Inference on trending panel data

Peter M. Robinson *, Carlos Velasco

London School of Economics and Universidad Carlos III de Madrid, Spain

**Abstract**

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

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

A considerable literature has developed around the theme of nonstationary panel data with individual and temporal effects. The nonstationarity can have both stochastic and deterministic origins. Unit roots and local-to-unit roots, in an autoregressive setting, have been the main representations of stochastic trends, while individual and temporal effects are often modelled in a separable, additive way, so that temporal effects are common over the cross-section and individual ones stay constant over time, and explanatory variables, including linear and other deterministic trends, can also feature. See for example Anderson and Hsiao (1981), Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003) for estimation of autoregressive panels when both \( N \) and \( T \) are large, and Arellano and Hahn (2006) and Moon et al. (2015) for surveys on associated inference issues; see also Hsiao (2014, Chapter 10). Autoregressive models are also commonly adopted in the nonstationary time series and cointegration literature. However, the latter has also developed fractional time series modelling, which can nest \( I(1) \) behaviour in a continuum of nonstationary possibilities, with the degree of nonstationarity characterized by a memory parameter, which can be estimated from the data. This approach has the advantages of flexibility and of yielding standard asymptotics, and possible local efficiency of inference, unlike the nonstandard theory which usually emerges from an autoregressive setting. Likewise, the fixed-design nonparametric regression literature suggests a modelling of deterministic trends which is less prone to specification error than parametric functions. Robinson and Velasco (2015) employed fractional models with individual effects in a panel data setting, but with no provision for other time-trending effects.
features such as deterministic trends. Another aspect of their work was the assumption of cross-sectional independence and homoscedasticity, conditional on individual effects, which is increasingly seen as restrictive, and sometimes replaced by factor modelling, or spatial modelling.

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

We consider an observable array \( \{y_{it}\}, i = 1, \ldots, N, t = 0, 1, \ldots, T \). The vectors of \( N \) cross-sectional observations \( y_t = (y_{1t}, \ldots, y_{Nt})' \), the prime denoting transposition, are assumed to be generated by the semiparametric model

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

The parametric aspect of (1) is due to the \( (p + 1) \times 1 \) parameter vector \( \theta_0 \), which is known only to lie in a given compact subset \( \Theta \) of \( \mathbb{R}^{p+1} \). In (1) \( L \) is the lag operator, and for any \( \theta \in \Theta \) and each \( t \geq 0 \), \( \lambda_t (L; \theta) \) is a scalar function given by

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

where the \( \lambda_j (\theta) \) are given functions to be defined subsequently. The unobservable vectors \( \varepsilon_t \) have elements with zero mean and are uncorrelated and homoscedastic across \( t \). Thus, (1) is an autoregressive representation for the \( y_t - \alpha_t \) with initial condition at \( t = 0 \), where the number of terms in (2) increases with \( t \). The assumptions on \( \lambda_t (L; \theta) \) that we will impose are aimed at covering fractionally integrated autoregressive moving average (FARIMA) sequences \( y_t - \alpha_t \), with unknown memory parameter that can lie in either the stationary or nonstationary regions.

Other aspects of (1) are nonparametric. We do not require uncorrelatedness or homoscedasticity across the elements of \( \varepsilon_t \), allowing it to have covariance matrix that is un restricted apart from remaining positive definite with increasing \( N \), thereby to reflect possible cross-sectional correlation and heteroscedasticity of a nonparametric nature, though it can also be assumed to be diagonal, to reflect a lack of cross-sectional correlation but the possibility of nonparametric heteroscedasticity. The vectors \( \alpha_t \) consist of stochastic or deterministic unobservable individual effects that can time-trend in a nonparametric way, and in a manner that can vary across elements of the vector, where with \( N \) increasing the familiar incidental parameters problem arises. Our allowance for temporally varying individual effects and cross-sectional correlation and heteroscedasticity of innovations, and our relaxation of temporal independence of innovations to martingale difference structure, extend the scope of the model of Robinson and Velasco (2015) to a practically significant degree, but as a consequence reduce the range of memory parameter values covered and limits the degree to which \( N \) can increase relative to \( T \). These limitations arise from the need to control bias due to the trending of \( \alpha_t \) that our first differencing approach is unable to eliminate completely. Note that our detailed specification of \( \alpha_t \) implies that its elements depend also on \( T \), while to cover some possible specifications of \( \varepsilon_t \) (which allow for cross-sectional dependence) we allow its elements to depend also on \( N \), and thence on \( T \); thus the \( y_{it} \) depend also on \( T \), though as is common we suppress reference to this fact in our notation.

The main goal of the paper is to justify statistical inference on the unknown parameter vector \( \theta_0 \). Detailed regularity conditions, later employed in establishing the properties of consistency and asymptotic normality, are described in the following section. In Section 3 an estimate of \( \theta_0 \), based on initial first-differencing of (1), is shown to be consistent and asymptotically normal, and feasible statistical inference is justified. In Section 4 a more refined estimate, with some advantages, is similarly analysed. Section 5 employs a spatial model to illustrate relative efficiency. A Monte Carlo study of finite-sample performance is reported in Section 6. Section 7 contains some final comments. The proofs of theorems are included in an Appendix.

2. Theoretical setting

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

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

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

which has the structure

\[
\lambda (L; \theta) = \Delta^\delta \psi (L; \xi),
\]
where \( \delta \) is a scalar, \( \xi \) is a \( p \times 1 \) vector, \( \theta = (\delta, \xi) \) and the functions \( \Delta^\delta \) and \( \psi (L; \xi) \) are described as follows. Defining the difference operator \( \Delta = 1 - L \), \( \Delta^\delta \) has the expansion

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

for non-integer \( \delta > 0 \), while for integer \( \delta = 0, 1, \ldots, \pi_j (\delta) = 1(j = 0, 1, \ldots, \delta)(-1)^{\delta}(\delta - 1) \cdots (\delta - j + 1)/j! \), taking \( 0/0 = 1 \) and 1 \((.)\) to be the indicator function; \( \psi (L; \xi) \) is a known function of its arguments such that for complex-valued \( x \), \( |\psi (x; \xi)| \neq 0, |x| \leq 1 \) and is continuously differentiable in \( \xi \), and in the expansion

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

the coefficients \( \psi_j (\xi) \) satisfy

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

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

\[
\lambda_j (\theta) = \sum_{k=0}^{j} \pi_{j-k} (\delta) \dot{\psi}_k (\xi), \quad j \geq 0. \tag{4}
\]

It is assumed that (3) holds for all \( \xi \in \mathcal{S} \) with \( c (\psi) \) satisfying

\[
\inf_{\mathcal{S}} c (\xi) = c^* > 0. \tag{5}
\]

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

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

and

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

where the prime denotes transposition and

\[
\chi_j (\xi) = (\chi_{ij} (\xi), \chi_{ij}^\prime (\xi))^\prime, \quad \chi_{ij} (\xi) = -j^{-1}, \quad \chi_{ij} (\xi) = \sum_{k=1}^{j} \phi_k (\xi) \psi_{j-k} (\xi).
\]

Then define the \( (p + 1) \times (p + 1) \) matrix

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

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

First, we relax their assumption of independence and identity of distribution of the unobservable elements \( \varepsilon_t \) of the vectors \( \varepsilon_T = (\varepsilon_{1}, \ldots, \varepsilon_{NT})^\prime \) across \( t \) to martingale difference structure for levels and squares/products: in particular, almost surely, for constants \( \sigma_{ij} \) (not depending on \( t \)),

\[
E \left( \varepsilon_{it} \varepsilon_{jt} \mid \varepsilon_{klt} = 0, \varepsilon_{klt} \leq N \right) = 0, \quad 1 \leq i, j \leq N, 0 \leq t \leq T, \tag{6}
\]

\[
E \left( \varepsilon_{it} \varepsilon_{jt} - \sigma_{ij} \right) E_{klt} = 0, \quad 1 \leq i, j \leq N, 0 \leq t \leq T \tag{7}
\]
where (6) implies \( E (\varepsilon_{it}) = 0 \). Recall from Section 1 that the \( \varepsilon_{it} \) can be regarded as depending also on \( N \), and thence \( T \), and thus this is true of the \( \sigma_{it0} \) (which we stress are constant over \( t \).) We impose also the moment condition

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

and the fourth cumulant condition

\[
\sup_i \sum_{t,j,k=1}^N \left| \text{cum} \left( \varepsilon_{it}, \varepsilon_{jt}, \varepsilon_{kt}, \varepsilon_{lt} \right) \right| = O(N), \tag{9}
\]

both of which are automatically satisfied if \( \varepsilon_{it} \) is Gaussian.

