Persistence Heterogeneity Testing in Panels with Interactive Fixed Effects

Yunus Emre Ergemen and Carlos Velasco

CREATES Research Paper 2018-11

Persistence Heterogeneity Testing in Panels with Interactive Fixed Effects

Yunus Emre Ergemen[*]   Carlos Velasco[†]

March 9, 2018

Abstract

We consider large N, T panel data models with fixed effects, a common factor allowing for cross-section dependence, and persistent data and shocks, which are assumed fractionally integrated. In a basic setup, the main interest is on the fractional parameter of the idiosyncratic component, which is estimated in first differences after factor removal by projection on the cross-section average. The pooled conditional-sum-of-squares estimate is $\sqrt{NT}$ consistent but the normal asymptotic distribution might not be centered, requiring the time series dimension to grow faster than the cross-section size for correction. We develop tests of homogeneity of dynamics, including the degree of integration, that have no trivial power under local departures from the null hypothesis of a non-negligible fraction of cross-section units. A simulation study shows that our estimates and test have good performance even in moderately small panels.

Mathematics Subject Classification: 60G22, 62F05, 62F10, 62F12, 62H25, 62M10

JEL Classification: C22, C23

Keywords: Fractional integration, panel data, factor models, long memory, homogeneity test.

1. Introduction

The classical approach in the panel data literature for modeling persistence uses autoregressive (AR) models under different panel specifications. For example, Arellano and Bond (1991) and Alvarez and Arellano (2003) develop asymptotics for their estimators based on AR models characterized by homogeneous slope parameters under large N and fixed or increasing T. Pesaran (2007) uses a dynamic panel setup to test for panel unit roots allowing for slope heterogeneity when N and T are both large. Han and Phillips (2010) obtain inference in an AR(1) panel based on generalized method of moments (GMM) estimation with strong instruments when the homogeneous AR parameter is near unity.

There is a recent literature on large N, T panel data models with long-range dependence constituting an alternative to AR specifications, see e.g. Robinson and Velasco (2015), Ergemen and Velasco (2017) and Ergemen (2017). Like models with AR dynamics, these models nest $I(1)$ behaviour, but smoothly and thus the estimates of long-range dependence parameters are asymptotically normal unlike the nonstandard asymptotics under nonstationary AR specifications. This leads to chi-squared test statistics for memory parameter estimates and, for instance, panel unit-root testing can be readily performed based on the memory estimates.

Although classical AR panel literature uses both homogeneous and heterogeneous AR parameter specifications in modeling persistence, little or no justification is provided for these choices. In fractional panel data literature, Ergemen and Velasco (2017) and Ergemen (2017) allow memory parameters under their setups to vary across cross-section units, providing valid arguments that this allows for a greater flexibility in suiting the presentation of data as well as describing the dynamics of different units more accurately rather than restricting all units to have the same dynamics. However, a formal testing procedure for persistence homogeneity has not yet been provided in the fractional panel data literature either. Developing such a test is important because when there is no statistically significant discrepancy between the integration orders of different cross-section units, it is preferable to employ pooled memory estimates that enjoy faster ($\sqrt{NT}$) convergence rates and avoid the curse of dimensionality as N increases.

To fill in this gap, this paper develops a testing framework for persistence homogeneity when interactive fixed effects are also present. In doing so, we first present a rigorous treatment for a panel data model that allows for fractionally integrated long-range dependence in both idiosyncratic shocks and a common-factor structure that accounts for cross-section dependence. In the model, persistence is described by a memory or fractional integration parameter, constituting an alternative to dynamic autoregressive (AR) panel data models. The setup we consider requires that both the number of cross section units, N, and the length of the time series, T, grow in the asymptotics, departing from the case of multivariate time series (with N fixed) or short panels (with T fixed). Our setup differs from Hassler et al. (2011) and Robinson and Velasco (2015) in that we model cross-section dependence employing an unobservable common factor structure that can be serially correlated and display long-range dependence, which makes the model more general by introducing cross-section

dependence without further structural impositions on the idiosyncratic shocks.

Using a type-II fractionally integrated panel data model with fixed effects and cross-section dependence modeled through a common factor dependence, we allow for long-range persistence through this factor and the integrated idiosyncratic shock. The model assumes a common set of parameters for the dynamics of the idiosyncratic component of all cross-sectional units. We deal with the fixed effects and the unobservable common factor through first differencing and projection on the cross-section average of the differenced data as a proxy for the common factor, respectively. Then, estimation of the memory parameter is based on a pooled conditional sum of squares (CSS) criterion function of the projection residuals which produces estimates asymptotically equivalent to Gaussian ML estimates. We require to impose conditions on the rate of growth of N and T to control for the projection error and for an initial condition bias induced by first differencing of the type-II fractionally integrated error terms, so that our pooled estimate can achieve the $\sqrt{NT}$ convergence rate. We nevertheless discuss bias correction methods that relax the restriction that T should grow substantially faster than N in the joint asymptotics, which would not affect the estimation of the heterogeneous model. We also propose a heterogenous model whose estimates, adapting the same CSS criterion to have heterogeneous dynamics, are shown to be asymptotically normal for each cross section unit as T gets larger.

We then develop LM tests for dynamics homogeneity that are similar to the slope homogeneity tests of Su and Chen (2013) in panel data models with interactive fixed effects and that allow for scale heterogeneity. In particular, these tests include the null of homogenous degree of integration and have no trivial power under local departures from the null hypothesis of a non-negligible fraction of cross-section units.

We explore the performance of our estimates and tests via Monte Carlo experiments, which indicate that our inference methods have good small-sample properties in a variety of scenarios.

Next section details the model and necessary assumptions. Section 3 explains the estimation strategy for both homogeneous and heterogeneous parameters, and discusses the asymptotic behaviour of the estimates. Section 4 shows the details of persistence homogeneity testing along with the asymptotics. Section 5 presents Monte Carlo simulations. Finally, Section 6 concludes the paper.

Throughout the paper, we use the notation $(N, T)_j$ to denote joint cross-section and time-series asymptotics, $\rightarrow_p$ to denote convergence in probability and $\rightarrow_d$ to denote convergence in distribution. All mathematical proofs and technical lemmas are collected in the appendix.

2. The Model

In this section, we detail a type-II fractionally integrated panel data model with fixed effects and cross-section dependence and list our assumptions. We consider that the observable y_{it}

satisfy

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

for $t = 0, 1, \ldots, T$, $i = 1, \ldots, N$, where $\varepsilon_{it} \sim iid(0, \sigma^2)$ are idiosyncratic shocks; $\theta_0 \in \Theta \subset \mathbb{R}^{p+1}$ is a $(p+1) \times 1$ parameter vector; L is the lag operator and for any $\theta \in \Theta$ and for each $t \geq 0$,

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

truncates $\lambda\left(L;\theta\right) = \lambda_\infty\left(L;\theta\right)$. We assume that $\lambda\left(L;\theta\right)$ has this 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, $\theta = (\delta, \xi')'$. Here $\Delta = 1 - L$, so that the fractional filter Δ^δ has the expansion

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

and denote the truncated version as $\Delta_t^\delta = \sum_{j=0}^{t-1} \pi_j(\delta) L^j$, with $\Gamma\left(-\delta\right) = (-1)^\delta \infty$ for $\delta = 0, 1, \ldots$, $\Gamma\left(0\right)/\Gamma\left(0\right) = 1$; $\psi\left(L;\xi\right)$ is a known function such that for complex-valued x, $\left|\psi\left(x;\xi\right)\right| \neq 0$, $|x| \leq 1$ and in the expansion

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

the coefficients $\psi_j\left(\xi\right)$ satisfy

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

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

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

behaves asymptotically as $\pi_j(\delta)$,

$$\lambda_j\left(\theta\right) = \psi\left(1;\xi\right) \pi_j(\delta) + O\left(j^{-\delta-2}\right), \quad \text{as } j \to \infty,$$

see Robinson and Velasco (2015), where

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

so the value of δ_0 determines the asymptotic stationarity $(\delta_0 < 1/2)$ or nonstationarity $(\delta_0 \geq 1/2)$ of $y_{it} - \alpha_i - \gamma_i f_t$ and $\psi(L;\xi)$ describes short memory dynamics.

The α_i are unobservable fixed effects, γ_i unobservable factor loadings and f_t is the unobservable common factor that is assumed to be an $I(\varrho)$ process, where we treat ϱ as a nuisance parameter. This way the model incorporates heterogeneity through α_i as well as γ_i and also introduces account cross-section dependence by means of the factor structure, $\gamma_i f_t$, which was not considered in Robinson and Velasco (2015). When we write (1) as

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

where $1(\cdot)$ is the indicator function, the memory of the observed y_{it} is $\max\{\delta_0, \varrho\}$, where f_t could be the major source of persistence in data. The model could be complemented with the presence of incidental trends and other exogenous or endogenous observable regressor series, see Ergemen and Velasco (2017) and Ergemen (2017).

The model can be reorganized in terms of the variable $\Delta_t^{\delta_0} y_{it}$ for $i = 1, \ldots, N$, and $t = 1, \ldots, T$ and when $\psi(L;\xi_0) = 1 - \xi_0 L$ corresponds to a finite AR(1) polynomial as

$$\Delta_t^{\delta_0} y_{it} = (1 - \xi_0)\Delta_t^{\delta_0}\alpha_i + \xi_0 \Delta_t^{\delta_0} y_{it-1} + \gamma_i(1 - \xi_0 L)\Delta_t^{\delta_0} f_t + \varepsilon_{it},$$

which is then easily comparable to a standard dynamic AR(1) panel data model with cross-section dependence, e.g. that of Han and Phillips (2010),

$$y_{it} = (1 - \rho)\alpha_i + \rho y_{it-1} + \gamma_i f_t + \varepsilon_{it}.$$

In both models, error terms are iid, and there are fixed effects (so long as $\delta_0 \neq 1$, $\xi_0 \neq 1$ and $\rho \neq 1$). However, autoregressive panel data models can only cover a limited range of persistence levels, just $I(0)$ or $I(1)$ series depending on whether $|\rho| < 1$ or $\rho = 1$. On the other hand, the fractional model (1) covers a wide range of persistence levels depending on the values of δ_0 and ϱ, including the unit root case and beyond. In addition, (1) accounts for persistence in cross-section dependence depending on the degree of integration of $\Delta_t^{\delta_0} f_t$.

We are interested in conducting inference on θ, in particular on δ. For the analysis in this paper we require that both N and T increase simultaneously due to presence of the unobserved common factor and the initial condition term in the fractional difference operator, unlike in Robinson and Velasco (2015), who only require T to grow in the asymptotics, while N could be constant or diverging simultaneously with T. In the first part of the paper we assume a common vector parameter, including a common integration parameter δ, for all cross-section units $i = 1, \ldots, N$. While the fractional integration parameter may as well be allowed to be heterogeneous, our approach is geared towards getting a pooled estimate for the entire panel exploiting potential efficiency gains. Further, this pooling has to control for potential distortions due to common factor elimination, that, as well as fixed effects removal, lead to some bias in the asymptotic distribution of parameter estimates, cf. Robinson and Velasco (2015).

We use the following assumptions throughout the paper:

Assumption A.

A.1. The idiosyncratic shocks, ε_{it}, $i = 1, 2, \ldots, N$, $t = 0, 1, 2, \ldots, T$ are independently and identically distributed both across i and t with zero mean and variance σ^2, and have a finite fourth-order moment, and $\delta_0 \in (0, 3/2)$.

A.2. The $I(\varrho)$ common factor is $f_t = \Delta_t^{-\varrho} z_t^f$, $\varrho < 3/2$, where $z_t^f = \varphi^f(L) v_{t-k}^f$ with $\varphi^f(s) = \sum_{k=0}^{\infty} \varphi_k^f s^k$, $\sum_{k=0}^{\infty} k|\varphi_k^f| < \infty$, $\varphi^f(s) \neq 0$ for $|s| \leq 1$, and $v_t^f \sim iid(0, \sigma_f^2)$, $E|v_t^f|^4 < \infty$.

A.3. ε_{it} and f_t are independent of the factor loadings γ_i, and are independent of each other for all i and t.

A.4. Factor loadings γ_i are independently and identically distributed across i, $\sup_i E|\gamma_i| < \infty$, and $\bar{\gamma} = N^{-1} \sum_{i=1}^{N} \gamma_i \neq 0$.

A.5. For $\xi \in \Xi$, $\psi(x; \xi)$ is differentiable 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, and (3) holds for all $\xi \in \Xi$ with $c(\xi)$ satisfying

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

Assumption A.1 implies that the idiosyncratic errors $\lambda^{-1}(L; \theta) \varepsilon_{it}$, are fractionally integrated with asymptotically stationary increments, $\delta_0 < 3/2$, which will be exploited by our projection technique. The homoskedasticity assumption on idiosyncratic shocks, ε_{it}, is not restrictive since y_{it} are still heteroskedastic as α_i and γ_i vary in each cross section.

By Assumption A.2, the common factor f_t is a zero mean fractionally integrated $I(\varrho)$ linear process, with the $I(0)$ increments possibly displaying short-range serial dependence but with positive and smooth spectral density at all frequencies. The zero mean assumption is not restrictive since we are allowing for fixed effects α_i which are not restricted in any way. Restrictions similar to Assumption A.2 have been used under different setups in e.g. Hualde and Robinson (2011) and Nielsen (2014). Under Assumption A.2, the range of persistence for the common factor covers unit root and beyond, making the model a powerful tool for several practical problems. Although we treat ϱ as a nuisance parameter, in empirical applications this parameter could be estimated based on the cross-section average of the observed series using semiparametric estimates, e.g. with a local Whittle approach. Assumption A.3 and A.4 are standard identifying conditions in one-factor models as also used in e.g. Pesaran (2006) and Bai (2009). In particular, the condition on $\bar{\gamma}$ is related to Assumption 5(b) of Pesaran (2006) and used to guarantee that our projection to remove factors works in finite samples.

Assumption A.5 ensures that $\psi(L; \xi)$ is smooth for $\xi \in \Xi$, and the weights ψ_j lead to short-memory dynamics as is also assumed by Robinson and Velasco (2015), where the parameter space Ξ can depend on stationarity and invertibility restrictions on $\psi(L; \xi)$.

3. Parameter Estimation

Bai (2009) and Pesaran (2006), among many others, study the estimation of panel data

6

models with cross-section dependence. Bai (2009) estimates the slope parameter in an interactive fixed effects model where the regressors and the common factor are stationary and idiosyncratic shocks exhibit no long-range dependence. Likewise, Pesaran (2006) estimates the slope parameter in a multifactor panel data model where covariates are $I(0)$. In this section we focus on the estimation of the parameter vector θ that describes the idiosyncratic dynamics of data, including the degree of integration.

In our estimation strategy, we first project out the unobserved common structure using sample averages of first-differenced data as proxies, where the fixed effects are readily removed by differencing. We then use a pooled conditional-sum-of-squares (CSS) estimation on first differences based on the remaining errors after projection.

First-differencing (1) to remove α_i, we get

$$\Delta y_{it} = \gamma_i \Delta f_t + \Delta \lambda_t^{-1}(L; \theta_0)\, \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, 2, \ldots, T,$$

where we denote by θ_0 the true parameter vector, and then Δy_{it} is projected on the cross-section average $\Delta \bar{y}_t = N^{-1} \sum_{i=1}^{N} \Delta y_{it}$ as (non-scaled) proxies for Δf_t with the projection coefficient $\hat{\phi}_i$ given by

$$\hat{\phi}_i = \frac{\sum_{t=1}^{T} \Delta \bar{y}_t \Delta y_{it}}{\sum_{t=1}^{T} (\Delta \bar{y}_t)^2},$$

which we assume can be computed for every i with $\sum_{t=1}^{T} (\Delta \bar{y}_t)^2 > 0$. Then we compute the residuals

$$\varepsilon_{it}(\theta) = \lambda_{t-1}\left(L; \theta^{(-1)}\right) \left(\Delta y_{it} - \hat{\phi}_i \Delta \bar{y}_t\right), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

where $\theta^{(-1)} = (\delta - 1, \xi')'$ adapts to the previous differencing initial step.

Then we denote by $\hat{\theta}$ the estimate of the unknown true parameter vector θ_0,

$$\hat{\theta} = \arg \min_{\theta \in \Theta} L_{N,T}(\theta),$$

where we assume Θ is compact and $L_{N,T}$ is the CSS of the projection residuals after fractional differencing

$$L_{N,T}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}(\theta)^2,$$

which is the relevant part of the concentrated (out of σ^2) Gaussian likelihood for $\varepsilon_{it}(\theta)$.

Note that after the first-differencing transformation to remove α_i, there is a mismatch between the sample available ($t = 1, 2, \ldots, T$) and the length of the filter $\lambda_{t-1}\left(L; \theta^{(-1)}\right)$ that can be applied to it, with the filter $\Delta \lambda_t^{-1}(L; \theta_0)$ that generates the data, since for instance

$$\lambda_{t-1}\left(L; \theta^{(-1)}\right) \Delta \lambda_t^{-1}(L; \theta_0)\, \varepsilon_{it} = \lambda_t(L; \theta)\, \lambda_t^{-1}(L; \theta_0)\, \varepsilon_{it} - \lambda_t\left(\theta^{(-1)}\right) \varepsilon_{i0},$$

because $\lambda_t\left(L; \theta^{(-1)}\right) \Delta = \lambda_t(L; \theta)$, $t = 0, 1, \ldots$. Even when $\theta = \theta_0$, all residuals involve ε_{i0}, i.e. the initial condition, which is reflected in a bias term of $\hat{\theta}$ as in Robinson and Velasco (2015).

7

The estimates are only implicitly defined and entail optimization over $\Theta = \mathcal{D} \times \Xi$, where Ξ is a compact subset of $\mathbb{R}^p$ and $\mathcal{D} = [\underline{\delta}, \overline{\delta}]$, with $0 < \underline{\delta} < \overline{\delta} < 3/2$. We aim to cover a wide range of values of $\delta \in \mathcal{D}$ with our asymptotics, c.f. Nielsen (2014) and Hualde and Robinson (2011), but there are interactions with other model parameters that might require to restrict the set $\mathcal{D}$ reflecting some a priori knowledge on the true value of δ or to introduce further assumptions on N and T. In particular, and departing from Robinson and Velasco (2015), it is essential to consider the interplay of ϱ and δ_0, i.e. the memories of the unobservable common factor and of the idiosyncratic shocks, respectively, since projection on cross-section averages of first differenced data is assuming that Δf_t is (asymptotically) stationary, but possibly with more persistence than the idiosyncratic components.

Then, for the asymptotic analysis of the estimate of θ, we further introduce the following assumptions.

Assumption B. The lower bound $\underline{\delta}$ of the set $\mathcal{D}$ satisfies

$$\max\left\{\varrho, \delta_0\right\} - 1/2 < \underline{\delta} \leq \delta_0. \tag{6}$$

Assumption B indicates that if the set $\mathcal{D}$ is quite informative on the lower possible value of δ_0 and this is not far from ϱ, the CSS estimate is consistent irrespective of the relationship between N and T, as we show in our first result.

Theorem 1. *Under Assumptions A and B, $\theta_0 \in \Theta$, and as $(N,T)_j \to \infty$,*

$$\hat{\theta} \to_p \theta_0.$$

Although the sufficient condition in Assumption B may seem restrictive, the lower bound could be adapted accordingly to meet the distance requirement from ϱ and δ_0 using information on the whereabouts of these parameters. This assumption may be relaxed at the cost of restricting the relative rates of growth of N and T in the asymptotics. In the technical appendix, we provide more general conditions that are implied by Assumption B to prove this result.

A similar result of consistency for CSS estimates is provided by Hualde and Robinson (2011) and Nielsen (2014) for fractional time series models and in Robinson and Velasco (2015) for fractional panels without common factors. Note that the theorem only imposes that both N and T grow jointly, but there is no restriction on their rate of growth when (6) holds. This contrasts with the results in Robinson and Velasco (2015), where only T was required to grow and N could be fixed or increasing in the asymptotics. An increasing T therein is required to control for the initial condition contribution due to first differencing for fixed effects elimination, as is needed here, but projection on cross-section averages for factor removal further requires that both N and T grow.

Next, we establish the asymptotic distribution of the parameter estimates, for which we assume that $\psi(L; \xi)$ is twice continuously differentiable for all $\xi \in \Xi$ with $\dot{\psi}_t(L; \xi) =$

$(d/d\xi)\psi_t(L;\xi)$ where it is assumed that $\left|\dot\psi_t(L;\xi)\right| = O\left(\exp(-c(\xi)j)\right)$. In establishing the asymptotic behaviour, the most delicate part is formulating the asymptotic bias. The initial condition (IC) bias of $(NT)^{1/2}\left(\hat\theta - \theta_0\right)$ is proportional to $T^{-1}\nabla_T(\theta_0)$, where

$$\nabla_T(\theta_0) = -\sum_{t=1}^T \tau_t(\theta_0)\left\{\dot\tau_t(\theta_0) - \chi_t(\xi_0)\right\}$$

where $\tau_t(\theta) = \lambda_t\left(\theta^{(-1)}\right) = \lambda_t(L;\theta)\,1 = \sum_{j=0}^t \lambda_j(\theta)$, $\dot\tau_t(\theta) = (\partial/\partial\theta)\tau_t(\theta)$ and χ_t is defined by

$$\chi(L;\xi) = \frac{\partial}{\partial\theta}\log\lambda(L;\theta) = (\log\Delta,(\partial/\partial\xi')\log\psi(L;\xi))' = \sum_{j=1}^\infty \chi_j(\xi)\,L^j.$$

The term $\nabla_T(\theta_0)$, depending only on the unknown θ_0 and T, also found in Robinson and Velasco (2015), appears because of the data-index mismatch that arises due to time truncation for negative values and first differencing.

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. For the asymptotic distribution analysis we further require the following conditions.

Assumption C.

C.1. As $(N,T)_j \to \infty$,
$$\frac{N}{T}\log^2 T + \frac{T}{N^3} \to 0.$$

C.2. $\max\{1/4, \varrho - 1/2\} < \delta_0 \le \min\{5/4, 5/2 - \varrho\}$.

The next result shows that the fractional integration parameter estimate is asymptotically normal and efficient at the $\sqrt{NT}$ convergence rate.

Theorem 2. *Under Assumptions A, B and C, $\theta_0 \in Int(\Theta)$, as $(N,T)_j \to \infty$,*

$$(NT)^{1/2}\left(\hat\theta - \theta_0 - T^{-1}B^{-1}(\xi_0)\nabla_T(\theta_0)\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right),$$

where $\nabla_T(\theta_0) = O(T^{1-2\delta_0}\log T\,1\{\delta_0 < \frac{1}{2}\} + \log^2 T\,1\{\delta_0 = \frac{1}{2}\} + 1\{\delta_0 > \frac{1}{2}\})$.

Corollary 1. *Under Assumptions of Theorem 2,*

$$(NT)^{1/2}\left(\hat\theta - \theta_0\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_0)\right)$$

for $\delta_0 > \frac{1}{2}$, and this also holds when $\delta_0 \in \left(\frac{1}{3}, \frac{1}{2}\right)$ if additionally, as $(N,T)_j \to \infty$, $NT^{1-4\delta_0}\log^2 T \to 0$, and when $\delta_0 = \frac{1}{2}$ if $NT^{-1}\log^4 T \to 0$.

These results parallel Theorem 5.3 in Robinson and Velasco (2015) additionally using Assumption C to control for the projection errors and requiring N to grow with T to remove the cross-sectionally averaged error terms, while the range of allowed values of δ_0 is limited in the same way. Assumption C.1 basically requires that T grows faster than N, but slower than N^3, so that different projection errors are not dominating to achieve the $\sqrt{NT}$ rate of convergence. This last restriction is milder than the related conditions that impose $TN^{-2} \to 0$ for slope estimation, e.g. Pesaran (2006), but we also need T to grow faster than N to control the initial condition bias.

