is*

$$\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(\chi_t^2(L)\,\Delta_{t+1}^{\delta_0}\alpha_i+\chi_t(L)^2\,\varepsilon_{it}\right)\left(\Delta_t^{\delta_0}\alpha_i+\varepsilon_{it}\right)$$
$$+\left(\chi_t(L)\,\Delta_t^{\delta_0}\alpha_i+\chi_t(L)\,\varepsilon_{it}\right)\times\left(\chi_t(L)\,\Delta_t^{\delta_0}\alpha_i+\chi_t(L)\,\varepsilon_{it}\right).$$

By arguments similar to those previously used this differs by $o_p(1)$ from

$$\frac{2}{T}\sum_{t=0}^{T}\left\{E\left(\chi_t(L)\,\varepsilon_{it}\right)^2+O_p\left(\left|\left(\chi_t(L)^2\,\tau_t^0\right)\tau_t^0\right|+\left(\chi_t(L)\,\tau_t^0\right)^2\right)\right\}$$
$$=\frac{2}{T}\sum_{t=0}^{T}\left\{\sum_{j=1}^{t}\frac{\sigma_0^2}{j^2}+O_p\left(\frac{\log^2 t}{t^{2\delta_0}}\right)\right\}\to 2\sigma_0^2\frac{\pi^2}{6}.\quad\square$$

Proof of Theorem 4.2. By the mean value theorem, $0=(\partial/\partial\theta)\,L_T^F\left(\widehat{\theta_T^F}\right)=E\left\{(\partial/\partial\theta)\,L_T^F(\theta_0)\right\}+\left[(\partial/\partial\theta)\,L_T^F(\theta_0)-E\left\{(\partial/\partial\theta)\,L_T^F(\theta_0)\right\}\right]+M_{TF}\left(\widehat{\theta}_F-\theta_0\right)$, where M_{TF} has a similar interpretation to M_{TU}. Writing $\tilde{\varepsilon}_{it}(\theta)=\lambda_t(L;\theta)\left(y_{it}-\hat{\alpha}_i(\theta)\right)$, we have from (22) and (17), $a_{iT}^0=a_{iT}(\theta_0)=\sum_{t=0}^{T}\tau_t^0\varepsilon_{it}$ and thus $\tilde{\varepsilon}_{it}(\theta_0)=\varepsilon_{it}-\tau_t^0\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}/S_{\tau\tau T}^0$. Also, since $(\partial/\partial\theta)\,\tilde{\varepsilon}_{it}(\theta)=(\partial/\partial\theta)\,\lambda_t(L;\theta)\,v_{it}-(\partial/\partial\theta)\,(a_{iT}(\theta)/S_{\tau\tau T}(\theta))\,\tau_t(\theta)-(a_{iT}(\theta)/S_{\tau\tau T}(\theta))\dot{\tau}_t(\theta)$, we have, using the orthogonality, across $t=0,1,\ldots,T$, of $\tilde{\varepsilon}_{it}(\theta)$ to $\tau_t(\theta)$,

$$(NT)^{\frac{1}{2}}\frac{\partial}{\partial\theta}L_T^F(\theta_0)=2\,(NT)^{-\frac{1}{2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(f_{it}-\frac{\dot{\tau}_t^0}{S_{\tau\tau T}^0}\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)$$
$$\times\left(\varepsilon_{it}-\frac{\tau_t^0}{S_{\tau\tau T}^0}\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)\tag{49}$$
$$=2w_T-\frac{2\,(NT)^{-\frac{1}{2}}}{S_{\tau\tau T}^0}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tau}_t^0\varepsilon_{it}\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}$$
$$-\frac{2\,(NT)^{-\frac{1}{2}}}{S_{\tau\tau T}^0}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t^0 f_{it}\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}$$
$$+\frac{2\,(NT)^{-\frac{1}{2}}}{S_{\tau\tau T}^{02}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t^0\dot{\tau}_t^0\left(\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)^2.\tag{50}$$

The second term in (50) has mean and covariance matrix respectively

$$-E\left(\frac{2\,(NT)^{-\frac{1}{2}}}{S_{\tau\tau T}^0}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tau}_t^0\varepsilon_{it}\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)=-\left(\frac{N}{T}\right)^{\frac{1}{2}}2\sigma_0^2\frac{S_{\tau\dot{\tau}T}^0}{S_{\tau\tau T}^0},$$

$$\frac{4}{NT}S_{\tau\tau T}^{0-2}\sum_{t=1}^{T}\sum_{u=1}^{T}\dot{\tau}_t^0\dot{\tau}_u^{0\prime}\sum_{r=0}^{T}\sum_{s=0}^{T}\tau_r^0\tau_s^0\sum_{i=1}^{N}E\left(\varepsilon_{ir}\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\right.$$
$$\left.-E\left(\varepsilon_{is}\varepsilon_{it}\right)E\left(\varepsilon_{ir}\varepsilon_{iu}\right)\right)$$
$$=\frac{4\sigma_0^4}{T}S_{\tau\tau T}^{0-2}\sum_{t=1}^{T}\dot{\tau}_t^0\dot{\tau}_t^{0\prime}\sum_{s=0}^{T}\tau_s^{02}$$
$$+\frac{4\sigma_0^4}{T}S_{\tau\tau T}^{0-2}\left(\sum_{t=1}^{T}\dot{\tau}_t^0\tau_t^0\right)\left(\sum_{t=1}^{T}\dot{\tau}_t^0\tau_t^0\right)'+O\left(T^{-1}\right)$$
$$=\frac{4\sigma_0^4}{T}S_{\tau\tau T}^{0-1}S_{\dot{\tau}\dot{\tau}T}^0+\frac{4\sigma_0^4}{T}S_{\tau\tau T}^{0-2}S_{\tau\dot{\tau}T}^0S_{\tau\dot{\tau}T}^{0\prime}+O\left(T^{-1}\right)=o(1)$$

as $T\to\infty$, where $S_{\dot{\tau}\dot{\tau}T}^0=\sum_{t=1}^{T}\dot{\tau}_t^0\dot{\tau}_t^{0\prime}$. Since $E\left(\varepsilon_{is}f_{it}\right)=E\left(\varepsilon_{is}\sum_{j=0}^{t-1}\chi_{t-j}^0\varepsilon_{ij}\right)=\sigma_0^2\chi_{t-s}^0 1\,(0\le s<t)$ for $t\ge 2$, the next term in (50) has mean

$$-\left(\frac{N}{T}\right)^{\frac{1}{2}}\frac{2}{S_{\tau\tau T}^0}E\left(\sum_{t=1}^{T}\tau_t^0 f_{it}\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)=-\left(\frac{N}{T}\right)^{\frac{1}{2}}\frac{2\sigma_0^2 S_{\tau\dot{\tau}T}^0}{S_{\tau\tau T}^0},$$

since, from Lemma 3 in Appendix C, the expectation is $\sigma_0^2\sum_{t=1}^{T}\tau_t^0\sum_{s=0}^{t-1}\tau_s^0\chi_{t-s}^0=\sigma_0^2\sum_{t=1}^{T}\tau_t^0\dot{\tau}_t^0$. The norm of its covariance matrix is dominated by the term in the top left hand corner,

