





Estimation of Fractionally Integrated Panels with Fixed Effects and Cross-Section Dependence

Yunus Emre Ergemen and Carlos Velasco

CREATES Research Paper 2015-35

Department of Economics and Business Economics Aarhus University Fuglesangs Allé 4 DK-8210 Aarhus V Denmark Email: oekonomi@au.dk Tel: +45 8716 5515

Estimation of Fractionally Integrated Panels with Fixed Effects and Cross-Section Dependence^{*} (Extended Draft)

Yunus Emre Ergemen[†]

Carlos Velasco[‡]

CREATES, Aarhus University

Universidad Carlos III de Madrid

August 17, 2015

Abstract

We consider large N, T panel data models with fixed effects, common factors allowing 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-ofsquares 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. Generalizing the basic setup to include covariates and heterogeneous parameters, we propose individual and common-correlation estimates for the slope parameters, while error memory parameters are estimated from regression residuals. The two parameter estimates are \sqrt{T} consistent and asymptotically normal and mutually uncorrelated, irrespective of possible cointegration among idiosyncratic components. A study of small-sample performance and an empirical application to realized volatility persistence are included.

JEL Classification: C22, C23 Keywords: Fractional cointegration, factor models, long memory, realized volatility.

^{*}We thank Manuel Arellano, Jesus Gonzalo, Javier Hualde, Bent Nielsen, Peter M. Robinson, Abderrahim Taamouti, the participants in the 67^{th} Econometric Society European Meeting and CREATES Symposium on Long Memory 2013for helpful comments and discussions. Financial support from the Spanish Plan Nacional de I+D+I (ECO2012-31748) is gratefully acknowledged. The authors also acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

[†]CREATES and Department of Economics and Business Economics, Aarhus University. Fuglesangs Allé 4, Building 2628, 8210 Aarhus V, DENMARK. ergemen@econ.au.dk.

[‡]Corresponding Author. Department of Economics, Universidad Carlos III de Madrid, Calle Madrid, 126, 28903 Getafe (Madrid), SPAIN. Carlos.Velasco@uc3m.es.

1. Introduction

In macroeconomics and finance, variables are generally presented in the form of panels describing dynamic characteristics of different units such as countries or assets. Some of these macroeconomic panels include GDP, interest, inflation and unemployment rates while in finance, it is standard to use a panel data approach in portfolio performance evaluations. Panel data analyses lead to more robust inference under correct specification since they allow for cross sections to be interacting with each other while also accounting for individual crosssection characteristics. Recent research in panel data theory has mainly focused on dealing with unobserved fixed effects and cross-section dependence in stationary weakly dependent panels, for instance, Pesaran (2006) proposes estimation of a general panel data model where all variables are I(0). The research on nonstationary panel data theory is also abundant. However, those papers which both contain nonstationarity and allow for fixed effects and cross-section dependence are limited to the unit-root case. For example, Kapetanios et al. (2011) extend the study by Pesaran (2006) to panels where observables and factors are integrated I(1) processes while regression errors are I(0). Furthermore, Bai and Ng (2004) and Bai (2010) propose unit-root testing procedures when idiosyncratic shocks and the common factor are both I(1). Similarly, Moon and Perron (2004) propose the use of dynamic factors for unit-root testing for panels with cross-section dependence.

In the same way that many economic time series, such as aggregate output, real exchange rates, equity volatility, asset and stock market realized volatility, have been shown to exhibit long-range dependence of non-integer orders, panel data models should also be able to accommodate such behaviour. However, the study of panel data models with fractional integration characteristics has been completely neglected until very recently, and only a few papers study fractional panels. Hassler et al. (2011) propose a test for the memory parameter under a fractionally integrated panel setup with multiple time series. Robinson and Velasco (2015) propose several estimation techniques for a type-II (i.e. time truncated) fractionally integrated panel data model with fixed effects.

In this paper, we consider panel data models where we allow for fractionally integrated long-range dependence in both idiosyncratic shocks and a set of common factors. In these models 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 (a) 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; (b) our model including covariates allows for, but does not require, fractional cointegration identifying long-run relationships between the unobservable idiosyncratic components of the observed time series.

Using a type-II fractionally integrated panel data model with fixed effects and crosssection dependence modelled through a common factor dependence, we allow for long-range persistence through this factor and the integrated idiosyncratic shock. We analyze two models in turn. The basic model assumes a common set of parameters for the dynamics of the idiosyncratic component of all cross-sectional units in absence of covariates. 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.

Once we include covariates in the second model, we can extend the study to cointegrating relationships since we allow the covariates to exhibit long-range persistence as well. The general model with covariates that we present in Section 4 can be seen as an extended version of the setup of Robinson and Hidalgo (1997) and Robinson and Hualde (2003) to panel data models and of Pesaran (2006) to nonstationary systems with possible cointegration among idiosyncratic components of observed variables, where endogeneity of covariates is driven by the common factor structure independent of those idiosyncratic components. However observed time series can display the same memory level due to dependence on a persistent common factor thereby leading to spurious regressions, the error term in the regression equation could be less integrated than the idiosyncratic shocks of covariates, leading to an unobservable cointegrating relationship which can only be disclosed by previously projecting out the factor structure.

To estimate possibly heterogeneous slope and memory parameters, we use a CSS criterion, where individual time series are now projected on (fractionally) differenced cross-section averages of the dependent variable and regressors, leading to GLS type of estimates for the slope parameter. We show that both individual slope and fractional integration parameter estimates are \sqrt{T} consistent, and asymptotically normally distributed. The slope estimates have an asymptotic Gaussian distribution irrespective of the possible cointegration among idiosyncratic components of the observables, which are assumed independent of the regression errors, though observables are not.

We explore the performance of our estimation method via Monte Carlo experiments, which indicate that our estimation method has good small-sample properties. Last but not least, we present an application on industry-level realized volatilities using the general model. We analyze how each industry realized volatility is related to a composite market realized volatility measure. We identify several cointegrating relationships between industry and market realized volatilities, which may have direct implications for policy and investment decisions.

Next section details the first model and necessary assumptions. Section 3 explains the estimation strategy, and discusses the asymptotic behaviour of the first model. Section 4 details the general model where covariates and heterogeneity in the parameters are introduced, and details the projection method. Section 5 presents Monte Carlo studies for both models. Section 6 contains an application on the systematic macroeconomic risk, employing industry-level realized volatility analysis. Finally, Section 7 concludes the paper.