Condition C.2 is only a sufficient condition basically requiring that the overall memory, $\varrho + \delta_0$, be not too large so that common factor projection with first-differenced data works well, especially if N grows relatively fast with respect to T, and that ϱ is not much larger than δ_0, so the common factor distortion can be controlled for. We relax these sufficient conditions in the technical appendix to prove our results.

The asymptotic centered normality of the uncorrected estimates further requires that $\delta_0 > \frac{1}{3}$ in view of Assumption C.1, so it is interesting for statistical inference purposes to explore a bias correction. Let $\tilde{\theta}$ be the fractional integration parameter estimate with IC bias correction constructed by plugging in the uncorrected estimate $\hat{\theta}$,

$$\tilde{\theta} = \hat{\theta} - T^{-1}B^{-1}\left(\hat{\xi}\right)\nabla_T(\hat{\theta}).$$

The next result shows that the bias-corrected estimate is asymptotically centered and efficient at the $\sqrt{NT}$ convergence rate.

Corollary 2. *Under Assumptions of Theorem 2,*

$$(NT)^{1/2}\left(\tilde{\theta} - \theta_0\right) \to_d \mathcal{N}\left(0, B^{-1}\left(\xi_0\right)\right).$$

Bias correction cannot relax the lower bound restriction on the true fractional integration parameter δ_0, but eliminates some further restrictions on N and T though still requires Assumption C.1 which implies the restrictions of Theorem 5.2 of Robinson and Velasco (2015) for a similar result in the absence of factors.

3.2 Estimation of a Heterogeneous Model

Although a panel data approach allows for efficient inference under a homogeneous setup, it may be restrictive from an empirical perspective. Most of the time, the applied econometrician is interested in understanding how each cross-section unit behaves while accounting for dependence between these units. We therefore consider the heterogeneous version of (1) with the same prescribed properties as

$$\lambda_t\left(L; \theta_{i0}\right)\left(y_{it} - \alpha_i - \gamma_i f_t\right) = \varepsilon_{it}, \tag{7}$$

where θ_{i0} may change for each cross-section unit. This type of heterogeneous modelling is well motivated in country-specific analyses of economic unions and asset-specific analyses

of portfolios where cross-section correlations are permitted and generally the interest is in obtaining inference for a certain unit rather than for the panel.

Under the heterogeneous setup, just like in the homogeneous case, the common factor structure is asymptotically replaced by the cross-section averages of the first-differenced data under the sufficient conditions given in Assumption C. The asymptotic behaviour of the heterogeneous estimates can be easily derived from the results obtained in Theorems 1 and 2 taking $N = 1$ as follows. Now, denote

$$\hat{\theta}_i = \arg\min_{\theta \in \Theta_i} L^*_{i,T}(\theta),$$

with Θ_i defined as before, $\mathcal{D}_i = \left[\underline{\delta}_i, \overline{\delta}_i\right] \subset (0, 3/2)$, and

$$L^*_{i,T}(\theta) = \frac{1}{T}\varepsilon_i(\theta)\varepsilon_i(\theta)',$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, and

$$\varepsilon_{it}(\theta_i) = \lambda_{t-1}\left(L; \theta_i^{(-1)}\right)\left(\Delta y_{it} - \hat{\phi}_i \Delta \bar{y}_t\right).$$

We have the following results replacing δ_0, $\underline{\delta}$ and $\overline{\delta}$ in Assumptions A.1, A.5, B and C.2 with δ_{i0}, $\underline{\delta}_i$ and $\overline{\delta}_i$, respectively. We denote these conditions as A_i, B_i and C_i, and assume them to hold for all i.

Theorem 3. *Under Assumptions A_i and B_i, $\theta_{i0} \in \Theta_i$, and as $(N, T)_j \to \infty$,*

$$\hat{\theta}_i \to_p \theta_{i0},$$

and under Assumptions A_i, B_i and C_i, $\theta_{i0} \in Int(\Theta_i)$, as $(N, T)_j \to \infty$,

$$T^{1/2}\left(\hat{\theta}_i - \theta_{i0}\right) \to_d \mathcal{N}\left(0, B^{-1}(\xi_{i0})\right).$$

An increasing N is still needed here, as in the homogeneous setting, since the projection errors arising due to factor removal require that $N \to \infty$. However the asymptotic theory is made easier due to the convergence rate being just $\sqrt{T}$ now, with which the initial-condition (IC) bias asymptotically vanishes for all values of $\delta_{i0} \in \mathcal{D}$, without any restriction on the relative rate of growth of N and T.

4. Homogeneity Test

In both AR and fractional panel data literature, the use of homogeneous versus heterogeneous dynamic parameters is a rather arbitrary choice in theoretical modeling of persistence. While it can be argued that allowing for heterogeneity in dynamics is a more flexible way

of modeling persistence in applied research, this can cause efficiency losses when the true dynamics are homogeneous. For example, Ergemen (2017) analyzes debt and GDP series for OECD countries under heterogeneity of dynamics, however both series are estimated to be $I(1)$ series, in which case considering homogeneous dynamics appear to be a more efficient approach.

Bearing in my mind such possibilities and the lack of formal justifications for the heterogeneity of dynamics in the literature, we develop a formal testing procedure in this section. We propose a test of the model (1) against its heterogeneous version (7) similar to the LM test developed in Su and Chen (2013) for slope homogeneity in panel data models with interactive fixed effects.

In order to set up the testing framework, we use the following auxiliary heterogeneous model

$$\lambda_t \left(L; \theta_0 + \vartheta_{i0} \right) \left(y_{it} - \alpha_i - \gamma_i f_t \right) = \varepsilon_{it},$$

where ϑ_i denote the deviations from a common value θ_0 and the null hypothesis of homogeneity corresponds to

$$H_0 : \vartheta_{10} = \cdots = \vartheta_{N0} = 0.$$

To estimate the deviations we use the CSS loss function

$$\tilde{L}_{N,T}(\vartheta_1, \ldots, \vartheta_N) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}_{it}(\vartheta_i)^2$$

where

$$\tilde{\varepsilon}_{it}(\vartheta_i) = \lambda_{t-1} \left(L; \hat{\theta}^{(-1)} + \vartheta_i \right) \left(\Delta y_{it} - \hat{\phi}_i \Delta \bar{y}_t \right), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

with $\hat{\theta}$ defined in Section 3 and to test the null hypothesis H_0 against

$$H_1 : \text{ at least one } \vartheta_{j0} \neq 0,$$

we consider LM tests, evaluating under H_0 the derivatives of $\tilde{L}_{N,T}$,

$$
\begin{aligned}
\frac{\partial}{\partial \vartheta_i} \tilde{L}_{N,T}(\vartheta_1, \ldots, \vartheta_N) &= \frac{2}{NT} \sum_{t=1}^{T} \tilde{\varepsilon}_{it}(\vartheta_i) \frac{\partial}{\partial \vartheta_i} \tilde{\varepsilon}_{it}(\vartheta_i) \\
&= \frac{2}{NT} \sum_{t=1}^{T} \tilde{\varepsilon}_{it}(\vartheta_i) \left\{ \frac{\partial}{\partial \vartheta_i} \lambda_{t-1} \left(L; \hat{\theta}^{(-1)} + \vartheta_i \right) \right\} \lambda_{t-1}^{-1} \left(L; \hat{\theta}^{(-1)} \right) \varepsilon_{it}(\hat{\theta}),
\end{aligned}
$$

i.e.

$$\frac{\partial}{\partial \vartheta_i} \tilde{L}_{N,T}(0, \ldots, 0) = \frac{2}{NT} \sum_{t=1}^{T} \varepsilon_{it}(\hat{\theta}) \chi_{t-1} \left(L; \hat{\theta}^{(-1)} \right) \varepsilon_{it}(\hat{\theta}).$$

Then we can check that

$$
\begin{aligned}
\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_i'} \tilde{L}_{N,T}(0, \ldots, 0) &= \frac{2}{NT} \sum_{t=1}^{T} \varepsilon_{it}(\hat{\theta}) \left\{ \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_i'} \lambda_{t-1} \left(L; \hat{\theta}^{(-1)} \right) \right\} \left(\Delta y_{it} - \hat{\phi}_i \Delta \bar{y}_t \right) \\
&\quad + \frac{2}{NT} \sum_{t=1}^{T} \chi_{t-1} \left(L; \hat{\theta}^{(-1)} \right) \varepsilon_{it}(\hat{\theta}) \chi_{t-1}' \left(L; \hat{\theta}^{(-1)} \right) \varepsilon_{it}(\hat{\theta}),
\end{aligned}
$$

12

and that the cross derivatives with respect to ϑ_i and ϑ_j, $i \neq j$, are identically zero, so we can set up the following LM type statistic

$$
\begin{aligned}
LM_{NT} & = \frac{1}{T} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \varepsilon_{it}(\hat{\theta}) \dot{\varepsilon}'_{it}(\hat{\theta}) \right) B(\hat{\theta})^{-1} \left(\sum_{t=1}^{T} \varepsilon_{it}(\hat{\theta}) \dot{\varepsilon}'_{it}(\hat{\theta}) \right) \\
& = \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{it}(\hat{\theta}) \varepsilon_{is}(\hat{\theta}) \dot{\varepsilon}'_{it}(\hat{\theta}) B(\hat{\theta})^{-1} \dot{\varepsilon}_{is}(\hat{\theta}),
\end{aligned}
$$

where $\dot{\varepsilon}_{it}(\hat{\theta}) = \chi_{t-1} \left(L; \hat{\theta}^{(-1)} \right) \varepsilon_{it}(\hat{\theta})$ and we replace the second derivative of the CSS function by $TB(\hat{\theta})$ as an approximation to the limit of $\sum_{t=1}^{T} \dot{\varepsilon}_{it}(\hat{\theta}) \dot{\varepsilon}'_{it}(\hat{\theta})$ for each i, ignoring the contributions from second derivatives of $\lambda_{t-1}(L; \theta)$. If a common parameter estimate $\hat{\theta}$ describes properly the dynamics of all cross-sections, then we expect, for all i, $\varepsilon_{it}(\hat{\theta})$ to be approximately uncorrelated with $\dot{\varepsilon}_{it}(\hat{\theta})$, which only depends on past observations of $\varepsilon_{it}(\hat{\theta})$. However, as the number of restrictions increases with N, we consider instead a properly centred and scaled statistics in the joint asymptotics,

$$
J_{NT} = \left(N^{-1/2} LM_{NT} - B_{NT} \right) / \sqrt{V_{NT}}
$$

as in Su and Chen (2013) where

$$
B_{NT} = \frac{1}{TN^{1/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2(\hat{\theta}) \dot{\varepsilon}_{it}(\hat{\theta}) B(\hat{\theta})^{-1} \dot{\varepsilon}'_{it}(\hat{\theta})
$$

$$
V_{NT} = \frac{2}{T^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\sigma}_i^4 \left[\sum_{s=1}^{T} \dot{\varepsilon}'_{it}(\hat{\theta}) B(\hat{\theta})^{-1} \dot{\varepsilon}_{is}(\hat{\theta}) \right]^2
$$

with individual estimates $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}^2(\hat{\theta})$ to account for possible heterogeneity in the variances across cross-section units, where V_{NT} can be replaced by

$$
V_{NT}^* = \frac{2(p+1)}{N} \sum_{i=1}^{N} \hat{\sigma}_i^8,
$$

or, under cross-section (unconditional) homoskedasticity, by

$$
\bar{V}_{NT} = \frac{2\hat{\sigma}^4}{T^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\sum_{s=1}^{T} \dot{\varepsilon}'_{it}(\hat{\theta}) B(\hat{\theta})^{-1} \dot{\varepsilon}_{is}(\hat{\theta}) \right]^2
$$

or just $\bar{V}_{NT}^* = 2(p+1)\hat{\sigma}^8$ with $\hat{\sigma}^2 = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2(\hat{\theta})$.

We introduce a variant of Assumption A for our asymptotic analysis.

Assumption A*.

All other conditions in Assumption A hold and Assumption A.1 is replaced by

A.1*.   The idiosyncratic shocks, ε_{it}, $i = 1, 2, \ldots, N$, $t = 0, 1, 2, \ldots, T$ are independently and identically distributed across t and are independent across i, with zero mean and variance σ_i^2, and have a finite 12th-order moment uniformly in i, with

$$\bar{\sigma}^{(8)} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^8 < \infty, \tag{8}$$

and $\delta_0 \in (0, 3/2)$.

Assumption A* removes the identical distribution condition across i, in particular allowing for heterogeneity across i in the variance, and it can be shown that under Assumption A* and the rest of conditions of Corollary 1, $(NT)^{-1/2}(\hat{\theta} - \theta_0) = O_p(1)$ and is still asymptotically normal but with a different asymptotic variance, increased by the factor $\bar{\sigma}^{(4)} / (\bar{\sigma}^{(2)})^2 > 1$ compared to the homoskedastic case, with $\bar{\sigma}^{(4)}$ and $\bar{\sigma}^{(2)}$ defined as described in (8). It also requires an increased number of moments to account for the structure of LM_{NT}.

Theorem 4. *Under the assumptions of Corollary 1 and Assumption A*, as $(N, T)_j \to \infty$, under H_0,*
$$J_{NT} \to_d \mathcal{N}(0, 1).$$

Then the null hypothesis is rejected when J_{NT} is large compared to the corresponding critical value from the standard normal distribution. In practice we would expect that V_{NT} and $\bar{V}_{NT}$ perform better than V_{NT}^* or $\bar{V}_{NT}^*$ under the null in finite samples, since they resemble more closely the actual variability in LM_{NT}, as the asymptotic approximations on which V_{NT}^* and $\bar{V}_{NT}^*$ are based might need moderate values of N and T to be accurate.

Similar tests can be developed to test homogeneity in some subset of θ (while maintaining homogeneity in the remaining dimensions). For that we just need to replace $\dot{\varepsilon}_{it}(\hat{\theta})$ in the definition LM_{NT}, B_{NT} and V_{NT} by the corresponding coordinates of the $(p+1)$ vector $\chi_{t-1}\left(L; \hat{\theta}^{(-1)}\right) \varepsilon_{it}(\hat{\theta})$ to be tested, and then, possibly, replace $(p+1)$ in the definitions of V_{NT} or $\bar{V}_{NT}$ by the number of homogeneity restrictions tested.

To investigate the power properties of our LM test we consider the following sequence of local alternatives
$$H_{1,TN} : \vartheta_{i0} = \zeta N^{-1/4} T^{-1/2} \ \text{ for } i \in \mathcal{I}_N$$

for some $(p+1)$-vector $\zeta \neq 0$ and some index set $\mathcal{I}_N$ satisfying for $n = 2, 4$,

$$\bar{\sigma}_1^{(n)} = \lim_{N \to \infty} \frac{1}{N} \sum_{i \in \mathcal{I}_N} \sigma_i^n \in (0, \infty),$$

which indicates that a non-negligible proportion of cross-section units display dynamics local to those defined by the common value θ_0. The next theorem shows that our homogeneity test has no trivial power under these local alternatives, which converge as in Su and Chen (2013) to the null almost at an optimal rate defined by the rate of convergence of the pooled estimate $\hat{\theta}$.

14

Theorem 5. *Under the assumptions of Theorem 4, as $(N,T)_j \to \infty$, under $H_{1,TN}$,*

$$J_{NT} \to_d \frac{\bar{\sigma}_1^{(4)} \zeta' B_0 \zeta}{\left(2\left(p+1\right)\bar{\sigma}^{(8)}\right)^{1/2}} + \mathcal{N}\left(0,1\right).$$

The positive drift depends naturally on the value of ζ, on the parameters that are not constant across i (through B_0), and on the relative variability of the cross-section units whose dynamics depart from H_0 (through the ratio $\bar{\sigma}_1^{(4)} / \left(\bar{\sigma}^{(8)}\right)^{1/2}$). In case of homoskedasticity, we find trivially that this drift is just $\upsilon \zeta' B_0 \zeta / \left(2\left(p+1\right)\right)^{1/2}$ where $\upsilon = \lim_{N\to\infty} N^{-1} \sum_{i\in\mathcal{I}_N} 1 \in (0,1)$, and for a pure fractional model the drift amounts to $\upsilon \zeta_\delta^2 \pi^2 / \left(6\sqrt{2}\right)$.

5. Monte Carlo Simulations

In this section we carry out a Monte Carlo experiment to study the small-sample performance of the memory estimates and the homogeneity test in the simplest case in which there are no idiosyncratic short memory dynamics, $\xi = 0$, no heteroskedasticity and persistence depends only on the value of δ_0. We draw the idiosyncratic shocks $\varepsilon_{i,t}$ as standard normal and the factor loadings γ_i from $U(-0.5, 1)$ not to restrict the sign. We then generate serially correlated common factors $f_t = 0.5 f_{t-1} + \Delta_t^{-\varrho} z_t^f$, based on the *iid* shocks z_t^f drawn as standard normals and then fractionally integrated to the order ϱ. The individual effects α_i are left unspecified since they are removed via first differencing in the estimation, and projections are based on the first-differenced data. We focus on different cross-section and time-series sizes, N and T, as well as different values of δ_0 and ϱ, and different ζ values leading to departures from the null to determine the power in the testing case. Simulations are based on 1,000 replications.

5.1. Memory Estimation

We investigate the finite-sample properties of our estimate of δ_0. In this case, we set $N = 10, 20$ and $T = 50, 100$ for values of $\delta_0 = 0.3, 0.6, 0.9, 1, 1.1, 1.4$ thus covering a heavily biased stationary case, a slightly nonstationary case, near-unit-root cases and finally a quite nonstationary case, respectively.

We report total biases containing initial-condition and projection biases as well as carry out bias correction based on estimated memory values to obtain projection biases for $\varrho = 0.4, 1$. As is clear in Table 1, when the factors are less persistent ($\varrho = 0.4$), the estimate is heavily biased for the stationary case of $\delta_0 = 0.3$ while the bias gets considerably smaller around the unit-root case. Noticeably, the bias becomes negative when $\delta_0 \geq 0.6$ for several (N, T) combinations. Better results in terms of bias are obtained with increasing T. Expectedly, when the factors have a unit root, the estimate of δ_0 contains a larger bias in the stationary ($\delta_0 = 0.3$) and in the moderately nonstationary ($\delta_0 = 0.6$) cases because the

idiosyncratic shocks are dominated by a more persistent common factor. Biases for other memory values are also exacerbated due to factor persistence increase except for the very high persistent case $\delta_0 = 1.4$. Bias correction works reasonably well when $\varrho = 0.4$ although benefits are limited for $\varrho = 1$. While there is a monotonically decreasing pattern for increasing δ_0 in terms of bias both for the total bias and bias-corrected cases, magnitudes of biases increase when δ_0 leaves the neighbourhood of unity.

Table 1 also reports the root mean square errors (RMSE), which indicate that performance increases with increasing δ_0, T and NT. Standard errors are dominated by bias in terms of contribution to RMSE. Table 2 shows the empirical coverage of 95% confidence intervals of δ_0 based on the asymptotics of our estimate. For $\varrho = 0.4, 1$, the true fractional parameter is poorly covered when $\delta_0 \leq 0.6$. Bias correction in these cases improves the results reasonably. For near-unit-root cases, the estimate achieves the most accurate coverage, especially by comparison with intervals based on estimates of $\delta_0 = 1.4$ and $\delta_0 \leq 0.6$.

5.2. Homogeneity Test

To determine the size and power properties of our test, we consider $N = 5, 10$ to study a factor-augmented multivariate time-series setup and a small panel setup, respectively, and $T = 50, 100$ to show the usefulness of our test in empirical studies that use moderately small time series. We also consider the pure fractional case in which both the common factor and the idiosyncratic errors are pure fractional processes as well as the case in which the common factor, generated as in the previous subsection, is serially correlated.

In terms of size, the test performs well for $\delta_0 < 1$ while it may be oversized when $\delta_0 \geq 1$, as Table 3 shows. An increasing N drives down the size of the test while an increasing T does not affect the size much. When $\delta_0 \geq 1$, an increasing ϱ exacerbates the oversize problem of the test, which is quite pronounced when the factor is serially correlated.

To check the power properties of the test, we consider $\zeta = 0.1N^{1/4}T^{1/2}, 0.5N^{1/4}T^{1/2}$, corresponding to local departure values of 0.1 and 0.5 under $H_{1,TN}$, and for ease of exposition, we analyze the case in which ϑ_{i0} exhibit departures from the null for all i. Table 4 shows that when the departures from the null are by 0.1, the test has low power for $\delta_0 \leq 0.6$ although for other δ_0 values, the power is reasonably high attaining the maximum when $\delta_0 = 1$ in most cases. Increasing N, T help in terms of power improvement. When $\zeta = 0.5N^{1/4}T^{1/2}$ in Table 5, power of the test is quite high in all cases except when $\delta_0 = 0.3$. These results also reflect the findings from the previous setting and even more so, demonstrate that the test is consistent even for relatively small panels.

6. Final Comments

We have considered large N, T panel data models with fixed effects and cross-section dependence where the idiosyncratic shocks and common factors are allowed to exhibit long-range

dependence. Our methodology for memory estimation consists in conditional-sum-of-squares estimation on the first differences of defactored variables, where projections are carried out on the sample means of differenced data. We then develop tests for homogeneity of dynamics based on deviations from our homogeneous estimates. While Monte Carlo experiments indicate satisfactory results, our methodology can be extended in the following directions: (a) Different estimation techniques, such as fixed effects and GMM, can be used under our setup as in Robinson and Velasco (2015); (b) The idiosyncratic shocks may be allowed to feature spatial dependence providing further insights in empirical analyses. Design of persistence homogeneity tests in a setting that allows for spatial dependence, though possibly challenging, could make testing procedures more robust.

7. Technical Appendix

We prove our results under more general conditions that are implied by Assumptions B and C allowing for some trade off between the choice of $\underline{\delta}$ and the asymptotic relationship between N and T. The weaker counterpart of Assumption B is as follows.

Assumption B*.

B*.1. $\delta_0 - 1 < \underline{\delta}/2$ and $\varrho - 1 < \underline{\delta}/2$.

B*.2. If $\varrho - \underline{\delta} > \frac{1}{2}$, as $(N, T)_j \to \infty$,

$$N^{-2} T^{2(\varrho - \underline{\delta}) - 1} \to 0$$

B*.3. If $\delta_0 - \underline{\delta} \geq \frac{1}{2}$, as $(N, T)_j \to \infty$,

$$N^{-1} T^{2(\delta_0 - 2\underline{\delta}) - 1} \rightarrow 0$$
$$N^{-1} \left(1 + T^{2(\delta_0 + \varrho - 1) - 4\underline{\delta}}\right) \left(\log T + T^{2(\varrho - 1) + 2(\delta_0 - 1) - 1}\right) \rightarrow 0.$$

7.1 Proof of Theorem 1

The projection parameter from the projection of Δy_{it} on its cross-section averages, $\Delta \bar{y}_t$, can be written as

$$\hat{\phi}_i = \frac{\sum_{t=1}^{T} \Delta \bar{y}_t \Delta y_{it}}{\sum_{t=1}^{T} (\Delta \bar{y}_t)^2} = \frac{\gamma_i}{\bar{\gamma}} + \eta_i \tag{9}$$

where

$$\eta_i = \frac{\sum_{t=1}^{T} \Delta \bar{y}_t \Delta \lambda_t^{-1} (L; \theta_0) \left(\varepsilon_{it} - \frac{\gamma_i}{\bar{\gamma}} \bar{\varepsilon}_t\right)}{\sum_{t=1}^{T} (\Delta \bar{y}_t)^2}$$

is the projection error. The conditional sum of squares then can be written as

$$L_{N,T}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\lambda_t^0 (L; \theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t\right) - \tau_t(\theta)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) - \eta_i \bar{\gamma} \lambda_{t-1} (L; \theta) f_t\right)^2 \tag{10}$$

where

$$\lambda_t^0 (L;\theta) = \lambda_t (L;\theta) \lambda_t^{-1} (L;\theta_0) = \sum_{j=1}^{t} \lambda_j^0 (\theta) L^j.$$

and in (10) the first term is the (corrected) usual idiosyncratic component, the second term is the initial condition term, and the third term is the projection error component.