$$\frac{4}{NT}S_{\tau\tau T}^{0-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\sum_{u=0}^{T}\sum_{r=0}^{T}\sum_{k=1}^{u}j^{-1}k^{-1}\tau_r^0\tau_s^0\tau_t^0\tau_u^0$$
$$\times\sum_{i=1}^{N}\left\{E\left(\varepsilon_{i,t-j}\varepsilon_{is}\varepsilon_{i,u-k}\varepsilon_{ir}\right)-E\left(\varepsilon_{i,t-j}\varepsilon_{is}\right)E\left(\varepsilon_{i,u-k}\varepsilon_{ir}\right)\right\}$$
$$=\frac{4\sigma_0^4}{T}S_{\tau\tau T}^{0-2}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{u=0}^{T}j^{-1}|j+t-u|^{-1}\tau_t^0\tau_u^0$$
$$+\frac{4\sigma_0^4}{T}S_{\tau\tau T}^{0-2}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{u=1}^{T}\sum_{k=1}^{u}j^{-1}k^{-1}\tau_{t-j}^0\tau_t^0\tau_{u-k}^0\tau_u^0+O\left(T^{-1}\right),$$

which is $O\left(T^{-\delta_0}\log T+T^{-1}\log^2 T\right)=o(1)$. The last term in (50) has expectation

$$\frac{2\,(NT)^{-\frac{1}{2}}}{S_{\tau\tau T}^{02}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t^0\dot{\tau}_t^0 E\left(\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)^2$$
$$=\left(\frac{N}{T}\right)^{\frac{1}{2}}\frac{2\sigma_0^2}{S_{\tau\tau T}^{02}}\left(\sum_{t=1}^{T}\tau_t^0\dot{\tau}_t^0\right)\left(\sum_{t=0}^{T}\tau_t^{02}\right)=\left(\frac{N}{T}\right)^{\frac{1}{2}}2\sigma_0^2\frac{S_{\tau\dot{\tau}T}^0}{S_{\tau\tau T}^0},$$

and its covariance matrix is

$$\frac{4S_{\tau\tau T}^{0-4}}{NT}\sum_{i=1}^{N}\left[E\left(\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)^4-\left\{E\left(\sum_{s=0}^{T}\tau_s^0\varepsilon_{is}\right)^2\right\}^2\right]$$
$$\times\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_t^0\dot{\tau}_t^0\tau_s^0\dot{\tau}_s^{0\prime}=o(1),$$

as $T \to \infty$, since, as is readily shown, the fourth moment is $O\left(S^{02}_{\tau\tau T}\right)$. Overall, we deduce that $(NT)^{\frac{1}{2}} \frac{\partial}{\partial\theta} L^F_T(\theta_0) + 2\sigma_0^2 (N/T)^{\frac{1}{2}} S^0_{\tau\dot{\tau}T}/S^0_{\tau\tau T} \to_d \mathcal{N}(0, \sigma_0^4 B(\xi_0))$. Using similar techniques as before, the probability limit of the second derivative term is $2\sigma_0^2 B(\xi_0)$, and the result follows. □

Proof of Theorem 4.3. We start with an analogous development as before. We have $z_{it}(\theta_0) = \varepsilon_{it} - \tau_t^0 \varepsilon_{i0}$, and $(\partial/\partial\theta) z_{it}(\theta_0) = f_{it} - \dot{\tau}_t^0 \varepsilon_{i0}$. Thus $(NT)^{\frac{1}{2}} (\partial/\partial\theta) L^D_T(\theta_0)$ is

$$\frac{2}{(NT)^{\frac{1}{2}}} \sum_{i=1}^N \sum_{t=1}^T \left(\frac{\partial}{\partial\theta} z_{it}(\theta_0)\right) z_{it}(\theta_0)$$
$$= \frac{2}{(NT)^{\frac{1}{2}}} \sum_{i=1}^N \sum_{t=1}^T \left(f_{it} - \dot{\tau}_t^0 \varepsilon_{i0}\right)\left(\varepsilon_{it} - \tau_t^0 \varepsilon_{i0}\right). \quad (51)$$

Expanding (51),

$$(NT)^{\frac{1}{2}} \frac{\partial}{\partial\theta} L^D_T(\theta_0) = 2w_T - \frac{2}{(NT)^{\frac{1}{2}}} \sum_{i=1}^N \varepsilon_{i0} \sum_{t=1}^T \tau_t^0 f_{it}$$
$$- \frac{2}{(NT)^{\frac{1}{2}}} \sum_{i=1}^N \varepsilon_{i0} \sum_{t=1}^T \dot{\tau}_t^0 \varepsilon_{it} + \frac{2}{(NT)^{\frac{1}{2}}} \sum_{i=1}^N \varepsilon_{i0}^2 \sum_{t=1}^T \dot{\tau}_t^0 \tau_t^0$$

reveals the same asymptotically $\mathcal{N}(0, \sigma_0^4 B(\xi_0))$ term w_T, while, noting that $f_{it} = \sum_{j=1}^{t-1} \chi_j \varepsilon_{i,t-j} + \chi_t \varepsilon_{i0}$ and employing similar arguments to before and $\delta_0 > \frac{1}{4}$ it is readily seen that the remaining part of (51) differs by $o_p(1)$ from its expectation, which is $2(N/T)^{\frac{1}{2}} \sigma_0^2 \left(S_{\tau\dot{\tau}T}(\theta_0) - \sum_{t=1}^T \tau_t^0 \chi_t\right) = 2(N/T)^{\frac{1}{2}} \sigma_0^2 \left(S_{\tau\dot{\tau}T}(\theta_0) - S_{\tau\chi T}(\theta_0)\right)$, since the second term has variance matrix

$$O\left(T^{-1} \sum_{t=1}^T \sum_{s=1}^T \tau_t^0 \tau_s^0 \sum_{j=1}^{t\wedge s} \left\|\chi_j \chi_{|t-s|+j}\right\| + T^{-1}\left(\sum_{t=1}^T \left\|\tau_t^0 \chi_t\right\|\right)^2\right)$$
$$= o(1),$$

that of the third one is $O\left(T^{-1} \sum_{t=1}^T \left\|\dot{\tau}_t^0 \dot{\tau}_t^{0\prime}\right\|\right) = o(1)$, while that of the last term is $4\sigma_0^2 T^{-1} \left(\sum_{t=1}^T \dot{\tau}_t^0 \tau_t^0\right)\left(\sum_{t=1}^T \dot{\tau}_t^0 \tau_t^0\right)' = 4\sigma_0^2 T^{-1} S_{\tau\dot{\tau}T}(\theta_0) S_{\tau\dot{\tau}T}(\theta_0)'$, whose norm is

$$O\left(T^{-1}\left(T^{1-2\delta_0}\log T + T^{-\delta_0} + 1 + 1\left(\delta_0 = \frac{1}{2}\right)\log T\right)^2\right)$$
$$= O\left(T^{1-4\delta_0}\log^2 T\right) + o(1) = o(1)$$

because $\delta_0 > \frac{1}{4}$. The probability limit of the second derivative term is obtained much as before. □