Secondly, we relax the assumption in Robinson and Velasco (2015) of homoscedasticity and lack of correlation across elements of the vectors \( \varepsilon_t \), in allowing \( E \varepsilon_t \varepsilon_t' = \Sigma_{0N} \), \( t = 0, 1, \ldots, T \), for an \( N \times N \) matrix \( \Sigma_{0N} = (\sigma_{ij0}) \) which is assumed to stay positive definite and have bounded elements with increasing \( N \) but is otherwise unknown, to reflect possible cross-sectional correlation and heteroscedasticity, or else is restricted to be diagonal, to reflect an assumed lack of cross-sectional correlation, but the possibility of heteroscedasticity. Since \( N \) is allowed to increase, \( \Sigma_{0N} \) can in either case be regarded as nonparametric. Specifically, we assume that, if \( N \) is regarded as increasing

\[
\lim_{N \to \infty} (\| \Sigma_{0N} \| + \| \Sigma_{0N}^{-1} \| ) < \infty, \tag{10}
\]

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

\[
\sigma_0^2 = \lim_{N \to \infty} \frac{1}{N} \text{tr} (\Sigma_{0N}), \tag{11}
\]

and of

\[
\kappa_0 = \lim_{N \to \infty} \frac{1}{N} \text{tr} (\Sigma_{0N}^2), \tag{12}
\]

while if \( N \) is regarded as fixed:

\[
\| \Sigma_{0N} \| + \| \Sigma_{0N}^{-1} \| < \infty, \quad \sigma_0^2 = \frac{1}{N} \text{tr} (\Sigma_{0N}), \quad \kappa_0 = \frac{1}{N} \text{tr} (\Sigma_{0N}^2). \]

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

\[
\lim_{N \to \infty} \sup_i (\sigma_{i00} + \sigma_{i0}^{-1}) < \infty. \]

Our final extension allows the \( \alpha_i = (\alpha_{i1}, \ldots, \alpha_{iN})' \) in (1) to vary with \( t \), being vectors of unobserved nonparametric (possibly stochastic, and cross-sectionally dependent) trending individual effects, such that

\[
\alpha_{it} = \alpha_i (t/T),
\]

for functions \( \alpha_i (u) \) which satisfy the (possibly stochastic) Lipschitz condition

\[
\sup_{N \geq 1} \sup_{1 \leq i \leq N, \{u, u+v\} \in [0, 1]} |\alpha_i (u) - \alpha_i (u + v)| = O_p (|v|) \quad \text{as} \quad v \to 0. \tag{13}
\]

Here there is no exogeneity requirement on \( \alpha_i (u) \), with no restriction on possible dependence with the \( \varepsilon_{it} \). However we find that if a strong exogeneity condition is imposed, and a related condition to (13) is added, we can slightly increase the range of values of \( \sigma_{0}^0 \) covered and relax the restrictions on increase of \( N \) with \( T \). In particular, almost surely,

\[
E (\varepsilon_{it} | A_N ) = 0, \quad 1 \leq i \leq N, \quad 0 \leq t \leq T, \quad N \geq 1, \quad T \geq 1,
\]

\[
E (\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij0} | A_N ) = 0, \quad 1 \leq i \leq N, \quad 0 \leq t \leq T, \quad N \geq 1, \quad T \geq 1, \tag{14}
\]

where

\[
A_N = \{ \alpha_i (u) - \alpha_i (u + v), u, u + v \in [0, 1], 1 \leq i \leq N \},
\]

and also

\[
\sup_{N \geq 1} \sup_{1 \leq i \leq N, \{u, u+v\} \in [0, 1]} E (\alpha_i (u) - \alpha_i (u + v))^2 = O (v^2) \quad \text{as} \quad v \to 0. \tag{15}
\]
Note that (14) is weaker than assuming the \( \varepsilon_{it} \) are independent of \( A_N \), and weaker also than the corresponding moment restriction that conditions on \( \{ \alpha_t (u), \ u \in [0, 1], \ 1 \leq i \leq N \} \). Note also that neither (13) nor (15) implies the other, though they are closely related.

To illustrate these various conditions, suppose \( \alpha_t (u) \) has the formal separable infinite series representation

\[
\alpha_t (u) = \beta_t + \sum_{j=1}^{\infty} \beta_j g_j (u), \ i \geq 1,
\]

where \( \beta_i \) is completely unrestricted (as it can be fully differenced out), the functions \( g_j (u) \) are nonstochastic, and the \( \beta_j \) can be random variables. Then since

\[
|\alpha_t (u) - \alpha_t (u + v)| \leq \left( \sum_{j=1}^{\infty} \beta_j^2 \sum_{j=1}^{\infty} (g_j (u) - g_j (u + v))^2 \right)^{1/2}
\]

a sufficient condition for (13) is

\[
\sup_{N \geq 1} \sup_{1 \leq i \leq N} \sum_{j=1}^{\infty} \beta_j^2 = O_p (1), \quad \sup_{N \geq 1} \sup_{1 \leq i \leq N} \sup_{(u,u+v) \in [0,1]} \sum_{j=1}^{\infty} (g_j (u) - g_j (u + v))^2 = O \left( v^2 \right), \quad \text{as } v \rightarrow 0,
\]

while the latter and the requirements that, for all \( N, \ T \geq 1, \sup_{N \geq 1} \sup_{1 \leq i \leq N} E \left( \sum_{j=1}^{\infty} \beta_j^2 \right) < \infty, \) and \( \{ \beta_{ij}, \ 1 \leq i \leq N, \ j \geq 1 \} \) is independent of \( \{ \varepsilon_{it}, \ 1 \leq i \leq N, \ 0 \leq t \leq T \}, \ N \geq 1, \ T \geq 1, \) are collectively sufficient for (14) and (15). An example of (16) are polynomials in \( t \), with degree and (possibly stochastic) coefficients allowed to vary with \( i \). In any case the trendind individual effects are nonparametric, as in fixed-design nonparametric regression (see e.g. Cai (2007)), though here they can be stochastic, and also the fact that there are \( N \) of them, where \( N \) (like \( T \)) will be regarded as increasing in our asymptotic theory, lends a further nonparametric aspect.

We can compare our approach with the familiar one (see e.g. Hsiao (2014)) in which our \( \alpha_{it} \) is replaced by the addition of a separable individual effect and time effect, namely \( \beta_i + \gamma_t \). Then first differencing, which is employed in the current paper, eliminates the individual effect \( \beta_i \) completely, but not the time effect \( \gamma_t \). With unrestricted \( \gamma_t \) this structure is not covered by our assumptions on \( \alpha_{it} \) and would entail significant complications to the methods in the paper, and strengthened conditions in order to achieve asymptotic distributional results, see also remarks in Section 7. If, however we take \( \gamma_t = \gamma (t/T) \), for a Lipschitz continuous function \( \gamma(u) \), we obtain \( \gamma_t = \gamma (t/T - \gamma) \), which becomes small as \( T \) diverges. We obtain the same error bound by differencing our \( \alpha_{it} \) given condition (13), but clearly they afford more generality than the additive \( \beta_i + \gamma (t/T) \), since our time trends can vary over individuals. It is thus possible that differenced estimation of \( \theta_0 \) that ignores the \( \alpha_{it} \) is more or less robust to their presence, and the extent of this will be examined below. The more specialized structure \( \beta_i + \gamma (t/T) \) was also employed by Robinson (2012), but there the focus was on estimating the nonparametric function \( \gamma (u) \). Note that the dependence on \( T \) of \( \alpha_t \), and also of \( \varepsilon_{it} \) (as noted above) in (1) implies that the \( y_t \) form a triangular array but our notation suppresses reference to this fact.

3. First difference estimates

We consider in the current section one of the estimates of \( \theta_0 \) proposed by Robinson and Velasco (2015) in their more specialized model, with our goal being to achieve the same property of consistency as if \( \alpha_{it} \) were absent from (1) or constant over \( t \) and the elements of \( \varepsilon_{it} \) were cross-sectionally uncorrelated and homoscedastic, and asymptotic normality with the same convergence rate as before but with a generally different asymptotic variance matrix, and also to justify feasible large sample inference.

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

\[
\Delta y_t = \Delta u_t + \Delta \alpha_t = \Delta u_t + O_p(T^{-1}), \quad \text{as } T \rightarrow \infty,
\]

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

\[
v_t = \lambda^{-1}_t (L; \theta_0) \Delta \alpha_t, \ t = 0, \ldots, T.
\]

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

\[
Z (\theta) = (z_1 (\theta), \ldots, z_T (\theta)),
\]

where

\[
z_t (\theta) = \lambda_t (L; \theta^{(-1)}) (\Delta y_t), \ t = 1, \ldots, T,
\]

Please cite this article in press as: Robinson P.M., Velasco C., Inference on trending panel data. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.06.003.
with \( \theta^{(-1)} = (\delta - 1, \xi')' \). From (27) of Robinson and Velasco (2015) and (17)

\[
z_t(\theta) = \lambda_t(L; \theta) v_t - \tau_t(\theta) \varepsilon_0 + \lambda_t(L; \theta^{(-1)}) \Delta a_t
\]

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

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

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

where

\[
L_T^D(\theta) = \frac{1}{NT} \text{tr} \left( Z(\theta) Z(\theta)' \right),
\]

\[
b_T^D(\theta) = -B^{-1}(\xi) \left( S_{\tau\tau}(\theta) - S_{\tau X}(\theta) \right),
\]

\[
S_{\tau\tau}(\theta) = \sum_{t=1}^{T} \tau_t(\theta) \tau_t(\theta)', \quad S_{\tau X}(\theta) = \sum_{t=1}^{T} \tau_t(\theta) \chi_t(\xi),
\]

where \( \hat{\tau}_t(\delta) = (\partial/\partial \delta) \pi_t(\delta) \) and

\[
\hat{\tau}_t(\theta) = \frac{\partial}{\partial \theta} \tau_t(\theta) = \left[ \sum_{k=0}^{t-k} \hat{\tau}_k(\delta) \sum_{j=0}^{t-k} \psi_j(\xi) \sum_{k=0}^{t-k} \psi_j(\xi) \right]',
\]

The basic objective function (22) is of conditional sum of squares type, but corrupted by the second and third components on the right hand side of (21), whose presence will be accounted for in the theoretical development, in particular by the term \( T^{-1} b_T^D \) in the definition of \( \hat{\theta}_T^D \), which corrects for the bias induced by the second term of (21) depending on the initial condition \( \varepsilon_0 \).