Following Hualde and Robinson (2011) we give the proof for the most general case where possibly $\underline{\delta} \leq \delta_0 - 1/2$. Additionally, the common factor in our model is $I(\varrho)$ by Assumption A.2. While δ may take arbitrary values from $[\underline{\delta}, \bar{\delta}] \subseteq (0, 3/2)$, ensuring uniform convergence of $L_{N,T}(\theta)$ requires the study of cases depending on $\delta_0 - \delta$, while controlling the distance $\varrho - \delta$. We analyze these separately in the following.

In analyzing the idiosyncratic component and the initial condition component, we closely follow Hualde and Robinson (2011). For $\epsilon > 0$, define $Q_\epsilon = \{\theta : |\delta - \delta_0| < \epsilon\}$, $\overline{Q}_\epsilon = \{\theta : \theta \notin Q_\epsilon, \delta \in \mathcal{D}\}$. For small enough ϵ,

$$Pr(\hat{\theta} \in \overline{Q}_\epsilon) \leq Pr\left(\inf_{\Theta \in \overline{Q}_\epsilon} S_{N,T}(\theta) \leq 0\right)$$

where $S_{N,T}(\theta) = L_{N,T}(\theta) - L_{N,T}(\theta_0)$. In the rest of the proof, we will show that $L_{N,T}(\theta)$, and thus $S_{N,T}(\theta)$, converges in probability to a well-behaved function when $\delta_0 - \delta < 1/2$ and diverges when $\delta_0 - \delta \geq 1/2$. In order to analyze the asymptotic behaviour of $S_{N,T}(\delta)$ in the neighborhood of $\delta = \delta_0 - 1/2$, a special treatment is required. For arbitrarily small $\zeta > 0$, such that $\zeta < \delta_0 - 1/2 - \underline{\delta}$, let us define the disjoint sets $\Theta_1 = \{\theta : \underline{\delta} \leq \delta \leq \delta_0 - 1/2 - \zeta\}$, $\Theta_2 = \{\theta : \delta_0 - 1/2 - \zeta < \delta < \delta_0 - 1/2\}$, $\Theta_3 = \{\theta : \delta_0 - 1/2 \leq \delta \leq \delta_0 - 1/2 + \zeta\}$ and $\Theta_4 = \{\theta : \delta_0 - 1/2 + \zeta < \delta \leq \bar{\delta}\}$, so $\Theta = \cup_{k=1}^4 \Theta_k$. Then we will show

$$Pr\left(\inf_{\theta \in \overline{Q}_\epsilon \cap \Theta_k} S_{N,T}(\delta) \leq 0\right) \to 0 \quad \text{as } (N,T)_j \to \infty, \ k = 1, \dots, 4. \tag{11}$$

We write $L_{N,T}(\theta)$ in (10) as

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \left(\lambda_t^0 (L;\theta) (\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t)\right)^2 + \tau_t^2(\theta)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0)^2 + \eta_i^2 \bar{\gamma}^2 (\lambda_{t-1}(L;\theta) f_t)^2 \right.$$

$$- 2\eta_i \bar{\gamma} (\lambda_{t-1}(L;\theta) f_t) \lambda_t^0(L;\theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t\right) + 2\eta_i \bar{\gamma} (\lambda_{t-1}(L;\theta) f_t) * \tau_t(\theta)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0)$$

$$\left. -2 \lambda_t^0(L;\theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t\right) * \tau_t(\theta)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) \right\}.$$

The projection error component in the conditional sum of squares,

$$\sup_{\theta \in \Theta} \left| \bar{\gamma}^2 \frac{1}{N} \sum_{i=1}^{N} \eta_i^2 \frac{1}{T} \sum_{t=1}^{T} (\lambda_{t-1}(L;\theta) f_t)^2 \right| = o_p(1), \tag{12}$$

18

because the left hand side is $O_p(T^{2\varrho+2\delta_0-6}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2})+O_p(T^{4\varrho+2(\delta_0-\underline{\delta})-7}+T^{2(\varrho-\underline{\delta}-1)}\log T$
$+N^{-1}T^{2(\varrho-\underline{\delta})+4\delta_0-7}+T^{2(\varrho-\underline{\delta})-1}N^{-2})=o_p(1)$ uniformly in $\theta\in\Theta$ by $\bar{\gamma}^2\to_p E[\gamma_i]^2$, Lemmas 1 and 2(a) and Assumption B*.2 since $2\varrho+\delta_0-\underline{\delta}<7/2$, $\varrho-\underline{\delta}<1$ and $\varrho+2\delta_0-\underline{\delta}<7/2$, are implied by Assumption B*.1.

Similarly,

$$\sup_{\theta\in\Theta}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t^2(\theta)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0)^2\right| = \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\tau_t^2(\theta)\right|\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{i0}^2-2\hat{\phi}_i\varepsilon_{i0}\bar{\varepsilon}_0+\hat{\phi}_i^2\bar{\varepsilon}_0^2\right)\right| \tag{13}$$
$$= O_p\left(T^{-2\underline{\delta}}+T^{-1}\log T\right)O_p(1)=o_p(1),$$

because $\underline{\delta}>0$, $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}^2+\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2=O_p(1)$, $\bar{\varepsilon}_0=O_p\left(N^{-1/2}\right)$ and Cauchy-Schwarz inequality, see Lemma 1, and therefore we find for the cross term corresponding to the sum of squares in (12) and (13) that

$$\sup_{\theta\in\Theta}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\bar{\gamma}\lambda_{t-1}(L;\theta)f_t*\tau_t(\theta)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0)\right|=o_p(1)$$

uniformly in δ by (12), (13) and Cauchy-Schwarz inequality.

The other cross terms involving usual fractional residuals $\lambda_t^0(L;\theta)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right)$ are also uniformly $o_p(1)$ for $\theta\in\Theta_4$ using Cauchy-Schwarz inequality and that the part of the conditional sum of squares in $(\lambda_t^0(L;\theta)(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t))^2$ converges uniformly in this set. Lemmas 3 and 4 show that these cross terms are also uniformly $o_p(1)$ for $\theta\in\Theta_1\cup\Theta_2\cup\Theta_3$ under the assumptions of the theorem. Then to show (11) we only need to analyze the terms in $(\lambda_t^0(L;\theta)(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t))^2$ for Θ_k, $k=1,\ldots,4$ as in Hualde and Robinson (2011).

Proof for $k=4$. We show that

$$\sup_{\theta\in\Theta_4}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(\lambda_t^0(L;\theta)(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t))^2-\sigma^2\sum_{j=0}^{\infty}\lambda_j^0(\theta)^2\right]\right|=o_p(1), \tag{14}$$

analyzing the idiosyncratic term, ε_{it}, and the cross-section averaged term, $\hat{\phi}_i\bar{\varepsilon}_t$, separately. For the idiosyncratic term, we first show following Hualde and Robinson (2011),

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_t^0(L;\theta)\varepsilon_{it}\right)^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=0}^{t}\lambda_j^0(\theta)\bar{\varepsilon}_{it-j}\right)^2$$
$$\to_p \sigma^2\sum_{j=0}^{\infty}\lambda_j^0(\theta)^2,$$

uniformly in δ by Assumption 1 as $(N,T)_j\to\infty$ since $-1/2+\zeta<\delta-\delta_0$ for some $\zeta>0$. Since the limit is uniquely minimized at $\theta=\theta_0$ as it is positive for all $\theta\neq\theta_0$, (11) holds for $k=4$ if (14) holds and the contribution of cross-section averaged term, $\hat{\phi}_i\bar{\varepsilon}_t$, is negligible.

19

To check (14) we show

$$\sup_{\theta \in \Theta_4} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(\sum_{j=0}^{t} \lambda_j^0 (\theta) \, \varepsilon_{it-j} \right)^2 - E \left(\sum_{j=0}^{t} \lambda_j^0 (\theta) \, \varepsilon_{it-j} \right)^2 \right] \right| = o_p(1),$$

where the term in absolute value is

$$\frac{1}{T} \sum_{j=0}^{T} \lambda_j^0 (\theta)^2 \frac{1}{N} \sum_{i=1}^{N} \sum_{l=0}^{T-j} (\varepsilon_{il}^2 - \sigma^2)$$

$$+ \frac{2}{T} \sum_{j=0}^{T-1} \lambda_j^0 (\theta) \, \lambda_k^0 (\theta) \frac{1}{N} \sum_{i=1}^{N} \sum_{l=k-j+1}^{T-j} \varepsilon_{il} \varepsilon_{il-(k-j)} = (a) + (b). \qquad (15)$$

Then,

$$E \sup_{\Theta_4} |(a)| \le \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{j=0}^{T} \sup_{\Theta_4} \lambda_j^0 (\theta)^2 \, E \left| \sum_{l=0}^{T-j} (\varepsilon_{il}^2 - \sigma^2) \right| \right).$$

Uniformly in j, $Var(N^{-1} \sum_{i=1}^{N} \sum_{l=0}^{T-j} \varepsilon_{il}^2) = O(N^{-1}T)$, so using $-1/2 + \zeta < \delta - \delta_0$,

$$\sup_{\Theta_4} |(a)| = O_p \left(N^{-1/2} T^{-1/2} \sum_{j=1}^{\infty} j^{-2\zeta-1} \right) = O_p(N^{-1/2} T^{-1/2}).$$

By summation by parts, the term (b) is equal to

$$\frac{2\lambda_{T-1}^0 (\theta)}{T} \sum_{j=0}^{T-1} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=j+1}^{T} \sum_{l=k-j+1}^{T-j} \lambda_j^0 (\theta) \, \varepsilon_{il} \varepsilon_{il-(k-j)}$$

$$- \frac{2}{T} \sum_{j=0}^{T-1} \lambda_j^0 (\theta) \sum_{k=j+1}^{T} \left[\lambda_{k+1}^0 (\theta) - \lambda_k^0 (\theta) \right] \frac{1}{N} \sum_{i=1}^{N} \sum_{r=j+1}^{k} \sum_{l=r-j+1}^{T-j} \varepsilon_{il} \varepsilon_{il-(r-j)}$$

$$= (b_1) + (b_2).$$

Then, using that $Var \left(N^{-1} \sum_{i=1}^{N} \sum_{k=j+1}^{T} \sum_{l=k-j+1}^{T-j} \left\{ \varepsilon_{il} \varepsilon_{il-(k-j)} \right\} \right) = O \left(N^{-1} T^2 \right)$ uniformly in i and j,

$$E \sup_{\Theta_4} |(b_1)| \le T^{-\zeta-3/2} \sum_{j=1}^{T} j^{-\zeta-1/2} Var \left(\sum_{k=j+1}^{T} \sum_{l=k-j+1}^{T-j} \left\{ \varepsilon_{il} \varepsilon_{il-(k-j)} \right\} \right)^{1/2} \le N^{-1/2} T^{-2\zeta},$$

while

$$E \sup_{\Theta_4} |(b_2)| \le T^{-1} \sum_{j=1}^{T} j^{-\zeta-1/2} \sum_{k=j+1}^{T} k^{-\zeta-3/2} Var \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{r=j+1}^{k} \sum_{l=r-j+1}^{T-j} \left\{ \varepsilon_{il} \varepsilon_{il-(r-j)} \right\} \right)^{1/2}$$

$$\le N^{-1/2} T^{-1/2} \sum_{j=1}^{T} j^{-\zeta-1/2} \sum_{k=j+1}^{T} k^{-\zeta-3/2} (k-j)^{1/2} \le K N^{-1/2} T^{-2\zeta},$$

20

and therefore $(b) = O_p(N^{-1/2}T^{-2\zeta}) = o_p(1)$.

Next, we deal with the terms carrying $\bar{\varepsilon}_t$ in the LHS of (14). We write

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\phi}_i^2 \left(\lambda_t^0 (L; \theta) \bar{\varepsilon}_t \right)^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_i^2 \frac{1}{T} \sum_{t=1}^{T} \left(\lambda_t^0 (L; \theta) \bar{\varepsilon}_t \right)^2. \qquad (16)$$

The average in i is $O_p(1)$ by Lemma 1, while the sum in t in the lhs (16) satisfies for θ^* with first component $\theta_{(1)}^* = \zeta - \frac{1}{2}$,

$$\frac{1}{T} \sum_{t=1}^{T} \left(\lambda_t^0 (L; \theta) \bar{\varepsilon}_t \right)^2 = O_p \left(\frac{\sigma^2}{N} \sum_{j=0}^{\infty} \lambda_j^0 (\theta^*)^2 \right) = O_p \left(N^{-1} \right) = o_p(1)$$

as $N \to \infty$, uniformly in $\theta \in \Theta_4$ as $T \to \infty$, and (16) is at most $O_p(N^{-1}) = o_p(1)$ uniformly in $\theta \in \Theta_4$.

Finally, the cross-term due to the square on the lhs of (14) is asymptotically negligible by Cauchy-Schwarz inequality. So we have proved (14), and therefore we have proved (11) for $k = 4$.

Proof for $k = 3, 2$. The uniform convergence for the idiosyncratic component for the proof of (11) follows as in Hualde and Robinson (2011), since the average in $i = 1, \ldots, N$ adds no additional complication as in the case $k = 4$. The treatment for the cross-section averaged term and the cross-product term follows from the same steps as the idiosyncratic term as well as the results we derived for $k = 4$ using $\frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_i^2 = O_p(1)$ and that $\bar{\varepsilon}_t$ has variance σ^2/N.

Proof for $k = 1$. Noting that

$$L_{N,T}^*(\theta) := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \left(\lambda_t^0 (L; \theta) (\varepsilon_{it} - \phi_i \bar{\varepsilon}_t) \right)^2 \geq \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \left(\sum_{t=1}^{T} \lambda_t^0 (L; \theta) (\varepsilon_{it} - \phi_i \bar{\varepsilon}_t) \right)^2,$$

we write

$$Pr \left(\inf_{\Theta_1} L_{N,T}^*(\theta) > K \right) \geq Pr \left(T^{2\zeta} \inf_{\Theta_1} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T^{\delta_0 - \delta + 1/2}} \sum_{t=1}^{T} \lambda_t^0 (L; \theta) (\varepsilon_{it-j} - \phi_i \bar{\varepsilon}_{t-j}) \right)^2 > K \right)$$

since $\delta - \delta_0 \leq -1/2 - \zeta$.

For arbitrarily small $\epsilon > 0$, we show

$$Pr \left(T^{2\zeta} \inf_{\Theta_1} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T^{\delta_0 - \delta + 1/2}} \sum_{t=1}^{T} \lambda_t^0 (L; \theta) (\varepsilon_{it-j} - \phi_i \bar{\varepsilon}_{t-j}) \right)^2 > K \right)$$

$$\geq Pr \left(\inf_{\Theta_1} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T^{\delta_0 - \delta + 1/2}} \sum_{t=1}^{T} \lambda_t^0 (L; \theta) (\varepsilon_{it-j} - \phi_i \bar{\varepsilon}_{t-j}) \right)^2 > \epsilon \right) \to 1 \text{ as } (N, T)_j \to \infty.$$

Define $h_{i,T}^{(1)}(\delta) = T^{-\delta_0+\delta-1/2}\lambda_t^0(L;\theta)\,\varepsilon_{it-j} = T^{-1/2}\sum_{j=1}^{T}\frac{\lambda_j^0(\theta)}{T^{\delta_0-\delta}}\varepsilon_{it-j}$ and
$h_T^{(2)}(\delta) = T^{-\delta_0+\delta-1/2}\lambda_t^0(L;\theta)\,\bar{\varepsilon}_{t-j} = T^{-1/2}\sum_{j=1}^{T}\frac{\lambda_j^0(\theta)}{T^{\delta_0-\delta}}\bar{\varepsilon}_{t-j}$. By the weak convergence results
in Marinucci and Robinson (2000), for each $i = 1,\dots,N$,

$$h_{i,T}^{(1)}(\delta) \Rightarrow \lambda_\infty^0(1;\theta)\int_0^1 \frac{(1-s)^{\delta_0-\delta}}{\Gamma(\delta_0-\delta+1)}\delta B_i(s)$$

as $(N,T)_j \to \infty$, where $B_i(s)$ is a scalar Brownian motion, $i = 0,\dots,N$, and by $\Rightarrow$ we mean convergence in the space of continuous functions in Θ_1 with uniform metric. Tightness and finite dimensional convergence follows from the fractional invariance property presented in Theorem 1 in Hosoya (2005) as well as $\sup_{iT} E\left[h_{i,T}^{(1)}(\delta)^2\right] < \infty$. Similarly, $N^{1/2}h_T^{(2)}(\delta)$ is weakly converging to the same functional of $B_0(s)$. Then, as $(N,T)_j \to \infty$, following the discussions for double-index processes in Phillips and Moon (1999) and $\frac{1}{N}\sum_{i=1}^{N}\phi_i^2 = O_p(1)$,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^{\delta_0-\delta+1/2}}\sum_{t=1}^{T}\lambda_t^0(L;\theta)\,(\varepsilon_{it-j}-\phi_i\bar{\varepsilon}_{t-j})\right)^2 \to_p \quad \lambda_\infty^0(1;\theta)^2\,\mathrm{Var}\left(\int_0^1 \frac{(1-s)^{\delta_0-\delta}}{\Gamma(\delta_0-\delta+1)}\delta B(s)\right)$$

$$= \frac{\sigma^2\lambda_\infty^0(1;\theta)^2}{(2(\delta_0-\delta)+1)\,\Gamma^2(\delta_0-\delta+1)},$$

uniformly in $\theta \in \Theta_1$, where

$$\inf_{\Theta_1}\lambda_\infty^0(1;\theta)^2\,\mathrm{Var}\left(\int_0^1 \frac{(1-s)^{\delta_0-\delta}}{\Gamma(\delta_0-\delta+1)}\delta B(s)\right) = \frac{\sigma^2}{(2(\delta_0-\underline{\delta})+1)\,\Gamma^2(\delta_0-\underline{\delta}+1)} > 0,$$

so that

$$Pr\left(\inf_{\Theta_1}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^{\delta_0-\delta+1/2}}\sum_{t=1}^{T}\lambda_t^0(L;\theta)\,(\varepsilon_{it-j}-\phi_i\bar{\varepsilon}_{t-j})\right)^2 > \epsilon\right) \to 1 \quad \text{as } (N,T)_j \to \infty$$

and (11) follows for $i = 1$ as ϵ is arbitrarily small. $\square$

7.2 Other Proofs in Sections 3 and 4

We use the following more general conditions that are implied by Assumption C in our proofs.

Assumption C*.

C*.1. As $(N,T)_j \to \infty$,

$$\frac{N}{T}\log^2 T + \frac{T}{N^3} \to 0.$$

C*.2. As $(N,T)_j \to \infty$,

$$N\left(T^{4(\varrho+\delta_0)-11}\log^2 T + N^{-2}T^{8\delta_0-11}\right)\log^2 T \to 0$$
$$N\left(T^{2(\varrho+\delta_0-3)} + T^{-2\delta_0}\log^2 T\right)\log^2 T \to 0.$$

22

C*.3. As $(N, T)_j \to \infty$,

$$N^{-1} T^{1-2\delta_0} \to 0.$$

Proof of Theorem 2. We first analyze the first derivative of $L_{N,T}(\theta)$ evaluated at $\theta = \theta_0$,

$$
\frac{\partial}{\partial \theta} L_{N,T}(\theta)|_{\theta=\theta_0} = \frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ -\eta_i \bar{\gamma} \lambda_{t-1}(L; \theta_0) f_t - \tau_t(\theta_0) \left(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0 \right) + \varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right\}
$$
$$
\times \left\{ -\eta_i \bar{\gamma} \chi_{t-1}(L; \xi_0) \lambda_{t-1}(L; \theta_0) f_t - \dot{\tau}_t(\theta_0) \left(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0 \right) + \chi_t(L; \xi_0) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) \right\},
$$

where $\chi_t(L; \xi_0) \varepsilon_{it} = \chi_{t-1}(L; \xi_0) \varepsilon_{it} + \chi_t(\xi_0) \varepsilon_{i0}$.

In open form with the $(NT)^{1/2}$ normalization,

$$
\sqrt{NT} \frac{\partial}{\partial \theta} L_{N,T}(\theta)|_{\theta=\theta_0} = \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_i^2 \bar{\gamma}^2 \lambda_{t-1}(L; \theta_0) f_t * \chi_{t-1}(L; \xi_0) \lambda_{t-1}(L; \theta_0) f_t \tag{17}
$$

$$
+ \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tau_t(\theta_0) \dot{\tau}_t(\theta_0) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0)^2 \tag{18}
$$

$$
+ \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_i \bar{\gamma} \lambda_{t-1}(L; \theta_0) f_t * \dot{\tau}_t(\theta_0) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) \tag{19}
$$

$$
- \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_i \bar{\gamma} \lambda_{t-1}(L; \theta_0) f_t * \chi_t(L; \xi_0) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) \tag{20}
$$

$$
+ \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_i \bar{\gamma} \chi_{t-1}(L; \xi_0) \lambda_{t-1}(L; \theta_0) f_t * \tau_t(\theta_0) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) \tag{21}
$$

$$
- \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tau_t(\theta_0) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) * \chi_t(L; \xi_0) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) \tag{22}
$$

$$
- \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_i \bar{\gamma} \chi_{t-1}(L; \xi_0) \lambda_{t-1}(L; \theta_0) f_t * \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) \tag{23}
$$

$$
- \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \dot{\tau}_t(\theta_0) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0)(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t) \tag{24}
$$

$$
+ \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t) * \chi_t(L; \xi_0) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right). \tag{25}
$$

The term (17) is asymptotically negligible, since with Lemmas 1 and 2 and $\varrho - \delta_0 < \frac{1}{2}$, we find that

$$
\frac{2 \bar{\gamma}^2 \sqrt{N}}{\sqrt{T}} \frac{1}{N} \sum_{i=1}^{N} \eta_i^2 \sum_{t=1}^{T} \lambda_{t-1}(L; \theta_0) f_t \chi_{t-1}(L; \xi_0) \lambda_{t-1}(L; \theta_0) f_t
$$
$$
= O_p(N^{1/2} T^{-1/2}) O_p(T^{2\varrho + 2\delta_0 - 6} + N^{-1} T^{4\delta_0 - 6} + T^{-1} \log T + N^{-2}) O_p(T),
$$

which is $o_p(1)$ under Assumption C*.1-2.

In (18), we can directly take the expectation of the main term to get the bias term stemming from the initial condition,

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\dot{\tau}_t(\theta_0)E\left[\varepsilon_{i0}^2\right] = 2\sigma^2\left(\frac{N}{T}\right)^{1/2}\sum_{t=1}^{T}\tau_t(\theta_0)\dot{\tau}_t(\theta_0),$$

which is $O\left(N^{1/2}\left(T^{-1/2}+T^{1/2-2\delta_0}\log^2 T\right)\right) = O\left(N^{1/2}T^{-1/2}\right)$, with variance

$$\frac{2}{NT}\sum_{i=1}^{N}Var\left[\varepsilon_{i0}^2\right]\left(\sum_{t=1}^{T}\tau_t(\theta_0)\dot{\tau}_t(\theta_0)\right)^2 = O\left(T^{-1}+T^{1-4\delta_0}\log^4 T\right) = o(1)$$

since $\delta_0 > 1/4$, as $(N,T)_j \to \infty$, while

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\dot{\tau}_t(\theta_0)\hat{\phi}_i^2\bar{\varepsilon}_0^2 = \frac{2}{\sqrt{NT}}N\bar{\varepsilon}_0^2\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2\sum_{t=1}^{T}\tau_t(\theta_0)\dot{\tau}_t(\theta_0)$$

$$= O_p\left((TN)^{-1/2}\left(1+T^{1-2\delta_0}\log^2 T\right)\right) = o_p(1)$$

because $\delta_0 > 1/4$, and by Cauchy-Schwarz inequality the cross term is of order

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

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

if $\delta_0 > 1/4$.