Proof of Theorem 4.4. We have $(\partial/\partial\theta) L^P_T(\theta_0) = \widehat{\sigma}^2_T(\theta_0)(\partial/\partial\theta)|\Omega_T(\theta_0)|^{\frac{1}{T}} + |\Omega_T(\theta_0)|^{\frac{1}{T}} (\partial/\partial\theta)\widehat{\sigma}^2_T(\theta_0)$. Thus, denoting $\Omega^j_T(\theta) = (\partial/\partial\theta_j)\Omega_T(\theta)$, where θ_j is the jth element of θ,

$$\frac{\partial}{\partial\theta_j} L^P_T(\theta_0) = |\Omega_T(\theta_0)|^{\frac{1}{T}} \left\{\frac{\widehat{\sigma}^2_T(\theta_0)}{T} trace\left(\Omega_T^{-1}(\theta_0)\Omega^j_T(\theta_0)\right)\right.$$
$$\left. - \frac{1}{NT}\sum_{i=1}^N \mathbf{z}'_{iT}(\theta_0)\Omega_T^{-1}(\theta_0)\Omega^j_T(\theta_0)\Omega_T^{-1}(\theta_0)\mathbf{z}_{iT}(\theta_0)\right\}$$
$$+ \frac{2}{NT}|\Omega_T(\theta_0)|^{\frac{1}{T}} \sum_{i=1}^N \dot{\mathbf{z}}^{j\prime}_{iT}(\theta_0)\Omega_T^{-1}(\theta_0)\mathbf{z}_{iT}(\theta_0),$$

where $\dot{\mathbf{z}}^j_{iT}(\theta)$ is the jth row of $\dot{\mathbf{z}}'_{iT}(\theta_0) = (\partial/\partial\theta)\mathbf{z}_{iT}(\theta)$. The term in braces has expectation zero because $\widehat{\sigma}^2_T(\theta_0)$ equals $(NT)^{-1}$ times

$$\sum_{i=1}^N \sum_{t=1}^T z_{it}^2(\theta_0) - \frac{1}{S^0_{\tau\tau T}} \sum_{i=1}^N \left(\sum_{t=1}^T \tau_t^0 z_{it}(\theta_0)\right)^2$$
$$= \sum_{i=1}^N \sum_{t=1}^T \left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)^2 - \frac{1}{S^0_{\tau\tau T}} \sum_{i=1}^N \left(\sum_{t=1}^T \tau_t^0\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\right)^2,$$

so $\widehat{\sigma}^2_T(\theta_0)$ has expectation

$$\sigma_0^2\left(1 + \frac{S^0_{\tau\tau T} - 1}{T}\right) - \frac{\sigma_0^2}{TS^0_{\tau\tau T}}\left(\left(S^0_{\tau\tau T} - 1\right) + \left(S^0_{\tau\tau T} - 1\right)^2\right) = \sigma_0^2,$$

and because

$$E\left(\frac{1}{NT}\sum_{i=1}^N \mathbf{z}'_{iT}(\theta_0)\Omega_T^{-1}(\theta_0)\Omega^j_T(\theta_0)\Omega_T^{-1}(\theta_0)\mathbf{z}_{iT}(\theta_0)\right)$$
$$= \frac{\sigma_0^2}{T} trace\left(\Omega_T^{-1}(\theta_0)\Omega^j_T(\theta_0)\right).$$

On the other hand,

$$\frac{2}{NT}\sum_{i=1}^N \dot{\mathbf{z}}'_{iT}(\theta_0)\Omega_T^{-1}(\theta_0)\mathbf{z}_{iT}(\theta_0) = \frac{2}{NT}\sum_{i=1}^N \left(\sum_{t=1}^T \dot{z}_{it}(\theta_0) z_{it}(\theta_0)\right.$$
$$\left. - \frac{1}{S^0_{\tau\tau T}}\left(\sum_{t=1}^T \dot{z}_{it}(\theta_0)\tau_t^0\right)\left(\sum_{t=1}^T z_{it}(\theta_0)\tau_t^0\right)\right).$$

Now

$$\frac{2}{NT}\sum_{i=1}^N \sum_{t=1}^T \dot{z}_{it}(\theta_0) z_{it}(\theta_0) = \frac{2}{NT}\sum_{i=1}^N \sum_{t=1}^T \left(f_{it} - \dot{\tau}_t^0\varepsilon_{i0}\right)\left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)$$

has expectation

$$\frac{2}{NT}\sum_{i=1}^N \sum_{t=1}^T \left(\dot{\tau}_t^0\tau_t^0 E\varepsilon_{i0}^2 - \tau_t^0 E\left(f_{it}\varepsilon_{i0}\right)\right) = \frac{2\sigma_0^2}{T}\left(S^0_{\tau\dot{\tau}T} - \sum_{t=1}^T \tau_t^0\chi_t^0\right)$$
$$= \frac{2\sigma_0^2}{T}\left(S^0_{\tau\dot{\tau}T} - S^0_{\tau\chi T}\right),$$

while

$$\frac{2}{NTS^0_{\tau\tau T}}\sum_{i=1}^N \left(\sum_{t=1}^T \dot{z}_{it}(\theta_0)\tau_t^0\right)\left(\sum_{t=1}^T z_{it}(\theta_0)\tau_t^0\right)$$
$$= \frac{2}{NTS^0_{\tau\tau T}}\sum_{i=1}^N \sum_{t=1}^T \left(f_{it} - \dot{\tau}_t^0\varepsilon_{i0}\right)\tau_t^0 \sum_{t=1}^T \left(\varepsilon_{it} - \tau_t^0\varepsilon_{i0}\right)\tau_t^0$$

has expectation

$$\frac{2}{TS^0_{\tau\tau T}}\left(E\left(\sum_{t=1}^T f_{it}\tau_t^0 \sum_{s=1}^T \tau_s^0\varepsilon_{is}\right) - \left(S^0_{\tau\tau T} - 1\right)\sum_{t=1}^T \tau_t^0 E\left(f_{it}\varepsilon_{i0}\right)\right.$$
$$\left. + \sigma_0^2 S^0_{\tau\dot{\tau}T}\left(S^0_{\tau\tau T} - 1\right)\right)$$
$$= \frac{2}{TS^0_{\tau\tau T}}\left(\sum_{t=1}^T \tau_t^0 E\left(\sum_{j=1}^t \chi_j^0\varepsilon_{i,t-j}\sum_{s=1}^T \varepsilon_{is}\tau_s^0\right)\right.$$
$$\left. - \sigma_0^2\left(S^0_{\tau\tau T} - 1\right)\sum_{t=1}^T \tau_t^0\chi_t^0 + \sigma_0^2 S^0_{\tau\dot{\tau}T}\left(S^0_{\tau\tau T} - 1\right)\right)$$
$$= \frac{2}{TS^0_{\tau\tau T}} E\left(\sum_{t=0}^{T-1}\left(\sum_{j=1}^{T-t}\tau_{t+j}^0\chi_j^0\right)\varepsilon_{it}\sum_{s=1}^T \varepsilon_{is}\tau_s^0 - \sigma_0^2\left(S^0_{\tau\tau T} - 1\right)S^0_{\tau\chi T}\right.$$