Finally define

\[
\mu_0 = \kappa_0/\alpha_0^4.
\]

**Theorem 3.1.** Under (1) and (2), and the conditions of Section 2 apart from (14) and (15), as \( T \to \infty \),

\[
\hat{\theta}_T^D \to \rho_0.
\]

If also \( \frac{1}{4} < \delta_0 \leq 1 \) and \( NT^{-1-2k_0}\log^2 T \to 0 \), or if \( \delta_0 > 1 \) and \( NT^{-1} \to 0 \), or if (14) and (15) are imposed with \( \frac{1}{4} < \delta_0 \leq \frac{1}{2} \) and \( NT^{-1-4k_0}\log^4 T \to 0 \), as \( T \to \infty \),

\[
(NT)^{\frac{1}{4}} \left( \hat{\theta}_T^D - \theta_0 \right) \to \mathcal{N}(0, \mu_0 B^{-1}(\varepsilon_0)).
\]

The consistency (24) requires no further restriction on \( \delta_0 \), beyond \( \delta_0 \in D \), and no restriction on the rate of increase of \( N \) with \( T \). However, comparing with Robinson and Velasco’s (2015) Theorem 5.2 for their much more special model, we restrict \( \delta_0 \) and \( N \) in order to achieve asymptotic normality (25), in particular requiring \( N = o(T) \) and \( \delta_0 \) to take nonstationary values when the \( \alpha_t \) are not exogenous; this is due to bias produced by temporal variation in individual effects. Of course the results hold if \( N \) stays fixed as \( T \to \infty \). Note that without imposing (14) and (15) we can cover only the nonstationary region \( \delta_0 > \frac{1}{2} \), while if we do impose (14) and (15) we can cover also the stationary values \( \frac{1}{4} < \delta_0 < \frac{1}{2} \) as well as the boundary point \( \delta_0 = \frac{1}{2} \).

In order to base statistical inference on Theorem 3.1 we estimate \( B(\varepsilon_0) \) by \( B(\hat{\varepsilon}_T^D) \), with \( \hat{\varepsilon}_T^D \) denoting the final \( p \) elements of \( \hat{\theta}_T^D \), and estimate \( \mu_0 \) by

\[
\hat{\mu}_T = N \text{tr} \left( \Sigma_N(\hat{\theta}_T^D) / \text{tr} \left( \Sigma_N(\hat{\theta}_T^D) \right) \right),
\]

where

\[
\Sigma_N(\theta) = \frac{1}{T} Z(\theta) Z(\theta)'.
\]

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

\[
\tilde{\Sigma}_N(\theta) = \frac{1}{T} \text{diag} \left( Z(\theta) Z(\theta)' \right).
\]
Theorem 3.2. Under the conditions of Theorem 3.1, as $T \to \infty$,
\begin{align}
B_\varepsilon(\hat{\varepsilon}_T^D) & \to_p B(\varepsilon_0), \\
\hat{\mu}_T^D & \to_p \mu_0,
\end{align}
(29)
with $\hat{\mu}_T^D$ given by (26) where $\bar{\Sigma}_N(\hat{\theta}_T^D)$ is defined either by (27) or, if $\sigma_{ij} = 0$ for all $i \neq j$, (28), $(NT)^{1/2} (B_\varepsilon(\hat{\varepsilon}_T^D) / \hat{\mu}_T^D)^{1/2} (\hat{\theta}_T^D - \theta_0)$ converges in distribution to a vector of independent standard normal random variables.

4. Pseudo maximum likelihood estimate based on first differences

Our next estimate is the difference pseudo maximum likelihood estimate (PMLE) of Robinson and Velasco (2015). Define the $T \times T$ matrix, $\Omega_T(\theta) = (\omega_{st}(\theta))$, $\omega_{st}(\theta) = 1$ $(s = t) + \tau_s(\theta) \tau_t(\theta)$, so that $\Omega_T(\theta_0)$ is proportional to the exact covariance matrix of the vector $(e_{11} - \tau_{11}(\theta_0)e_{10}, \ldots, e_{TT} - \tau_{TT}(\theta_0)e_{00})$, cf (21). Unlike in the difference estimate of the previous section we thus allow for the initial value effect in our estimation, though as there we attempt to incorporate cross-sectional correlation or heteroscedasticity only in studetnization, not in the point estimation of $\theta_0$, and we have to contend in the theory with the $O_p(T^{-1})$ error of differencing the $a_{it}$ in (21).

We estimate $\theta_0$ by

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

where $\Theta$ is as defined in the previous section and

$$
L_T^D(\theta) = |\Omega_T(\theta)|^{-1/2} \theta_T^{-2}(\theta),
$$
(31)
in which

$$
\theta_T(\theta) = \frac{1}{NT} \text{tr} \left( Z(\theta) \Omega_T^{-1}(\theta) Z(\theta) \right).
$$

Theorem 4.1. Under (1) and (2), and the conditions of Section 2 apart from (14) and (15), as $T \to \infty$,

$$
\hat{\theta}_T \to_p \theta_0.
$$
(32)

If also $\frac{1}{T} \leq \delta_0 \leq 1$ and $NT^{-1} \log^2 T \to 0$, or if $\delta_0 > 1$ and $NT^{-1} \to 0$, or if (14) and (15) are imposed with $\frac{3}{8} \leq \delta_0 \leq \frac{1}{2}$ and $NT^{3-8\delta_0} \log T \to 0 + NT^{-1} \log^2 T \to 0$, as $T \to \infty$,

$$
(NT)^{1/2} (\hat{\theta}_T - \theta_0) \to_{dN} (0, \mu_0B^{-1}(\delta_0)).
$$
(33)

The cost of allowing temporal variation in individual effects is thus somewhat greater than with the previous estimates, the higher lower bound of $\frac{1}{T}$ for $\delta_0$ (compared to $\frac{1}{2}$ in Theorem 3.1, which is implicitly required for Theorem 4.4 of Robinson and Velasco (2015)) under (14) and (15) resulting from the need to handle additional terms, due to $(\partial / \partial \theta) \theta_T^{-2}(\theta)$, which are not involved in $(\partial / \partial \theta) L_T^D(\theta_0)$ (see the derivations near the end of the proof of Theorem 4.1).

Now define

$$
\hat{\varepsilon}_T^D = N \text{tr} \left( \bar{\Sigma}_N(\theta_T^D) \right) / tr^2 \left( \bar{\Sigma}_N(\theta_T^D) \right)
$$
(34)

and denote by $\hat{\varepsilon}_T^D$ the final $p$ elements of $\hat{\theta}_T$.

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

$$
B_\varepsilon(\hat{\varepsilon}_T^D) \to_p B(\varepsilon_0),
$$
$$
\hat{\mu}_T^D \to_p \mu_0,
$$

with $\hat{\mu}_T^D$ given by (34) where $\bar{\Sigma}_N(\hat{\theta}_T^D)$ is defined either using (27) or, if $\sigma_{ij} = 0$ for all $i \neq j$, (28), $(NT)^{1/2} (B_\varepsilon(\hat{\varepsilon}_T^D) / \hat{\mu}_T^D)^{1/2} (\hat{\theta}_T^D - \theta_0)$ converges in distribution to a vector of independent standard normal random variables.

5. Inefficiency of estimation

In the limiting covariance matrix $\mu_0B^{-1}(\varepsilon_0)$ in Theorems 3.1 and 4.1, the factor $\mu_0 = 1$ when the $e_{it}$ are homoscedastic and uncorrelated across $i$, but in general $\mu_0 \geq 1$ so heteroscedasticity and/or cross-sectional correlation inflates the variance matrix in the limiting distribution by a scalar factor, relative to the outcome of Robinson and Velasco (2015). Note also that $\hat{\theta}_T^D \geq 1$ whether it is based on either of the estimates (27) or (28) of $\Sigma_{ON}$. 

Please cite this article in press as: Robinson P.M., Velasco C., Inference on trending panel data. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.06.003.
The potential inefficiency of our estimates, or equivalently the degree of invalidity of the inference rules which assume homoscedasticity and lack of correlation across \( i \), can be examined by considering a specific model for \( \varepsilon_t \). The spatial moving average model is defined by

\[
e_{it} = (I_N + \rho W) \eta_t,
\]

where \( I_N \) is the \( N \times N \) identity matrix, \( W \) is an \( N \times N \) user-chosen ‘spatial weight’ matrix with zero diagonal elements, taken here to be symmetric and normalized such that \( \|W\| = 1 \), and correspondingly the scalar \( \rho \) satisfies \( |\rho| < 1 \), while the elements of \( \eta_t \) are mutually uncorrelated with common variance \( \sigma^2 \). In general all elements of \( W \) can depend on \( N \), to allow for possible normalizations, such as on row and/or column sums, motivating our allowance for all elements of \( \varepsilon_t \) to depend on \( N \). Thus

\[
tr(\Sigma_{ON}) = \sigma^2 tr((I + \rho W)^2) = \sigma^2(N + \rho^2 tr(W^2)),
\]

\[
tr(\Sigma_{ON}^2) = \sigma^4 tr((I + \rho W)^4) = \sigma^4(N + 6 \rho^2 tr(W^2) + 4 \rho^3 tr(W^3) + \rho^4 tr(W^4)),
\]

and so

\[
\frac{Ntr(\Sigma_{ON}^2)}{(tr(\Sigma_{ON}))^2} = \frac{N(N + 6 \rho^2 tr(W^2) + 4 \rho^3 tr(W^3) + \rho^4 tr(W^4))}{(N + \rho^2 tr(W^2))^2} = 1 + \frac{4 \rho^2 N tr(W^2) + 4 \rho^3 N tr(W^3) + \rho^4 \left(N tr(W^4) - tr(W^2)^2\right)}{(N + \rho^2 tr(W^2))^2} \geq 1 + \frac{4 \rho^2 N tr(W^2) + \rho^4 tr(W^3)}{(N + \rho^2 tr(W^2))^2}. \tag{36}
\]

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

\[
W = I_r \otimes B_s, \ B_s = (s - 1)^{-1}(1,1', \ldots, 1),
\]

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

\[
B_s^2 = (s - 1)^{-2}(s - 2, 1,1', \ldots, 1), \ B_s^3 = (s - 1)^{-3}(s^2 - 3s + 3, 1,1', \ldots, 1),
\]

we have

\[
tr(W^2) = r(s - 1)^{-2}(s - 2 + s) = N(s - 1)^{-1},
\]

\[
tr(W^3) = r(s - 1)^{-3}(s(s^2 - 3s + 3) - s) = N(s - 2)(s - 1)^{-2},
\]

and the lower bound (36) becomes

\[
1 + \frac{4 \rho^2((s - 1)^{-1} + \rho(s - 2)(s - 1)^{-2})}{(1 + \rho^2(s - 1)^{-1})^2} = 1 + \frac{4 \rho^2(s - 1 + \rho(s - 2))}{(s - 1 + \rho^2)^2}. \tag{38}
\]

A lower bound for \( \mu_0 \) is obtained by letting \( N \rightarrow \infty \), and if \( s \rightarrow \infty \) this tends to 1, but if \( s \) stays fixed the bound exceeds 1, and again increases with \( \rho^2 \).