We show that (19) is $o_p(1)$ considering the contribution of

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta_0)f_t\dot{\tau}_t(\theta_0)\varepsilon_{i0}$$

whose absolute value is bounded, using Lemmas 1 and 2(c), and $\varrho-\delta_0 < \frac{1}{2}$, by

$$2\sqrt{NT}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}^2\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\right)^{1/2}\left|\frac{1}{T}\sum_{t=1i}^{T}\lambda_{t-1}(L;\theta_0)f_t\dot{\tau}_t(\theta_0)\right|$$

$$= O_p\left((NT)^{1/2}\left(T^{2(\varrho+\delta_0-3)}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2}\right)^{1/2}T^{-1}\right)$$

$$+ O_p\left((NT)^{1/2}\left(T^{2(\varrho+\delta_0-3)}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2}\right)^{1/2}\left\{\begin{array}{c}T^{\varrho-2\delta_0-1/2}+T^{-\delta_0/2-1/2}\\+T^{-\delta_0}+T^{\varrho-3\delta_0/2-1}\end{array}\right\}\log T\right)$$

$$= O_p\left(N^{1/2}\left(T^{\varrho+\delta_0-3}+T^{-1/2}\log T+N^{-1}\right)T^{\varrho-2\delta_0}\log T\right)$$

$$+ O_p\left(N^{1/2}T^{\varrho+\delta_0-3}T^{-\delta_0/2}\log T\right)$$

$$+ O_p\left(N^{1/2}T^{\varrho-3\delta_0/2-1}\log T+N^{-1/2}T^{\varrho-3\delta_0/2-1/2}+N^{1/2}T^{-\delta_0}\log^2 T\right)+o_p(1)$$

24

which, using $\varrho - \delta_0 < \frac{1}{2}$, is

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

which is $o_p(1)$ by Assumption C*.

For (20), we consider the contribution of

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta_0)f_t * \chi_t(L;\xi_0)\varepsilon_{it}$$

whose absolute value is bounded by

$$2\sqrt{NT}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}(L;\theta_0)f_t * \chi_t(L;\xi_0)\varepsilon_{it}\right)^2\right)^{1/2}$$

$$= O_p\left((NT)\left(T^{2\varrho+2\delta_0-6} + T^{-1}\log T + N^{-1}T^{4\delta_0-6} + N^{-2}\right)T^{-1}\right)^{1/2}$$

$$= O_p\left(N\left(T^{2\varrho+2\delta_0-6} + T^{-1}\log T + N^{-1}T^{4\varrho-6}\log T + N^{-2}\right)\right)^{1/2} = o_p(1)$$

by using Assumptions C*.1-2, because, uniformly in i, using $\varrho - \delta_0 < \frac{1}{2}$,

$$E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}(L;\theta_0)f_t * \chi_t(L;\xi_0)\varepsilon_{it}\right)^2\right]$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\lambda_{t-1}(L;\theta_0)f_t * \chi_t(L;\xi_0)\varepsilon_{it} * \lambda_{r-1}(L;\theta_0)f_r * \chi_r(L;\xi_0)\varepsilon_{ir}\right]$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\lambda_{t-1}(L;\theta_0)f_t * \lambda_{r-1}(L;\theta_0)f_r\right]E\left[\chi_t(L;\xi_0)\varepsilon_{it} * \chi_r(L;\xi_0)\varepsilon_{ir}\right]$$

$$= O\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{t}\left(|t-r|^{2(\varrho-\delta_0)-2} + |t-r|^{\varrho-\delta_0-2}\right)\log t\right)$$

$$= O\left(T^{-1}\log T + T^{2(\varrho-\delta_0-1)}\right) = O\left(T^{-1}\log T\right)\blacksquare.$$

Then the term (20) is $o_p(1)$ because the factor depending on $\hat{\phi}_i\chi_t(L;\xi_0)\bar{\varepsilon}_t$ could be dealt with similarly using Cauchy-Schwarz inequality and Lemma 1.

The proof that the term (21) is $o_p(1)$ could be dealt with exactly as when bounding (19), while the proof that the term (23) is $o_p(1)$ could be dealt with in a similar but easier way than (20).

The leading term of (24), depending on $\varepsilon_{i0}\varepsilon_{it}$,

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tau}_t(\theta_0)(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0)(\varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t),$$

25

has zero mean and variance proportional to

$$\frac{1}{T}\sum_{t=1}^{T}\dot{\tau}_t(\theta_0)^2 = O\left(T^{-1} + T^{-2\delta_0}\right) = o\left(1\right)$$

so it is negligible and the same can be concluded for the other terms depending on $\hat{\phi}_i$.

The behaviour of the main term in (22) is given in Lemma 5 and that of (25) in Lemma 6 and, combining the plims of (18) and (22), we obtain the definition of $\nabla_T(\delta)$.

Then collecting the results for all terms (17) to (25) we have found that

$$\sqrt{NT}\frac{\partial}{\partial\theta}L_{N,T}(\theta)|_{\theta=\theta_0} \to_d \left(\frac{N}{T}\right)^{1/2}\sum_{t=1}^{T}\{\tau_t(\theta_0)\dot{\tau}_t(\theta_0) - \tau_t(\theta_0)\chi_t(\theta_0)\} + \mathcal{N}\left(0, 4B\left(\xi_0\right)\right).$$

Finally we analyze the second derivative of $L_{N,t}(\theta)$ evaluated at $\theta = \theta_0$, $(\partial^2/\partial\theta\partial\theta')L_{N,T}(\theta)|_{\theta=\theta_0}$, which equals

$$\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\eta_i\bar{\gamma}\chi_{t-1}(L;\xi_0)\lambda_{t-1}(L;\theta_0)f_t - \dot{\tau}_t(\theta_0)\left(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0\right) + \chi_t(L;\xi_0)\left(\varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t\right)\right\}$$

$$\times\left\{-\eta_i\bar{\gamma}\chi_{t-1}(L;\xi_0)\lambda_{t-1}(L;\theta_0)f_t - \dot{\tau}_t(\theta_0)\left(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0\right) + \chi_t(L;\xi_0)\left(\varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t\right)\right\}'$$

$$+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\eta_i\bar{\gamma}\lambda_{t-1}(L;\theta_0)f_t - \tau_t(\theta_0)\left(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0\right) + \varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t\right\}$$

$$\times\left\{-\eta_i\bar{\gamma}b_t^0(L)\lambda_{t-1}(L;\theta_0)f_t - \ddot{\tau}_t(\theta_0)\left(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0\right) + b_t^0(L)\left(\varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t\right)\right\},$$

where $b_t^0(L) = \dot{\chi}_t(L;\xi_0) + \chi_t(L;\xi_0)\chi_t(L;\xi_0)'$, $\dot{\chi}_t(L;\xi) = (\partial/\partial\theta')\chi_t(L;\xi)$ and $\ddot{\tau}_t(\theta) = (\partial^2/\partial\theta\partial\theta')\tau_t(\theta)$. Using the same techniques as in the proof of Theorem 1, as N and T get larger, only the term on $\chi_t(L;\xi_0)\varepsilon_{it}\chi_t(L;\xi_0)'\varepsilon_{it}$ in the first element of the rhs contributes to the probability limit, see the proof of Theorem 5.2 in Robinson and Velasco (2015). In the second part of the expression, all terms are asymptotically negligible by using the same arguments as in the convergence in distribution of the score, obtaining as $N \to \infty$ and $T \to \infty$,

$$\frac{\partial^2}{\partial\theta\partial\theta'}L_{N,T}(\theta)|_{\theta=\theta_0} \to_p 2\sigma^2 B\left(\xi_0\right).$$

Lemma 7 shows the convergence of the Hessian $L_{N,T}(\theta)$ evaluated at $\hat{\theta}$ to that evaluated at θ_0, and the proof is then complete. $\square$

Proof of Corollary 1. The result is a direct consequence of Theorem 2.

Proof of Corollary 2. Follows from Theorem 2 as the proofs of Theorems 5.1 and 5.2 in Robinson and Velasco (2015).

Proof of Theorem 3. These are simple consequences of the results from Theorems 1 and 2, taking $N = 1$, where the rate of convergence is just $\sqrt{T}$ now so that the asymptotic IC bias is removed for any $\delta_{i0} \in \mathcal{D}$. $\square$

Proof of Theorem 4. The key point is to write

$$N^{-1/2}\left(LM_{NT} - N^{1/2}B_{NT}\right) = \frac{2N^{-1/2}}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it}(\hat{\theta})\varepsilon_{is}(\hat{\theta})\dot{\varepsilon}'_{it}(\hat{\theta})B(\hat{\theta})^{-1}\dot{\varepsilon}_{is}(\hat{\theta})$$

and approximate this by

$$\sum_{t=1}^{T}z_{NT,t}, \quad \text{where } z_{NT,t} = \frac{2N^{-1/2}}{T}\sum_{i=1}^{N}\varepsilon_{it}\sum_{s=1}^{t-1}\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is} \tag{26}$$

by Lemma 10, where $B_0 = B(\xi_0)$, with $z_{NT,t}$ being a Martingale Difference, so that

$$\mathrm{Var}\left[\sum_{t=1}^{T}z_{NT,t}\right] = \sum_{t=1}^{T}\mathrm{Var}\left[z_{NT,t}\right] = \sum_{t=1}^{T}E\left[z_{NT,t}^2\right]$$

and fully expanding matrix calculations using $\dot{\varepsilon}_{it} = \chi_t(L)\varepsilon_{it} = \sum_{j=1}^{t}\chi_j\varepsilon_{it-j}$, with $\chi_j = \left(\chi_j^1, \ldots, \chi_j^{p+1}\right)'$,

$$
\begin{aligned}
\mathrm{Var}\left[z_{NT,t}\right] &= \frac{4}{NT^2}\sum_{i=1}^{N}\sigma_i^2 E\left[\left(\sum_{s=1}^{t-1}\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}\right)^2\right] \\
&= \frac{4}{NT^2}\sum_{i=1}^{N}\sigma_i^2\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}E\left[\varepsilon_{is}\varepsilon_{ir}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{ir}\right] \\
&= \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^2\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}E\left[\varepsilon_{is}\varepsilon_{ir}\dot{\varepsilon}_{it}^{a}\dot{\varepsilon}_{is}^{b}\dot{\varepsilon}_{it}^{c}\dot{\varepsilon}_{ir}^{d}\right] \\
&= \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^2\sum_{s=1}^{t-1}E\left[\varepsilon_{is}^2\right]E\left[\dot{\varepsilon}_{it}^{a}\dot{\varepsilon}_{it}^{c}\right]E\left[\dot{\varepsilon}_{is}^{b}\dot{\varepsilon}_{is}^{d}\right](1 + o(1)) \\
&= \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^8\sum_{j=1}^{t}\chi_j^{a}\chi_j^{c}\sum_{s=1}^{t-1}\sum_{k=1}^{s}\chi_k^{b}\chi_k^{d}(1 + o(1))
\end{aligned}
$$

(omitting lower order contributions from other pairwise moments and higher order cumulant

combinations) which is

$$\frac{4}{NT^2} \sum_{a,b,c,d=1}^{p+1} B_{a,b}^{-1} B_{c,d}^{-1} \sum_{i=1}^{N} \sigma_i^8 \sum_{j=1}^{t} \chi_j^a \chi_j^c \sum_{k=1}^{t-1} (t-k) \chi_k^b \chi_k^d (1 + o(1))$$

$$= \frac{4t}{NT^2} \sum_{i=1}^{N} \sigma_i^8 trace \left\{ B_0^{-1} B_0 B_0^{-1} B_0 \right\} (1 + o(1))$$

$$= \frac{4t}{NT^2} \sum_{i=1}^{N} \sigma_i^8 (p+1) (1 + o(1))$$

since $\sum_{k=1}^{t-1} (t-k) \chi_k^b \chi_k^d = t \sum_{k=1}^{t-1} \chi_k^b \chi_k^d (1 + o(1))$ as $t \to \infty$, because $\chi_k^b = O(k^{-1})$, so that

$$\mathrm{Var} \left[\sum_{t=1}^{T} z_{NT,t} \right] \to \frac{4(p+1)}{NT^2} \sum_{i=1}^{N} \sigma_i^8 \sum_{t=1}^{T} t \to 2\bar{\sigma}^8 (p+1),$$

as $(N,T)_j \to \infty$, so that by Lemma 9

$$N^{-1/2} LM_{NT} - B_{NT} \to_d N \left(0, 2\bar{\sigma}^8 (p+1) \right).$$

Then it remains to show that

$$V_{NT} \quad \to_p \quad \frac{2}{T^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_i^4 \left[\sum_{s=1}^{T} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \right]^2$$

$$\to_p \quad \frac{2}{T^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_i^4 \sum_{s=1}^{T} \sum_{r=1}^{T} E \left[\dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{ir} \right]$$

$$\to \quad \frac{4}{T^2 N} \sum_{i=1}^{N} \sigma_i^8 \sum_{t=1}^{T} t * trace \left\{ B_0^{-1} B_0 B_0^{-1} B_0 \right\}$$

$$\to \quad 2\bar{\sigma}^{(8)} (p+1),$$

using Lemma 11, and the theorem follows. $\square$

Proof of Theorem 5. The proof consists of four steps, where we omit most details already in other proofs.

Step 1. $\hat{\theta} - \theta_0 \to_p 0$, which follows by an essentially similar proof as Theorem 1, the ϑ_{i0} contribution for $i \in \mathcal{I}_N$ being negligible.

Step 2. $\hat{\theta} - \theta_0 = O_p \left(N^{-1/4} T^{-1/2} \right)$, in particular for $\upsilon = \bar{\sigma}_1^{(2)} / \bar{\sigma}^{(2)} \in (0,1)$,

$$\hat{\theta} - \theta_0 = \upsilon \zeta N^{-1/4} T^{-1/2} + O_p \left(N^{-1/2} T^{-1/2} \right)$$

using the same Taylor expansion as in Theorem 2, with stochastic remainder of order $O_p \left(N^{-1/2} T^{-1/2} \right)$, but additional (average) drift $\upsilon \zeta N^{-1/4} T^{-1/2}$ obtained by a similar argument as in step 4 below.

28

Step 3. Equivalence of analysis of J_{NT} evaluated under $\hat{\theta}$ and evaluated under θ_0 (for all cross-section units) with error $\hat{\theta} - \theta_0 = O_p\left(N^{-1/4}T^{-1/2}\right)$, similar to Lemma 10 where now the contribution of the term (37) which depends directly on $\hat{\theta} - \theta_0$ is $O_p\left(N^{1/4}T^{-1/2}\log^3 T\right) = o_p(1)$, but there is a new drift term, so that the right hand side of

$$\varepsilon_{it}(\hat{\theta}) - \varepsilon_{it}(\theta_0) = \left(\hat{\theta} - \theta_0\right)\chi_t\left(L;\hat{\theta}^*\right)\left(\varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t\right) - \tau_t(\hat{\theta}^*)(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0) - \eta_i\bar{\gamma}\lambda_{t-1}\left(L;\hat{\theta}\right)f_t$$

for $\left\|\hat{\theta}^* - \theta\right\| \leq \left\|\hat{\theta} - \theta_0\right\|$ is still negligible, but for $\theta_{NT} = \theta_0 + \zeta N^{-1/4}T^{-1/2}$ and for

$$d_{it,NT} = N^{-1/4}T^{-1/2}\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{it}1_{\{i\in\mathcal{I}_N\}}$$

we have that

$$
\begin{aligned}
\varepsilon_{it}(\theta_0) &= \lambda_t\left(L;\theta_0\right)\lambda_t^{-1}\left(L;\theta_{NT}\right)\varepsilon_{it} \\
&= \varepsilon_{it} + d_{it,NT}
\end{aligned}
$$

plus a remainder for $i \in \mathcal{I}_N$ of order $N^{-1/2}T^{-1}$ depending on a lag polynomial of ε_{it} with coefficients $\dot{\chi}_t = O\left(t^{-1}\log t\right)$ as $t \to \infty$ uniformly in N and T.

Step 4. Evaluation of the drift generated by $d_{it,NT}$. We have that now $N^{-1/2}\left(LM_{NT} - N^{1/2}B_{NT}\right)$ can be approximated by $\sum_{t=1}^{T}\bar{z}_{NT,t}$ where

$$\bar{z}_{NT,t} = \frac{2N^{-1/2}}{T}\sum_{i=1}^{N}(\varepsilon_{it} + d_{it,NT})\sum_{s=1}^{t-1}(\varepsilon_{is} + d_{is,NT})\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is} \tag{27}$$

and where the term in $\varepsilon_{it}\varepsilon_{is}$ is the same as $\sum_{t=1}^{T}z_{NT,t}$ leading to the asymptotic normal distribution, while the term in $\varepsilon_{it}d_{is,NT}$ keeps the martingale structure and is negligible, while that in $d_{it,NT}d_{is,NT}$,

$$\frac{2N^{-1}}{T^2}\sum_{i\in\mathcal{I}_N}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{it}\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is},$$

converges to its expectation,

$$\frac{2}{NT^2}\sum_{i\in\mathcal{I}_N}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left[\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{it}\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}\right]$$

$$=\frac{2}{NT^2}\sum_{i\in\mathcal{I}_N}\sum_{a,b=1}^{p+1}\sum_{j_1,...,j_4}\zeta'\chi_{j_1}\left(\theta_0\right)\zeta'\chi_{j_2}\left(\theta_0\right)\chi_{j_3}^{a}\left(\theta_0\right)\chi_{j_4}^{b}\left(\theta_0\right)B_0^{a,b}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left[\varepsilon_{it-j_1}\varepsilon_{is-j_2}\varepsilon_{it-j_3}\varepsilon_{is-j_4}\right]$$

$$=\frac{2}{NT^2}\sum_{i\in\mathcal{I}_N}\sum_{a,b=1}^{p+1}B_0^{a,b}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{j_1,...,j_4}\zeta'\chi_{j_1}\left(\theta_0\right)\zeta'\chi_{j_2}\left(\theta_0\right)\chi_{j_3}^{a}\left(\theta_0\right)\chi_{j_4}^{b}\left(\theta_0\right)$$

$$\times\left\{E\left[\varepsilon_{it-j_1}\varepsilon_{it-j_3}\right]E\left[\varepsilon_{is-j_2}\varepsilon_{is-j_4}\right]+E\left[\varepsilon_{it-j_1}\varepsilon_{is-j_2}\right]E\left[\varepsilon_{it-j_3}\varepsilon_{is-j_4}\right]+E\left[\varepsilon_{it-j_1}\varepsilon_{is-j_4}\right]E\left[\varepsilon_{is-j_2}\varepsilon_{it-j_3}\right]+\kappa_4\left[\varepsilon_{it}\right.\right.$$

$$=\frac{2}{NT^2}\sum_{i\in\mathcal{I}_N}\sigma_i^4 1_{\{i\in\mathcal{I}_N\}}\sum_{a,b=1}^{p+1}B_0^{a,b}\sum_{t=1}^{T}\sum_{j_1=1}^{t-1}\zeta'\chi_{j_1}\left(\theta_0\right)\chi_{j_1}^{a}\left(\theta_0\right)\sum_{s=1}^{t-1}\sum_{j_2=1}^{s}\zeta'\chi_{j_2}\left(\theta_0\right)\chi_{j_2}^{b}\left(\theta_0\right)+o\left(1\right)$$

$$=\frac{2}{NT^2}\sum_{i\in\mathcal{I}_N}\sigma_i^4 1_{\{i\in\mathcal{I}_N\}}\sum_{a,b=1}^{p+1}B_0^{a,b}\sum_{t=1}^{T}\sum_{j_1=1}^{t-1}\zeta'\chi_{j_1}\left(\theta_0\right)\chi_{j_1}^{a}\left(\theta_0\right)\sum_{k=1}^{t-1}\left(t-k\right)\zeta'\chi_{k}\left(\theta_0\right)\chi_{k}^{b}\left(\theta_0\right)+o\left(1\right)$$

$$=\bar{\sigma}_1^{(4)}\zeta'B_0 B_0^{-1}B_0\zeta+o\left(1\right)$$

$$=\bar{\sigma}_1^{(4)}\zeta'B_0\zeta+o\left(1\right)$$

taking only into account only the first pair of expectations since the other terms contribute a smaller order, while the term in $\varepsilon_{is}d_{it,NT}$,

$$\frac{2N^{-3/4}}{T^{3/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{it}\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}$$

converges to zero, because its expectation is

$$\frac{2N^{-3/4}}{T^{3/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left[\zeta'\chi_t\left(L;\theta_0\right)\varepsilon_{it}\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}\right]$$

$$\leq K\frac{N^{1/4}}{T^{3/2}}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\left(t-s\right)_+^{-2}=O\left(N^{1/4}T^{-1/2}\right)=o\left(1\right),$$

because $E\left[\varepsilon_{is}\dot{\varepsilon}_{is}\right]=0$. Finally, using similar arguments, we still obtain $V_{NT}\to_p 2\left(p+1\right)\sigma^{(8)}$ as under H_0, and the theorem follows. $\square$

8. Lemmas

Lemma 1. *Under Assumptions A, as $(N, T)_j \to \infty$,*

$$\frac{1}{N} \sum_{i=1}^{N} \eta_i^2 = O_p(T^{2\varrho + 2\delta_0 - 6} + T^{-1} \log T + N^{-1} T^{4\delta_0 - 6} + N^{-2})$$

$$\frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_i^2 = O_p(1).$$

Proof of Lemma 1. We only prove the first statement, since the second one is an easy consequence of the first one, (9) and $\bar{\gamma}^2 \to_p (E[\gamma_i])^2 > 0$ and $E[\gamma_i^2] < \infty$. Write

$$\frac{1}{N} \sum_{i=1}^{N} \eta_i^2 = \frac{\frac{1}{NT^2} \sum_{t=1}^{T} \sum_{t'=1}^{T} \Delta \bar{y}_t \Delta \bar{y}_{t'} \sum_{i=1}^{N} \lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) (\varepsilon_{it} - \frac{\gamma_i}{\bar{\gamma}} \bar{\varepsilon}_t) \lambda_{t'}^{-1} \left(L; \theta_0^{(-1)} \right) (\varepsilon_{it'} - \frac{\gamma_i}{\bar{\gamma}} \bar{\varepsilon}_{t'})}{\left(\frac{1}{T} \sum_{t=1}^{T} (\Delta \bar{y}_t)^2 \right)^2}.$$

The denominator converges to a positive constant term because

$$\frac{1}{T} \sum_{t=1}^{T} (\Delta \bar{y}_t)^2 = \bar{\gamma}^2 \frac{1}{T} \sum_{t=1}^{T} (\Delta f_t)^2 + \frac{1}{T} \sum_{t=1}^{T} (\lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \bar{\varepsilon}_t)^2 + 2\bar{\gamma} \frac{1}{T} \sum_{t=1}^{T} \Delta f_t \lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \bar{\varepsilon}_t$$

and by Assumptions A.3 and 4, satisfies as $(N, T)_j \to \infty$,

$$\frac{1}{T} \sum_{t=1}^{T} (\Delta \bar{y}_t)^2 \to_p E(\gamma_i)^2 \sigma_{\Delta f_t}^2, \quad \sigma_{\Delta f_t}^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\left[(\Delta f_t)^2 \right],$$

since $\varrho < 2/3$ and the second term is negligible due to cross section averaging, because it has expectation

$$\frac{1}{TN^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E\left[\lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{it} \lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{jt} \right]$$

$$= \frac{\sigma^2}{TN} \sum_{t=1}^{T} E\left[\sum_{j=0}^{t} \lambda_j^{-1} \left(\theta_0^{(-1)} \right)^2 \right] = O\left(N^{-1} \right) = o\left(1 \right)$$

and variance

$$\frac{1}{T^2 N^4} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t'=1}^{T} \sum_{i'=1}^{N} \sum_{j'=1}^{N}$$

$$\times \left(E\left[\lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{it} \lambda_{t'}^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{i't'} \right] E\left[\lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{jt} \lambda_{t'}^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{j't'} \right] \right.$$

$$\left. + E\left[\lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{it} \lambda_{t'}^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{j't'} \right] E\left[\lambda_t^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{jt} \lambda_{t'}^{-1} \left(L; \theta_0^{(-1)} \right) \varepsilon_{i't'} \right] \right)$$

$$= O\left(T^{-2} N^{-4} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t'=1}^{T} \left\{ |t - t'|_+^{2(\delta_0 - 1) - 1} \right\}^2 \right) = O\left(T^{-1} N^{-2} \left(T^{4(\delta_0 - 1) - 1} + \log T \right) \right) = o\left(1 \right),$$

31

and the third is $o_p(1)$ because of Cauchy-Schwarz inequality.