$$\left.+\sigma_0^2 S^0_{\tau\dot{\tau}T}\left(S^0_{\tau\tau T}-1\right)\right)$$
$$=\frac{2\sigma_0^2}{TS^0_{\tau\tau T}}\left(\sum_{t=1}^{T-1}\tau_t^0\sum_{j=1}^{T-t}\tau^0_{t+j}\chi^0_j-\left(S^0_{\tau\tau T}-1\right)S^0_{\tau\chi T}\right.$$
$$\left.+S^0_{\tau\dot{\tau}T}\left(S^0_{\tau\tau T}-1\right)\right)$$
$$=\frac{2\sigma_0^2}{TS^0_{\tau\tau T}}\left(S^0_{\tau\dot{\tau}T}-S^0_{\tau\chi T}-\left(S^0_{\tau\tau T}-1\right)S^0_{\tau\chi T}+S^0_{\tau\dot{\tau}T}\left(S^0_{\tau\tau T}-1\right)\right)$$
$$=\frac{2\sigma_0^2}{T}\left(S^0_{\tau\dot{\tau}T}-S^0_{\tau\chi T}\right)$$

because

$$\sum_{t=1}^{T-1}\tau_t^0\sum_{j=1}^{T-t}\tau^0_{t+j}\chi^0_j=\sum_{t=2}^{T}\tau_t^0\sum_{j=1}^{t-1}\tau^0_{t-j}\chi^0_j=\sum_{t=2}^{T}\tau_t^0\left(\sum_{j=1}^{t}\tau^0_{t-j}\chi^0_j-\tau^0_0\chi^0_t\right)=\sum_{t=1}^{T}\tau_t^0\dot{\tau}_t^0-\sum_{t=1}^{T}\tau_t^0\chi_t^0=S^0_{\tau\dot{\tau}T}-S^0_{\tau\chi T},$$

from Lemma 3 in Appendix C. It follows that $E\left(\partial/\partial\theta\right)L_T^P\left(\theta_0\right)=0$. By similar means to those used before it may be shown that $(NT)^{1/2}\left\{\left(\partial/\partial\theta\right)L_T^P\left(\theta_0\right)-E\left(\partial/\partial\theta\right)L_T^P\left(\theta_0\right)\right\}=2w_T+o_p(1)\to_d\mathcal{N}(0,4\sigma_0^2 B\left(\xi_0\right))$. We again omit the details of the convergence of the second derivative term. □

Proof of Theorem 5.1. We have $(NT)^{\frac12}\left(\widetilde{\theta}_T^F-\theta_0\right)=(NT)^{\frac12}\left(\widehat{\theta}_T^F-\theta_0-T^{-1}b_T^F\left(\theta_0\right)\right)-(N/T)^{\frac12}\left(b_T^F\left(\widehat{\theta}_T^F\right)-b_T^F\left(\theta_0\right)\right)$. It suffices to show that the second term on the right is $o_p(1)$. By the mean value theorem, $b_T^F\left(\widehat{\theta}_T^F\right)-b_T^F\left(\theta_0\right)=\left(\partial/\partial\theta'\right)b_T^F\left(\theta^*\right)\left(\widehat{\theta}_T^F-\theta_0\right)$. Looking just at the case $\psi\left(L;\xi\right)\equiv1$ for brevity,

$$\frac{\partial}{\partial\theta'}b_T^F\left(\theta\right)=-\frac{6}{\pi^2}\left(S_{\tau\tau T}(\theta)\frac{\partial}{\partial\delta}S_{\tau\dot{\tau}T}(\theta)-S_{\tau\dot{\tau}T}(\theta)\frac{\partial}{\partial\delta}S_{\tau\tau T}(\theta)\right)S^{-2}_{\tau\tau T}(\theta)$$
$$=O\left(\left|\frac{\partial}{\partial\delta}S_{\tau\dot{\tau}T}(\theta)\right|S^{-1}_{\tau\tau T}(\theta)+\left(S_{\tau\dot{\tau}T}(\theta)\right)^2S^{-2}_{\tau\tau T}(\theta)\right)$$
$$=O\left(\log^2T1\left(\delta\le\frac12\right)+1\left(\delta>\frac12\right)\right)$$

from Lemma 1, with the same orders holding more generally. Since $\widehat{\theta}_T^F-\theta_0=O_p(\left\|b_T^F\left(\theta_0\right)\right\|/T+(NT)^{-\frac12})$, where $b_T^F\left(\theta\right)=O(\log T1(\delta\le\frac12)+1(\delta>\frac12))$,

$$(N/T)^{\frac12}\left(b_T^F\left(\widehat{\theta}_T^F\right)-b_T^F\left(\theta_0\right)\right)=O_p\left((N/T)^{\frac12}\log^2T(\log T/T+(NT)^{-\frac12})\right)=O_p\left(N^{\frac12}T^{-\frac32}\log^3T+\log^2T/T\right),\quad\delta_0\le\frac12,$$
$$(N/T)^{\frac12}\left(b_T^F\left(\widehat{\theta}_T^F\right)-b_T^F\left(\theta_0\right)\right)=O_p\left((N/T)^{\frac12}(T^{-1}+(NT)^{-\frac12})\right)=O_p\left(N^{\frac12}T^{-\frac32}+T^{-1}\right),\quad\delta_0>\frac12,$$

and these are $o(1)$ under the stated conditions. □

Proof of Theorem 5.2. As in the previous proof, $(NT)^{\frac12}\left(\widetilde{\theta}_T^D-\theta_0\right)=(NT)^{\frac12}\left(\widehat{\theta}_T^D-\theta_0-T^{-1}b_T^D\left(\theta_0\right)\right)-(N/T)^{\frac12}\left(b_T^D\left(\widehat{\theta}_T^D\right)-b_T^D\left(\theta_0\right)\right)$, with $b_T^D\left(\widehat{\theta}_T^D\right)-b_T^D\left(\theta_0\right)=\left(\partial/\partial\theta'\right)b_T^D\left(\theta^*\right)\left(\widehat{\theta}_T^D-\theta_0\right)$ and

$$\frac{\partial}{\partial\theta'}b_T^D\left(\theta\right)=O\left(\left\|\frac{\partial}{\partial\theta}S_{\tau\dot{\tau}T}(\theta)\right\|+\left\|\frac{\partial}{\partial\theta}S_{\tau\chi T}(\theta)\right\|\right)=O\left(\frac{\log^2T}{T^{2\delta}-1}1\left(\delta<\frac12\right)+\log^3T1\left(\delta=\frac12\right)+1\left(\delta>\frac12\right)\right),$$

applying Lemma 1 again. Since $\widehat{\theta}_T^D-\theta_0=O_p(\left\|b_T^D\left(\theta_0\right)\right\|/T+(NT)^{-\frac12})$, where $b_T^D\left(\theta\right)=O(T^{1-2\delta}\log T1(\delta<\frac12)+\log^2T1(\delta=\frac12)+1(\delta>\frac12))$,