Throughout the paper, we use the notation $(N, T)_j$ to denote joint cross-section and timeseries 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 Basic 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, ..., T, i = 1, ..., 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

 $t \ge 0$,

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

truncates $\lambda(L;\theta) = \lambda_{\infty}(L;\theta)$. We assume that $\lambda(L;\theta)$ 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 $\Gamma(-\delta) = (-1)^{\delta} \infty$ for $\delta = 0, 1, \ldots, \Gamma(0) / \Gamma(0) = 1$; $\psi(L; \xi)$ is a known function such that for complex-valued x, $|\psi(x; \xi)| \neq 0, |x| \leq 1$ and in the expansion

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

the coefficients $\psi_j(\xi)$ satisfy

$$\psi_0(\xi) = 1, \ |\psi_j(\xi)| = O(\exp(-c(\xi)j)),$$
(3)

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

$$\lambda_{j}(\theta) = \sum_{k=0}^{j} \pi_{j-k}(\delta) \psi_{k}(\xi), \quad j \ge 0,$$
(4)

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

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

see Robinson and Velasco (2015), where

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

so the value of δ_0 determines the asymptotic stationarity ($\delta_0 < 1/2$) or nonstationarity ($\delta_0 \ge 1/2$) 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

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

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 Section 4.

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

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

which is then easily comparable to a standard dynamic AR(1) panel data model with crosssection 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, ..., 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, ..., N, t = 0, 1, 2, ..., 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

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

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. Although there is no developed theory for fractionally integrated factor models in the literature, 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 shortmemory 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 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

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

where we denote by θ_0 the true parameter vector, and then Δy_{it} is projected on the crosssection 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

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

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, ..., T) and the length of the filter $\lambda_{t-1}(L; \theta^{(-1)})$ that can be applied to it, with the filter $\Delta \lambda_t^{-1}(L; \theta_0)$ that generates the data, since for instance

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

because $\lambda_t (L; \theta^{(-1)}) \Delta = \lambda_t (L; \theta)$, $t = 0, 1, \dots$ 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).

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

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

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 $|\dot{\psi}_t(L;\xi)| = O(\exp(-c(\xi)j))$. 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

$$abla_T(heta_0) = -\sum_{t=1}^T au_t(heta_0) \left\{ \dot{ au}_t(heta_0) - \chi_t(\xi_0) \right\}$$

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

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

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 $\{1/4, \varrho - 1/2, \varrho/2 - 1/12\} < \delta_0 \le \min\{5/4, 5/2 - \varrho\}$.

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

Corollary 1. Under Assumptions of Theorem 2,

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

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 $TN^{-2} \rightarrow$ 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, $\rho + \delta_0$, 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

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

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

$$L_{i,T}^*(\theta) = \frac{1}{T} \boldsymbol{\varepsilon}_i(\theta) \boldsymbol{\varepsilon}_i(\theta)',$$

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 $(N,T)_j \to \infty$,

$$T^{1/2}\left(\hat{\theta}_{i}-\theta_{i0}\right) \rightarrow_{d} \mathcal{N}\left(0,B^{-1}\left(\xi_{i0}\right)\right)$$

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. The Model with Covariates

In order to be able to fully understand how panel variables that exhibit long-range dependence behave, it is essential to not only allow for fractionally integrated shocks but also include covariates that may be persistent, possibly including cointegrated systems with endogenous regressors. In this section, we propose a heterogeneous panel data model with fixed effects and cross-section dependence where shocks that hit both the dependent variable and covariates may be persistent, and covariates are allowed to be endogenous through this unobserved common factor.

For i = 1, ..., N and t = 0, 1, ..., T, the model that generate the observed series y_{it} and X_{it} is given by

$$y_{it} = \alpha_i + \beta'_{i0} X_{it} + \gamma'_i f_t + \lambda_t^{-1} (L; \theta_{i0}) \varepsilon_{it},$$

$$X_{it} = \mu_i + \Gamma'_i f_t + e_{it}$$
(7)

where X_{it} is $k \times 1$, unobserved f_t is $m \times 1$ with k, m fixed, and γ_i , Γ_i are vectors of factor loadings. The variates α_i and μ_i are covariate-specific fixed effects, and $f_t \sim I(\varrho)$ and $e_{it} \sim I(\vartheta_i)$ with elements satisfying Assumption A.2 where ϱ and ϑ_i are nuisance parameters, and the constant parameters θ_{i0} and β_{i0} are the objects of interest. We later use a random coefficient model for β_{i0} to study the properties of a mean-group type estimate for the average value of β_{i0} .

In the factor models of Pesaran (2006) and Bai (2009) the possible endogenous covariates are I(0), so they can only address cases in which there is no long-range dependence in the panel. Kapetanios et al. (2011) study a model where factors and regressors are I(1) processes while errors are stationary I(0) series. Our approach, on the other hand, is specifically geared towards general nonstationary behaviour in panels and addresses estimation of both cointegrating and non-cointegrating relationships among idiosyncratic terms. We do not explicitly include the presence of observable common factors and time trends in the equations for y_{it} and X_{it} , but these could be incorporated and treated easily by our estimation methods as we later discuss.

We introduce the following regularity conditions that generalize Assumption A to model the system in (7).

Assumption D

D.1. The idiosyncratic shocks, ε_{it} , i = 1, 2, ..., N, t = 1, 2, ..., T are independently distributed across i and identically and independently distributed across t with zero mean and variance σ_i^2 , and have a finite fourth-order moment, and $\delta_{i0} \in (0, 3/2)$.

D.2. The common factor satisfies $f_t = \Delta_t^{-\varrho} z_t^f$, $\varrho < 3/2$, where $z_t^f = \Phi_k^f(L) v_{t-k}^f$ with $\Phi^f(s) = \sum_{k=0}^{\infty} \Phi_k^f s^k$, $\sum_{k=0}^{\infty} k \left\| \Phi_k^f \right\| < \infty$, $\det(\Phi^f(s)) \neq 0$ for $|s| \leq 1$ and $v_t^f \sim iid(0, \Omega_f)$, $\Omega_f > 0$, $E \left\| v_t^f \right\|^4 < \infty$, and the idiosyncratic shocks e_{it} are independent in i and satisfy $e_{it} = \Delta_t^{-\vartheta_i} z_{it}^e$, $\sup_i \vartheta_i < 3/2$, where $z_{it}^e = \Phi_{ik}^e(L) v_{it-k}^e$ with $\Phi_i^e(s) = \sum_{k=0}^{\infty} \Phi_{ik}^e s^k$, $\sup_i \sum_{k=0}^{\infty} k \left\| \Phi_{ik}^e \right\| < \infty$, $\det(\Phi_i^e(s)) \neq 0$ for $|s| \leq 1$ and $v_{it}^e \sim iid(0, \Omega_{ie})$, $\Omega_{ie} > 0$, $\sup_{i,t} E \left\| v_{it}^e \right\|^4 < \infty$.

D.3. The covariate-specific idiosyncratic shocks, e_{it} , the idiosyncratic error terms, ε_{it} , and the unobservable common factor, f_t , are all pairwise independent and independent of γ_i and Γ_i , which are also independent in *i*.