6. Monte Carlo simulations

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

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

We consider different choices of \( N, T \) and \( T_0 \), in particular we employ three basic values of \( NT \), namely 100, 200, 400 with two combinations of \( N \) and \( T \) for each, to account for relatively short (\( T = 10, 12 \)) and moderate time series (\( T = 20, 25, 50 \)), and also \( NT = 96 \) for \( T = 12 \). The range of values of \( N \) thus varies from 8 through 20 from the smallest to the largest sample size. The values of \( T_0 \) include a stationary one (\( T_0 = 0.3 \)), which our theorems will predict will be the most problematic from the point of view of bias, a moderately non-stationary one (\( T_0 = 0.6 \)), a value close to the unit root (\( T_0 = 0.9 \)), and a
more nonstationary one ($\delta_0 = 1.2$). For FARIMA models we consider two autoregressive parameter values, $\xi = 0.5, 0.8$. Optimizations were carried out using the Matlab function fmincon with $D = [0.1, 1.5]$ and $\Sigma = [-0.95, 0.95]$, and the results are based on 10,000 independent replications.

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

The results in Tables 1–8 concern the pure fractional case, $\theta = \delta$. We first consider the case without trends, where constant fixed effects are exactly removed by first differencing. Table 1 provides a bias comparison of the estimates of $\delta_0$. In general bias is not affected by cross-sectional correlation compared with the results in Robinson and Velasco (2015), who did not allow for such, and typically reduces with increase of $\delta_0$ and $T$ as expected. Bias of the difference estimate $\hat{\delta}_D$ for $\delta_0 = 0.3$ can be of an order of magnitude larger than in the other cases for the smallest $T$ for a given $NT$, despite bias-correction having a large beneficial effect (results without bias correction are not reported here). The PMLE $\hat{\delta}_P$ does much better in this difficult setup, but the difference for larger $\delta_0$ is much smaller. MSE results in Table 2 confirm the consistency of estimates, no clear superiority of any of the two estimates apart from the bias effect, and increase in the variance of estimates with cross-sectional correlation (that is, with increasing $\rho$). The empirical size of the properly studentized t-test is sensitive to the parameter and sample size values, though they converge with increasing $NT$ to the nominal value. In general, performance tends to deteriorate with the larger $\rho$ and smaller $\delta_0$, and PMLE-based tests do better than difference ones. Tests not using consistent estimates of $\mu_0$ under cross-sectional correlation are systematically very oversized in all cases, cf. Table 4.

We repeat the experiment in Tables 5–8 for $\rho = 0.8$ but with a linear trend specified as $a_t(w) = \beta_tw$ (cf (16)), where the $\beta_t$ are generated independently from the $\chi^2(0, \gamma^2)$ distribution with $\gamma = 1.3$ (and our estimates are of course invariant to fixing a temporally constant component of the individual effects at zero). The larger value of $\gamma$ can generate relatively large trends that can dominate the behaviour of the time series as shown by the bias and MSE results in Tables 5 and 6, respectively. However, for the smaller $\gamma$, first differencing seems to account properly for the heterogeneous deterministic component, though bias is substantially increased, as is MSE, for the small values of $NT$ and $T$. In this case PMLE-based t-tests perform in a similar way as in the absence of trend for not too small $T$ and $NT$ and $\delta_0 \geq 0.6$.

In Tables 9–14 we consider the results for the FARIMA model, again for only one value of $\rho$, 0.9. The estimation of $\delta_0$ is substantially affected by the autoregressive short run dynamic component, and the bias in Table 9 can change sign with the value of $\delta_0$ and of $T$, for $\delta_0 = 0.5$, the results improving in most cases with increasing $\delta_0$, while bias is always positive for the largest $\delta_0$, 0.8, so persistence is incorporated in the estimation of $\delta_0$ in finite samples. In general, the PMLE dominates
bias-corrected difference estimates again. There is an overall increase in variability in Table 10 compared to Table 2, since both parameter estimates are highly correlated. Estimation results for $\xi_0$ in Tables 11–12 are parallel to the ones for $\delta_0$: large bias for small NT and $\delta_0$, negative bias in all cases for $\xi_0 = 0.8$, but no clear pattern for $\xi_0 = 0.5$, and MSE decreasing with
Table 7
Empirical Size (%) Corrected 5% t-test. $I(\delta_0), \rho = 0.9$, linear trend $\beta_i (t/T), \beta_i \sim N(0, \gamma^2)$.

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\delta_T$</th>
<th>$\delta_P$</th>
<th>$\delta_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0$</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>10 100</td>
<td>26.91</td>
<td>10.15</td>
</tr>
<tr>
<td>12</td>
<td>96</td>
<td>28.41</td>
<td>12.21</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>32.34</td>
<td>5.74</td>
</tr>
<tr>
<td>25</td>
<td>200</td>
<td>32.77</td>
<td>10.30</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>41.13</td>
<td>6.84</td>
</tr>
<tr>
<td>50</td>
<td>400</td>
<td>37.46</td>
<td>9.52</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\delta_T$</th>
<th>$\delta_P$</th>
<th>$\delta_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0$</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>10 100</td>
<td>37.04</td>
<td>19.42</td>
</tr>
<tr>
<td>12</td>
<td>96</td>
<td>34.93</td>
<td>19.47</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>61.65</td>
<td>30.31</td>
</tr>
<tr>
<td>25</td>
<td>200</td>
<td>36.18</td>
<td>15.08</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>61.28</td>
<td>24.91</td>
</tr>
<tr>
<td>50</td>
<td>400</td>
<td>39.53</td>
<td>12.57</td>
</tr>
</tbody>
</table>

Table 9
Empirical bias $\hat{\delta} \times 100$. FARIMA($\xi_0$, $\delta_0$), $\rho = 0.9$.

<table>
<thead>
<tr>
<th>$\xi_0$</th>
<th>$\delta_T$</th>
<th>$\delta_P$</th>
<th>$\delta_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0$</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>0.5</td>
<td>10 100</td>
<td>5.00</td>
<td>−3.24</td>
</tr>
<tr>
<td>12</td>
<td>96</td>
<td>1.42</td>
<td>−5.08</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>8.78</td>
<td>−2.63</td>
</tr>
<tr>
<td>25</td>
<td>200</td>
<td>−3.61</td>
<td>−7.81</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>4.58</td>
<td>−4.96</td>
</tr>
<tr>
<td>50</td>
<td>400</td>
<td>−4.22</td>
<td>−6.48</td>
</tr>
</tbody>
</table>

$NT$, $T$ and $\delta_0$. The feasible asymptotic inference results reported in Tables 13 and 14 confirm previous ideas on the need for consistent estimation of $\mu_0$ to account for cross-sectional correlation, though now oversizing is more severe for the larger values of $T$ across the whole range of values of $\delta_0$ and $\xi_0$.

7. Final comments

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

1. The incidental trend problem has been considered in integrated and nearly integrated (with an autoregressive local-to-unit root) panel models by Moon and Phillips (1999, 2000, 2004) and Phillips and Sul (2007), where the availability of an increasing number of cross-sections may allow consistent estimation of the local to unity parameter after bias correction. Moon et al. (2007) established optimal tests of an autoregressive panel unit root in this context, where the rate of convergence to the null hypothesis depends on the particular specification of the incidental parameters.

Please cite this article in press as: Robinson P.M., Velasco C., Inference on trending panel data. Journal of Econometrics (2018). https://doi.org/10.1016/j.jeconom.2018.06.003.
2. To follow up on remarks in the penultimate paragraph of Section 1 and just after (23), bias correction in panel autoregressions has also been pursued to achieve a centred asymptotic distribution with \((NT)^{1/2}\) convergence rate, mainly for fixed effects maximum likelihood estimation in the presence of lagged endogenous variables, see e.g. Hahn and Kuersteiner.
Table 14

Empirical Size (%) Uncorrected 5% Wald-test. FARIMA($\xi_0$, $\delta_0$), $\rho = 0.9$.

<table>
<thead>
<tr>
<th>$\delta_0$ :</th>
<th>$\alpha_T^0$</th>
<th>$\alpha_T^P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_0 = 0.5$</td>
<td>$\xi_0 = 0.5$</td>
<td>$\xi_0 = 0.8$</td>
</tr>
<tr>
<td>$\xi_0 = 0.5$</td>
<td>$\xi_0 = 0.8$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.8$</td>
<td>$\xi_0 = 0.8$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.9$</td>
<td>$\xi_0 = 0.9$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 1.2$</td>
<td>$\xi_0 = 1.2$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.3$</td>
<td>$\xi_0 = 0.3$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.6$</td>
<td>$\xi_0 = 0.6$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.9$</td>
<td>$\xi_0 = 0.9$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 1.2$</td>
<td>$\xi_0 = 1.2$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.3$</td>
<td>$\xi_0 = 0.3$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.6$</td>
<td>$\xi_0 = 0.6$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 0.9$</td>
<td>$\xi_0 = 0.9$</td>
<td></td>
</tr>
<tr>
<td>$\xi_0 = 1.2$</td>
<td>$\xi_0 = 1.2$</td>
<td></td>
</tr>
</tbody>
</table>

(2002, 2011) and references in the survey of Arellano and Hahn (2006). The source of bias in our first difference based method differs from theirs. With our fractional stochastic trend, bias is basically driven by the truncated infinite autoregressive representation of our fractional filter, rather than by their finite autoregression, so it can be described as an initial condition problem instead of regressor endogeneity induced after fixed effects estimation. Our bias correction, though, is similar in spirit to that of Hahn and Kuersteiner (2002), since bias depends only on the true value of parameters (and of $T$), and it can be (iteratively) estimated, but does not require a specific $N/T$ limit and covers trends beyond linear ones.