$$(N/T)^{\frac12}\left(b_T^D\left(\widehat{\theta}_T^D\right)-b_T^D\left(\theta_0\right)\right)=O_p\left((N/T)^{\frac12}T^{1-2\delta_0}\log^2T\left(T^{-2\delta_0}\log T+(NT)^{-\frac12}\right)\right)=O_p\left(N^{\frac12}T^{\frac12-4\delta_0}\log^3T+T^{-2\delta_0}\log^2T\right),\quad\delta_0<\frac12,$$
$$\left(\frac{N}{T}\right)^{\frac12}\left(b_T^D\left(\widehat{\theta}_T^D\right)-b_T^D\left(\theta_0\right)\right)=O_p\left(\left(\frac{N}{T}\right)^{\frac12}\log^3T\left(\frac{\log^2T}{T}+\left(\frac{1}{NT}\right)^{1/2}\right)\right)=O_p\left(N^{\frac12}\frac{\log^5T}{T^{\frac32}}+\frac{\log^3T}{T}\right),\quad\delta_0=\frac12,$$
$$(N/T)^{\frac12}\left(b_T^D\left(\widehat{\theta}_T^D\right)-b_T^D\left(\theta_0\right)\right)=O_p\left((N/T)^{\frac12}\left(T^{-1}+(NT)^{-1/2}\right)\right)=O_p\left(N^{\frac12}T^{-\frac32}+T^{-1}\right),\quad\delta_0>\frac12,$$

which are $o(1)$ under the stated conditions. □

Appendix B. Proofs of propositions

Proof of Proposition 1. Define $\nu\left(L;\theta\right)=\lambda\left(L;\theta\right)\lambda^{-1}\left(L;\theta_0\right)=\Delta^{\delta-\delta_0}\psi\left(L;\theta\right)\phi\left(L;\theta_0\right)=\sum_{j=0}^{\infty}\nu_j\left(\theta\right)L^j$, and note that $\lambda_t\left(L;\theta\right)v_{it}=\sum_{j=0}^{t}\nu_j\left(\theta\right)v_{i,t-j}$ and

$$\sup_{\Theta}\left|\nu_j\left(\theta\right)\right|\le K\sup_{D}j^{\delta_0-\delta-1}\le Kj^{\delta_0-\underline{\delta}-1}\le Kj^{-1/2-\zeta}\tag{52}$$

for some $\zeta>0$. For $\eta>0$ let $N_\eta=\{\theta:\|\theta-\theta_0\|\le\eta\}$, $\overline{N}_\eta=\left\{\theta:\theta\notin N_\eta,\theta\in\Theta\right\}$. Writing $\ell_T\left(\theta\right)=L_T\left(\theta\right)-L_T\left(\theta_0\right)$, $P(\widetilde{\theta}\in\overline{N}_\eta)\le P\left(\inf_{\overline{N}_\eta}\ell_T\left(\theta\right)\le0\right)\le P\left(\sup_{\Theta}\left|V_T\left(\theta\right)\right|\ge\inf_{\overline{N}_\eta}U\left(\theta\right)\right)$, where $V_T\left(\theta\right)=U\left(\theta\right)+L_T\left(\theta_0\right)-L_T\left(\theta\right)$ with

$$U\left(\theta\right)=\sigma_0^2\sum_{j=1}^{\infty}\nu_j^2\left(\theta\right)=\sigma_0^2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\nu\left(e^{i\lambda};\theta\right)\right|^2d\lambda-1\right)=\sigma_0^2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|1-e^{i\lambda}\right|^{2(\delta-\delta_0)}\left|\frac{\psi\left(e^{i\lambda};\xi\right)}{\psi\left(e^{i\lambda};\xi_0\right)}\right|^2d\lambda-1\right).$$

Because $U\left(\theta\right)$ is continuous, vanishes if and only if $\theta=\theta_0$, and, from Hualde and Robinson (2011), is otherwise positive, $\inf_{\overline{N}_\eta}U\left(\theta\right)>0$. It thus remains to show that $\sup_{\Theta}\left|V_T\left(\theta\right)\right|\to_p0$. We have $V_T\left(\theta\right)=U\left(\theta\right)+A_T(\theta_0)-A_T(\theta)+B_T(\theta_0)-B_T(\theta)$, so in view of (29) it suffices to show that $\sup_{\Theta}\left|A_T(\theta)-A_T(\theta_0)-U\left(\theta\right)\right|\to_p0$. Now $A_T(\theta)-A_T(\theta_0)-U\left(\theta\right)$ is

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\nu_j^2(\theta)\sum_{t=0}^{T-j}\left(\varepsilon_{it}^2-\sigma_0^2\right)+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{k=0}^{j-1}\nu_j(\theta)\nu_k(\theta)\varepsilon_{i,t-j}\varepsilon_{i,t-k}-\frac{\sigma_0^2}{T}\sum_{j=1}^{T}\left(j-1\right)\nu_j^2(\theta)-\sigma_0^2\sum_{j=T+1}^{\infty}\nu_j(\theta),$$

where

$$\sup_{\Theta}\left|\sum_{i=1}^{N}\sum_{j=1}^{T}\nu_j^2(\theta)\sum_{t=0}^{T-j}\left(\varepsilon_{it}^2-\sigma_0^2\right)\right| \le \sum_{i=1}^{N}\sup_{\Theta}\left|\sum_{j=1}^{T}\nu_j^2(\theta)\sum_{t=0}^{T-j}\left(\varepsilon_{it}^2-\sigma_0^2\right)\right|, \quad (53)$$

$$\sup_{\Theta}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{k=0}^{j-1}\nu_j(\theta)\nu_k(\theta)\varepsilon_{i,t-j}\varepsilon_{i,t-k}\right| \le \sum_{i=1}^{N}\sup_{\Theta}\left|\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{k=0}^{j-1}\nu_j(\theta)\nu_k(\theta)\varepsilon_{i,t-j}\varepsilon_{i,t-k}\right|. \quad (54)$$

It follows from the proof of Theorem 1 of Hualde and Robinson (2011), constancy of the $\nu_j(\theta)$ across i, and identity of distribution of ε_{it} across i, that the i-summands in the right sides of (53) and (54) are $o_p(1)$ uniformly in i, whence (53) + (54) = $o_p(NT)$. Finally $\sup_\Theta T^{-1}\sum_{j=0}^{T}(j-1)\nu_j^2(\theta)+\sup_\Theta\sum_{j=T+1}^{\infty}\nu_j^2(\theta)\to 0$ from (52). □