D.4. Rank $(\overline{\mathbf{C}}_N) = m \leq k+1$, where the matrix $\overline{\mathbf{C}}_N$ is

$$\overline{\mathbf{C}}_{N} = \begin{pmatrix} \overline{\beta'_{0} \Gamma'}_{N} + \overline{\gamma'}_{N} \\ \overline{\Gamma'}_{N} \end{pmatrix}$$

with $\overline{\gamma}_N = N^{-1} \sum_{i=1}^N \gamma_i$, $\overline{\Gamma}_N = N^{-1} \sum_{i=1}^N \Gamma_i$, $\overline{\beta'_0 \Gamma'}_N = N^{-1} \sum_{i=1}^N \beta'_{i0} \Gamma'_i$.

Assumption D.1 relaxes the identical distribution condition across i in Assumption A.1,

in particular allowing for each equation error to have different persistence and variance. Assumption D.2 states that the factor series and the regressor idiosyncratic terms are multivariate integrated nonsingular linear processes of orders ρ and ϑ_i , respectively, where the I(0)innovations of f_t are not collinear. We assume that all components of these vectors are of the same integration order to simplify conditions and presentation, though some heterogeneity could be allowed at the cost of making notation much more complex.

Assumption D.3 is a standard condition and does not restrict covariates to be exogenous, because as long as $\Gamma_i \neq 0$ and $\gamma_i \neq 0$, endogeneity will be present. Furthermore, this could be relaxed by assuming $E(X \otimes \varepsilon) = 0$ and finite higher order moments, but this would require more involved derivations and no further insights.

Assumption D.4 introduces a rank condition that simplifies derivations and requires that $k+1 \ge m$. It is possible that some of our results hold if this condition is dropped, but at the cost of introducing more technical assumptions and derivations, see e.g. Pesaran (2006) and Kapetanios et al. (2011). This condition facilitates the identification of the *m* factors using the k + 1 cross section averages of the observables and still allows for cointegration among idiosyncratic elements of each unit.

Under the given set of assumptions, we perform the estimation in first differences to remove fixed effects. For i = 1, ..., N and t = 1, ..., T, the first-differences model, including only asymptotically stationary variables, is

$$\Delta y_{it} = \beta'_{i0} \Delta X_{it} + \gamma'_i \Delta f_t + \Delta \lambda_t^{-1} (L; \theta_0) \varepsilon_{it}, \qquad (8)$$

$$\Delta X_{it} = \Gamma'_i \Delta f_t + \Delta e_{it}.$$

The estimation we propose for each β_{i0} is in essence a GLS estimation after prewhitening by means of fractional δ^* differencing, where δ^* is a sufficiently large differencing parameter chosen by the econometrician that could be a noninteger (thus extending Bai and Ng (2004)'s method based on first differencing), because if we write

$$\Delta_{t-1}^{\delta^*-1}\Delta y_{it} = \beta_{i0}' \Delta_{t-1}^{\delta^*-1} \Delta X_{it} + \gamma_i' \Delta_{t-1}^{\delta^*-1} \Delta f_t + \Delta_{t-1}^{\delta^*-1} \Delta \lambda_t^{-1} \left(L; \theta_0\right) \varepsilon_{it},$$

the idiosyncratic error term is approximately $\Delta_t^{\delta^*-\delta_{i0}}\psi(L;\xi_0)\varepsilon_{it} \approx I(0)$ when $\delta^* \approx \delta_{i0}$. Adapting Pesaran (2006), we remove the factor structure by projecting the transformed model on the fractionally differenced cross-section averages, possibly using a different δ^* for each equation in order to match the corresponding persistence level. The general intuition is that to control strong persistence, enough differencing is needed in absence of knowledge on the true value of δ_{i0} , e.g. setting $\delta^* = 1$ and working with first differences as in Section 3. This policy requires that all variables in (8) are (asymptotically) stationary and bears the implicit assumption that variables have persistence around the unit root, while allowing δ_{i0} to be smaller, implying a cointegration relationship between the idiosyncratic terms of y_{it} , $\lambda_t^{-1}(L;\theta_0) \varepsilon_{it} \sim I(\delta_{i0})$, and of X_{it} , $e_{it} \sim I(\vartheta_i)$, when $\vartheta_i > \delta_{i0}$. In case of the presence of incidental linear trends, it would be possible to work with second differences of data, which would remove exactly them at the cost of introducing slightly modified initial conditions for the fractional differences of observed data.

Denote $\mathbf{y}_i = (y_{i1}, \dots, y_{iT}), \mathbf{X}_i = (X_{i1}, \dots, X_{iT}), \mathbf{F} = (f_1, \dots, f_T), \mathbf{E}_i = (e_{i1}, \dots, e_{iT})$ and $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$. We can write down the model in first differences as

$$\Delta \mathbf{y}_{i} = \beta_{i0}^{\prime} \Delta \mathbf{X}_{i} + \gamma_{i}^{\prime} \Delta \mathbf{F} + \Delta \lambda_{t}^{-1} (L; \theta_{0}) \varepsilon_{i}$$

$$\Delta \mathbf{X}_{i} = \Gamma_{i}^{\prime} \Delta \mathbf{F} + \Delta \mathbf{E}_{i}.$$

Then, the projection matrix can be denoted by

$$\begin{split} \bar{\mathcal{W}}_T &= \bar{\mathcal{W}}_T(\delta^*) &= \mathbf{I}_T - \bar{\mathbf{H}}(\delta^*)(\bar{\mathbf{H}}(\delta^*)'\bar{\mathbf{H}}(\delta^*))^- \bar{\mathbf{H}}(\delta^*)'\\ \bar{\mathbf{H}}(\delta^*) &= \begin{pmatrix} \bar{\mathbf{y}}(\delta^*)\\ \bar{\mathbf{X}}(\delta^*) \end{pmatrix}' \end{split}$$

where $(\cdot)^-$ denotes generalized inverse, $\overline{\mathcal{W}}_T$ is the $T \times T$ projection matrix, and $\overline{\mathbf{H}}(\delta^*)$ is the $T \times (k+1)$ matrix of fractionally differenced cross-section averages with

$$\bar{\mathbf{y}}(\delta^*) := \frac{1}{N} \sum_{j=1}^{N} \mathcal{Y}_j(\delta^*), \quad \mathcal{Y}_j = \mathcal{Y}_j(\delta^*) = \Delta^{\delta^* - 1} \Delta \mathbf{y}_j$$
$$\bar{\mathbf{X}}(\delta^*) := \frac{1}{N} \sum_{j=1}^{N} \mathcal{X}_j(\delta^*), \quad \mathcal{X}_j = \mathcal{X}_j(\delta^*) = \Delta^{\delta^* - 1} \Delta \mathbf{X}_j$$

Denote $\mathcal{F} = \mathcal{F}(\delta^*) = \Delta^{\delta^*-1}(\Delta \mathbf{F})'$ and introduce the infeasible projection matrix on unobserved factors

$$\mathcal{W}_f = I_T - \mathcal{F}(\mathcal{F}'\mathcal{F})^- \mathcal{F}'.$$

Adapting Pesaran (2006), under the rank conditions in Assumptions D.2 and D.4, as $(N, T)_j \rightarrow \infty$, we have that

$$\overline{\mathcal{W}}_T \mathcal{F} \approx \mathcal{W}_f \mathcal{F} = 0.$$

That is, both projections can be used interchangeably for factor removal in the asymptotics as long as the rank condition holds. Along this line, the possibility of including observed factors in the covariates as in Pesaran (2006) should also be noted just by enlarging $\bar{\mathbf{H}}(\delta^*)$ with an appropriately fractionally differenced version of such factors. Introducing such observed factors would not alter any of the results since they would also be entirely removed by projection, and, similarly a constant could be added to project out the contribution of the differences of individual linear trends.