3. As an alternative to our nonparametric estimation of the factor $\mu_0$ that inflates the limiting covariance matrix we could invest in a parametric model for the covariance matrix of $\xi_t$, such as a factor model (see e.g. Ergemen and Velasco (2017)) or a spatial model (cf. the discussion in Section 5). With a correct specification improved finite-sample properties are likely to result, though a misspecified parameterization would lead to inconsistent estimation of $\mu_0$ and thus invalidate inferences based on Theorems 3.2 and 4.2.

4. Point estimates of $\delta_0$ that use either our nonparametric estimates of $\Sigma_{0N}$, or parametric ones such as those just described, to correct for cross-sectional correlation/heteroscedasticity to the extent of being asymptotically efficient in a Gaussian context, can be constructed. Their investigation would be worthwhile but likely entail a considerable amount of further work and possibly some further restrictions on $\delta_0$ and $T$.

5. Our imposition of the same dynamics over the cross-section affords parameter estimation with rate $(NT)^{\frac{1}{2}}$, but it is possible to allow for variation over the cross section. In particular, Hassler et al. (2011) developed tests in a panel with such fractional structure, units, but without allowing for individual effects and keeping $N$ fixed as $T \to \infty$. In our context where $N$ can increase it would be possible to keep the number of time series parameters fixed by assuming they are constant within finitely-many known cross-sectional subsets, over which the parameters can vary.

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

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

Acknowledgements

We are grateful for the comments of two referees. Financial support from the Ministerio de Economía y Competitividad (Spain), grants ECO2012-31748, ECO2014-57007p and MDM 2014-0431, and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444) is gratefully acknowledged by the second author.

Appendix

Proof of Theorem 3.1

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

$$
\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} \nu_j^2 (\theta) \sum_{t=0}^{T-j} (\epsilon_{it}^2 - \sigma_{ii}) + \frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{k=0}^{T-t} \nu_t (\theta) \nu_{t'} (\theta) \epsilon_{i,t-j} \epsilon_{i,t-k} \\
+ \frac{1}{NT} \left( \sum_{i=1}^{N} \sum_{j=1}^{T} (T - j + 1) \nu_j^2 (\theta) \right) + \frac{\sigma^2_o}{T} \sum_{j=1}^{T} (j - 1) \nu_j^2 (\theta) - \sigma^2_o \sum_{j=T+1}^{\infty} \nu_j^2 (\theta)
$$

(39)

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

$$
\sup_{\theta} \left| \sum_{i=1}^{N} \sum_{j=1}^{T} \nu_j^2 (\theta) \sum_{t=0}^{T-j} (\epsilon_{it}^2 - \sigma_{ii}) \right| \leq \sum_{i=1}^{N} \sup_{\theta} \left| \sum_{j=1}^{T} \nu_j^2 (\theta) \sum_{t=0}^{T-j} (\epsilon_{it}^2 - \sigma_{ii}) \right|,
$$

which is uniformly $o_p (NT)$ much as in RV using also the proof of Theorem 1 of Hualde and Robinson (2011) (6) and (8). The second term in (39) is uniformly $o_p (1)$ in much the same way, and the remaining terms are uniformly $o(1)$ from (11) and (52) of RV. Since from (21) $L_T^2 (\theta) = \sum_{t=1}^{T} \| \lambda_t (L; \theta) v_t - \tau_t (\theta) \epsilon_{00} + \lambda_{t-1} (L; \theta^{(-1)}) \Delta \alpha_t \| ^2$ it remains to show that the term in $\Delta \alpha_t$ contributes negligibly. First, we have

$$
\sup_{\theta} \left( \sum_{t=1}^{T} \| \lambda_{t-1} (L; \theta^{(-1)}) \Delta \alpha_t \| ^2 \right) \leq \sup_{\theta} \| \Delta \alpha_t \| ^2 \sum_{t=1}^{T} \left( \sup_{\theta} \sum_{j=0}^{T-t-j} \| \lambda_j (\theta^{(-1)}) \| ^2 \right) \leq O_p \left( NT^{-2} \sum_{t=1}^{T} \left( t^{1-\delta} + 1 \right)^2 \right)
$$

(40)

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

$$
\sup_{\theta} \left( \sum_{t=1}^{T} \left( \lambda_{t-1} (L; \theta^{(-1)}) \Delta \alpha_t \right) \lambda_t (L; \theta) v_t \right) \leq \left( \sup_{\theta} \sum_{t=1}^{T} \| \lambda_t (L; \theta) v_t \| ^2 \right) \left( \sup_{\theta} \sum_{t=1}^{T} \left( \lambda_{t-1} (L; \theta^{(-1)}) \Delta \alpha_t \| ^2 \right) ^{1/2} \leq O_p \left( (NT)^{1/2} \right) = O_p (NT),
$$

(41)

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

$$
\sup_{\theta} \left( \sum_{t=1}^{T} \left( \lambda_{t-1} (L; \theta^{(-1)}) \Delta \alpha_t \right) \epsilon_0 \right) \leq \left( \sup_{\theta} \sum_{t=1}^{T} \| \epsilon_0 \| ^2 \right) \left( \sup_{\theta} \sum_{t=1}^{T} \left( \lambda_{t-1} (L; \theta^{(-1)}) \Delta \alpha_t \| ^2 \right) ^{1/2} \leq O_p \left( (NT)^{1/2} \right) = O_p (NT).
$$

(42)

Thus consistency of arg min$_{\theta \in \Theta} L_T^2 (\theta)$ is established, and thence straightforwardly (24), using smoothness properties of $b_T^0$.

The latter are also used, along with (24), in proving (25), as in RV’s Theorem 5.2, which incidentally required weaker conditions on $\delta_0$ and $N$ than ours, so we do not repeat the details. But we have to extend RV’s Theorem 4.3 on arg min$_{\theta \in \Theta} L_T^2 (\theta) - b_T^0 (\theta_0)$ to our setting, and this requires first extending the CLT for scores in Proposition 2 of RV. We can write their $w_T$ as

$$
w_T = \frac{1}{2 (NT)^{1/2}} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} A_T (\theta_0) = \frac{1}{(NT)^{1/2}} \sum_{t=1}^{T} \sum_{t'=0}^{T-t-1} X_{t-j} (\theta_0) \epsilon'_j \epsilon_t.
$$

Please cite this article in press as: Robinson P.M., Velasco C., Inference on trending panel data. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.06.003.
Thus

\[
Ew'Tw'_T = \frac{1}{NT} \sum_{t=1}^T \sum_{j=0}^T \sum_{k=0}^T \sum_{s=0}^T \chi_{t-j} (\theta_0) \epsilon'_t \epsilon'_s \epsilon'_k \chi'_{t-j} (\theta_0)
\]

\[
= \frac{1}{NT} \sum_{t=1}^T \sum_{j=0}^T \sum_{s=0}^T \chi_{t-j} (\theta_0) \epsilon'_t \Sigma \epsilon'_s \epsilon'_k \chi'_{t-j} (\theta_0)
\]

\[
= \frac{\text{tr} (\Sigma^2 \epsilon_0)}{NT} \sum_{t=1}^T \sum_{j=0}^T \sum_{s=0}^T \chi_{t-j} (\theta_0) \chi'_{t-j} (\theta_0).
\]

Then given (12) and much as in RV,

\[
w_T \rightarrow_{dN} (0, \kappa_0 B (\theta_0)).
\]

Apart from the extra terms discussed below the score and Hessian are handled much as in RV, where the latter has probability limit \(2\sigma^2 \epsilon_0 (\theta_0)\), with \(\sigma^2\) as in (11).

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

\[
(NT)^{1/2} \frac{\partial}{\partial \theta} L_T^0 (\theta_0) = \frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( f_t - \epsilon_0 \hat{\epsilon}_t^0 \right) + \frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( f_t - \epsilon_0 \hat{\epsilon}_t^0 \right) (\Delta \alpha_i)^{0} \left( L \right) \left( \epsilon_t - \tau_0 \epsilon_0 \right) + \Delta \alpha_i \lambda_{t-1}^0 \left( L \right).
\]

where

\[
\tau_i^0 = \tau_i^0 (\theta_0), \quad \hat{\epsilon}_t^0 = \hat{\epsilon}_t (\theta_0), \quad \frac{\partial \tau_i}{\partial \theta} = \frac{\partial \tau_i (\theta)}{\partial \theta},
\]

\[
\lambda_{t-1}^0 \left( L \right) = \lambda_{t-1} \left( L, \theta_0 \right), \quad \hat{\lambda}_{t-1}^0 \left( L \right) = \hat{\lambda}_{t-1} \left( L, \theta_0 \right),
\]

\[
\hat{\lambda}_{t-1} \left( L, \theta_0 \right) = \frac{\partial \lambda_{t-1} \left( L, \theta_0 \right)}{\partial \theta}.
\]

The additional terms contributing to the asymptotic bias of \( (NT)^{1/2} \frac{\partial}{\partial \theta} L_T^0 (\theta_0) \) that are not covered in RV are thus

\[
\frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( \Delta \alpha_i \hat{\lambda}_{t-1}^0 \left( L \right) \right) \left( \epsilon_t - \tau_0 \epsilon_0 \right) + \frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( f_t - \epsilon_0 \hat{\epsilon}_t^0 \right) \left( \Delta \alpha_i \lambda_{t-1}^0 \left( L \right) \right)
\]

\[
+ \frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( \Delta \alpha_i \hat{\lambda}_{t-1}^0 \left( L \right) \right) \left( \Delta \alpha_i \lambda_{t-1}^0 \left( L \right) \right).
\]