Proof of Proposition 2. The left side of (32) can be written $w_T = T^{-\frac{1}{2}}\sum_{t=0}^{T}r_{tT}$, where $r_{tT}=N^{-\frac{1}{2}}\sum_{i=1}^{N}\varepsilon_{it}\sum_{j=1}^{t}\chi_j^0\varepsilon_{i,t-j}$, and our notation stresses the possibility that N increases with T. Denoting by F_{t-1} the σ-field of events generated by $\{\varepsilon_{is}, i\ge 1, s<t\}$, $\varkappa'\left\{E\left(\sum_{t=0}^{T}r_{tT}\right)\left(\sum_{t=0}^{T}r_{tT}\right)\right\}^{-\frac{1}{2}}\sum_{t=0}^{T}r_{tT}\to_d \mathcal{N}(0,1)$ for any vector $\varkappa$ such that $\|\varkappa\|=1$ since (see e.g. Brown (1971)), $E\left(r_{tT}\,|F_{t-1}\right)=0,\ t\ge 1$ (by serial independence of the ε_{it}), $E\|r_{tT}\|^4=O\Big(N^{-2}\sum_{h=1}^{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{t}j^{-2}\right)^2\Big)\le K$ (since $\|r_{tT}\|$ is dominated by the contribution from the first element of χ_j^0) and $T^{-1}\sum_{t=0}^{T}\Big\{E\left(\|\varkappa' r_{tT}\|^2|F_{t-1}\right)-E\left(\|\varkappa' r_{tT}\|^2\right)\Big\}\to_p 0$, since its left side has mean zero and variance

$$\frac{1}{T^2}\sum_{t=0}^{T}\left(E\left\{E\left(\|\varkappa' r_{tT}\|^2|F_{t-1}\right)^2\right\}-\left\{E\left(\|\varkappa' r_{tT}\|^2\right)\right\}^2\right) = O((NT)^{-1}),$$

because, looking just at the case $\varkappa$ is null apart from its first element, which is 1, $E\left\{E\left(\|\varkappa' r_{tT}\|^2|F_{t-1}\right)^2\right\}$ is

$$\sigma_0^4\frac{1}{N^2}\sum_{h=1}^{N}\sum_{i=1}^{N}E\left\{\left(\sum_{j=1}^{t}\frac{\varepsilon_{h,t-j}}{j}\right)^2\left(\sum_{k=1}^{t}\frac{\varepsilon_{i,t-k}}{k}\right)^2\right\} = \sigma_0^4\left\{\sum_{j=1}^{t}j^{-2}\right\}^2+O\left(\frac{1}{N^2}\sum_{i=1}^{N}\left\{\left(\sum_{j=1}^{t}\frac{1}{j^2}\right)^2+\sum_{j=1}^{t}\frac{1}{j^4}\right\}\right) = \left\{E\left(\|\varkappa' r_{tT}\|^2\right)\right\}^2+O(N^{-1}).\quad □$$

Appendix C. Technical lemmas

Lemma 1. *As $T\to\infty$,*

$$S_{\tau\tau T}(\theta) = \left\{\frac{\psi(1;\xi)^2T^{1-2\delta}}{(1-2\delta)\Gamma(1-\delta)^2}+O(1)\right\}1\left(\delta<\frac{1}{2}\right)+\left\{\frac{\psi(1;\xi)^2}{\pi}\log T+O(1)\right\}1\left(\delta=\frac{1}{2}\right)+\left\{\sum_{j=0}^{\infty}\tau_j^2(\theta)+O(T^{1-2\delta})\right\}1\left(\delta>\frac{1}{2}\right);$$

$$S_{\tau\dot\tau T}(\theta) = \begin{bmatrix}-\psi(1;\xi)\log T\\ \dot\psi(1;\xi)\end{bmatrix}\times\left[\left\{\frac{\psi(1;\xi)T^{1-2\delta}}{(1-2\delta)\Gamma(1-\delta)^2}+O(T^{1-2\delta})\right\}1\left(\delta<\frac{1}{2}\right)+\left\{\frac{\psi(1;\xi)}{\pi}\log T+O(1)\right\}1\left(\delta=\frac{1}{2}\right)\right]+\left\{\sum_{j=1}^{\infty}\tau_j(\theta)\dot\tau_t(\theta)+O(T^{1-2\delta}\log T)\right\}1\left(\delta>\frac{1}{2}\right);$$

$$S_{\tau mT}(\theta)=\left\{\sum_{j=1}^{\infty}\frac{\tau_j(\theta)}{j}+O(T^{-\delta})\right\}1(\delta>0);$$

$$\frac{\partial}{\partial\theta}S_{\tau\dot\tau T}(\theta) = O\left(T^{1-2\delta}\log^2 T1\left(\delta<\frac{1}{2}\right)+\log^3 T1\left(\delta=\frac{1}{2}\right)+1\left(\delta>\frac{1}{2}\right)\right);$$

$$\frac{\partial}{\partial\theta}S_{\tau\chi T}(\theta)=O(1)1(\delta>0).$$

Proof of Lemma 1. For $\delta\le\frac{1}{2}$, $\left|\sum_{j=1}^{T}t^{-2\delta}-\int_1^T x^{-2\delta}dx\right|\le K$ and from (10) and (11), $S_{\tau\tau T}(\delta)=1+\psi(1;\xi)^2\Gamma(1-\delta)^{-2}\sum_{t=1}^{T}t^{-2\delta}(1+O(t^{-1}))$. Thus, since $\int_1^T x^{-2\delta}dx=(1-2\delta)^{-1}T^{1-2\delta}+O(1)$, $\delta\le\frac{1}{2}$, $\int_1^T x^{-1}dx=\log T$, the approximations of $S_{\tau\tau T}(\theta)$ for $\delta\le\frac{1}{2}$ are readily checked, whereas for $\delta>\frac{1}{2}$, $S_{\tau\tau T}(\theta)=\sum_{j=0}^{\infty}\tau_j^2(\theta)-\sum_{j=T+1}^{\infty}\tau_j^2(\theta)=\sum_{j=0}^{\infty}\tau_j^2(\theta)+O(T^{1-2\delta})$ from (10). Next, since

$$\dot\tau_j(\theta)=\left(-\tau_j(\theta)\psi(1;\xi)\{\log j+O(1)\},\tau_j(\theta)\dot\psi'(1;\xi)\times\left\{1+O\left(j^{-1}\right)\right\}\right)',\quad \text{as } j\to\infty, \quad (55)$$

and, for $\delta\le\frac{1}{2}$, $\left|\sum_{t=1}^{T}t^{-2\delta}\log t-\int_1^T x^{-2\delta}\log x dx\right|\le K$, where $\int_1^T x^{-2\delta}\log x dx=(1-2\delta)^{-1}T^{1-2\delta}(\log T+O(1))$, $\delta<\frac{1}{2}$, $\int_1^T x^{-1}\log x dx=\frac{1}{2}\log^2 T$, where $\int_1^T x^{-2\delta}\log x dx=(1-2\delta)^{-1}T^{1-2\delta}(\log T+O(1))$, $\delta<\frac{1}{2}$, $\int_1^T x^{-1}\log x dx=\frac{1}{2}\log^2 T$, the approximations of the components of $S_{\tau\dot\tau T}(\theta)$ with $\delta\le\frac{1}{2}$ may be checked, whereas that for $\delta>\frac{1}{2}$ follows because (10) and (55) imply $\sum_{j=T+1}^{\infty}\tau_j(\theta)\dot\tau_t(\theta)=O(T^{1-2\delta}\log T)$. The remaining results follow similarly and straightforwardly. □

Lemma 2. *Uniformly in i, as $T\to\infty$,*

$$E\left[\sup_{\theta\in\Theta}|a_{iT}(\theta)|^2\right]=O\left(T^{\delta_0-\underline{\delta}+\max(\frac{1}{2}-\underline{\delta},0)}\log^2 T\right)^2.$$