The (preliminary) estimate of β_{i0} for some fixed δ^* is given by

$$\hat{eta}_i(\delta^*) := \left(\mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{X}'_i\right)^{-1} \mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{Y}'_i$$

where the following identification condition is satisfied.

Assumption D.5. $\mathcal{X}_i \overline{\mathcal{W}}_T \mathcal{X}'_i$ and $\mathcal{X}_i \mathcal{W}_f \mathcal{X}'_i$ are full rank for all $i = 1, \ldots, N$.

Note that choosing $\delta^* \geq 1$, so that $\vartheta_i + \delta_{i0} - 2\delta^* < 1$ for all possible values of ϑ_i and δ_{i0} , guarantees that all detrended variables are asymptotically stationary and that sample moments converge to population limits as $(N, T)_j \to \infty$. This, together with the identifying conditions in Assumption D lead to the consistency of $\hat{\beta}_i(\delta^*)$, as we show in the next theorem. This does not require further restrictions on the rate on which both N and T diverge, just that δ^* is not smaller than one. This approach is similar to the choice of working with first differences in Bai and Ng (2004) when trying to estimate the common factors from I(1) nonstationary data by principal components although using δ^* provides greater flexibility extending Bai and Ng (2004)'s method based on first differencing.

Theorem 4. Under Assumption D, $\delta^* \ge 1$, as $(N, T)_j \to \infty$,

$$\hat{\beta}_i(\delta^*) \to_p \beta_{i0}.$$

We next analyze the asymptotic distribution of $\hat{\beta}_i(\delta^*)$ when δ^* is large enough so that aggregate memory of the idiosyncratic regression error term and regressor component is as small as desired. Define for $\delta^* \geq 1$,

$$\Sigma_{ie}(j) = \sum_{k=0}^{\infty} \Phi_{ik}^{e} \left(\delta^{*} - \vartheta_{i}\right) \mathbf{\Omega}_{ie} \Phi_{ij+k}^{e} \left(\delta^{*} - \vartheta_{i}\right)', \quad j = 0, 1, \dots,$$

 $\Sigma_{ie}(j) = \Sigma_{ie}(-j)', j < 0$, where the weights $\Phi_{ik}^e(\delta^* - \vartheta_i) = \sum_{j=0}^k \Phi_{ik-j}^e \pi_j(\delta^* - \vartheta_i)$ incorporate the prewhitening effect, and for $\vartheta_i + \delta_{i0} - 2\delta^* < 1/2$ (which can be guaranteed by taking $\delta^* > 5/4$), define

$$\mathbf{\Sigma}_{i0} = \sum_{j=-\infty}^{\infty} \mathbf{\Sigma}_{ie} \left(j
ight) \zeta_{i0} \left(j
ight),$$

where $\zeta_{i0}(j) = \sum_{k=0}^{\infty} \lambda_k^{-1} (\delta_{i0} - \delta^*, \xi_{i0}) \lambda_{k+|j|}^{-1} (\delta_{i0} - \delta^*, \xi_{i0}), \quad j = 0, \pm 1, \dots$

Setting $\delta^* = 1$ could be enough to obtain asymptotically normal estimates of β_{i0} if we further restrict the aggregate memory as in the next condition. Set

$$\vartheta_{max} = \max_{i} \vartheta_i, \ \delta_{max} = \max_{i} \delta_{i0}$$

Assumption E. $\delta^* > 5/4$, or $\delta^* \ge 1$ and $\vartheta_{max} + \delta_{max} - 2\delta^* < 1/2$, $\max\{\delta_{max}, \vartheta_{max}\} < 11/8$ and $\max\{\varrho + \delta_{max}, \varrho + \vartheta_{max}\} < 11/4$.

This condition could be dispensed with if we allow N to grow faster than T in the asymptotics, while the condition $T/N^2 \rightarrow 0$ as used by Pesaran (2006) for weakly dependent series is also needed in our analysis. There is no requirement on the distribution of values of δ_i across individuals.

Let

$$\boldsymbol{\Upsilon}_{\beta_{i}} = \sigma_{i}^{2} \boldsymbol{\Sigma}_{ie}^{-1} \left(0 \right) \boldsymbol{\Sigma}_{i0} \boldsymbol{\Sigma}_{ie}^{-1} \left(0 \right)$$

Theorem 5. Under Assumptions D and E, and if $T/N^2 \to 0$ as $(N,T)_j \to \infty$, then

$$\sqrt{T}\left(\hat{\beta}_i(\delta^*) - \beta_{i0}\right) \to_d \mathcal{N}(0, \Upsilon_{\beta_i}).$$

Note that when $\delta^* = \delta_{i0}$ and $\psi(L;\xi) = 1$, $\Upsilon_{\beta_i} = \sigma_i^2 \Sigma_{ie}^{-1}(0)$, so the theorem shows in this case the estimate $\hat{\beta}_i(\delta^*)$ is effectively an efficient GLS estimate and the asymptotic variance of $\hat{\beta}_i(\delta^*)$ simplifies in the usual way, not depending on the dynamics of the error term. The rate of convergence is \sqrt{T} for the range of allowed memory parameters (or if δ^* is large enough as described in Assumption 5), irrespective of possible cointegration among idiosyncratic terms, as the GLS estimate is designed in terms of approximately independent regressor and error time series after factor removal. Consistent estimates of the asymptotic variance of $\hat{\beta}_i(\delta^*)$ could be designed adapting the methods of Robinson and Hidalgo (1997) and Robinson (2005) in terms of projected observations to eliminate factors and an estimate of δ_{i0} or the residual series.

4.2. Estimation of Dynamic Parameters

We now turn to individual long and short memory parameter estimation. In the treatment of the basic model, we proved consistency of the parameter estimates for the heterogeneous case in subsection 3.2. Similarly, denote

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

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

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

where

$$\varepsilon_i(\theta) = \lambda \left(L; \delta_i - \delta^*, \xi \right) \left(\tilde{\mathbf{y}}_i(\delta^*) - \hat{\beta}_i(\delta^*)' \tilde{\mathbf{X}}_i(\delta^*) \right)$$

and the vectors of observations $\tilde{\mathbf{y}}_i = \mathcal{Y}_i \overline{\mathcal{W}}_T$ and $\tilde{\mathbf{X}}_i = \mathcal{X}_i \overline{\mathcal{W}}_T$ and the least squares coefficients $\hat{\beta}_i(\delta^*)$ are obtained after projection of \mathcal{Y}_i and \mathcal{X}_i on both $\bar{\mathbf{y}}(\delta^*)$ and $\bar{\mathbf{X}}(\delta^*)$ for a given δ^* . The next assumption requires that $\underline{\delta}$ is not very small compared to the other memory parameters, implying that they can not be very different if we require that δ_{i0} belong to the set \mathcal{D} so that they are also bounded from above.