In the numerator, it suffices to focus on the dominating term ε_{it} of the error term $\varepsilon_{it} - \frac{\gamma_i}{\bar\gamma}\bar\varepsilon_t$, since $\bar\varepsilon_t = O_p(N^{-1/2})$ and $\bar\gamma \to_p E(\gamma_i) \neq 0$ by Assumption A.4. Then,

$$\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta\bar y_t \Delta\bar y_{t'}\sum_{i=1}^{N}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'} \tag{28}$$

$$= \frac{\bar\gamma^2}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta f_t \Delta f_{t'}\sum_{i=1}^{N}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}$$

$$+ \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_t\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_{t'}\sum_{i=1}^{N}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}$$

$$+ \frac{2\bar\gamma}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta f_t\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_{t'}\sum_{i=1}^{N}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}.$$

Up to the factor $\bar\gamma^2$, the expectation of the first term in (28), which is positive, is, using the independence of f_t and ε_{it} and Assumption A.3,

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(\Delta f_t \Delta f_{t'}\right)E\left(\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right).$$

The expectations above for all $t \neq t'$ are, cf. Lemma 8,

$$E\left(\Delta f_t \Delta f_{t'}\right) = O\left(|t-t'|_+^{2(\varrho-1)-1} + |t-t'|_+^{\varrho-2}\right)$$

$$E\left(\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right) = O\left(|t-t'|_+^{2(\delta_0-1)-1} + |t-t'|_+^{\delta_0-2}\right)$$

where $|a|_+ = \max\{|a|,1\}$ and bounded for $t=t'$ because $\max\{\varrho,\delta_0\} < 2/3$, so that Δf_t and $\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}$ are asymptotically stationary. Then, this term is

$$O_p\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{t}|t-t'|_+^{2\varrho+2\delta_0-6} + |t-t'|_+^{\varrho+\delta_0-4}\right) = O_p\left(T^{2\varrho+2\delta_0-6} + T^{-1}\log T\right).$$

The expectation of the second term in (28), which is also positive, is

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left[\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_t\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_{t'}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right]$$

$$= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}E\left[\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{jt}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{kt'}\lambda_t^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right]$$

$$= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{a=1}^{t}\sum_{b=1}^{t'}\sum_{c=1}^{t}\sum_{d=1}^{t'}\tau_a^0\tau_b^0\tau_c^0\tau_\delta^0 E\left[\varepsilon_{jt-a}\varepsilon_{kt'-b}\varepsilon_{it-c}\varepsilon_{it'-d}\right],$$

32

where $\tau_a^0 = \tau_a(\theta_0) = \lambda_a^{-1}\left(\theta_0^{(-1)}\right)$ and the expectation can be written using the indicator function $1\{\cdot\}$ as

$$
\begin{aligned}
= \; & E\left[\varepsilon_{jt-a}\varepsilon_{kt'-b}\right] E\left[\varepsilon_{it-c}\varepsilon_{it'-d}\right] 1\{t - a = t' - b\} 1\{t - c = t' - d\} 1\{j = k\} \\
& + E\left[\varepsilon_{jt-a}\varepsilon_{it'-d}\right] E\left[\varepsilon_{kt'-b}\varepsilon_{it-c}\right] 1\{t - a = t' - d\} 1\{t' - b = t - c\} 1\{j = i = k\} \\
& + E\left[\varepsilon_{jt-a}\varepsilon_{it-c}\right] E\left[\varepsilon_{kt'-b}\varepsilon_{it'-d}\right] 1\{t - a = t - c\} 1\{t' - b = t' - d\} 1\{j = i = k\} \\
& + \kappa_4\left[\varepsilon_{it}\right] \varrho\{t - a = t' - b = t - c = t' - d\} 1\{j = k = i\}.
\end{aligned}
$$

This leads to four different types of contributions, the first type being

$$
\begin{aligned}
& \frac{\sigma^4}{NT^2} \sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{a=1}^{t\wedge t'}\sum_{c=1}^{t\wedge t'} \tau_a^0 \tau_{a+|t-t'|}^0 \tau_c^0 \tau_{c+|t-t'|}^0 \\
= \; & O\left(\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{t} |t-t'|_+^{4(\delta_0-1)-2} + |t-t'|_+^{2\delta_0-4}\right) = O\left(N^{-1}\left(T^{-1}\log T + T^{4(\delta_0-1)-2}\right)\right),
\end{aligned}
$$

proceeding as in Lemma 8. The second type is

$$
\frac{\sigma^4}{N^2T^2} \sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{a=1}^{t\wedge t'}\sum_{c=1}^{t\wedge t'} \tau_a^0 \tau_{a+|t-t'|}^0 \tau_c^0 \tau_{c+|t-t'|}^0 = O\left(N^{-2}\left(T^{-1}\log T + T^{4(\delta_0-1)-2}\right)\right)
$$

and the third one is, using that $\left(\tau_a^0\right)^2 = \pi_a^2(1-\delta_0) \sim a^{2\delta_0-4}$ and $\delta_0 < 3/2$,

$$
\frac{\sigma^4}{N^2T^2} \sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{a=1}^{t}\sum_{b=1}^{t'} \left(\tau_a^0\right)^2 \left(\tau_b^0\right)^2 = O\left(N^{-2}\right).
$$

The final fourth type involving fourth order cumulants is

$$
\frac{\kappa_4}{N^2T^2} \sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{a=1}^{t\wedge t'} \left(\tau_a^0 \tau_{a+|t-t'|}^0\right)^2 = O\left(\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} |t-t'|_+^{2\delta_0-4}\right) = O\left(N^{-1}T^{-1}\right),
$$

The third term in (28) can be bounded using Cauchy-Schwarz inequality and the Lemma follows compiling all the contributions. $\square$

Lemma 2. *Under Assumptions A and B, as $T \to \infty$,*

(a) $\quad \displaystyle\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}(\lambda_{t-1}(L;\theta)f_t)^2\right| = O_p\left(1 + T^{2(\varrho-\delta)-1}\right)$

(b) $\quad \displaystyle\frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}(L;\theta_0)f_t * \chi_{t-1}(L;\xi_0)\lambda_{t-1}(L;\theta_0)f_t = O_p\left(1 + T^{2(\varrho-\delta_0)-1}\log T\right)$

(c) $\quad \displaystyle\frac{1}{T}\sum_{t=1}^{T}\dot\tau_{t-1}(\theta_0)\lambda_{t-1}(L;\theta_0)f_t = O_p\left(T^{-1} + \left\{T^{2(\varrho-2\delta_0)-1} + T^{-\delta_0-1} + T^{-2\delta_0} + T^{2(\varrho-\delta_0-1)-\delta_0}\right\}^{1/2}\log T\right).$

Proof of Lemma 2. To prove (a) note that by the triangle inequality,

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} (\lambda_{t-1}(L;\theta) f_t)^2 \right| \leq \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \left\{ (\lambda_{t-1}(L;\theta) f_t)^2 - E\left[(\lambda_{t-1}(L;\theta) f_t)^2\right] \right\} \right| \quad (29)$$

$$+ \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} E\left[(\lambda_{t-1}(L;\theta) f_t)^2\right] \right|.$$

Under Assumption 2, we have

$$\lambda_{t-1}(L;\theta) f_t = \psi(L;\xi) \Delta_{t-1}^{\delta-\varrho} z_t = \sum_{j=0}^{t-1} \lambda_j(\delta - \varrho;\xi) z_{t-j} = \sum_{j=0}^{\infty} c_j v_{t-j},$$

where $c_j = c_j(\delta - \varrho, \xi) = \sum_{k=0}^{j} \varphi_k^f \lambda_{j-k}(\delta - \varrho, \xi) \sim c j^{\varrho-\delta-1}$ as $j \to \infty$ under Assumption A.2.

First, notice that uniformly in $\theta \in \Theta$

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} E\left[(\lambda_{t-1}(L;\theta) f_t)^2\right] \right| = \sup_{\theta \in \Theta} \left| \frac{\sigma_v^2}{T} \sum_{t=1}^{T} \sum_{j=0}^{t} c_j^2 \right| \leq \sup_{\theta \in \Theta} \left| \frac{K}{T} \sum_{t=1}^{T} \left(1 + t^{2(\varrho-\delta)-1}\right) \right| = O(1 + T^{2(\varrho-\underline{\delta})-1}),$$

while the first term on the lhs of (29) is

$$\frac{1}{T} \sum_{j=1}^{T-1} c_j^2 \sum_{l=1}^{T-j} (v_l^2 - \sigma_v^2) + \frac{2}{T} \sum_{j=0}^{T-2} \sum_{k=j+1}^{T-1} c_j c_k \sum_{l=k-j+1}^{T-j} v_l v_{l-(k-j)} = (a) + (b),$$

say. Then, with $\gamma_v(j) = E[v_0 v_j]$,

$$E \sup_{\Theta} |(a)| \leq \frac{1}{T} \sum_{j=0}^{T-1} \sup_{\Theta} c_j^2 E \left| \sum_{l=1}^{T-j} (v_l^2 - \gamma_v(j)) \right|.$$

Uniformly in j, $Var(\sum_{l=1}^{T-j} v_l^2) = O(T)$, so

$$\sup_{\Theta} |(a)| = O_p \left(T^{-1/2} \sum_{j=1}^{T-1} j^{2(\varrho-\underline{\delta})-2} \right) = O_p(T^{-1/2} \log T + T^{2(\varrho-\underline{\delta})-3/2}).$$

Next, using summation by parts, we can express (b) as

$$\frac{2c_{T-1}}{T} \sum_{j=0}^{T-2} c_j \sum_{k=j+1}^{T-1} \sum_{l=k-j+1}^{T-j} \left\{ v_l v_{l-(k-j)} - \gamma_v(j-k) \right\}$$

$$+ \frac{2}{T} \sum_{j=0}^{T-2} c_j \sum_{k=j+1}^{T-2} (c_{k+1} - c_k) \sum_{r=j+1}^{k} \sum_{l=r-j+1}^{T-j} \left\{ v_l v_{l-(r-j)} - \gamma_v(j-r) \right\} = (b_1) + (b_2).$$

34

Uniformly in j,

$$Var\left(\sum_{k=j+1}^{T-1}\sum_{l=k-j+1}^{T-j} v_l v_{l-(k-j)}\right) = O(T^2),$$

so,

$$
\begin{aligned}
E\sup_{\Theta}|(b_1)| &\leq KT^{-1}T^{\varrho-\underline{\delta}-1}\sum_{j=0}^{T} j^{\varrho-\underline{\delta}-1}\left\{Var\left(\sum_{k=j+1}^{T-1}\sum_{l=k-j+1}^{T-j} v_l v_{l-(k-j)}\right)\right\}^{1/2}\\
&= O(T^{2(\varrho-\underline{\delta})-1}+T^{\varrho-\underline{\delta}-1}\log T)
\end{aligned}
$$

where K is some arbitrarily large positive constant. Similarly,

$$
\begin{aligned}
E\sup_{\Theta}|(b_2)| &\leq KT^{-1}\sum_{j=0}^{T} j^{\varrho-\underline{\delta}-1}\sum_{k=j+1}^{T} k^{\varrho-\underline{\delta}-2}\left\{Var\left(\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{T-j} v_l v_{l-(r-j)}\right)\right\}^{1/2}\\
&= O(T^{2(\varrho-\underline{\delta})-1}+T^{\varrho-\underline{\delta}-1}\log T+1)
\end{aligned}
$$

since

$$Var\left(\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{T-j} v_l v_{l-(r-j)}\right) \leq K(k-j)(T-j).$$

The proof of (b) of the Lemma is similar but simpler than that of (a) and is omitted.

To prove (c) of the Lemma, note that $T^{-1}\sum_{t=1}^{T}\lambda_{t-1}(L;\theta_0)f_t\dot{\tau}_t(\theta_0)$ has zero mean and variance

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}\dot{\tau}_t(\theta_0)\dot{\tau}_r(\theta_0)E\left[\lambda_{t-1}(L;\theta_0)f_t\lambda_{r-1}(L;\theta_0)f_r\right]. \tag{30}$$

When $0 \leq \varrho - \delta_0 \leq 1$, $|E[\lambda_{t-1}(L;\theta_0)f_t * \lambda_{r-1}(L;\theta_0)f_r]| \leq K|t-r|_+^{2(\varrho-\delta_0)-1}$ and using that $|\dot{\tau}_t(\theta_0)| \leq Kt^{-\delta_0}\log t$, (30) is

$$
\begin{aligned}
&O\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{t}(tr)^{-\delta_0}\log t\log r|t-r|_+^{2(\varrho-\delta_0)-1}\right)\\
&= O\left(\frac{1}{T^2}\sum_{t=1}^{T}t^{-\delta_0}\log^2 t\left\{t^{-\delta_0}\left(t^{2(\varrho-\delta_0)}+\log t\right)+\left(t^{1-\delta_0}+\log t\right)\left(t^{2(\varrho-\delta_0)-1}+1\right)\right\}\right)\\
&= O\left(T^{-2}\right)+O\left(T^{-1-\delta_0}\left\{T^{-\delta_0}\left(T^{2(\varrho-\delta_0)}+\log T\right)+\left(T^{1-\delta_0}+1\right)\left(T^{2(\varrho-\delta_0)-1}+1\right)\right\}\right)\log^2 T\\
&= O\left(T^{-2}\right)+O\left(\left\{T^{-1-2\delta_0}\left(T^{2(\varrho-\delta_0)}+\log T\right)+\left(T^{1-\delta_0}+1\right)\left(T^{2(\varrho-\delta_0-1)-\delta_0}+T^{-1-\delta_0}\right)\right\}\right)\log^2 T\\
&= O\left(T^{-2}\right)+O\left(\left\{T^{2(\varrho-2\delta_0)-1}+T^{-1-2\delta_0}\log T+T^{-1-\delta_0}+T^{-2\delta_0}+T^{2(\varrho-2\delta_0-1)+1}+T^{2(\varrho-\delta_0-1)-\delta_0}\right\}\right)\log^2 T\\
&= O\left(T^{-2}\right)+O\left(\left\{T^{2(\varrho-2\delta_0)-1}+T^{-1-\delta_0}+T^{-2\delta_0}+T^{2(\varrho-\delta_0-1)-\delta_0}\right\}\log^2 T\right).
\end{aligned}
$$

When $\varrho - \delta_0 < 0$, $|E[\lambda_{t-1}(L;\theta_0)f_t * \lambda_{r-1}(L;\theta_0)f_r]| \leq K|t-r|_+^{\varrho-\delta_0-1}$, see Lemma 8, so (30)

is

$$O\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{t}(tr)^{-\delta_0}\log t\log r|t-r|_+^{\varrho-\delta_0-1}\right)$$

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

and the result follows. $\square$

Lemma 3. *Under the assumptions of Theorem 1, as* $(N,T)_j\to\infty$,

$$\sup_{\theta\in\Theta_1\cup\Theta_2\cup\Theta_3}\left|\frac{\bar\gamma}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\left(\varepsilon_{it}-\hat\phi_i\bar\varepsilon_t\right)\right|=o_p(1).$$

Proof of Lemma 3. For $\theta\in\Theta_1\cup\Theta_2\cup\Theta_3$, since $\bar\gamma\to_p E[\gamma_i]=O_p(1)$ as $N\to\infty$, we only need to consider

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\left(\varepsilon_{it}-\hat\phi_i\bar\varepsilon_t\right)$$

$$=\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\varepsilon_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\hat\phi_i\bar\varepsilon_t,$$

where the first term is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\varepsilon_{it}$$

$$=\ \frac{1}{T^{-1}\sum_t(\Delta\bar y_t)^2}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\Delta\bar y_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\left(\varepsilon_{ir}-\frac{\gamma_i}{\bar\gamma}\bar\varepsilon_r\right)*\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\varepsilon_{it}$$

$$=\ \frac{1}{T^{-1}\sum_t(\Delta\bar y_t)^2}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\left(\bar\gamma\Delta f_r+\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_r\right)\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\left(\varepsilon_{ir}-\frac{\gamma_i}{\bar\gamma}\bar\varepsilon_r\right)*\lambda_{t-1}(L;\theta)f_t*\lambda_t^0(L;\theta)\varepsilon_{it}.$$

Next $\bar\gamma^{-1}=O_p(1)$ as $N\to\infty$ and $\frac{1}{T^{-1}\sum_t(\Delta\bar y_t)^2}=O_p(1)$ as $T\to\infty$, cf. proof of Lemma 1, while

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\left(\bar\gamma\Delta f_r+\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_r\right)\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\left(\varepsilon_{ir}-\frac{\gamma_i}{\bar\gamma}\bar\varepsilon_r\right)\lambda_{t-1}(L;\theta)f_t\lambda_t^0(L;\theta)\varepsilon_{it} \qquad (31)$$

$$=\ \frac{\bar\gamma}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\Delta f_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{t-1}(L;\theta)f_t\lambda_t^0(L;\theta)\varepsilon_{it}$$

$$+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{t-1}(L;\theta)f_t\lambda_t^0(L;\theta)\varepsilon_{it}$$

$$-\frac{1}{NT^2\bar\gamma}\sum_{i=1}^{N}\gamma_i\sum_{t=1}^{T}\sum_{r=1}^{T}\bar\gamma\Delta f_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_r\lambda_{t-1}(L;\theta)f_t\lambda_t^0(L;\theta)\varepsilon_{it}$$

$$-\frac{1}{NT^2\bar\gamma}\sum_{i=1}^{N}\gamma_i\sum_{t=1}^{T}\sum_{r=1}^{T}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar\varepsilon_r\lambda_{t-1}(L;\theta)f_t\lambda_t^0(L;\theta)\varepsilon_{it}.$$

The first term on the rhs of (31) can be written as $\bar{\gamma}$ times

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{t}\sum_{k=0}^{t}\lambda_j\left(\delta-\varrho,\xi\right)\lambda_k^0\left(\theta\right)z_{t-j}\varepsilon_{it-k}\frac{1}{T}\sum_{r=1}^{T}\Delta f_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}$$

which using Lemma 8 and $|a|_+ = \max\{|a|, 1\}$ has expectation

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\Delta f_r\lambda_{t-1}\left(L;\theta\right)f_t\right]E\left[\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\right]$$

$$= O\left(\begin{array}{c}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}\left(|t-r|_+^{2(\varrho-1)-\underline{\delta}}+|t-r|_+^{\varrho-1-\underline{\delta}}+|t-r|_+^{\varrho-2}\right)\\\times\left(|t-r|_+^{2(\delta_0-1)-\underline{\delta}}+|t-r|_+^{\delta_0-1-\underline{\delta}}+|t-r|_+^{\delta_0-2}\right)\end{array}\right)$$

$$= o\left(1\right)$$

uniformly in $\theta \in \Theta_1 \cup \Theta_2 \cup \Theta_3$, since all exponents in $|t-r|_+$ are negative under Assumptions A and B*.1, so that we can write its centered version as

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{t}\sum_{k=0}^{t}\lambda_j\left(\delta-\varrho,\xi\right)\lambda_k^0\left(\theta\right)A_{i,t-j,t-k}$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{t}\lambda_j\left(\delta-\varrho,\xi\right)\lambda_j^0\left(\theta\right)A_{i,t-j,t-j}$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{t}\sum_{k\neq j}\lambda_j\left(\delta-\varrho,\xi\right)\lambda_k^0\left(\theta\right)A_{i,t-j,t-k}$$

$= (a) + (b)$, say, where

$$A_{i,t-j,t-k} = z_{t-j}\varepsilon_{it-k}\frac{1}{T}\sum_{r=1}^{T}\Delta_r^{1-\varrho}z_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}-\frac{1}{T}\sum_{r=1}^{T}E\left[z_{t-j}\varepsilon_{it-k}\Delta_r^{1-\varrho}z_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right].$$

Then

$$E\sup_{\delta}|\left(a\right)| \leq \frac{1}{T}\sum_{j=0}^{T}\sup_{\delta}\left|\lambda_j\left(\delta-\varrho,\xi\right)\lambda_j^0\left(\theta\right)\right|E\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{\ell=1}^{T-j}A_{i,\ell,\ell}\right|,$$

where

$$Var\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{\ell=1}^{T-j}A_{i,\ell,\ell}\right] = \frac{1}{N}Var\left[\sum_{\ell=1}^{T-j}A_{i,\ell,\ell}\right]$$

with

$$Var\left[\sum_{\ell=1}^{T-j}A_{i,\ell,\ell}\right] = \sum_{\ell=1}^{T-j}Var\left[A_{i,\ell,\ell}\right] + \sum_{\ell=1}^{T-j}\sum_{\ell'\neq\ell}Cov\left[A_{i,\ell,\ell}, A_{i,\ell',\ell'}\right].$$