Proof of Lemma 2. We have $\lambda_t(L;\theta)v_{it}=\sum_{j=0}^{t}\nu_j(\theta)\varepsilon_{i,t-j}$, where $\nu_j(\theta)$ was defined in the proof of Proposition 1. Thus $a_{iT}(\theta)=\sum_{t=0}^{T}\tau_t(\theta)\sum_{j=0}^{t}\nu_j(\theta)\varepsilon_{i,t-j}=\sum_{t=0}^{T}c_t(\theta)\varepsilon_{it}$, where $c_t(\theta)=\sum_{j=0}^{T-t}\tau_{t+j}(\theta)\nu_j(\theta)$. By summation-by-parts $a_{iT}(\theta)=\sum_{t=0}^{T-1}\left(c_t(\theta)-c_{t+1}(\theta)\right)\sum_{s=0}^{t}\varepsilon_{is}+c_T(\theta)\sum_{t=0}^{T}\varepsilon_{it}$. Now $|c_T(\theta)|=|\tau_T(\theta)|\le KT^{-\delta}$, while $c_t(\theta)-c_{t+1}(\theta)=\sum_{j=0}^{T-t-1}\left(\tau_{t+j}(\theta)-\tau_{t+j+1}(\theta)\right)\nu_j(\theta)+\tau_T(\theta)\nu_{T-t}(\theta)$, so

$$|c_t(\theta)-c_{t+1}(\theta)| \le K\sum_{j=1}^{T-t-1}(t+j)^{-\delta-1}j^{\delta_0-\delta-1}+\frac{K}{T^\delta}(T-t)^{\delta_0-\delta-1} \le \frac{K}{t^{\delta+1}}\sum_{j=1}^{T-t-1}j^{\delta_0-\delta-1}+K\frac{(T-t)^{\delta_0-\delta-1}}{T^\delta}. \quad (56)$$

For $\delta < \delta_0$ (56) is bounded by $Kt^{-\delta-1}(T-t)^{\delta_0-\delta} + KT^{-\delta}(T-t)^{\delta_0-\delta-1}$, which is bounded by $Kt^{-\delta-1}T^{\delta_0-\delta}$ for $t \le T/2$, and by $KT^{-\delta}(T-t)^{\delta_0-\delta-1}$ for $t \ge T/2$. For $\delta = \delta_0$ (56) is bounded by $Kt^{-\delta_0-1}\log T$ for $t \le T/2$, and by $KT^{-\delta_0-1}\log T + KT^{-\delta_0}(T-t)^{-1}$ for $t \ge T/2$. For $\delta > \delta_0$ (56) is bounded by $Kt^{-\delta-1}$ for $t \le T/2$, and by $KT^{-\delta-1} + KT^{-\delta}(T-t)^{\delta_0-\delta-1}$ for $t \ge T/2$. Then we can write

$$\sup_{\theta\in\Theta}|a_{iT}(\theta)|^2$$
$$\le 2\sum_{t=0}^{T-1}\sum_{r=0}^{T-1}\sup_{\theta\in\Theta}|c_t(\theta)-c_{t+1}(\theta)|\sup_{\theta\in\Theta}|c_r(\theta)-c_{r+1}(\theta)|$$
$$\times\left|\sum_{s=0}^{t}\varepsilon_{is}\sum_{\ell=0}^{r}\varepsilon_{i\ell}\right| + 2\sup_{\theta\in\Theta}|c_T^2(\theta)|\left|\sum_{t=0}^{T}\varepsilon_{it}\right|^2,$$

Now $E\left|\sum_{t=0}^{T}\varepsilon_{it}\right|^2 = O(T)$ and using Cauchy–Schwarz inequality $E\left|\sum_{s=0}^{t}\varepsilon_{is}\sum_{\ell=0}^{r}\varepsilon_{i\ell}\right| = O\left(t^{1/2}r^{1/2}\right)$. Thus, uniformly in i for $\delta < \delta_0$, $E\left[\sup_{\theta\in\Theta\cap\{\delta<\delta_0\}}|a_{iT}(\theta)|^2\right]$ is

$$O\Bigg(\sup_{\delta<\delta_0}T^{2(\delta_0-\delta)}\sum_{t=1}^{[T/2]}\sum_{r=1}^{[T/2]}t^{-\delta-\frac{1}{2}}r^{-\delta-\frac{1}{2}}$$
$$+\sup_{\delta<\delta_0}\frac{1}{T^{2\delta}}\sum_{t=[T/2]}^{T-1}\sum_{r=[T/2]}^{T-1}(T-t)^{\delta_0-\delta-1}(T-r)^{\delta_0-\delta-1}t^{\frac{1}{2}}r^{\frac{1}{2}}$$
$$+\sup_{\delta<\delta_0}T^{1-2\delta}\Bigg)$$
$$=O\left(\sup_{\delta<\delta_0}T^{2(\delta_0-\delta)}\left\{T^{1-2\delta}1\left(\delta<\frac{1}{2}\right)+\log^2 T1\left(\delta=\frac{1}{2}\right)\right.\right.$$
$$\left.\left.+1\left(\delta>\frac{1}{2}\right)\right\}+T^{2(\delta_0-2\underline{\delta})+1}+T^{1-2\underline{\delta}}\right)$$
$$=O\left(T^{2(\delta_0-2\underline{\delta})+1}+T^{2(\delta_0-\underline{\delta})}+\log^2 T1\left(\underline{\delta}=\frac{1}{2}\right)\right)$$
$$=O\left(T^{\delta_0-\underline{\delta}+\max(\frac{1}{2}-\underline{\delta},0)}\log T\right)^2 \quad \text{for } \delta<\delta_0;$$
$$O\left(\left\{\sum_{t,r=1}^{[T/2]}(tr)^{-\delta_0-\frac{1}{2}}+T^{-2\delta_0}\sum_{t,r=[T/2]}^{T-1}(T-t)^{-1}(T-r)^{-1}(tr)^{\frac{1}{2}}\right.\right.$$
$$\left.\left.+T^{1-2\delta_0}\right\}\log^2 T\right)$$
$$=O\left(\log T1\left(\delta_0>\frac{1}{2}\right)+\log^2 T1\left(\delta_0=\frac{1}{2}\right)+T^{\frac{1}{2}-\delta_0}\log T\right)^2$$
$$=O\left(T^{\max(\frac{1}{2}-\delta_0,0)}\log T1\left(\delta_0\neq\frac{1}{2}\right)+\log^2 T1\left(\delta_0=\frac{1}{2}\right)\right)^2$$
$$=O\left(T^{\max(\frac{1}{2}-\delta_0,0)}\log^2 T\right)^2 \quad \text{for } \delta=\delta_0;$$

uniformly; and for $\delta > \delta_0$,

$$O\Bigg(\sup_{\delta>\delta_0}\Bigg\{\sum_{t=1}^{[T/2]}t^{-\delta-1/2}+T^{-\delta}\sum_{t=[T/2]}^{T-1}\left((T-t)^{\delta_0-\delta-1}\right)t^{1/2}$$
$$+T^{-\delta-1}T^{3/2}+T^{1/2-\delta}\Bigg\}\Bigg)^2$$
$$=O\left(1(\delta_0>1/2)+\log T1(\delta_0=1/2)+T^{1/2-\delta_0}1(\delta_0<1/2)\right)^2$$
$$=O\left(T^{\max(1/2-\delta_0,0)}1(\delta_0\neq 1/2)+\log T1(\delta_0=1/2)\right)^2$$
$$=O\left(T^{\max(1/2-\delta_0,0)}\log T\right)^2 \quad \text{for } \delta>\delta_0.$$