Denoting \( \lambda_j^0 = \lambda_j \left( \theta_0 \right) \), we have \( \hat{\lambda}_j^0 = \frac{3}{10} \lambda_j = O \left( \left| \lambda_j^0 \right| \log j \right) = O \left( j^{-\delta_0} \log t \right) \) and thus

\[
\left\| \frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( \Delta \alpha_i \hat{\lambda}_{t-1}^0 \left( L \right) \right) \left( \epsilon_t - \tau_0 \epsilon_0 \right) \right\|
\]

\[
\leq \sup_{i, t} \left\| \Delta \alpha_i \right\| \frac{2}{(NT)^{1/2}} \sum_{t=1}^T \left( \sum_{j=0}^{t-1} \left\| \hat{\lambda}_j^0 \right\| \right) \left\| \epsilon_t - \tau_0 \epsilon_0 \right\|
\]

\[
= O_p \left( NT^{-1/2} \left( T \log T \right) \right)
\]

\[
= O_p \left( T^{-3/2} \left( T^{-1/2} + T \log T \right) \right)
\]

which is \( o_p (1) \) as \( NT^{1/2} \left( T \log T \right) \rightarrow 0 \), with \( \delta_0 > 1/2 \). The second term in (43) can be bounded similarly, given \( E \left| f_t - \epsilon_0 \hat{\epsilon}_t^0 \right| = O \left( \sqrt{N} \right) \). The third is

\[
O_p \left( NT^{-2} \left( T \log T \right) \right)
\]

\[
= O_p \left( N^{1/2} T^{-3/2} \left( T^{-1/2} + T \log T \right) \right)
\]

which is \( o_p (1) \) since \( NT^{1/2} \left( T \log T \right) \rightarrow 0 \).
Now impose (14) and (15) and allow $\delta_0 > \frac{1}{4}$. Let $K$ denote a generic finite constant. From (14) the first term in (43) has zero mean and, using (10), variance bounded by

$$
\frac{K}{T} \left\| \Sigma_{0N} \right\| \sum_{t=1}^{T} \left( t^{1-\delta_0} + \log t \right)^2 \log^2 t 
+ \frac{K}{T^3} \left\| \Sigma_{0N} \right\| \left( \sum_{t=1}^{T} \left( t^{1-\delta_0} + \log t \right) t^{-\delta_0} \log t \right)^2
=
O \left( T^{-3} \left( T^{3-2\delta_0} + T \log T \right) \log^2 T \right) + O \left( T^{-3} \left( T^{2(1-\delta_0)} + \log T \right)^2 \log^2 T \right)
=
O \left( T^{-2} \left( T^{1-4\delta_0} + T^{-1/2} \log^2 T \right) \log T \right),
$$

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

$$
\left\| E \left[ \left( \hat{\lambda}^0_{t-1} (L) \Delta \alpha^0_t \right) \left( \hat{\lambda}^0_{t-1} (L) \Delta \alpha^0_t \right) \right]\right\| \leq E \left\| \hat{\lambda}^0_{t-1} (L) \Delta \alpha^0_t \right\|^2
= O \left( E \left\| \Delta \alpha^0_t \right\|^2 \left( \sum_{j=0}^{t-1} j^{-\delta_0} \log j \right) \right)
= O \left( T^{-2} \left( t^{1-\delta_0} + 1 \right)^2 \log^2 t \right).
$$

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

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

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

**Proof of Theorem 3.2**

Given Theorem 3.1 and continuity of $b (\xi)$ it suffices to prove (30). We give the proof only for (27) because that for (28) is simpler. From (11), $\hat{\rho}^p_1 - \rho_0$ differs by $o (1)$ from

$$
Ntr \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) \right) / tr^2 \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) \right) - Ntr \left( \Sigma^2_{0N} \right) / tr^2 \left( \Sigma_{0N} \right),
$$

which is

$$
Ntr \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) - \Sigma^2_{0N} \right) / tr^2 \left( \Sigma_{0N} \right)
- Ntr \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) \right) \left( tr^2 \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) \right) / tr^2 \left( \Sigma_{0N} \right) \right) / \left( tr^2 \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) \right) \right)
$$

so the result follows on showing that

$$
tr \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) - \Sigma_{0N} \right) = o_p (N), \tag{44}
$$

$$
tr \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) - \Sigma_{0N} \right) = o_p (N), \tag{45}
$$

noting that (44) implies that $tr \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) \right) / N$ has a positive, finite probability limit. We have

$$
\tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) - \Sigma_{0N} = \left( \tilde{\Sigma}^2_N \left( \hat{\rho}^p_1 \right) - \tilde{\Sigma}^2_N \left( \hat{\theta}_0 \right) \right) + \left( \tilde{\Sigma}^2_N \left( \hat{\theta}_0 \right) - \Sigma_{0N} \right).
$$

Now

$$
\tilde{\Sigma}^2_N \left( \hat{\theta}_0 \right) = \frac{1}{T} \sum_{t=1}^{T} (\epsilon_t + s_t) (\epsilon_t + s_t)',
$$

where

$$
A_T = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t \epsilon_t', \quad R_T = \frac{1}{T} \sum_{t=1}^{T} (s_t s_t' + \epsilon_t s_t'),
$$

Please cite this article in press as: Robinson P.M., Velasco C. Inference on trending panel data. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.06.003.
with \( s_t = \lambda_{t-1} \left( L; \theta_0^{(-1)} \right) \Delta \alpha_t - \tau_t (\theta_0) \varepsilon_0 \). Thus

\[
\text{tr} \left( \widetilde{\Sigma}_N (\theta_0) - \Sigma_{ON} \right) = \text{tr} \left( A_T - \Sigma_{ON} + R_T \right).
\]

Since

\[
E \left( \varepsilon_t^2 - \sigma_{0t} \right) \left( \varepsilon_{jt}^2 - \sigma_{0jt} \right) = 2 \sigma_{0jt} + \text{cum} \left( \varepsilon_{it}, \varepsilon_{jt}, \varepsilon_{jt}, \varepsilon_{jt} \right),
\]

we have

\[
E \text{tr}^2 \left( A_T - \Sigma_{ON} \right) = E \left( \frac{1}{T} \sum_{t=1}^{T} \left( \| \varepsilon_t \|^2 - \text{tr} \left( \Sigma_{ON} \right) \right) \right)^2
\]

\[
= \frac{1}{T^2} E \left( \sum_{t=1}^{T} \left( \sum_{i=1}^{N} (\varepsilon_{it}^2 - \sigma_{0it}) \right) \left( \sum_{j=1}^{N} (\varepsilon_{jt}^2 - \sigma_{0jt}) \right) \right)
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \left( 2 \sigma_{0jt} + \text{cum} \left( \varepsilon_{it}, \varepsilon_{jt}, \varepsilon_{jt}, \varepsilon_{jt} \right) \right)
\]

\[
= \frac{2}{T} \text{tr} \left( \Sigma_{ON}^2 \right) + \frac{1}{T} \sum_{t=1}^{T} \text{cum} \left( \varepsilon_{it}, \varepsilon_{jt}, \varepsilon_{jt}, \varepsilon_{jt} \right)
\]

\[
= O \left( \frac{N}{T} \right) = o (N^2)
\]

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

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

where

\[
\frac{1}{T} \sum_{t=1}^{T} \| s_t \|^2 \leq \frac{2}{T} \sum_{t=1}^{T} \left\| \lambda_t \left( \theta_0^{(-1)} \right) \varepsilon_0 \right\|^2 + \frac{2}{T} \sum_{t=1}^{T} \left\| \lambda_{t-1} \left( L; \theta_0^{(-1)} \right) \Delta \alpha_t \right\|^2
\]

\[
= o_p \left( N \right), \text{ as } T \to \infty,
\]

as in (40)–(42). Also

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

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

\[
\text{tr} \left( \widetilde{\Sigma}_N (\bar{\theta}_T^p) - \widetilde{\Sigma}_N (\theta_0) \right) = \frac{1}{T} \sum_{t=1}^{T} \left( \| z_t (\bar{\theta}_T^p) \|^2 - \| z_t (\theta_0) \|^2 \right)
\]

\[
= \frac{2}{T} \sum_{t=1}^{T} z_t' (\theta_0) \left( \frac{\partial z_t (\bar{\theta})}{\partial \theta'} (\bar{\theta}_T^p - \theta_0) \right)
\]

\[
= O_p \left( \left( \frac{1}{NT^3} \sum_{t=1}^{T} \| z_t (\theta_0) \|^2 \sum_{t=1}^{T} \left\| \frac{\partial z_t (\bar{\theta})}{\partial \theta'} \right\|^2 \right)^{1/2} \right).
\]

From [21],

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

and thus

\[
\sum_{t=1}^{T} \| z_t (\theta_0) \|^2 = O_p \left( NT \right).
\]
Thus from Theorem 3.1 and with $\delta$ dominates, and we have
\[
\left\| \frac{\partial \lambda_t}{\partial \theta} - v_t \right\| \leq K \sum_{j=1}^{t} (\log j) j^{\theta_0 - \delta - 1} \left( \sum_{i=1}^{N} \epsilon_i^2 - \ell \right)^{1/2} = O_p \left( N^{1/2} \left( t^{\theta_0 - \delta} + \log^2 t \right) \right),
\]
while
\[
\hat{t}_t(\theta) = O \left( (\log t) t^{-\delta} \right), \quad \left\| \hat{\lambda}_{t-1} - \theta^{(-1)} \right\| \Delta \alpha_t = O_p \left( \frac{N^{1/2} \left( t^{\epsilon - \delta_0} \right)}{T} \right).
\]

Thus from Theorem 3.1 and with $T$ large enough and any $\epsilon > 0$
\[
\left\| \frac{\partial z_t(\theta)}{\partial \theta^j} \right\| = O_p \left( N^{1/2} \left( t^\epsilon + t^{\epsilon - \delta_0} \right) \right) = O_p \left( N^{1/2} t^\epsilon \right),
\]
uniformly in $t$, and so
\[
\sum_{t=1}^{T} \left\| \frac{\partial z_t(\theta)}{\partial \theta^j} \right\|^2 = O_p \left( N T^{1 + 2\epsilon} \right).
\]