Now $Var\left[A_{i,\ell,\ell}\right]$ is

$$\frac{1}{T^2}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}E\left[z_\ell^2\Delta_r^{1-\varrho}z_r\Delta_{r'}^{1-\varrho}z_{r'}\varepsilon_{i\ell}^2\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\\ -E\left[z_\ell\varepsilon_{i\ell}\Delta_r^{1-\varrho}z_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[z_\ell\varepsilon_{i\ell}\Delta_{r'}^{1-\varrho}z_{r'}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\end{array}\right\}$$

$$=\frac{1}{T^2}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}E\left[z_\ell^2\Delta^{1-\varrho}z_r\Delta^{1-\varrho}z_{r'}\right]E\left[\varepsilon_{i\ell}^2\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\\ -E\left[\varepsilon_{i\ell}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[z_\ell\Delta_r^{1-\varrho}z_r\right]E\left[\varepsilon_{i\ell}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]E\left[z_\ell\Delta_{r'}^{1-\varrho}z_{r'}\right]\end{array}\right\}$$

$$=\frac{1}{T^2}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}\left(E\left[z_\ell^2\right]E\left[\Delta^{1-\varrho}z_r\Delta^{1-\varrho}z_{r'}\right]+2E\left[z_\ell\Delta^{1-\varrho}z_r\right]E\left[z_\ell\Delta^{1-\varrho}z_{r'}\right]\right)\\ \times\left(E\left[\varepsilon_{i\ell}^2\right]E\left[\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]+2E\left[\varepsilon_{i\ell}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[\varepsilon_{i\ell}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\right)\\ -E\left[\varepsilon_{i\ell}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[z_\ell\Delta^{1-\varrho}z_r\right]E\left[\varepsilon_{i\ell}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]E\left[z_\ell\Delta^{1-\varrho}z_{r'}\right]\end{array}\right\}$$

and $\sum_{\ell=1}^{T-j}Var\left[A_{i,\ell,\ell}\right]$ is, using Lemma 8 (omitting the logarithmic terms),

$$O\left(\frac{1}{T^2}\sum_{\ell=1}^{T-j}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}\left(|r-r'|_+^{2(\varrho-1)-1}+|r-r'|_+^{\varrho-2}+|r-\ell|^{\varrho-2}|r'-\ell|^{\varrho-2}\right)\\ \times\left(|r-r'|_+^{2(\delta_0-1)-1}+|r-r'|_+^{\delta_0-2}+|r-\ell|^{\delta_0-2}|r'-\ell|^{\delta_0-2}\right)\end{array}\right\}\right)$$

$$=O\left(\log T+T^{2(\varrho-1)+2(\delta_0-1)-1}\right),$$

while using a similar argument

$$Cov\left[A_{i,\ell,\ell},A_{i,\ell',\ell'}\right]$$

$$=\frac{1}{T^2}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}E\left[z_\ell z_{\ell'}\Delta_r^{1-\varrho}z_r\Delta_{r'}^{1-\varrho}z_{r'}\varepsilon_{i\ell}\varepsilon_{i\ell'}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\\ -E\left[z_\ell\varepsilon_{i\ell}\Delta_r^{1-\varrho}z_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[z_{\ell'}\varepsilon_{i\ell'}\Delta_{r'}^{1-\varrho}z_{r'}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\end{array}\right\}$$

$$=\frac{1}{T^2}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}E\left[z_\ell z_{\ell'}\Delta^{1-\varrho}z_r\Delta^{1-\varrho}z_{r'}\right]E\left[\varepsilon_{i\ell}\varepsilon_{i\ell'}\Delta_{r+1}^{1-\delta_0}\varepsilon_{ir}\Delta_{r+1}^{1-\delta_0}\varepsilon_{ir'}\right]\\ -E\left[\varepsilon_{i\ell}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[z_\ell\Delta_r^{1-\varrho}z_r\right]E\left[\varepsilon_{i\ell'}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]E\left[z_{\ell'}\Delta_{r'}^{1-\varrho}z_{r'}\right]\end{array}\right\}$$

$$=\frac{1}{T^2}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}\left(E\left[z_\ell z_{\ell'}\right]E\left[\Delta^{1-\varrho}z_r\Delta^{1-\varrho}z_{r'}\right]+E\left[z_\ell\Delta^{1-\varrho}z_r\right]E\left[z_{\ell'}\Delta^{1-\varrho}z_{r'}\right]+E\left[z_{\ell'}\Delta^{1-\varrho}z_r\right]E\left[z_\ell\Delta^{1-\varrho}z_{r'}\right]\right)\\ \times\left(\begin{array}{c}E\left[\varepsilon_{i\ell}\varepsilon_{i\ell'}\right]E\left[\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\\ +E\left[\varepsilon_{i\ell}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[\varepsilon_{i\ell'}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\\ +E\left[\varepsilon_{i\ell'}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[\varepsilon_{i\ell}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]\end{array}\right)\\ -E\left[\varepsilon_{i\ell}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\right]E\left[z_\ell\Delta^{1-\varrho}z_r\right]E\left[\varepsilon_{i\ell'}\lambda_{r'}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir'}\right]E\left[z_{\ell'}\Delta^{1-\varrho}z_{r'}\right]\end{array}\right\}$$

and using Lemma 8 $\sum_{\ell=1}^{T-j}\sum_{\ell'\neq\ell}Cov\left[A_{i,\ell,\ell},A_{i,\ell',\ell'}\right]$ is

$$O\left(\frac{1}{T^2}\sum_{\ell=1}^{T-j}\sum_{\ell'=1}^{T-j}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}\left(\begin{array}{c}|\ell-\ell'|^{-2}\left(|r-r'|_+^{2(\varrho-1)-1}+|r-r'|_+^{\varrho-2}\right)\\ +|r-\ell|^{\varrho-2}|r'-\ell'|^{\varrho-2}+|r'-\ell|^{\varrho-2}|r-\ell'|^{\varrho-2}\end{array}\right)\\ \times\left(\begin{array}{c}|\ell-\ell'|^{-2}\left(|r-r'|^{2(\delta_0-1)-1}+|r-r'|_+^{\delta_0-2}\right)\\ +|r-\ell|^{\delta_0-2}|r'-\ell'|^{\delta_0-2}+|r-\ell'|^{\delta_0-2}|r'-\ell|^{\delta_0-2}\end{array}\right)\end{array}\right\}\right)$$

$$=O\left(\log T+T^{2(\varrho-1)+2(\delta_0-1)-1}\right).$$

Then, using $\left|\lambda_j\left(\delta - \varrho, \xi\right)\lambda_j^0\left(\theta\right)\right| \leq Cj^{\varrho + \delta_0 - 2\delta - 2}$,

$$
\begin{aligned}
E\sup_\delta\left|\left(a\right)\right| &\leq \frac{1}{T}\sum_{j=0}^{T}\sup_\delta\left|\lambda_j\left(\delta - \varrho, \xi\right)\lambda_j^0\left(\theta\right)\right| E\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{\ell=1}^{T-j}A_{i,\ell,\ell}\right| \\
&= O\left(N^{-1}\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)\left(T^{-2} + \sup_\delta T^{2(\varrho-1)+2(\delta_0-1)-4\delta}\right)\right)^{1/2} \\
&= o\left(1\right) + O\left(N^{-1}T^{4(\varrho-1)+4(\delta_0-1)-1-4\underline{\delta}}\right)^{1/2} = o\left(1\right)
\end{aligned}
$$

since $\delta_0 - 1 < \underline{\delta}/2$ and $\varrho - 1 < \underline{\delta}/2$, using Assumption B*.1.

For (b) a similar result is obtained using summation by parts as in the proof of the bound for (b_2) in Lemma 1. First, we can express $(b) = (b_1) + (b_2)$ with

$$
\begin{aligned}
(b_1) &= \frac{2\lambda_T^0\left(\theta\right)}{NT}\sum_{j=0}^{T-1}\lambda_j\left(\delta - \varrho, \xi\right)\sum_{k=j+1}^{T}\sum_{\ell=k-j+1}^{T-j}\sum_{i=1}^{N}A_{i,\ell,\ell-(k-j)} \\
(b_2) &= \frac{2}{NT}\sum_{j=0}^{T-1}\lambda_j\left(\delta - \varrho, \xi\right)\sum_{k=j+1}^{T-1}\left(\lambda_{k+1}^0\left(\theta\right) - \lambda_k^0\left(\theta\right)\right)\sum_{r=j+1}^{k}\sum_{\ell=r-j+1}^{T-j}\sum_{i=1}^{N}A_{i,\ell,\ell-(r-j)},
\end{aligned}
$$

so that we find that that $E\sup_\delta\left|(b_1)\right|$ is bounded by

$$
\begin{aligned}
&KT^{-1}T^{\delta_0-\underline{\delta}-1}\sum_{j=1}^{T}j^{\varrho-\underline{\delta}-1}TN^{-1/2}\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2} \\
\leq\ &KN^{-1/2}T^{\delta_0-\underline{\delta}-1}\left(1 + T^{\varrho-\underline{\delta}}\right)\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2} \\
\leq\ &K\left\{N^{-1}\left(T^{2(\delta_0-1)-2\underline{\delta}} + T^{2(\delta_0+\varrho-1)-4\underline{\delta}}\right)\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)\right\}^{1/2}
\end{aligned}
$$

which is $o\left(1\right)$ by using Assumptions B*.1-3 while $E\sup_\delta\left|(b_2)\right|$ is bounded by

$$
\begin{aligned}
&KT^{-1}N^{-1/2}\sum_{j=0}^{T-1}j^{\varrho-\underline{\delta}-1}\sum_{k=j+1}^{T-1}k^{\delta_0-\underline{\delta}-2}T\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2} \\
\leq\ &KT^{-1}N^{-1/2}\sum_{j=0}^{T-1}j^{\delta_0+\varrho-2\underline{\delta}-2}T\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2} \\
\leq\ &KN^{-1/2}\left(1 + T^{\varrho+\delta_0-2\underline{\delta}-1}\right)\left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2},
\end{aligned}
$$

which is $o\left(1\right)$ under Assumptions B*.1-3.

The bounds for the other terms on the rhs of (31) follow in a similar form, noting that the presence of cross section averages introduce a further $N^{-1/2}$ factor in the probability bounds. $\square$

Lemma 4. *Under the assumptions of Theorem 1, as $(N, T)_j \to \infty$,*

$$\sup_{\theta \in \Theta_1 \cup \Theta_2 \cup \Theta_3} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_t^0 (L; \theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) \tau_t(\theta)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) \right| = o_p(1).$$

Proof of Lemma 4. Opening the double product $\lambda_t^0 (L; \theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0)$ into four different terms, we study them in turn. First note that the expectation of

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_t^0 (L; \theta) \, \varepsilon_{it} \tau_t(\theta) \varepsilon_{i0} \tag{32}$$

is

$$\frac{\sigma^2}{T} \sum_{t=1}^{T} \tau_t(\theta) \lambda_t^0 (\theta) = O\left(T^{-1} \log T + T^{\delta_0 - 2\underline{\delta} - 1} \right) = o(1)$$

uniformly in δ by Assumption B*.3, so we can show that (32) is negligible by showing that

$$\sup_{\theta \in \Theta_1 \cup \Theta_2 \cup \Theta_3} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=0}^{t} \lambda_j^0 (\theta) \tau_t(\theta) \left\{ \varepsilon_{it-j} \varepsilon_{i0} - \sigma^2 (t = j) \right\} \right| = o_p(1).$$

The term inside the absolute value is

$$\frac{1}{T} \sum_{t=1}^{T} \lambda_t^0 (\theta) \tau_t(\theta) \frac{1}{N} \sum_{i=1}^{N} \left\{ \varepsilon_{i0}^2 - \sigma^2 \right\}$$

$$+ \frac{1}{T} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \lambda_j^0 (\theta) \tau_t(\theta) \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{it-j} \varepsilon_{i0}$$

where the first term is $O_p \left(N^{-1/2} \left(T^{-1} + T^{\delta_0 - 2\underline{\delta} - 1} \right) \right) = o_p(1)$, uniformly in δ, while the second can be written using summation by parts as

$$\frac{1}{T} \sum_{j=0}^{T} \sum_{k=j+1}^{T} \lambda_j^0 (\theta) \tau_k(\delta) \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{ik-j} \varepsilon_{i0}$$

$$= \frac{\tau_T(\delta)}{T} \sum_{j=0}^{T} \lambda_j^0 (\theta) \frac{1}{N} \sum_{i=1}^{N} \sum_{k=j+1}^{T} \varepsilon_{ik-j} \varepsilon_{i0}$$

$$- \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=0}^{T} \lambda_j^0 (\theta) \sum_{k=j+1}^{T} \left\{ \tau_{k+1}(\delta) - \tau_k(\delta) \right\} \frac{1}{N} \sum_{i=1}^{N} \sum_{r=j+1}^{k} \varepsilon_{ir-j} \varepsilon_{i0}$$

$$= (b_1) + (b_2).$$

Then,

$$E \sup_{\delta} |b_1| \leq K T^{-\underline{\delta} - 1} \sum_{j=0}^{T} j^{\delta_0 - \underline{\delta} - 1} N^{-1/2} (T - j)^{1/2}$$

$$\leq K T^{-\underline{\delta} - 1} \left(1 + T^{\delta_0 - \underline{\delta}} \right) N^{-1/2} T^{1/2} \leq K N^{-1/2} \left(T^{-\underline{\delta} - 1/2} + T^{\delta_0 - 2\underline{\delta} - 1/2} \right) = o(1),$$

by Assumption B*.3, because $\mathrm{Var}\left[N^{-1}\sum_{i=1}^{N}\sum_{k=j+1}^{T}\varepsilon_{ik-j}\varepsilon_{i0}\right] \leq KN^{-1/2}\left(T-j\right)^{1/2}$. Similarly,

$$
\begin{aligned}
E\sup_{\delta}|b_1| &\leq KT^{-1}\sum_{j=0}^{T}j^{\delta_0-\underline{\delta}-1}\sum_{k=j+1}^{T}k^{-\underline{\delta}-1}N^{-1/2}\left(k-j\right)^{1/2}\\
&\leq KT^{-1}\sum_{j=0}^{T}j^{\delta_0-\underline{\delta}-1}T^{-\underline{\delta}+1/2}N^{-1/2}\\
&\leq KN^{-1/2}\left(T^{-1}+T^{\delta_0-\underline{\delta}-1}\right)T^{-\underline{\delta}+1/2} \leq KN^{-1/2}\left(T^{-\underline{\delta}-1/2}+T^{\delta_0-2\underline{\delta}-1/2}\right) = o\left(1\right).
\end{aligned}
$$

The second term is

$$
-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\hat{\phi}_i\bar{\varepsilon}_t\tau_t(\theta)\varepsilon_{i0} = -\frac{1}{T}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\bar{\varepsilon}_t\tau_t(\theta)\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i\varepsilon_{i0} = o_p\left(1\right)
$$

because we can show that

$$
\sup_{\theta\in\Theta_1\cup\Theta_2\cup\Theta_3}\left|\frac{1}{T}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\bar{\varepsilon}_t\tau_t(\theta)\right| = o_p\left(1\right)
$$

using the same method as for bounding (32), while

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i\varepsilon_{i0} &= \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i}{\bar{\gamma}}\varepsilon_{i0} + \frac{1}{N}\sum_{i=1}^{N}\eta_i\varepsilon_{i0}\\
&= O_p\left(N^{-1/2}\right) + O_p(T^{2\varrho+2\delta_0-6}+T^{-1}+N^{-1}T^{4\delta_0-6}+N^{-2})^{1/2} = o_p\left(1\right)
\end{aligned}
$$

by Lemma 1 and Cauchy-Schwarz inequality.

The third term,

$$
-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\tau_t(\theta)\hat{\phi}_i\bar{\varepsilon}_0 = -\frac{\bar{\varepsilon}_0}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\tau_t(\theta)\left(\frac{\gamma_i}{\bar{\gamma}}+\eta_i\right)
$$

is negligible because, on the one hand

$$
\sup_{\theta\in\Theta_1\cup\Theta_2\cup\Theta_3}\left|\frac{\bar{\varepsilon}_0}{\bar{\gamma}NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\tau_t(\theta)\gamma_i\right| = o_p\left(1\right)
$$

because $\bar{\varepsilon}_0 = O_p\left(N^{-1/2}\right)$, $\bar{\gamma}^{-1} = O_p(1)$ and the average can be bounded as (32) since γ_i is independent of ε_{it}, which is zero mean, and on the other hand under Assumption B*,

$$
\left|\frac{\bar{\varepsilon}_0}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\tau_t(\theta)\eta_i\right| \leq |\bar{\varepsilon}_0|\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\right)^2\tau_t^2(\theta)\right|^{1/2}\left|\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\right|^{1/2} = o_p\left(1\right)
$$

41

because we can show that

$$\sup_{\theta\in\Theta_1\cup\Theta_2\cup\Theta_3}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\right)^2\tau_t^2(\theta)\right|=O_p\left(1+T^{2(\delta_0-2\underline{\delta})-1}\right)$$

using again the same methods, $|\bar{\varepsilon}_0|=O_p\left(N^{-1/2}\right)$ and $\left|\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\right|=O_p(T^{2\varrho+2\delta_0-6}+T^{-1}+N^{-1}T^{4\delta_0-6}+N^{-2})$ by Lemma 1.

Finally, the last term,

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\hat{\phi}_i^2\bar{\varepsilon}_t\tau_t(\theta)\bar{\varepsilon}_0 \; = \; \bar{\varepsilon}_0\frac{1}{T}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\bar{\varepsilon}_t\tau_t(\theta)\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2
$$

$$
= \; O_p\left(N^{-1/2}\right)o_p\left(1\right)O_p\left(1\right)=o_p\left(1\right),
$$

is also negligible, proceeding as before. $\square$

Lemma 5. *Under the conditions of Theorem 2,*

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\left(\varepsilon_{i0}-\phi_i\bar{\varepsilon}_0\right)*\chi_t\left(L;\xi_0\right)\left(\varepsilon_{it}-\phi_i\bar{\varepsilon}_t\right)=-2\sigma^2\left(\frac{N}{T}\right)^{1/2}\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(\xi_0\right)+o_p\left(1\right).$$

Proof of Lemma 5. The main term on the left hand side converges to its expectation,

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\tau_t(\theta_0)\varepsilon_{i0}*\chi_t\left(L;\xi_0\right)\varepsilon_{it}\right]=-2\sigma^2\left(\frac{N}{T}\right)^{1/2}\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(\xi_0\right),$$

since its variance is

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\tau_t(\theta_0)\tau_r(\theta_0)Cov\left[\varepsilon_{i0}*\chi_t\left(L;\xi_0\right)\varepsilon_{it},\varepsilon_{i0}*\chi_r\left(L;\xi_0\right)\varepsilon_{ir}\right]
$$

$$
= \; \frac{1}{T}\sum_{t=1}^{T}\tau_t(\theta_0)^2\left[\sigma^4\left(\sum_{j=1}^{t}j^{-2}+t^{-2}\right)+\{\kappa_4\}\right]
$$

$$
+\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{t}\tau_t(\theta_0)\tau_r(\theta_0)\left[\sigma^4\left(\sum_{j=0}^{t}j^{-1}\left(t-r+j\right)^{-1}+t^{-1}r^{-1}\right)+\kappa_4 t^{-2}1\left\{t=r\right\}\right]
$$

$$
= \; O\left(T^{-1}+T^{-2\delta_0}\right)+O\left(T^{-1}\sum_{t=1}^{T}\sum_{r=1}^{t}(rt)^{-\delta_0}\left(|t-r|_+^{-1}\log t+(tr)^{-1}\right)\right)
$$

$$
= \; O\left(T^{-1}+T^{-2\delta_0}\right)+O\left(T^{-1}\sum_{t=1}^{T}t^{-\delta_0}\left(t^{-\delta_0}\log^2 t+t^{-1}\log t\right)\right)=O\left(T^{-1}\log^4 T+T^{-2\delta_0}\log^2 T\right),
$$

42

which is $o\left(1\right)$ as $T\to\infty$, while for the other three terms, we can check in turn that

$$
\begin{aligned}
-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\varepsilon_{i0}\hat{\phi}_i\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t &= O_p\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i0}\hat{\phi}_i\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t\right) \\
&= O_p\left((T/N)^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}\hat{\phi}_i\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t\right) \\
&= O_p\left((T/N)^{-1/2}N^{-1/2}\left\{1+T^{1/2-\delta_0}\log^{1/2}T\right\}\right)
\end{aligned}
$$

which is $O_p\left(T^{-1/2}+T^{-\delta_0}\log^{1/2}T\right)=o_p\left(1\right)$ because

$$
\begin{aligned}
\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t &= O_p\left(N^{-1/2}\left\{\sum_{t=1}^{T}\tau_t(\theta_0)^2\log t\right\}^{1/2}\right) \\
&= O_p\left(N^{-1/2}\left\{1+T^{1/2-\delta_0}\log^{1/2}T\right\}\right),
\end{aligned}
$$

while

$$
\begin{aligned}
\left|\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\hat{\phi}_i\bar{\varepsilon}_0\chi_t\left(L;\xi_0\right)\varepsilon_{it}\right| &\le |\bar{\varepsilon}_0|\left|\frac{2}{N}\sum_{i=1}^{N}\hat{\phi}_iT^{-1/2}\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(L;\xi_0\right)\varepsilon_{it}\right| \\
&= O_p\left((NT)^{-1/2}\left\{1+T^{1/2-\delta_0}\log^{1/2}T\right\}\right)=o_p\left(1\right),
\end{aligned}
$$

using $\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i=O_p\left(1\right)$, $|\bar{\varepsilon}_0|=O_p\left(N^{-1/2}\right)$ and the same argument as for $N=1$, and finally

$$
\begin{aligned}
\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\bar{\varepsilon}_0\hat{\phi}_i^2\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t &= \sqrt{N}\bar{\varepsilon}_0\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2T^{-1/2}\sum_{t=1}^{T}\tau_t(\theta_0)\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t \\
&= O_p\left((NT)^{-1/2}\left\{\sum_{t=1}^{T}\tau_t(\theta_0)^2\log t\right\}^{1/2}\right) \\
&= O_p\left(N^{-1/2}\left\{T^{-1/2}+T^{-\delta_0}\log^{1/2}T\right\}\right)=o_p\left(1\right),
\end{aligned}
$$

and the proof is completed. $\square$

Lemma 6. *Under the conditions of Theorem 2,*

$$
\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right)\left[\chi_t\left(L;\xi_0\right)\varepsilon_{it}-\hat{\phi}_i\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t\right]\right\}\to_d\mathcal{N}\left(0,4B\left(\xi_0\right)\right).
$$

Proof of Lemma 6. The left hand side can be written as

$$
\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{it}\chi_t\left(L;\xi_0\right)\varepsilon_{it}-\varepsilon_{it}\hat{\phi}_i\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t-\hat{\phi}_i\bar{\varepsilon}_t\chi_t\left(L;\xi_0\right)\varepsilon_{it}+\hat{\phi}_i^2\bar{\varepsilon}_t*\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t\right\}
$$

$$
\tag{33}
$$

where Proposition 2 in Robinson and Velasco (2015) shows the asymptotic $\mathcal{N}(0, 4B(\xi_0))$ distribution of the first term as $(N, T)_j \to \infty$, and we now show that the remainder terms are negligible. Then the second term on (33) can be written as

$$\frac{2}{\sqrt{NT}} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \left\{ \frac{\gamma_i}{\bar{\gamma}} + \eta_i \right\} \chi_t(L; \xi_0) \varepsilon_{jt},$$

where $2(NT)^{-1/2} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \gamma_i \chi_t(L; \xi_0) \varepsilon_{jt}$ has zero expectation and variance proportional to

$$\frac{1}{NT} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{i'=1}^{N} \sum_{j'=1}^{N} \sum_{t'=1}^{T} E\left[\varepsilon_{it} \gamma_i \chi_t(L; \xi_0) \varepsilon_{jt} \varepsilon_{i't'} \gamma_{i'} \chi_{t'}(L; \xi_0) \varepsilon_{j't'} \right]$$

$$= \frac{1}{NT} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{i'=1}^{N} \sum_{j'=1}^{N} \sum_{t'=1}^{T} E\left[\gamma_i \gamma_{i'} \right] E\left[\varepsilon_{it} \chi_t(L; \xi_0) \varepsilon_{jt} \varepsilon_{i't'} \chi_{t'}(L; \xi_0) \varepsilon_{j't'} \right]$$

$$= \frac{1}{NT} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} E\left[\gamma_i^2 \right] E\left[\varepsilon_{it}^2 \right] E\left[\{ \chi_t(L; \xi_0) \varepsilon_{jt} \}^2 \right] = O(N^{-1}) = o(1)$$

so this term is $o_p(1)$ as $N \to \infty$. Then the other term depending on η_i is also negligible because, using Cauchy-Schwarz inequality,

$$\left| \frac{2}{\sqrt{NT}} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \eta_i \chi_t(L; \xi_0) \varepsilon_{jt} \right| \leq \frac{2}{\sqrt{NT}} \left(\frac{1}{N} \sum_{i=1}^{N} \eta_i^2 \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \chi_t(L; \xi_0) \varepsilon_{jt} \right)^2 \right)^{1/2}$$

$$= O_p\left((NT)^{-1/2} (T^{2\varrho + 2\delta_0 - 6} + T^{-1})^{1/2} (NT)^{1/2} \right)$$

$$= O_p\left((T^{2\varrho + 2\delta_0 - 6} + T^{-1})^{1/2} \right) = o_p(1)$$

because

$$E\left[\left(\sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \chi_t(L; \xi_0) \varepsilon_{jt} \right)^2 \right] = \sum_{j=1}^{N} \sum_{j'=1}^{N} \sum_{t=1}^{T} \sum_{t'=1}^{T} E\left[\varepsilon_{it} \varepsilon_{it'} \chi_t(L; \xi_0) \varepsilon_{jt} \chi_{t'}(L; \xi_0) \varepsilon_{j't'} \right]$$

$$= \sum_{j=1}^{N} \sum_{t=1}^{T} E\left[\varepsilon_{it}^2 \right] E\left[\{ \chi_t(L; \xi_0) \varepsilon_{jt} \}^2 \right] = O(NT).$$