The claimed bound is then readily assembled using $\underline{\delta} \le \delta_0$. □

Lemma 3. *For all* θ,

$$\sum_{j=1}^{t}\chi_j(\xi)\,\tau_{t-j}(\theta)=\dot{\tau}_t(\theta). \tag{57}$$

Proof of Lemma 3. We write $\pi_j=\pi_j(\delta)$, $\psi_j=\psi_j(\xi)$, $\dot{\psi}_j=\dot{\psi}_j(\xi)$, $\phi_j=\phi_j(\xi)$, $\tau_j=\tau_j(\theta)$, $\dot{\tau}_t=\dot{\tau}_t(\theta)$. The first element of $\dot{\tau}_t$ is $\sum_{k=1}^{t}\dot{\pi}_k\sum_{j=0}^{t-k}\psi_j$, which is the coefficient of L^t in the expansion of $\psi(L;\xi)(\partial/\partial\delta)\Delta^\delta$, where $\dot{\pi}_k$ is the coefficient of L^k in the expansion of $(\partial/\partial\delta)\Delta^\delta$. But also

$$\psi(L;\xi)\frac{\partial}{\partial\delta}\Delta^\delta=\psi(L;\xi)\Delta^\delta\log\Delta=-\sum_{j=0}^{\infty}\pi_jL^j\sum_{k=1}^{\infty}\frac{L^k}{k}\sum_{l=0}^{\infty}\psi_lL^l$$
$$=-\sum_{t=1}^{\infty}\sum_{k=1}^{t}\frac{1}{k}\left(\sum_{l=0}^{t-k}\sum_{j=0}^{l}\pi_j\psi_{l-j}\right)L^t$$
$$=-\sum_{t=1}^{\infty}\sum_{k=1}^{t}\frac{1}{k}\tau_{t-k}L^t=\sum_{t=1}^{\infty}\sum_{k=1}^{t}\chi_{1k}\tau_{t-k}L^t,$$

so the top elements of both sides of (57) are equal. The vector consisting of the remaining elements of the right side of (57) is $\sum_{k=0}^{t}\pi_k\sum_{j=0}^{t-k}\dot{\psi}_j$, whereas the left side is

$$\sum_{j=1}^{t}\chi_{2j}\tau_{t-j}=\sum_{s=0}^{t-1}\left(g\sum_{k=1}^{t-s}\phi_{t-s-k}\dot{\psi}_k\right)\sum_{k=0}^{s}\pi_k\sum_{l=0}^{s-k}\psi_l$$
$$=\sum_{s=0}^{t-1}\sum_{k=0}^{s}\pi_k\sum_{l=0}^{s-k}\psi_l\left(\phi_0\dot{\psi}_{t-s}+\cdots+\phi_{t-s-1}\dot{\psi}_1\right)$$
$$=\sum_{s=0}^{t-1}\left(\pi_0\left(\psi_0+\psi_1+\cdots+\psi_s\right)+\pi_1\left(\psi_0+\psi_1\right.\right.$$
$$\left.\left.+\cdots+\psi_{s-1}\right)+\cdots+\pi_s\phi_0\right)$$
$$\times\left(\phi_0\dot{\psi}_{t-s}+\cdots+\phi_{t-s-1}\dot{\psi}_1\right)$$
$$=\pi_0\psi_0\left(\phi_0\dot{\psi}_t+\cdots+\phi_{t-1}\dot{\psi}_1\right)+\left(\pi_0\left(\psi_0+\psi_1\right)\right.$$
$$\left.+\pi_1\psi_0\right)\left(\phi_0\dot{\psi}_{t-1}+\cdots+\phi_{t-2}\dot{\psi}_1\right)+\left(\pi_0\left(\psi_0\right.\right.$$
$$\left.\left.+\psi_1+\psi_2\right)+\pi_1\left(\psi_0+\psi_1\right)+\pi_2\psi_0\right)\left(\phi_0\dot{\psi}_{t-2}\right.$$
$$\left.+\cdots+\phi_{t-3}\dot{\psi}_1\right)+\cdots+\left(\pi_0\left(\psi_0+\psi_1+\cdots\right.\right.$$
$$\left.+\psi_{t-1}\right)+\pi_1\left(\psi_0+\psi_1+\cdots+\psi_{t-2}\right)+\cdots$$
$$\left.+\pi_{t-1}\phi_0\right)\dot{\psi}_1$$
$$=\dot{\psi}_t+\left(\pi_0\phi_1+\pi_0\left(\psi_0+\psi_1\right)+\pi_1\psi_0\right)\dot{\psi}_{t-1}$$
$$+\left(\pi_0\psi_0\phi_2+\left(\pi_0\left(\psi_0+\psi_1\right)+\pi_1\psi_0\right)\phi_1\right.$$
$$+\left(\pi_0\left(\psi_0+\psi_1+\psi_2\right)+\pi_1\left(\psi_0+\psi_1\right)\right.$$
$$\left.\left.+\pi_2\psi_0\right)\right)\dot{\psi}_{t-2}+\cdots+\dot{\psi}_1\left(\pi_0\psi_0\phi_{t-1}\right.$$
$$+\cdots\left(\pi_0\left(\psi_0+\psi_1+\cdots+\psi_{t-1}\right)+\pi_1\left(\psi_0+\psi_1\right.\right.$$
$$\left.\left.\left.+\cdots+\psi_{t-2}\right)+\cdots+\pi_{t-1}\right)\right)$$
$$=\sum_{k=0}^{t-1}\pi_k\sum_{j=1}^{t-k}\dot{\psi}_j,$$

since $\sum_{l=0}^{m}\psi_l^0\phi_{m-l}^0=1\,(m=0)$, which follows from the identity $\phi(L;\theta)\,\psi(L;\theta)\equiv 1$. □

References

Beran, J., 1995. Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. J. R. Stat. Soc. Ser. B. 57, 659–672.

Bloomfield, P., 1973. An exponential model for the spectrum of a scalar time series. Biometrika 60, 217–226.

Brown, B.M., 1971. Martingale central limit theorems. Ann. of Math. Stat. 42, 59–66.

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

Han, C., Phillips, P.C.B., 2010. GMM estimation for dynamic panels with fixed effects and strong instruments at unity. Econometric Theory 26, 119–151.

Hassler, U., Demetrescu, M., Tarcolea, A.L., 2011. Asymptotic normal tests for integration in panels with cross-dependent units. Adv. Stat. Anal. 95, 187–204.

Hsiao, C., 2014. Analysis of Panel Data, third ed. Cambridge University Press.

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

Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. Ann. Statist. 39, 3152–3181.

Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression". J. Econometrics 47, 67–84.

Robinson, P.M., 1994. Efficient tests of nonstationary hypotheses. J. Amer. Statist. Assoc. 89, 1420–1437.

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

Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. J. Amer. Statist. Assoc. 95, 1229–1243.