It follows that (46) is
\[
O_p \left( \frac{N T^{1+\epsilon}}{N^{1/2} T^{3/2}} \right) = O_p \left( N^{1/2} t^{\epsilon - 1/2} \right) = o_p \left( N \right),
\]
so $tr \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) = o_p \left( N \right).$ The proof of (44) is completed. Next, to establish (45) we have
\[
\left| tr \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) \right| = \left| tr \left( \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) \left( \Sigma_N (\hat{\theta}_T^N) + \Sigma_N (\theta_0) \right) \right) \right|
\leq \left( tr \left( \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) \right) \right)^{1/2} \left( tr \left( \left( \Sigma_N (\hat{\theta}_T^N) + \Sigma_N (\theta_0) \right) \right) \right)^{1/2},
\]
where
\[
tr \left( \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) \right) = tr \left( \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) + \Sigma_N (\theta_0) - \Sigma_N (\theta_0) \right)^2
\leq 2tr \left( \left( \Sigma_N (\hat{\theta}_T^N) - \Sigma_N (\theta_0) \right) \right)^2 + 2tr \left( \left( \Sigma_N (\theta_0) - \Sigma_N (\theta_0) \right) \right)^2.
\]
Now
\[
tr \left( \left( \Sigma_N (\theta_0) - \Sigma_N (\theta_0) \right) \right) = tr \left( (A_T - \Sigma_N + R_T)^2 \right)
\leq 2tr \left( (A_T - \Sigma_N)^2 \right) + tr \left( R_T^2 \right),
\]
where
\[
Etr \left( (A_T - \Sigma_N)^2 \right) = E \left( tr \left( \frac{1}{T} \sum_{i=1}^{T} (e_i \epsilon_i^t - \Sigma_N) \right) \right)^2
\leq \frac{1}{T^2} E \left( \sum_{i=1}^{T} \| e_i \|^4 \right) + tr \left( \sum_{i,j=1}^{T} (e_i \epsilon_i^t, \epsilon_i^t, \epsilon_i^t - \Sigma_N^2) \right)
\leq \frac{1}{T^2} \sum_{i=1}^{T} E \| e_i \|^4 + \sum_{i,j=1}^{N} \text{cum}(e_i, e_j, \epsilon_i^t, \epsilon_i^t)
\leq O \left( \frac{N}{T} \right).
\]
from (9) and (12). Thus \( tr \left( (A_T - \sum_{i=1}^{T} \rho_i) \right)^2 = o_p(N) \). From the proof of (44) it is readily seen that \( tr \left( K_T^2 \right) = o_p(N) \), to complete the proof of (45). □

**Proof of Theorem 4.1**

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

\[
\sum_{t=1}^{T} \frac{\left( \sum_{i=1}^{L} \tau_i (\theta) \lambda_{i-1} (L; \theta^{(-1)}) \Delta \alpha_{it} \right)^2}{S_{\tau \tau T}(\theta)},
\]

which comes from \( \left( Z_T^2(\theta) \right)^2 / S_{\tau \tau T}(\theta) \), where \( Z_T(\theta) = (z_{i1} (\theta), \ldots, z_{iT} (\theta))' \), \( \tau_i (\theta) = (\tau_{i1} (\theta), \ldots, \tau_{iT} (\theta))' \). We can bound (47) by

\[
\sup_{i,t} |\Delta \alpha_{it}|^2 \sup_{D,\Xi} \frac{\left| \sum_{t=1}^{T} \tau_i (\theta) \sum_{j=0}^{T} \lambda_j (\theta^{(-1)}) \right|^2}{S_{\tau \tau T}(\theta)}
= O_p \left( T^{-2} \sup_{D} \frac{\sum_{t=1}^{T} t^{-\delta} (t^{1-\delta} + 1)^2}{T^{1-2\delta} + 1} \right),
= O_p \left( T^{-2} \sup_{D} \frac{(T^{2(1-\delta)} + 1)^2}{T^{1-2\delta} + 1} \right)
= O_p \left( T^{-2} + 1 \right)
\]

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

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

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \lambda_{i-1}^0 (L) \Delta \alpha_{it}^2 + 2 \left( \varepsilon_{it} - \tau_i^0 \varepsilon_{it0} \right) \lambda_{i-1}^0 (L) \Delta \alpha_{it} \right\}
- \frac{1}{S_{\tau \tau T}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \sum_{i=1}^{L} \tau_i^0 \left( \lambda_{i-1}^0 (L) \Delta \alpha_{it} \right) + 2 \tau_i^0 \left( \varepsilon_{it} - \tau_i^0 \varepsilon_{it0} \right) \right\}.
\]

We have

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_{i-1}^0 (L) \Delta \alpha_{it}^2 = O_p \left( T^{-2} N \sum_{t=1}^{T} (t^{1-\delta_0} + 1)^2 \right)
= O_p \left( T^{-2} N (T^{3(1-\delta_0)} + T) \right)
= o_p \left( N (T^{-2} + T^{-1}) \right),
\]

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

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} 2 \left( \varepsilon_{it} - \tau_i^0 \varepsilon_{it0} \right) \lambda_{i-1}^0 (L) \Delta \alpha_{it} = O_p \left( T^{-1} N \sum_{t=1}^{T} (t^{1-\delta_0} + 1) \right)
= O_p \left( T^{-1} N (T^{2-\delta_0} + T) \right),
\]

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

\[
\frac{1}{S_{\tau \tau T}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tau_i^0 \left( \lambda_{i-1}^0 (L) \Delta \alpha_{it} \right) = O_p \left( S_{\tau \tau T}^{0-1} T^{-2} N \left( \sum_{t=1}^{T} t^{-\delta_0} (t^{1-\delta_0} + 1)^2 \right) \right)
= O_p \left( N (T^{2(1-\delta_0)} + \log T) \right)
= O_p \left( S_{\tau \tau T}^{0-1} N \left( T^{2-2\delta_0} + T^{-2} \log T \right) \right)
= O_p \left( N (T^{1-2\delta_0} + N T^{2(1-\delta_0)} + N T^{-2} \log^2 T) \right).
\]
so its contribution to \( \hat{\sigma}^2 (\theta_0) \) is \( O_p \left( T^{-2\delta_0} + T^{1-4\delta_0} + T^{-3}\log^2 T \right) \), as is that of

\[
\frac{2}{S_{zt}^2 T^2} \sum_{i=1}^N \left( t_{1t}^0 \left( \epsilon_{it}^0 - \tau_{1t}^0 \epsilon_{i0}^0 \right) + \lambda_{t-1}^0 \right) \Delta \alpha_{it}.
\]

Overall, the contribution of the extra terms in \( \hat{\sigma}^2 (\theta_0) \) to \( (NT)^{1/2} \left( \partial/\partial \theta \right) L_T^p (\theta_0) \) is

\[
O_p \left( (NT)^{1/2} \left( T^{-\delta_0} + T^{1-4\delta_0} + T^{-1} \right) \right) = O_p \left( N T \left( T^{-1-2\delta_0} + T^{3-8\delta_0} + T^{-1} \right) \right)^{1/2} = O_p (1),
\]

since \( NT^{-1-2\delta_0} + NT^{-1} \to 0 \) and \( \delta_0 > \frac{1}{2} \), and they do not contribute to the asymptotic bias. With the notation \( [a_i]^T \) meaning the \( T \times 1 \) column vector with \( i \)th element \( a_i \), with respect to the contribution of \( \frac{\partial}{\partial \theta} \hat{\sigma}^2 (\theta) \) to \( \left( \partial/\partial \theta \right) L_T^p (\theta_0) \) we have to consider the extra terms in \( z_{it} (\theta_0) = [z_{it} (\theta_0)]^T \) depending on \( \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right),

\[
\frac{2}{NT} \sum_{i=1}^N \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T \Omega_T^{-1} (\theta_0) \Omega_T^1 (\theta_0) \Omega_T^{-1} (\theta_0) \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T
\]

and those in \( \tilde{z}_{it} (\theta_0) = [\tilde{z}_{it} (\theta_0)]^T \) depending on \( \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right),

\[
\frac{2}{NT} \sum_{i=1}^N \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T \Omega_T^{-1} (\theta_0) \Omega_T^1 (\theta_0) \Omega_T^{-1} (\theta_0) \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T + \frac{2}{NT} \sum_{i=1}^N \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T \Omega_T^{-1} (\theta_0) \Omega_T^1 (\theta_0) \Omega_T^{-1} (\theta_0) \left[ \epsilon_{it}^0 - \tau_{1t}^0 \epsilon_{i0}^0 \right]^T.
\]