Assumption F. max $\{\delta_{\max}, \vartheta_{\max}, \varrho\} - \underline{\delta} < 1/2$ and $\max\{\delta_{\max}, \vartheta_{\max}\} < 5/4$.

Note that when $\delta_{i0} \in \mathcal{D}$ the conditions in Assumption F also imply $\vartheta_i - \delta_{i0} < 1/2$ because $\vartheta_i \leq \vartheta_{\max}$ and $\underline{\delta} \leq \delta_{i0}$, and also imply $\varrho - \delta_{i0} < 1/2$. The next theorem presents the consistency and asymptotic normality of the dynamics parameter estimates.

Theorem 6. Under the assumptions of Theorem 5 and Assumption F, $\theta_{i0} \in Int(\Theta)$ as $(N,T)_j \to \infty$,

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

Here Assumption F basically implies the sufficient conditions for Assumption B in terms of the lower bound $\underline{\delta}$, while taking $\delta^* \geq 1$ mirrors the approach of working with first differenced data as in Theorem 1. Note that Theorem 5 guarantees the \sqrt{T} consistency of $\hat{\beta}_i(\delta^*)$, which might be stronger than needed for the consistency of $\hat{\theta}_i$, but simplifies the proof. The asymptotic distribution of the dynamic parameter estimate is normal analogously to the result in Corollary 2, without the burden of the initial condition bias of Theorem 2 since the rate of consistency for each $\hat{\theta}_i$ is just \sqrt{T} . We finally show the efficiency of the feasible GLS slope estimate $\tilde{\beta}_i(\hat{\theta}_i)$ obtained by plugging in an estimate of the vector θ_{i0} , where $\hat{\theta}_i$ is \sqrt{T} consistent for θ_{i0} , with δ^* and δ_{i0} satisfying the restrictions in Assumption E. Note that this requires $\delta_{i0} \geq 1$ in a general set up where factors and the idiosyncratic component of regressors can have orders of integration arbitrarily close to 3/2. For that, define the following generalized prewhitened series,

$$\widehat{\mathcal{Y}}_{j} = \widehat{\mathcal{Y}}_{j}(\widehat{\theta}_{i}) = \lambda_{t-1} \left(L; \widehat{\theta}_{i}^{(-1)} \right) \Delta \mathbf{y}_{j}
\widehat{\mathcal{X}}_{j} = \widehat{\mathcal{X}}_{j}(\widehat{\theta}_{i}) = \lambda_{t-1} \left(L; \widehat{\theta}_{i}^{(-1)} \right) \Delta \mathbf{X}_{j}$$

for j = 1, ..., N, and their cross-section averages, $\hat{\mathbf{y}}(\hat{\theta}_i)$ and $\hat{\mathbf{X}}(\hat{\theta}_i)$, and the corresponding projection matrix $\widehat{\mathcal{W}}_T$ based on $\hat{\mathbf{H}}(\hat{\theta}_i) = \left(\hat{\mathbf{y}}(\hat{\theta}_i)' \ \hat{\mathbf{X}}(\hat{\theta}_i)' \right)$. Then the GLS estimate is

$$\widetilde{\beta}_i(\widehat{\theta}_i) := \left(\widehat{\mathcal{X}}_i \widehat{\mathcal{W}}_T \widehat{\mathcal{X}}'_i\right)^{-1} \widehat{\mathcal{X}}_i \widehat{\mathcal{W}}_T \widehat{\mathcal{Y}}'_i,$$

where the matrix $\widehat{\mathcal{X}}_i \widehat{\mathcal{W}}_T \widehat{\mathcal{X}}'_i$ is assumed full rank.

Let

$$ar{\Sigma}_{ie} = \sum_{k=0}^\infty ar{\Phi}^e_{ik} \mathbf{\Omega}_{ie} ar{\Phi}^{e\prime}_{ik}$$

be the asymptotic variance matrix of the idiosyncratic component of the prewhitened regressors $\widehat{\mathcal{X}}_{i}^{0} = \widehat{\mathcal{X}}_{i}(\theta_{i0})$ where the weights $\overline{\Phi}_{ik}^{e} = \sum_{j=0}^{k} \Phi_{ik-j}^{e} \lambda_{j} (\delta_{i0} - \vartheta_{i}, \xi_{i0})$ incorporate the prewhitening effect.

Theorem 7. Under the assumptions of Theorem 5 with $\delta^* = \delta_{i0}$ and $\hat{\theta}_i - \theta_{i0} = O_p(T^{-1/2})$,

$$\sqrt{T}\left(\tilde{\beta}_i(\hat{\theta}_i) - \beta_i\right) \to_d \mathcal{N}(0, \sigma_i^2 \bar{\Sigma}_{ie}^{-1}).$$

Consistent estimation of σ_i^2 can be conducted directly from the sample variance of residuals $\varepsilon_i(\hat{\theta}_i)$, while estimation of $\bar{\Sigma}_{ie}$ would require the sample second moment matrix of the projected and prewhitened series regressors, i.e. $\hat{\mathcal{X}}_i \widehat{\mathcal{W}}_T \hat{\mathcal{X}}'_i$. Further iterations to estimate θ can also be envisaged using the efficient $\tilde{\beta}_i(\hat{\theta}_i)$ instead of the preliminary $\hat{\beta}_i(\delta^*)$.

4.3. Estimation of Mean Effects

Given the panel data structure, in many cases there is an interest in estimating the average effect across all cross section units. The simplest estimate capturing average effects is the common correlation mean group estimate that averages all individual coefficients, possibly with a common δ^* ,

$$\hat{\beta}_{CCMG}(\delta^*) = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i(\delta^*).$$

Other possibilities such as the common correlation pooled estimate,

$$\hat{\beta}_{CCP}(\delta^*) := \left(\sum_{i=1}^N \mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{X}'_i\right)^{-1} \sum_{i=1}^N \mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{Y}'_i,$$

can be more in the spirit of the joint estimation of the memory parameter presented in Section 2. For the asymptotic analysis of the mean group estimate we consider a simple linear random coefficients model

$$\beta_{i0} = \beta_0 + w_i, \quad w_i \sim iid(0, \Omega_w),$$

where w_i is independent of all the other variables in the model. The asymptotic analysis of the pooled estimate requires further regularity conditions so it is left for future research.