The third term in (33) is also $o_p(1)$ since it can be written as

$$\frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \chi_t(L; \xi_0) \varepsilon_{it} \hat{\phi}_i \bar{\varepsilon}_t = \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \frac{\gamma_i}{\bar{\gamma}} + \eta_i \right\} \chi_t(L; \xi_0) \varepsilon_{it} \bar{\varepsilon}_t$$

where $2(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_i\chi_t(L;\xi_0)\varepsilon_{it}\bar{\varepsilon}_t$ has zero expectation and variance

$$\frac{2}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\sum_{t'=1}^{T}E\left[\gamma_i\gamma_{i'}\right]E\left[\chi_t(L;\xi_0)\varepsilon_{it}\varepsilon_{jt}\chi_t(L;\xi_0)\varepsilon_{i't'}\varepsilon_{j't'}\right]$$

$$=\frac{2}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E\left[\gamma_i^2\right]E\left[\varepsilon_{jt}^2\right]E\left[\{\chi_t(L;\xi_0)\varepsilon_{it}\}^2\right]=O\left(N^{-1}\right)$$

while

$$\left|\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\chi_t(L;\xi_0)\varepsilon_{it}\bar{\varepsilon}_t\right|\ \leq\ \frac{2N}{\sqrt{NT}}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\chi_t(L;\xi_0)\varepsilon_{it}\bar{\varepsilon}_t\right)^2\right)^{1/2}$$

$$=\ O_p\left(N^{1/2}T^{-1/2}(T^{2\varrho+2\delta_0-6}+T^{-1})^{1/2}\left(TN^{-1}\right)^{1/2}\right)$$

$$=\ O_p\left((T^{2\varrho+2\delta_0-6}+T^{-1})^{1/2}\right)=o_p(1)$$

because

$$E\left[\left(\sum_{t=1}^{T}\chi_t(L;\xi_0)\varepsilon_{it}\bar{\varepsilon}_t\right)^2\right]\ =\ \frac{1}{N^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{j=1}^{N}\sum_{j'=1}^{N}E\left[\chi_t(L;\xi_0)\varepsilon_{it}\varepsilon_{jt}\chi_{t'}(L;\xi_0)\varepsilon_{it'}\varepsilon_{j't'}\right]$$

$$=\ \frac{1}{N^2}\sum_{t=1}^{T}\sum_{j=1}^{N}E\left[\varepsilon_{jt}^2\right]E\left[\{\chi_t(L;\xi_0)\varepsilon_{it}\}^2\right]=O\left(TN^{-1}\right).$$

Finally, the fourth term in (33) is also negligible, since

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\phi}_i^2\bar{\varepsilon}_t\chi_t(L;\xi_0)\bar{\varepsilon}_t\ =\ \frac{2}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2\frac{1}{N}\sum_{a=1}^{N}\sum_{b=1}^{N}\sum_{t=1}^{T}\varepsilon_{at}\chi_t(L;\xi_0)\varepsilon_{bt}$$

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

since $N^{-1}\sum_{i=1}^{N}\hat{\phi}_i^2=O_p(1)$ and $N^{-1}\sum_{a=1}^{N}\sum_{b=1}^{N}\sum_{t=1}^{T}\varepsilon_{at}\chi_t(L;\xi_0)\varepsilon_{bt}$ is $O_p\left(T^{1/2}\right)$ because it has zero expectation and variance

$$\frac{1}{N^2}\sum_{a=1}^{N}\sum_{b=1}^{N}\sum_{a'=1}^{N}\sum_{b'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left[\varepsilon_{at}\varepsilon_{a't'}\chi_t(L;\xi_0)\varepsilon_{bt}\chi_{t'}(L;\xi_0)\varepsilon_{b't'}\right]$$

$$=\ \frac{1}{N^2}\sum_{a=1}^{N}\sum_{b=1}^{N}\sum_{t=1}^{T}E\left[\varepsilon_{at}^2\right]E\left[\{\chi_t(L;\xi_0)\varepsilon_{bt}\}^2\right]=O\left(T\right).\qquad\square$$

Lemma 7. *Under the assumptions of Theorem 2 and for $\theta\to_p\theta_0$,*

$$\ddot{L}_{N,T}(\theta)\to_p\ddot{L}_{N,T}(\theta_0).$$

Proof of Lemma 7. This follows as Theorem 2 of Hualde and Robinson (2011), using the same techniques as in the proof of Theorem 1 to bound uniformly the initial condition and projection terms in a neighborhood of θ_0. $\square$

Lemma 8. *Under Assumptions A and B*.1, for $\theta \in \Theta$, as $T \to \infty$,*

$$E\left[\Delta f_r \lambda_{t-1}\left(L;\theta\right) f_t\right]$$
$$= O\left(|t-r|_+^{2(\varrho-1)-\delta}\left(\log|t-r|_+\right)^{1\{\varrho=1\}+1\{\varrho=\delta\}} + |t-r|_+^{\varrho-1-\delta} + |t-r|_+^{\varrho-2}\right)$$

$$E\left[\lambda_{t-1}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{t-1}^0\left(L;\theta\right)\varepsilon_{it}\right]$$
$$= O\left(|t-r|_+^{2(\delta_0-1)-\delta}\left(\log|t-r|_+\right)^{1\{\delta_0=1\}+1\{\delta_0=\delta\}} + |t-r|_+^{\delta_0-1-\delta} + |t-r|_+^{\delta_0-2}\right),$$

where $|a|_+ = \max\{|a|, 1\}$, $1\{\cdot\}$ is the indicator function and

$$E\left[\Delta^{1-\varrho}z_r z_t\right] = O\left(|t-r|_+^{\varrho-2}\right)$$
$$E\left[\lambda_{t-1}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\varepsilon_{it}\right] = O\left(|t-r|_+^{\delta_0-2}\right).$$

Proof of Lemma 8. We only prove the statement for $E\left[\Delta f_r \lambda_{t-1}\left(L;\theta\right) f_t\right]$, since the rest follow similarly. Under Assumption A.2, if $t > r$

$$E\left[\Delta f_r \lambda_{t-1}\left(L;\theta\right) f_t\right] = E\left[\Delta_r^{1-\varrho}z_r \lambda_{t-1}\left(L;\delta-\varrho,\xi\right) z_t\right] = \sigma_v^2 \sum_{j=0}^r d_j\left(1-\varrho\right) c_{j+t-r}\left(\delta-\varrho\right),$$

where $d_j\left(a\right) = \sum_{k=0}^j \varphi_k^f \pi_{j-k}(a) \sim cj^{-a-1}$ and $c_j\left(a\right) = c_j\left(a,\xi\right) = \sum_{k=0}^j \varphi_k^f \lambda_{j-k}(a,\xi) \sim cj^{-a-1}$ as $j \to \infty$, $d_j\left(0\right) = \varphi_j^f$ and $\sum_{j=0}^\infty d_j\left(a\right) = \sum_{j=0}^\infty c_j\left(a\right) = 0$ if $a > 0$, $\xi \in \Xi$, so that the absolute value of the last expression is bounded by, $\varrho \geq 1$,

$$K\sum_{j=0}^r |d_j(1-\varrho)|\left(j+t-r\right)^{\varrho-\delta-1} \leq K\left(t-r\right)^{\varrho-\delta-1}\sum_{j=0}^{t-r}|d_j(1-\varrho)| + K\sum_{j=t-r+1}^r j^{2\varrho-\delta-3}$$
$$\leq K\left(t-r\right)^{\varrho-\delta-1}\left(t-r\right)^{\varrho-1}\left(\log|t-r|_+\right)^{1\{\varrho=1\}} + K\left(t-r\right)^{2(\varrho-1)-\delta}$$
$$= O\left(\left(t-r\right)^{2(\varrho-1)-\delta}\left(\log|t-r|_+\right)^{1\{\varrho=1\}}\right)$$

since $\varrho - 1 < \delta$, $\varrho < 3/2$ and $2(\varrho-1) - \delta < 0$ by Assumption B*.1, and $d_j(1-\varrho) \sim cj^{\varrho-2}$, $\varrho > 1$, while $d_j(0)$ is summable.

If $\varrho < 1$, then using summation by parts $E\left[\Delta f_r \lambda_{t-1}\left(L;\theta\right) f_t\right]$ is equal to

$$\sigma_v^2 \sum_{j=0}^{r-1}\left\{c_{j+t-r+1}\left(\delta-\varrho\right) - c_{j+t-r}\left(\delta-\varrho\right)\right\}\sum_{k=0}^j d_k\left(1-\varrho\right) + c_t\left(\delta-\varrho\right)\sum_{k=0}^r d_k\left(1-\varrho\right)$$
$$= O\left(\left(t-r\right)^{\varrho-\delta-2}\sum_{j=0}^{t-r} j^{\varrho-1} + \sum_{j=t-r}^{r-1} j^{2\varrho-3-\delta} + t^{\varrho-\delta-1}\right)$$
$$= O\left(\left(t-r\right)^{2(\varrho-1)-\delta} + \left(t-r\right)^{\varrho-\delta-1}\right),$$

46

using that $c_{j+t-r+1}(\delta - \varrho) - c_{j+t-r}(\delta - \varrho) = c_{j+t-r+1}(\delta - \varrho + 1)$ and $\sum_{k=0}^{r} d_k (1 - \varrho) \leq K$.

If $r > t$

$$E\left[\Delta f_r \lambda_{t-1}(L; \theta) f_t\right] = \sigma_v^2 \sum_{j=0}^{t} d_{j+r-t}(1 - \varrho) c_j(\delta - \varrho),$$

so that the absolute of the last expression is bounded by, $\varrho \geq \delta$,

$$
\begin{aligned}
K \sum_{j=0}^{t} (j + r - t)^{\varrho - 2} |c_j(\delta - \varrho)| \;\leq\;& K (r - t)^{\varrho - 2} \sum_{j=0}^{r-t} |c_j(\delta - \varrho)| + K \sum_{j=r-t+1}^{t} j^{2\varrho - \delta - 3} \\
\leq\;& K (r - t)^{\varrho - 2} (r - t)^{\varrho - \delta} \left(\log |t - r|_+\right)^{1\{\varrho = \delta\}} + K (r - t)^{2(\varrho - 1) - \delta} \\
=\;& O\left((r - t)^{2(\varrho - 1) - \delta} \left(\log |t - r|_+\right)^{1\{\varrho = \delta\}}\right).
\end{aligned}
$$

since $\varrho - 1 < \delta$ and $\varrho < 3/2$ and $c_j(\delta - \varrho) \sim c j^{\varrho - 1 - \delta}$, $\varrho > \delta$.

If $\varrho < \delta$, then using summation by parts $E\left[\Delta f_r \lambda_{t-1}(L; \theta) f_t\right]$ is equal to

$$
\begin{aligned}
&\sigma_v^2 \sum_{j=0}^{t-1} \left\{c_{j+r-t+1}(1 - \varrho) - c_{j+r-t}(1 - \varrho)\right\} \sum_{k=0}^{j} d_k(\delta - \varrho) + c_r(1 - \varrho) \sum_{k=0}^{t} d_k(\delta - \varrho) \\
=\;& O\left((r - t)^{\varrho - 3} \sum_{j=0}^{r-t} j^{\varrho - \delta} + \sum_{j=r-t}^{t-1} j^{2\varrho - 3 - \delta} + r^{\varrho - 2}\right) \\
=\;& O\left((r - t)^{2(\varrho - 1) - \delta} + (r - t)^{\varrho - 2}\right).
\end{aligned}
$$

Similarly, if $r = t$

$$E\left[\Delta f_t \lambda_{t-1}(L; \theta) f_t\right] = \sigma_v^2 \sum_{j=0}^{t} c_j(1 - \varrho) d_j(\delta - \varrho) = O(1),$$

as the absolute value of the last expression is bounded by $\sum_{j=0}^{r} j^{2(\varrho - 1) - \delta - 1} \leq K$, since $2(\varrho - 1) - \delta < 0$ by Assumption B*.1. $\square$

Lemma 9. *Under the Assumptions of Theorem 4,*

$$\sum_{t=1}^{T} z_{NT,t} \to_d N\left(0, 2\bar{\sigma}^8 (p + 1)\right).$$

where $z_{NT,t}$ and $\bar{\sigma}^2$ are defined in (26) and (8) respectively.

Proof of Lemma 9. Since $z_{NT,t}$ is a martingale difference sequence given the independence of ε_{it}, we have

$$\text{Var}\left[\sum_{t=1}^{T} z_{NT,t}\right] = \sum_{t=1}^{T} \text{Var}\left[z_{NT,t}\right] = \sum_{t=1}^{T} E\left[z_{NT,t}^2\right]$$

where, fully expanding matrix calculations using $\dot{\varepsilon}_{it} = \chi_t(L)\varepsilon_{it} = \sum_{j=1}^{t}\chi_j\varepsilon_{it-j}$, with $\chi_j = \left(\chi_j^1,\ldots,\chi_j^{p+1}\right)'$,

$$
\begin{aligned}
\operatorname{Var}\left[z_{NT,t}\right] &= \frac{4}{NT^2}\sum_{i=1}^{N}\sigma_i^2 E\left[\left(\sum_{s=1}^{t-1}\varepsilon_{is}\dot{\varepsilon}_{it}'B_0^{-1}\dot{\varepsilon}_{is}\right)^2\right] \\
&= \frac{4}{NT^2}\sum_{i=1}^{N}\sigma_i^2\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}E\left[\varepsilon_{is}\varepsilon_{ir}\dot{\varepsilon}_{it}'B_0^{-1}\dot{\varepsilon}_{is}\dot{\varepsilon}_{it}'B_0^{-1}\dot{\varepsilon}_{ir}\right] \\
&= \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^2\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}E\left[\varepsilon_{is}\varepsilon_{ir}\dot{\varepsilon}_{it}^a\dot{\varepsilon}_{is}^b\dot{\varepsilon}_{it}^c\dot{\varepsilon}_{ir}^d\right] \\
&= \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^2\sum_{s=1}^{t-1}E\left[\varepsilon_{is}^2\right]E\left[\dot{\varepsilon}_{it}^a\dot{\varepsilon}_{it}^c\right]E\left[\dot{\varepsilon}_{is}^b\dot{\varepsilon}_{is}^d\right](1+o(1)) \\
&= \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^8\sum_{j=1}^{t}\chi_j^a\chi_j^c\sum_{s=1}^{t-1}\sum_{k=1}^{s}\chi_k^b\chi_k^d(1+o(1)),
\end{aligned}
$$

(omitting lower order contributions from other pairwise moments and higher order cumulant combinations) which is

$$
\begin{aligned}
& \frac{4}{NT^2}\sum_{a,b,c,d=1}^{p+1}B_{a,b}^{-1}B_{c,d}^{-1}\sum_{i=1}^{N}\sigma_i^8\sum_{j=1}^{t}\chi_j^a\chi_j^c\sum_{k=1}^{t-1}(t-k)\chi_k^b\chi_k^d(1+o(1)) \\
={}& \frac{4t}{NT^2}\sum_{i=1}^{N}\sigma_i^8\operatorname{trace}\left\{B_0^{-1}B_0 B_0^{-1}B_0\right\}(1+o(1)) \\
={}& \frac{4t}{NT^2}\sum_{i=1}^{N}\sigma_i^8(p+1)(1+o(1))
\end{aligned}
$$

since $\sum_{k=1}^{t-1}(t-k)\chi_k^b\chi_k^d = t\sum_{k=1}^{t-1}\chi_k^b\chi_k^d(1+o(1))$ as $t\to\infty$, because $\chi_k^b = O(k^{-1})$. Therefore,

$$
\operatorname{Var}\left[\sum_{t=1}^{T}z_{NT,t}\right] \to \frac{4(p+1)}{NT^2}\sum_{i=1}^{N}\sigma_i^8\sum_{t=1}^{T}t \to 2\bar{\sigma}^8(p+1),
$$

as $(N,T)_j\to\infty$, and to apply a standard CLT, e.g. Brown (1971) it remains to show that

$$
\sum_{t=1}^{T}E\left[z_{NT,t}^2\big|\mathcal{I}_{N,t-1}\right] \to_p \operatorname{Var}\left[\sum_{t=1}^{T}z_{NT,t}\right] = \sum_{t=1}^{T}E\left[z_{NT,t}^2\right], \tag{34}
$$

where $\mathcal{I}_{N,t-1} = \sigma\left(\varepsilon_{it-1},\varepsilon_{it-2},\ldots;i=1,\ldots,N\right)$ and that the Lindeberg condition holds . For the first condition, note that

$$
E\left[z_{NT,t}^2\big|\mathcal{I}_{N,t-1}\right] = \frac{4}{NT^2}\sum_{i=1}^{N}\sigma_i^2\left(\sum_{s=1}^{t-1}\varepsilon_{is}\dot{\varepsilon}_{it}'B_0^{-1}\dot{\varepsilon}_{is}\right)^2
$$

and therefore, proceeding as in the proof of Theorem 4, the expectation of the lhs of (34) is

$$\frac{4}{NT^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sigma_i^2 E\left[\left(\sum_{s=1}^{t-1} \varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is}\right)^2\right] = \frac{4}{NT^2} \sum_{i=1}^{N} \sigma_i^2 \sum_{t=1}^{T} \sum_{s=1}^{t-1} \sum_{r=1}^{t-1} E\left[\varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \varepsilon_{ir} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{ir}\right]$$

$$= \frac{4}{NT^2} \sum_{i=1}^{N} \sigma_i^8 \sum_{t=1}^{T} t * trace\left\{B_0^{-1} B_0 B_0^{-1} B_0\right\} (1 + o(1))$$

$$= 2\bar{\sigma}^{(8)} (p+1) + o(1)$$

$$= \mathrm{Var}\left[\sum_{t=1}^{T} z_{NT,t}\right] + o(1)$$

while its variance is

$$\frac{16}{N^2 T^4} \sum_{i=1}^{N} \sigma_i^4 \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \mathrm{Cov}\left[\left(\sum_{s=1}^{t_1-1} \varepsilon_{is} \dot{\varepsilon}_{it_1}' B_0^{-1} \dot{\varepsilon}_{is}\right)^2, \left(\sum_{s=1}^{t_2-1} \varepsilon_{is} \dot{\varepsilon}_{it_2}' B_0^{-1} \dot{\varepsilon}_{is}\right)^2\right] \tag{35}$$

$$= \frac{16}{N^2 T^4} \sum_{i=1}^{N} \sigma_i^4 \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{s_1=1}^{t_1-1} \sum_{s_2=1}^{t_1-1} \sum_{s_3=1}^{t_2-1} \sum_{s_4=1}^{t_2-1} E\left[\varepsilon_{is_1} \dot{\varepsilon}_{it_1}' B_0^{-1} \dot{\varepsilon}_{is_1} \varepsilon_{is_2} \dot{\varepsilon}_{it_1}' B_0^{-1} \dot{\varepsilon}_{is_2} \varepsilon_{is_3} \dot{\varepsilon}_{it_2}' B_0^{-1} \dot{\varepsilon}_{is_3} \varepsilon_{is_4} \dot{\varepsilon}_{it_2}' B_0^{-1} \dot{\varepsilon}_{is_4}\right]$$

minus the contribution of $E\left[\left(\sum_{s=1}^{t-1} \varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is}\right)^2\right] E\left[\left(\sum_{s'=1}^{t'-1} \varepsilon_{is'} \dot{\varepsilon}_{it'}' B_0^{-1} \dot{\varepsilon}_{is'}\right)^2\right]$. The expectation can be expressed as a linear combination of products of second order moments and higher order cumulants of ε_{it} (noting that the elements of $\dot{\varepsilon}_{it}$ are linear combinations of $\dot{\varepsilon}_{is-j}$, $j = 1, 2, \ldots, t$), of which the ones providing the contributions with largest order of magnitude are those expressed as (the product of six) pairwise expectations among $\{\dot{\varepsilon}_{it_1}, \dot{\varepsilon}_{it_1}, \dot{\varepsilon}_{it_2}, \dot{\varepsilon}_{it_2}, \varepsilon_{is_1}, \dot{\varepsilon}_{is_1}, \ldots, \varepsilon_{is_4}, \dot{\varepsilon}_{is_4}\}$. All of them would imply at least 4 expectations with bounds of the type $|s_j - s_k|_+^{-1} \log T$ and/or $|t_h - s_j|_+^{-1} \log T$, $j, k = 1, \ldots, 4, j \neq k; h = 1, 2$, (because $\dot{\varepsilon}_{it_1}$ and $\dot{\varepsilon}_{it_2}$ appear twice each, and can be combined in two pairwise expectations of the form $E\left[\dot{\varepsilon}_{it_h}^2\right]$, considering $p = 0$ wlog, and because the product of expectations term, involving only 2 such bounds, is removed in the covariance calculation). All these combinations of 4 bounds imply that the contribution of at least 2 time summations out of 6 (in $\{t_1, t_2, s_1, \ldots, s_4\}$) can be bounded by $O\left(\log^4 T\right)$, while the contribution from the other 4 summations is $O\left(T^4\right)$, so that (35) can be bounded by

$$\frac{K}{N^2 T^4} \sum_{i=1}^{N} \sigma_i^4 T^4 \log^4 T = O\left(N^{-1} \log^4 T\right) = o(1)$$

as $(N, T)_j \to \infty$, and (34) holds.