Since \( \Omega_T^{-1} (\theta_0) = I_T - \frac{\theta_i^0}{\theta_t^0} \Omega_T^{-1} (\theta_0) = \theta_i^0 \Omega_T^{-1} (\theta_0) + \theta_t^0 \Omega_T^{-1} (\theta_0) \), where \( \theta_i^0 = \theta_i (\theta_0), \theta_t^0 = \left( \theta_{t_1}^0, \ldots, \theta_{t_T}^0 \right) \) and \( \theta_t^0 = \left( \partial/\partial \theta \right) \theta_t (\theta_0) = O \left( k^{-\delta_0} \log k \right) \) as \( k \to \infty \), the first term is

\[
\frac{2}{NT} \sum_{i=1}^N \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T \Omega_T^{-1} (\theta_0) \Omega_T^1 (\theta_0) \Omega_T^{-1} (\theta_0) \left[ \lambda_{t-1}^0 \left( \Delta \alpha_{it} \right) \right]^T
\]

\[
= O_p \left( T^{-3} \sum_{t} \left( t^{-1-\delta_0} + 1 \right) t^{-\delta_0} \sum_{t} \left( t^{-1-\delta_0} + 1 \right) t^{-\delta_0} \right)
\]

\[
+ O_p \left( T^{-3} \sum_{t} \left( t^{1-\delta_0} + 1 \right) t^{-\delta_0} \right)
\]

\[
= O_p \left( T^{-3} (T^{2(1-\delta_0)} + \log T)^2 \log T \right)
\]

\[
= O_p \left( T^{1-3\delta_0} \log T + T^{-3} \log^3 T \right),
\]

while using

\[
E \left[ \sum_{t=1}^T \epsilon_{1t} \right] \leq E \left[ \sum_{t=1}^T \epsilon_{1t}^0 \right]^{1/2} = [\sigma_0^2 (S_{zt}^0 T - 1)]^{1/2} = O \left( T^{1-\delta_0} + \log^2 T \right),
\]

and similarly

\[
E \left\| \sum_{t=1}^T \epsilon_{1t} \right\| = O \left( T^{1-\delta_0} \log T + \log T \right),
\]

\[
E \left\| \sum_{t=1}^T \epsilon_{1t}^0 \right\| = O \left( S_{zt}^0 \right) = O \left( T^{1-2\delta_0} \log T + \log^2 T \right),
\]
the second one is

\[
\frac{2}{NT} \sum_{i=1}^{N} \left[ \lambda^0_{t-1} (L) \Delta \alpha_{it} \right]^T \Omega^{-1} (\theta_0) \Omega^T (\theta_0) \Omega^{-1} (\theta_0) \left[ \epsilon_{it} - \epsilon_{t0} \right]_T
\]

\[
= O_p \left( T^{-2} \sum_t (t^{1-\delta_0} + 1) t^{-\delta_0} \left( T^{1-2\delta_0} \log T + \log^2 T \right) \right)
\]

\[
+ O_p \left( T^{-2} \sum_{t \neq T} (t^{1-\delta_0} + 1) t^{-\delta_0} \left( T^{1-2\delta_0} \log T + \log^2 T \right) \right)
\]

\[
= O_p \left( T^{-2} \left( T^{2(1-\delta_0)} + \log T \right) \left( T^{1-2\delta_0} \log T + \log^2 T \right) \right)
\]

\[
= O_p \left( T^{1-4\delta_0} \log T + T^{-2} \log^2 T \right).
\]

Their joint contribution to \((NT)^{1/2} \left( \partial / \partial \theta \right) L_p^p (\theta_0)\) is thus

\[
O_p \left( (NT)^{1/2} \left( T^{1-4\delta_0} \log T + T^{-2} \log^2 T \right) \right)
\]

\[
= O_p \left( (N \left( T^{3-8\delta_0} \log^2 T + T^{-3} \log^4 T \right))^{1/2} \right).
\]

which is \(O_p(1)\) under \(NT^{1-2\delta_0} \log^2 T + NT^{-1} \rightarrow 0\) and \(\delta_0 > \frac{1}{2}\). The extra terms due to the new element in \(\hat{z}_t (\theta_0) = f_{it} - \epsilon_{t0} + \lambda^0_{t-1} (L) \Delta \alpha_{it}\), where \(f_{it}\) is the ith row of \(f_t\), are

\[
\frac{2}{NT} \sum_{i=1}^{N} \left( \sum_{t=1}^{T} (\lambda^0_{t-1} (L) \Delta \alpha_{it}) \right) z_{it} (\theta_0) + \hat{z}_t (\theta_0) \left( \lambda^0_{t-1} (L) \Delta \alpha_{it} \right)
\]

\[
- \frac{2}{NT \delta_0 + T} \sum_{i=1}^{N} \left( \sum_{t=1}^{T} (\lambda^0_{t-1} (L) \Delta \alpha_{it}) \right) \left( \sum_{t=1}^{T} z_{it} (\theta_0) \tau^0_t \right)
\]

\[
+ \left( \sum_{t=1}^{T} \hat{z}_t (\theta_0) \tau^0_t \right) \left( \sum_{t=1}^{T} (\lambda^0_{t-1} (L) \Delta \alpha_{it}) \tau^0_t \right).
\]

(51)

Using similar methods as before, the first term is

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

Next, since

\[
E \left[ \sum_{t=1}^{T} z_{it} (\theta_0) \tau^0_t \right] = O \left( S^0_{t \neq T} \right) = O \left( T^{1-2\delta_0} + \log T \right),
\]

\[
E \left[ \sum_{t=1}^{T} \hat{z}_t (\theta_0) \tau^0_t \right] = O \left( S^0_{t \neq T} \log T \right),
\]

the other terms are

\[
O_p \left( \frac{1}{T^2 S^0_{t \neq T}} \left( \sum_{t=1}^{T} (t^{1-\delta_0} + 1) t^{-\delta_0} \log t \right) S^0_{t \neq T} + S^0_{t \neq T} \left( \sum_{t=1}^{T} (t^{1-\delta_0} + 1) t^{-\delta_0} \right) \right)
\]

\[
= O_p \left( \frac{1}{T^2} \left( \left( T^{2(1-\delta_0)} + \log T \right) \log T + \left( T^{2(1-\delta_0)} + \log T \right) \log T \right) \right)
\]

\[
= O_p \left( T^{-2\delta_0} + T^{-2} \log T \right).
\]

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

\[
O_p \left( (NT)^{1/2} \left( T^{2\delta_0} + T^{-2} \log T \right) \log T \right)
\]

\[
= O_p \left( \left( N \left( T^{1-4\delta_0} + T^{-3} \log^2 T \right) \log^2 T \right)^{1/2} \right),
\]

which is \(O_p(1)\) since \(NT^{1-2\delta_0} \log^2 T + NT^{-1} \rightarrow 0\) for \(\delta_0 > \frac{1}{2}\). This completes the proof when (14) and (15) are not imposed.
Now imposing (14) and (15) and allowing $\delta > \frac{3}{8}$, we have first to reanalyze two terms in $\hat{\sigma}^2_T(\theta_0)$ depending on $\Delta \alpha_i$ in the normalized score based on $L^D_T$, given in (48). Now

$$
\sum_{i=1}^{N} \sum_{t=1}^{T} 2 \left( \varepsilon_{it} - \tau_0 \varepsilon_{i0} \right) \lambda^0_{t-1} (L) \Delta \alpha_i
$$

has zero mean and variance bounded by

$$
K \frac{N ||\Sigma||}{T^2} \left( \sum_{t=1}^{T} (t^{1-\delta_0} + 1)^2 + \sum_{t=1}^{T} (t^{1-\delta_0} + 1) t^{-\delta_0} \right)^2
$$

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

$$
\sum_{i=1}^{N} \sum_{t=1}^{T} \left( \varepsilon_{it} - \tau_0 \varepsilon_{i0} \right) \lambda^0_{t-1} (L) \Delta \alpha_i
$$

has zero mean and variance bounded by

$$
\frac{2}{S^0_T \sum_{i=1}^{N} \sum_{t=1}^{T} (t^{1-\delta_0} + 1)^2 + \frac{KN}{T^2 S^0_T \sum_{i=1}^{N} \sum_{t=1}^{T} (t^{1-\delta_0} + 1) t^{-\delta_0} \right)^2}
$$

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

$$
O_p \left( \left( N T^{1-4\delta_0} \log^2 T + \log^2 T \right) \right)
$$

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

$$
\frac{2}{NT} \sum_{i=1}^{N} \left[ \lambda^0_{t-1} (L) \Delta \alpha_i \right]^{\top} \Omega^{-1}_T (\theta_0) \Omega_1 \Omega^{-1}_T (\theta_0) \left[ \varepsilon_{it} - \tau_0 \varepsilon_{i0} \right]^{\top},
$$

which has zero mean and variance

$$
O \left( \left( N^{-1} T^{-4} \sum_{t=1}^{T} (t^{1-\delta_0} + 1) t^{-\delta_0} \right) \right)
$$

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

$$
O_p \left( N^{-1} \left( T^{-2\delta_0} \log^2 T + T^{-4\delta_0} \log^4 T \right) \right)
$$

which is $O_p \left( 1 \right)$ since $NT^{1-2\delta_0} \log^2 T + NT^{-1} \log^2 T \to 0$ and $\delta_0 > \frac{3}{8}$.
has zero mean and variance
\[
O \left( N^{-1} T^{-4} \left( \sum_{t=1}^{T} (t^{1-\delta_0} + 1) \log T + \left( \sum_{t=1}^{T} (t^{1-\delta_0} + 1) t^{-\delta_0} \log T \right)^2 \right) \right)
\]
\[
= O \left( N^{-1} T^{-4} \left( (T^{1-\delta_0} + T) \log T + \left( (T^{2-2\delta_0} + \log T) \log T \right)^2 \right) \right)
\]
\[
= O \left( N^{-1} \left( (T^{1-\delta_0} + T^{-3}) \log T + \left( (T^{2-2\delta_0} + T^{-2} \log T) \log T \right)^2 \right) \right),
\]
and a similar result holds for the term depending on \((\bar{e}_{it} - \tau_i^0 \hat{e}_{it}) \left( \hat{\lambda}_{t-1} (L) \Delta \alpha_{it} \right)\), so the contribution to the standardized score of this first term is
\[
O_p \left( N \left( (T^{1-4\delta_0} + T^{-2}) \log^2 T \right)^{1/2} \right)
\]
\[
+ O_p \left( \left( (T^{-2\delta_0} + T^{-2}) \log^2 T \right)^{1/2} \right) + O \left( \left( (T^{1-4\delta_0} + T^{-3} \log^2 T) \log^2 T \right)^{1/2} \right),
\]
which is \(o_p(1)\) since \(NT^{1-4\delta_0} \log^3 T + NT^{-1} \log^2 T \to 0\) for \(\delta_0 > \frac{3}{8}\), while the second term is shown \(o_p(1)\) as before. □

**Proof of Theorem 4.2**

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

**References**


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