Theorem 8. Under Assumptions D and E, and $(T^{-1}\mathcal{X}_i\bar{\mathcal{W}}_T\mathcal{X}'_i)^{-1}$ having finite second order moments for all $i=1, \ldots, N$, as $(N,T)_j \to \infty$,

$$\sqrt{N}\left(\hat{\beta}_{CCMG}(\delta^*) - \beta_0\right) \to_d \mathcal{N}(0, \Omega_w).$$

This theorem extends previous results in Pesaran (2006) and Kapetanios et al. (2011) for I(0) and I(1) variables under similar conditions to D.5 based on original data, where now the rate of convergence is \sqrt{N} , and no restrictions are required in the rate of growth of N and T. Consistent estimates of the asymptotic variance can be proposed as in Pesaran (2006), since, asymptotically, variability only depends on the heterogeneity of the β_{i0} ,

$$\hat{\Omega}_w = \frac{1}{N} \sum_{i=1}^N \left(\hat{\beta}_i(\delta^*) - \hat{\beta}_{CCMG}(\delta^*) \right) \left(\hat{\beta}_i(\delta^*) - \hat{\beta}_{CCMG}(\delta^*) \right)'.$$

Similarly, the average effect can be estimated based on $\tilde{\beta}_i(\hat{\theta}_i)$ as

$$\tilde{\beta}_{CCMG}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^{N} \tilde{\beta}_i(\hat{\theta}_i), \quad \hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \dots, \hat{\theta}_N\right),$$

which is also asymptotically normally distributed and the asymptotic variance-covariance matrix can be estimated by

$$\tilde{\Omega}_w = \frac{1}{N} \sum_{i=1}^N \left(\tilde{\beta}_i(\hat{\theta}_i) - \tilde{\beta}_{CCMG}(\hat{\boldsymbol{\theta}}) \right) \left(\tilde{\beta}_i(\hat{\theta}_i) - \tilde{\beta}_{CCMG}(\hat{\boldsymbol{\theta}}) \right)'.$$

5. Monte Carlo Simulations

In this section we carry out a Monte Carlo experiment to study the small-sample performance of the slope and memory estimates in the simplest case where there are not short memory dynamics, $\xi = 0$, 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 based on the *iid* shocks 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 . Simulations are based on 1,000 replications.

5.1 Simulations for the Basic Model

In this first subsection we investigate the finite-sample properties of our estimate of δ_0 under the basic setup without covariates. 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 $\rho =$ 0.4, 1. As is clear in Table 1, when the factors are less persistent ($\rho = 0.4$), the estimate is heavily biased for the stationary case of $\delta_0 = 0.3$ while it 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 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 $\rho = 0.4$ although benefits are limited for $\rho = 1$. 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 $\rho = 0.4, 1$, 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 Simulations for the General Model

Based on the general model, we conduct a finite-sample study to check the accuracy of both slope and fractional parameter estimates. We draw the shocks and factor loadings and generate the common factor the same way we followed under the basic setup, while the idiosyncratic component of covariates follows a pure fractional process of memory ϑ . We investigate the performance for (N,T) = (10,50) and (N,T) = (20,100) for the parameter values $\delta_0 = 0.5, 0.75, 1; \ \vartheta = 0.75, 1, 1.25$, and $\varrho = 0.4, 1$, covering both cointegration (e.g. $\vartheta = 1.25$ and $\delta = 1$) and non-cointegration cases (e.g. $\vartheta = 1$ and $\delta = 1$). For projection of estimated factors based on prewhitened cross section averages, we take $\delta^* = 1$.

Tables 3 and 4 present biases and RMSE's for both slope and fractional parameter estimates for (N,T) = (10,50), (20,100), respectively. Biases of both common correlation pooled (CCP) and mean group (CCMG) estimates are very reasonable with biases of pooled estimates generally dominating those of MG estimates, particularly when $\rho = 1$. Biases of slope estimates become negative with their magnitudes increasing with NT for the two smallest values of ϑ . The pooled estimate of the fractional parameter suffers from large biases when δ_0 is small relative to ϑ or ρ due to the idiosyncratic shocks in the regression equation being dominated by other sources of persistence. As expected, biases in fractional parameter estimates decrease with δ_0 in all cases.

In terms of performance, slope estimates behave quite well both in cointegration and noncointegration cases implying that cointegration is not necessary for the estimation of slope in practice. However, for several cases standard errors of fractional parameter estimates are rather large, which can be explained by persistence distortions from the common factor and covariate shocks. Nevertheless, performances of both slope and fractional parameter estimates are clearly improving with δ_0 when $\vartheta = 0.75$, 1 and in all cases with NT. Efficiency gains of GLS type of estimates using $\hat{\delta}$ are very small, if any, for the MG estimate for all values of δ_0 , but for $\delta_0 < \delta^* = 1$ the CCP estimate behaviour can deteriorate substantially, so overdifferencing in the prewhitening step seems a safe recommendation in practice.

6. Fractional Panel Analysis of Realized Volatilities

The capital asset pricing model (CAPM) and its variations have long been used in finance to determine a theoretically appropriate required rate of return in a diversified portfolio, where estimating *beta* is essential as it measures the sensitivity of expected excess stock returns to expected excess market returns. While CAPM and other such models prove useful in an I(0) environment, they fail to provide valid inference for variables that exhibit fractional long-range dependence such as volatility.

In this application, we assess the sensitivity of industry realized volatilities to a market realized volatility measure. In particular, we estimate the betas for volatility under our general setup, which permits possible cointegrating relationships. Such relationships may have direct policy and investment implications since they enable to see which industries are susceptible to a potential market risk upheaval. Bearing in mind an economy as a portfolio of industries, we use our general model to get an idea about the systematic risk in an economy.

In order to calculate monthly realized volatility measures, we use daily average-valueweighted returns data spanning the time period 2000-2011 (T=144 months) from Kenneth French's Data Library for 30 industries in the U.S. economy. As for the composite market returns, we use a weighted average of daily returns of NYSE, NASDAQ and AMEX since the companies considered in industry returns trade in one of these markets. Using the composite index returns of NYSE, NASDAQ and AMEX, i.e. $r_{m,t}$, we calculate

$$RVM_t = \left(\sum_{s \in t}^{N_t} r_{m,s}^2\right)^{1/2}, \quad t = 1, 2, \dots, T,$$

where N_t is the number of trading (typically 22) days in a month. Next, for each industry,

we calculate

$$RVI_{i,t} = \left(\sum_{s \in t}^{N_t} e_{i,s}^2\right)^{1/2}, \quad t = 1, 2, \dots, T_s$$

where $e_{i,s} = r_{i,s} - r_{m,s}$, cf. Chauvet et al. (2012). Along this line, while jump-robust measures such as bipower variation could also be used, our main focus is to show that our general model is suited to address the empirical problem described herein.