To check Lindeberg condition we show the sufficient condition

$$\sum_{t=1}^{T} E\left[z_{NT,t}^4\right] \to 0, \tag{36}$$

as $(N, T)_j \to \infty$. Denoting $\mu_{i4} = E\left[\varepsilon_{it}^4\right] < \infty$, we write

$$\sum_{t=1}^{T} E\left[z_{NT,t}^4\right] = \frac{16}{N^2 T^4} \sum_{i=1}^{N} \mu_{i4} \sum_{t=1}^{T} E\left[\sum_{j=1}^{4} \sum_{s_j=1}^{t-1} \varepsilon_{is_j} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is_j}\right].$$

The expectation involves again second order moments and higher order cumulants of ε_{it} of which the ones providing the largest magnitude contributions are again those expressed as (products of six) pairwise expectations among $\{\dot{\varepsilon}_{it}, \dot{\varepsilon}_{it}, \dot{\varepsilon}_{it}, \dot{\varepsilon}_{it}, \varepsilon_{is_1}, \dot{\varepsilon}_{is_1}, \dots, \varepsilon_{is_4}, \dot{\varepsilon}_{is_4}\}$. All of them would imply at least 4 expectations with bounds of the type $|s_j - s_k|_+^{-1} \log T$ and/or $|t - s_j|_+^{-1} \log T$, $j, k = 1, \dots, 4, j \neq k$ (because $\dot{\varepsilon}_{it}$ appears 4 times and can be combined in two pairwise expectations $E\left[\dot{\varepsilon}_{it}^2\right]$ considering $p = 0$ wlog). These 4 bounds imply that the contribution of at least 2 time summations out of 5 (in $\{t, s_1, \dots, s_4\}$) can be bounded by $O\left(\log^4 T\right)$, and the contribution from the other 3 summations is $O\left(T^3\right)$. Then we can bound

$$\sum_{t=1}^{T} E\left[z_{NT,t}^4\right] \leq \frac{K}{NT^4} T^3 \log^4 T = O\left(N^{-1} T^{-1} \log^4 N\right) = o\left(1\right)$$

as $(N, T)_j \to \infty$, so (36) holds and the lemma follows. $\square$

Lemma 10. *Under the Assumptions of Theorem 4,*

$$\sum_{t=1}^{T} \left(z_{NT,t} - \hat{z}_{NT,t}\right) = o_p\left(1\right)$$

where $z_{NT,t}$ is defined in (26) and

$$\hat{z}_{NT,t} = \frac{2}{N^{1/2} T} \sum_{i=1}^{N} \varepsilon_{it}(\hat{\theta}) \sum_{s=1}^{t-1} \varepsilon_{is}(\hat{\theta}) \dot{\varepsilon}_{it}'(\hat{\theta}) B(\hat{\theta})^{-1} \dot{\varepsilon}_{is}(\hat{\theta}).$$

Proof of Lemma 10. Note that

$$\varepsilon_{it}(\hat{\theta}) = \lambda_t^0\left(L; \hat{\theta}\right)\left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t\right) - \tau_t(\hat{\theta}^*)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) - \eta_i \bar{\gamma} \lambda_{t-1}\left(L; \hat{\theta}\right) f_t$$

so that

$$\varepsilon_{it}(\hat{\theta}) - \varepsilon_{it} = \left(\hat{\theta} - \theta\right) \chi_t\left(L; \hat{\theta}^*\right)\left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t\right) - \tau_t(\hat{\theta}^*)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) - \eta_i \bar{\gamma} \lambda_{t-1}\left(L; \hat{\theta}\right) f_t$$

for $\left\|\hat{\theta}^* - \theta\right\| \leq \left\|\hat{\theta} - \theta\right\|$ and

$$\begin{aligned}
\dot{\varepsilon}_{it}(\hat{\theta}) - \dot{\varepsilon}_{it} &= \chi_t\left(L; \hat{\theta}\right) \varepsilon_{it}(\hat{\theta}) - \chi_t\left(L; \theta_0\right) \varepsilon_{it} \\
&= \chi_t\left(L; \hat{\theta}\right)\left(\varepsilon_{it}(\hat{\theta}) - \varepsilon_{it}\right) + \left(\chi_t\left(L; \hat{\theta}\right) - \chi_t\left(L; \theta_0\right)\right) \varepsilon_{it}
\end{aligned}$$

50

with $B(\hat{\theta})^{-1} = B_0^{-1} + O_p\left(\left\|\hat{\theta} - \theta\right\|\right).$

We analyze one of the terms in which the difference $z_{NT,t} - \hat{z}_{NT,t}$ can be decomposed, which is one of the critical ones since its analysis can not exploit the martingale difference structure that can be used to bound the other contributions. Following as in the proof of Theorem 1, we can write that

$$\frac{2N^{-1/2}}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\varepsilon_{it} - \varepsilon_{it}(\hat{\theta})\right) \sum_{s=1}^{t-1} \varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}$$

$$= \frac{2N^{-1/2}}{T} \left(\hat{\theta} - \theta\right) \sum_{i=1}^{N} \sum_{t=1}^{T} \chi_t\left(L;\hat{\theta}^*\right) \left(\varepsilon_{it} - \hat{\phi}_i\bar{\varepsilon}_t\right) \sum_{s=1}^{t-1} \varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is} \tag{37}$$

$$-\frac{2N^{-1/2}}{T} \sum_{i=1}^{N}(\varepsilon_{i0} - \hat{\phi}_i\bar{\varepsilon}_0) \sum_{t=1}^{T} \tau_t(\hat{\theta}^*) \sum_{s=1}^{t-1} \varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is} \tag{38}$$

$$-\frac{2N^{-1/2}}{T} \sum_{i=1}^{N} \eta_i \sum_{t=1}^{T} \bar{\gamma}\lambda_{t-1}\left(L;\hat{\theta}\right) f_t \sum_{s=1}^{t-1} \varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}. \tag{39}$$

Then

$$\frac{2N^{-1/2}}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \chi_t\left(L;\theta_0\right) \varepsilon_{it}\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}$$

has expectation bounded by

$$K\frac{N^{-1/2}}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \left(|t-s|_+^{-2} \log^2 t\right) = O\left(N^{1/2}\log^2 T\right)$$

because $E\left[\varepsilon_{is}B_0^{-1}\dot{\varepsilon}_{is}\right] = 0$, and variance bounded by

$$\frac{K}{T^2} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{s=1}^{t-1} \sum_{s'=1}^{t'-1} \left(|s-s'|_+^{-1} |t'-s|_+^{-1} |t'-s'|_+^{-1} \log^3 T\right) = O\left(\log^5 T\right)$$

by independence of cross-sections, so using $\hat{\theta} - \theta = O_p\left((NT)^{-1/2}\right)$, the term (37) can be showed to be $O_p\left(T^{-1/2}\log^3 T\right) = o_p(1)$, since the contribution of the factor $\hat{\phi}_i\bar{\varepsilon}_t$ can be treated similarly using Cauchy Schwarz inequality and that $E[\bar{\varepsilon}_t^2] = O\left(N^{-1}\right).$

Next,

$$\frac{2N^{-1/2}}{T} \sum_{i=1}^{N} \varepsilon_{i0} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \tau_t(\theta_0)\varepsilon_{is}\dot{\varepsilon}'_{it}B_0^{-1}\dot{\varepsilon}_{is}$$

has expectation bounded by

$$\frac{2N^{-1/2}}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{t-1} t^{-\delta_0} \left(s^{-1} |t-s|_+^{-1} \log t\right) = O\left(N^{1/2}T^{-1}\log^2 T\right) = o(1),$$

51

and variance bounded by

$$\frac{K}{T^2} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{s=1}^{t-1} \sum_{s'=1}^{t'-1} (tt')^{-\delta_0} \left(\begin{array}{c} t^{-1}s^{-1} \left\{ |t'-s'|_+^{-1} |s-s'|_+^{-1} \right\} \log^3 t \\ + s'^{-1}s^{-1} \left\{ |t-s|_+^{-1} |t'-s'|_+^{-1} + |t-t'|_+^{-1} |s-s'|_+^{-1} \right\} \log^3 t \end{array} \right) = O\left(T^{-1} \log^5 T\right)$$

so (38) is $o_p(1)$.

Finally, (39) is also $o_p(1)$ because

$$\left| \frac{2N^{-1/2}}{T} \sum_{i=1}^{N} \eta_i \sum_{t=1}^{T} \sum_{s=1}^{t-1} \lambda_{t-1}(L;\theta_0) f_t \varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \right|$$

$$\leq \frac{2N^{-1/2}}{T} \left| \sum_{i=1}^{N} \eta_i^2 \right|^{1/2} \left| \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \sum_{s=1}^{t-1} \lambda_{t-1}(L;\theta_0) f_t \varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \right)^2 \right|^{1/2}$$

$$= O_p\left(T^{-1} \left| N^{-1} \sum_{i=1}^{N} \eta_i^2 \right|^{1/2} \right) O_p\left(NT^2 \right)^{1/2}$$

$$= O_p\left(N^{1/2} \left| N^{-1} \sum_{i=1}^{N} \eta_i^2 \right|^{1/2} \right) = o_p(1)$$

using $2(\varrho - \delta_0) < 1$ and Assumption C2 with the bound of Lemma 1, because

$$E\left(\sum_{t=1}^{T} \sum_{s=1}^{t-1} \lambda_{t-1}(L;\theta_0) f_t \varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \right)^2$$

$$= \sum_{t=1}^{T} \sum_{s=1}^{t-1} \sum_{t'=1}^{T} \sum_{s'=1}^{t'-1} E\left[\lambda_{t-1}(L;\theta_0) f_t \lambda_{t'-1}(L;\theta_0) f_{t'} \right] E\left[\varepsilon_{is} \dot{\varepsilon}_{it}' B_0^{-1} \dot{\varepsilon}_{is} \varepsilon_{is'} \dot{\varepsilon}_{it'}' B_0^{-1} \dot{\varepsilon}_{is'} \right]$$

$$\leq K \sum_{t=1}^{T} \sum_{s=1}^{t-1} \sum_{t'=1}^{T} \sum_{s'=1}^{t'-1} \left(|t-t'|_+^{2(\varrho-\delta_0)-1} + |t-t'|_+^{\varrho-\delta_0-1} \right) \left(\begin{array}{c} |s-t'|_+^{-1} |t-s'|_+^{-1} |s-s'|_+^{-1} \\ + |t-t'|_+^{-1} |s-s'|_+^{-2} \end{array} \right)$$

$$= O\left(T^2\right). \quad \square$$

Lemma 11. *Under the Assumptions of Theorem 4,*

$$V_{NT}, V_{NT}^* \to_p 2(p+1) \bar{\sigma}^{(8)}$$

and under cross-section homoskedasticity, $\bar{V}_{NT}, \bar{V}_{NT}^* \to_p 2(p+1) \sigma^8$ as $(N,T)_j \to \infty$.

Proof of Lemma 11. Under conditional heteroskedasticity it follows from uniform convergence of $\hat{\sigma}_i^2$ to σ_i^2 and from convergence of $N^{-1}\sum_{i=1}^N \hat{\sigma}_i^8$ to $\bar{\sigma}^{(8)}$. To show those, write

$$
\begin{aligned}
\hat{\sigma}_i^2 - \sigma_i^2 &= \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}^2(\hat{\theta}) - \sigma_i^2 \\
&= \frac{1}{T}\sum_{t=1}^T \left(\varepsilon_{it}^2(\hat{\theta}) - \varepsilon_{it}^2\right) + \frac{1}{T}\sum_{t=1}^T \left(\varepsilon_{it}^2 - \sigma_i^2\right) \\
&= v_{1,i,N} + v_{2,i,N}, \quad \text{say,}
\end{aligned}
$$

where

$$
\sup_i |v_{2,i,N}| \leq \sum_{i=1}^N |v_{2,i,N}| \leq \left(N \sum_{i=1}^N |v_{2,i,N}|^2\right)^{1/2} = O_p\left(NT^{-1/2}\right)
$$

because $\mathrm{E}[v_{2,i,N}] = 0$ and $\mathrm{Var}[v_{2,i,N}] = O\left(T^{-1}\right)$ uniformly in i, while

$$
\begin{aligned}
\sup_i |v_{1,i,N}| &\leq \sum_{i=1}^N |v_{1,i,N}| \leq \left\|\hat{\theta} - \theta\right\| \sum_{i=1}^N \frac{1}{T}\sum_{t=1}^T \left(\left|\chi_t\left(L;\hat{\theta}^*\right)\varepsilon_{it}\right|^2 + \left|\chi_t\left(L;\hat{\theta}^*\right)\varepsilon_{it}\right||\varepsilon_{it}|\right) \\
&= O_p\left(N^{1/2}T^{-1/2}\log^2 T\right)
\end{aligned}
$$

because $\left\|\hat{\theta} - \theta\right\| = O_p\left(N^{-1/2}T^{-1/2}\right)$ and $\left|\chi_t\left(L;\hat{\theta}^*\right)1\right| \leq K\log t$ uniformly in $\hat{\theta}^*$, so that

$$
\sup_i \left|\hat{\sigma}_i^2 - \sigma_i^2\right| = o_p\left(1\right)
$$

and $\sup_i |\hat{\sigma}_i^n - \sigma_i^n| = o_p\left(1\right)$ for any finite n, and therefore

$$
\frac{1}{N}\sum_{i=1}^N \hat{\sigma}_i^8 - \bar{\sigma}^{(8)} = \frac{1}{N}\sum_{i=1}^N \hat{\sigma}_i^8 - \frac{1}{N}\sum_{i=1}^N \sigma_i^8 + o\left(1\right) \leq \sup_i \left|\hat{\sigma}_i^8 - \sigma_i^8\right| + o\left(1\right) = o_p\left(1\right).
$$

Under homoskedasticity we have that

$$
\hat{\sigma}^2 - \sigma^2 = \frac{1}{N}\sum_{i=1}^N \hat{\sigma}_i^2 - \frac{1}{N}\sum_{i=1}^N \sigma^2 \leq \sup_i \left|\hat{\sigma}_i^2 - \sigma_i^2\right| = o_p\left(1\right).
$$

and therefore $\hat{\sigma}^8 - \bar{\sigma}^{(8)} = o_p\left(1\right)$. Then the rest of the proof follows similarly as that of Lemma 10. $\square$

References

ALVAREZ, J. AND M. ARELLANO (2003): "The Time-series and Cross-section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71(4), 1121–1159.

ARELLANO, M. AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and An Application to Unemployment Equations," *Review of Economic Studies*, 58, 277–297.

BAI, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77(4), 1229–1279.

BROWN, B. (1971): "Martingale Central Limit Theorems," *Annals of Mathematical Statistics*, 42, 59–62.

ERGEMEN, Y. E. (2017): "System Estimation of Panel Data Models under Long-Range Dependence," *forthcoming in Journal of Business and Economic Statistics*.

ERGEMEN, Y. E. AND C. VELASCO (2017): "Estimation of Fractionally Integrated Panels with Fixed-Effects and Cross-Section Dependence," *Journal of Econometrics*, 196(2), 248–258.

HAN, C. AND P. PHILLIPS (2010): "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity," *Econometric Theory*, 26(01), 119–151.

HASSLER, U., M. DEMETRESCU, AND A. I. TARCOLEA (2011): "Asymptotic Normal Tests for Integration in Panels with Cross-Dependent Units," *Advances in Statistical Analysis*, 95, 187–204.

HOSOYA, Y. (2005): "Fractional Invariance Principle," *Journal of Time Series Analysis*, 26, 463–486.

HUALDE, J. AND P. M. ROBINSON (2011): "Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models," *The Annals of Statistics*, 39(6), 3152–3181.

MARINUCCI, D. AND P. ROBINSON (2000): "Weak Convergence of Multivariate Fractional Processes," *Stochastic Processes and their Applications*, 86, 103–120.

NIELSEN, M. Ø. (2014): "Asymptotics for the Conditional-Sum-of-Squares Estimator in Multivariate Fractional Time Series Models," *Journal of Time Series Analysis*, doi: 10.1111/jtsa.12100.

PESARAN, H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

——— (2007): "A Simple Panel Unit Root Test in the Presence of Cross-section Dependence," *Journal of Applied Econometrics*, 22(2), 265–312.

PHILLIPS, P. AND H. R. MOON (1999): "Linear Regression Limit Theory For Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

ROBINSON, P. M. AND C. VELASCO (2015): "Efficient Estimation of Fractionally Integrated Panel Data Models with Fixed Effects," *Journal of Econometrics*, 185, 435–452.

SU, l. AND Q. CHEN (2013): "Testing Homogeneity in Panel Data Models with Interactive Fixed Effects," *Econometric Theory*, 29, 1079–1135.

Table 1: Empirical bias and RMSE of $\hat{\delta}$ and $\tilde{\delta}$

		Uncorrected estimates, $\hat{\delta}$				Bias-corrected estimates, $\tilde{\delta} = \hat{\delta} - T^{-1}\nabla(\hat{\delta})$			
(N, T):		(10, 50)	(10, 100)	(20, 50)	(20, 100)	(10, 50)	(10, 100)	(20, 50)	(20, 100)
$\varrho = 0.4$:									
$\delta_0 = 0.3$	Bias	0.1672	0.1458	0.1787	0.1493	0.0066	0.0355	0.0322	0.0433
	RMSE	0.1761	0.1521	0.1838	0.1532	0.1104	0.0830	0.0869	0.0727
$\delta_0 = 0.6$	Bias	0.0485	0.0368	0.0536	0.0380	-0.0011	0.0076	0.0066	0.0094
	RMSE	0.0657	0.0484	0.0627	0.0438	0.0596	0.0388	0.0435	0.0279
$\delta_0 = 0.9$	Bias	-0.0019	-0.0024	0.0042	0.0018	-0.0078	-0.0054	-0.0009	-0.0009
	RMSE	0.0406	0.0286	0.0289	0.0192	0.0444	0.0301	0.0306	0.0199
$\delta_0 = 1.0$	Bias	-0.0120	-0.0096	-0.0049	-0.0042	-0.0126	-0.0099	-0.0052	-0.0043
	RMSE	0.0422	0.0302	0.0287	0.0196	0.0441	0.0309	0.0299	0.0201
$\delta_0 = 1.1$	Bias	-0.0209	-0.0159	-0.0125	-0.0092	-0.0182	-0.0144	-0.0095	-0.0075
	RMSE	0.0459	0.0332	0.0311	0.0216	0.0459	0.0329	0.0308	0.0212
$\delta_0 = 1.4$	Bias	-0.0549	-0.0400	-0.0402	-0.0291	-0.0474	-0.0361	-0.0326	-0.0252
	RMSE	0.0721	0.0528	0.0530	0.0380	0.0668	0.0499	0.0476	0.0351
$\varrho = 1$:									
$\delta_0 = 0.3$	Bias	0.3595	0.3718	0.3285	0.3346	0.3039	0.3435	0.2649	0.2995
	RMSE	0.3755	0.3856	0.3412	0.3474	0.3380	0.3649	0.2941	0.3209
$\delta_0 = 0.6$	Bias	0.1603	0.1652	0.1315	0.1309	0.1357	0.1526	0.1029	0.1153
	RMSE	0.1809	0.1833	0.1469	0.1461	0.1677	0.1755	0.1288	0.1357
$\delta_0 = 0.9$	Bias	0.0435	0.0478	0.0277	0.0299	0.0404	0.0463	0.0240	0.0281
	RMSE	0.0704	0.0663	0.0479	0.0440	0.0710	0.0662	0.0478	0.0434
$\delta_0 = 1.0$	Bias	0.0213	0.0273	0.0102	0.0149	0.0220	0.0277	0.0105	0.0152
	RMSE	0.0540	0.0471	0.0359	0.0302	0.0559	0.0480	0.0373	0.0308
$\delta_0 = 1.1$	Bias	0.0048	0.0128	-0.0023	0.0050	0.0082	0.0147	0.0010	0.0068
	RMSE	0.0462	0.0358	0.0317	0.0234	0.0480	0.0370	0.0326	0.0242
$\delta_0 = 1.4$	Bias	-0.0316	-0.0146	-0.0270	-0.0121	-0.0240	-0.0106	-0.0194	-0.0081
	RMSE	0.0547	0.0338	0.0416	0.0245	0.0509	0.0323	0.0372	0.0228

Table 2: Empirical coverage of 95% CI based on $\hat{\delta}$ and $\tilde{\delta}$

(N, T):	Uncorrected estimates, $\hat{\delta}$				Bias-corrected estimates, $\tilde{\delta} = \hat{\delta} - T^{-1}\nabla(\hat{\delta})$			
	(10, 50)	(10, 100)	(20, 50)	(20, 100)	(10, 50)	(10, 100)	(20, 50)	(20, 100)
$\varrho = 0.4$:								
$\delta_0 = 0.3$	3.90	0.60	0.10	0.00	48.30	42.90	41.70	33.00
$\delta_0 = 0.6$	68.00	66.00	46.00	43.20	76.90	79.80	75.20	77.30
$\delta_0 = 0.9$	91.80	92.00	91.50	92.90	89.90	90.50	90.40	91.90
$\delta_0 = 1.0$	91.10	90.80	92.30	93.10	89.90	89.90	90.90	92.50
$\delta_0 = 1.1$	87.70	86.40	89.60	89.90	87.90	87.20	89.70	90.30
$\delta_0 = 1.4$	63.40	62.70	61.00	68.30	68.90	66.90	70.00	72.10
$\varrho = 1$:								
$\delta_0 = 0.3$	0.00	0.00	0.00	0.00	5.90	1.40	4.70	0.70
$\delta_0 = 0.6$	13.90	5.90	9.20	11.10	25.90	11.40	23.90	28.70
$\delta_0 = 0.9$	70.60	55.30	73.70	61.40	70.60	55.50	74.70	77.70
$\delta_0 = 1.0$	81.90	72.70	85.70	78.80	80.50	72.20	84.90	78.10
$\delta_0 = 1.1$	87.50	83.90	89.80	87.40	85.80	82.50	89.10	86.20
$\delta_0 = 1.4$	79.50	86.30	75.60	84.30	83.40	87.60	82.40	87.60

Table 3: Size of the test based on $\hat{\delta}$ for 5% nominal size (in percentages)

		Pure fractional case				Serially correlated factor case			
(N, T):	(5, 50)	(5, 100)	(10, 50)	(10, 100)		(5, 50)	(5, 100)	(10, 50)	(10, 100)
$\varrho = 0.4$:									
$\delta_0 = 0.3$	5.30	5.50	3.70	3.10		5.10	5.10	3.70	3.30
$\delta_0 = 0.6$	5.90	6.60	4.10	4.00		5.60	6.20	4.00	3.80
$\delta_0 = 0.9$	5.50	5.90	3.60	3.80		5.20	6.00	3.60	4.00
$\delta_0 = 1.0$	6.60	8.40	6.30	6.70		6.90	8.70	5.80	6.20
$\delta_0 = 1.1$	13.30	15.40	15.30	17.50		14.40	15.00	16.70	17.20
$\varrho = 1$:									
$\delta_0 = 0.3$	5.20	5.10	4.20	3.30		5.60	6.00	5.00	3.30
$\delta_0 = 0.6$	5.30	6.20	3.80	4.00		5.70	5.40	4.20	3.50
$\delta_0 = 0.9$	5.60	6.10	4.20	4.70		6.60	7.20	7.60	6.50
$\delta_0 = 1.0$	8.90	9.50	7.90	7.20		10.30	11.80	14.10	11.50
$\delta_0 = 1.1$	14.30	18.50	18.80	21.10		16.60	19.30	23.10	25.00

Table 4: Power of the test (with $\zeta = 0.1 * N^{1/4}T^{1/2}$) based on $\hat{\delta}$ for 5% nominal size (in percentages)

(N, T):	Pure fractional case				Serially correlated factor case			
	(5, 50)	(5, 100)	(10, 50)	(10, 100)	(5, 50)	(5, 100)	(10, 50)	(10, 100)
$\varrho = 0.4:$								
$\delta_0 = 0.3$	13.90	26.90	8.30	23.80	11.70	18.50	7.30	14.60
$\delta_0 = 0.6$	6.90	8.10	5.30	7.10	7.30	8.90	5.10	7.10
$\delta_0 = 0.9$	53.80	67.90	76.80	90.80	64.40	79.90	82.60	94.80
$\delta_0 = 1.0$	81.80	89.30	91.00	93.40	89.10	93.20	93.50	95.50
$\delta_0 = 1.1$	87.00	93.30	95.80	96.80	84.20	89.60	93.70	95.10
$\varrho = 1:$								
$\delta_0 = 0.3$	9.70	13.00	8.00	11.90	13.80	16.80	11.00	16.70
$\delta_0 = 0.6$	8.00	9.20	6.60	7.40	18.70	23.60	14.80	17.60
$\delta_0 = 0.9$	68.00	84.10	82.90	94.20	79.10	90.70	86.10	90.50
$\delta_0 = 1.0$	89.90	93.90	94.50	97.50	88.50	95.00	95.10	97.50
$\delta_0 = 1.1$	82.50	89.10	91.10	94.00	69.90	78.10	82.70	86.40

Table 5: Power of the test (with $\zeta = 0.5 * N^{1/4}T^{1/2}$) based on $\hat{\delta}$ for 5% nominal size (in percentages)

(N, T):	Pure fractional case				Serially correlated factor case			
	(5, 50)	(5, 100)	(10, 50)	(10, 100)	(5, 50)	(5, 100)	(10, 50)	(10, 100)
$\varrho = 0.4:$								
$\delta_0 = 0.3$	56.60	74.80	58.90	78.80	62.00	72.20	65.10	73.70
$\delta_0 = 0.6$	92.50	94.10	98.00	98.90	92.60	94.80	98.10	98.90
$\delta_0 = 0.9$	84.10	89.00	92.30	96.50	90.10	95.90	96.30	99.50
$\delta_0 = 1.0$	97.30	99.60	99.60	99.80	98.80	100	99.80	100
$\delta_0 = 1.1$	99.20	99.30	99.70	100	98.80	99.00	99.30	99.80
$\varrho = 1:$								
$\delta_0 = 0.3$	94.80	96.70	97.70	98.60	87.20	90.20	91.80	93.30
$\delta_0 = 0.6$	87.90	89.90	94.60	96.40	80.50	86.20	84.60	86.60
$\delta_0 = 0.9$	90.00	97.40	97.50	99.40	95.60	99.30	98.40	99.50
$\delta_0 = 1.0$	99.00	99.90	100	100	99.00	100	99.80	99.80
$\delta_0 = 1.1$	98.80	99.10	99.50	98.80	96.30	97.10	98.50	98.70

Research Papers
2018

CREATES
Center for Research in Econometric
Analysis of Time Series