Figure 1 shows the behaviour of monthly industry realized volatilities and justifies a heterogeneous approach. Figure 2 shows the realized volatility in the composite average of NYSE, NASDAQ and AMEX, where especially closer to the spike there is a trending behaviour also shared by some of the industries as seen in Figure 1.

Observing that the volatility of volatility is time-varying, we scale each industry as well as the market realized volatility by their corresponding standard deviations. Then we estimate

$$RVI_{i,t} = \alpha_i + \beta_{i0}^0 RVM_t + \beta_{i0}X_{i,t} + \gamma_i'f_t + \Delta_{t+1}^{-\delta_i}v_{i,t},$$

where RVM_t , the $I(\vartheta)$ market realized volatility, is the observable common factor that is treated as a covariate; each $X_{i,t}$ is the average effect of I(0) industry-specific factors: bookto-market ratio and market capitalization, which are also covariates; f_t are $I(\varrho)$ unobservable common factors that are projected out as described in earlier sections so that possible cointegrating relationships can be disclosed between $RVI_{i,t}$ and RVM_t .

We obtain fractional integration degrees of market and industry realized volatilities resorting to local Whittle estimation, Robinson (1995), with bandwidth choices of $m = T^{0.6}, T^{0.7}$ corresponding to m = 20, 32, respectively, and refrain from adding more Fourier frequencies to avoid higher-frequency contamination. Table 5 collectively presents the local Whittle estimates of fractional integration values of the 30 U.S. industry realized volatilities as well as those of the composite market. For both bandwidth choices, the industry realized volatilities display heterogeneity lying above the nonstationarity bound. The market realized volatility is also nonstationary being integrated of an order around 0.6. The unobserved common factor has integration orders of $\rho = 0.71, 0.66$ for m = 20, 32, respectively, which we estimate based on the cross-section averages of the industry realized volatilities.

We use our general model to jointly estimate the fractional order of residuals (δ_i) and slope coefficients $(\beta_{i0}^0 \text{ and } \beta_{i0})$ based on the projections of first-differenced data $(\delta^* = 1)$ in order to be able to confirm and identify cointegrating relationships. Fama-French factors are known to be I(0) in finance, rendering cointegration possible only between the market and industry realized volatilities. Table 6 presents the fractional order of residuals, from which the cointegrating relationships are confirmed based on the results presented in Table 5.

The main criterion for cointegration in this setup is $\delta_i < \vartheta_i$ since the equality of realized volatility integration orders between industries and the market cannot be rejected in all but very few cases. Based on these two requirements together, cointegrating relationships are confirmed between the market realized volatility and the realized volatilities of all industries but Financial Services, Business Equipment and Telecommunications for m = 20. With the bandwidth of m = 32, more pronounced cointegrating relationships with the market realized volatility are indicated for the realized volatilities of all industries except Financial Services. Estimates of the cointegrating parameters and their robust standard errors calculated from Theorem 5 asymptotic covariance are reported in Table 7, from which it is obvious that the market realized volatility has a positive and significant effect on all industry realized volatilities with heterogeneous magnitudes while the average effect of industry characteristics (captured by Fama-French factors) display differences in behaviour across industries. Although for several industries slope parameters are estimated under non-cointegrating relationships, the finite-sample study in the previous section indicates that these estimates are still reliable.

This empirical study reveals that our general model can be used to assess the relationship between market and industry realized volatilities. In fact, other types of such nonstationarity assessment can be performed using our general model. Further studies may focus on estimating cointegrating vectors in-between industries to exactly identify the industries that could be safe to invest in during crises periods as well as to be able to foresee a potential crisis through the real sector.

7. 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. 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; (c) The independence assumption between the idiosyncratic shocks in the general model can be relaxed to allow for nonfactor endogeneity thereby leading to a cointegrated system analysis in the classical sense as in Ergemen (2015) who considers a less flexible modelization due to the lack of allowance of multiple covariates; (d) Panel unit-root testing can be readily performed using our methodology, but it could also be interesting to develop tests that can detect breaks in the general model parameters.

8. 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 $(N, T)_j \to \infty$,
 $T^{2(\varrho - \underline{\delta}) - 1} N^{-2} \to 0$

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

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

8.1 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

$$\eta_{i} = \frac{\sum_{t=1}^{T} \Delta \bar{y}_{t} \Delta \lambda_{t}^{-1} \left(L; \theta_{0}\right) \left(\varepsilon_{it} - \frac{\gamma_{i}}{\bar{\gamma}} \bar{\varepsilon}_{t}\right)}{\sum_{t=1}^{T} \left(\Delta \bar{y}_{t}\right)^{2}}$$

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

$$L_{N,T}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\lambda_t^0(L;\theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) - \tau_t(\theta) (\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) - \eta_i \bar{\gamma} \lambda_{t-1}(L;\theta) f_t \right)^2$$
(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 $[\underline{\delta}, \overline{\delta}] \subseteq (0, 3/2)$, 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 ,

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

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 $\delta_0 - \delta \ge 1/2$. 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 $\zeta < \delta_0 - 1/2 - \delta$, let us define the disjoint sets $\Theta_1 = \{\theta : \delta \le \delta \le \delta_0 - 1/2 - \zeta\}$, $\Theta_2 = \{\theta : \delta_0 - 1/2 - \zeta < \delta < \delta_0 - 1/2\}$, $\Theta_3 = \{\theta : \delta_0 - 1/2 \le \delta \le \delta_0 - 1/2 + \zeta\}$ and $\Theta_4 = \{\theta : \delta_0 - 1/2 + \zeta < \delta \le \delta\}$, so $\Theta = \cup_{k=1}^4 \Theta_k$. Then we will show

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

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

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\lambda_{t}^{0}\left(L;\theta\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)\right)^{2}+\tau_{t}^{2}\left(\theta\right)\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)^{2}+\eta_{i}^{2}\bar{\gamma}^{2}\left(\lambda_{t-1}\left(L;\theta\right)f_{t}\right)^{2}\right.\\\left.-\eta_{i}\bar{\gamma}\left(\lambda_{t-1}\left(L;\theta\right)f_{t}\right)\lambda_{t}^{0}\left(L;\theta\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)+\eta_{i}\bar{\gamma}\left(\lambda_{t-1}\left(L;\theta\right)f_{t}\right)*\tau_{t}\left(\theta\right)\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)\right.\\\left.-\lambda_{t}^{0}\left(L;\theta\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)*\tau_{t}\left(\theta\right)\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)\right\}.$$

The projection error component in the conditional sum of squares,

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

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 $\bar{\gamma}^2\rightarrow_p E[\gamma_i]^2$, Lemmas 1 and 2(a) and Assumption B*.2 since $\varrho-\underline{\delta}<1$, $2\varrho+\delta_0-\underline{\delta}<7/2$ and $\varrho+2\delta_0-\underline{\delta}<7/2$, are implied by Assumption B*.1.

Similarly,

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

because

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

uniformly in $\theta \in \Theta$ with $\underline{\delta} > 0$, using $\frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i0}^2 + \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_i^2 = O_p(1)$, $\bar{\varepsilon}_0 = O_p(N^{-1/2})$ 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)

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

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 $\theta \in \Theta_1$ using Cauchy-Schwarz inequality and that this part of the conditional sum of squares 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 $\Theta_k, k = 1, \ldots, 4$ as in Hualde and Robinson (2011).

Proof for k = 4. We show that

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

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),

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

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.

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

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

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

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

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

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

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

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\phi}_{i}^{2} \left(\lambda_{t}^{0} \left(L; \theta \right) \bar{\varepsilon}_{t} \right)^{2} = \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_{i}^{2} \frac{1}{T} \sum_{t=1}^{T} \left(\lambda_{t}^{0} \left(L; \theta \right) \bar{\varepsilon}_{t} \right)^{2}.$$
(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 $\theta^*_{(1)} = \zeta - \frac{1}{2}$,

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

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, ..., 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

$$L_{N,T}^{*}(\theta) := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \left(\lambda_{t}^{0}\left(L;\theta\right) \left(\varepsilon_{it} - \phi_{i}\bar{\varepsilon}_{t}\right) \right)^{2} \ge \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^{2}} \left(\sum_{t=1}^{T} \lambda_{t}^{0}\left(L;\theta\right) \left(\varepsilon_{it} - \phi_{i}\bar{\varepsilon}_{t}\right) \right)^{2},$$

we write

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

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

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

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

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

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, $i = 0, \ldots, N$, 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 $B_0(s)$. Then, as $(N,T)_j \to \infty$, following the discussions for doubleindex processes in Phillips and Moon (1999) and $\frac{1}{N}\sum_{i=1}^N \phi_i^2 = O_p(1)$,

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

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

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

so that

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

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

8.2 Other Proofs in Section 3

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

Assumption C^{*}.

 $\mathbf{C^*.1.} \text{ As } (N,T)_j \to \infty,$

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

 $\mathbf{C^*.2.} \text{ As } (N,T)_j \to \infty,$

$$N\left(T^{4(\varrho+\delta_0)-11}\log^2 T + T^{8\delta_0-11}\right)\log^2 T \to 0$$
$$N\left(T^{2(\varrho-2\delta_0)-1} + T^{\varrho-2\delta_0-1}\right)\log^2 T \to 0$$

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

$$N^{-1}T^{2(\varrho-2\delta_0)}\log^2 T \to 0.$$

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

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

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,

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

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

$$+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\bar{\gamma}\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}*\dot{\tau}_{t}(\theta_{0})(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0})$$
(19)

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\bar{\gamma}\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}*\chi_{t}\left(L;\xi_{0}\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)$$
(20)

$$+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\bar{\gamma}\chi_{t-1}(L;\xi_{0})\lambda_{t-1}(L;\theta_{0})f_{t}*\tau_{t}(\theta_{0})(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0})(21)$$

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{t}(\theta_{0})(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0})*\chi_{t}\left(L;\xi_{0}\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)$$
(22)

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\bar{\gamma}\chi_{t-1}\left(L;\xi_{0}\right)\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}*\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)$$
(23)

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tau}_{t}(\theta_{0})(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0})(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t})$$
(24)

$$+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t})*\chi_{t}\left(L;\xi_{0}\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right).$$
(25)

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

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

which is $o_{p}(1)$ under Assumption C^{*}.

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

$$\frac{2}{NT} \sum_{i=1}^{N} Var\left[\varepsilon_{i0}^{2}\right] \left(\sum_{t=1}^{T} \tau_{t}(\theta_{0})\dot{\tau}_{t}(\theta_{0})\right)^{2} = O\left(T^{-1} + T^{1-4\delta_{0}}\log^{4}T\right) = o\left(1\right)$$
since $\delta_0 > 1/4$, as $(N, T)_j \to \infty$, while

$$\frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tau_t(\theta_0) \dot{\tau}_t(\theta_0) \hat{\phi}_i^2 \bar{\varepsilon}_0^2 = \frac{2}{\sqrt{NT}} N \bar{\varepsilon}_0^2 \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_i^2 \sum_{t=1}^{T} \tau_t(\theta_0) \dot{\tau}_t(\theta_0) \\ = O_p \left((TN)^{-1/2} \left(1 + T^{1-2\delta_0} \log^2 T \right) \right) = O_p \left(1 \right)$$

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

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

if $\delta_0 > 1/4$.

We show that (19) is $o_{p}(1)$ considering the contribution of

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

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

$$2\sqrt{NT} \left(\frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i0}^{2} \frac{1}{N} \sum_{i=1}^{N} \eta_{i}^{2} \right)^{1/2} \left| \frac{1}{T} \sum_{t=1i}^{T} \lambda_{t-1} \left(L; \theta_{0} \right) f_{t} \dot{\tau}_{t}(\theta_{0}) \right|$$

$$= O_{p} \left((NT)^{1/2} \left(T^{2(\varrho+\delta_{0}-3)} + T^{-1} \log T + N^{-1} T^{4\delta_{0}-6} + N^{-2} \right)^{1/2} T^{-1} \right)$$

$$+ O_{p} \left((NT)^{1/2} \left(T^{2(\varrho+\delta_{0}-3)} + T^{-1} \log T + N^{-1} T^{4\delta_{0}-6} + N^{-2} \right)^{1/2} \left\{ T^{\varrho-2\delta_{0}-1/2} + T^{-\delta_{0}/2-1/2} \right\} \log T \right)$$

$$= O_{p} \left(N^{1/2} \left(T^{2(\varrho+\delta_{0}-3)} + T^{-1} \log T + N^{-2} \right)^{1/2} T^{\varrho-2\delta_{0}} \log T \right)$$

$$+ O_{p} \left(N^{1/2} T^{\varrho+\delta_{0}-3} T^{-\delta_{0}/2} \log T \right) + o_{p} \left(1 \right)$$

which is $o_p(1)$ by Assumptions C*.1-2.

For (20), we consider the contribution of

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

whose absolute value is bounded by

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

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

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

$$E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}*\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\right)^{2}\right]$$

$$=\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}*\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}*\lambda_{r-1}\left(L;\theta_{0}\right)f_{r}*\chi_{r}\left(L;\xi_{0}\right)\varepsilon_{ir}\right]$$

$$=\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}*\lambda_{r-1}\left(L;\theta_{0}\right)f_{r}\right]E\left[\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}*\chi_{r}\left(L;\xi_{0}\right)\varepsilon_{ir}\right]$$

$$=O\left(\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{r=1}^{t}|t-r|^{2(\varrho-\delta_{0})-2}\log t\right)$$

$$=O\left(T^{-1}+T^{2(\varrho-\delta_{0}-1)}\log T\right)=O\left(T^{-1}\right).$$

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 $\varepsilon_{i0}\varepsilon_{it}$,

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tau}_{t}(\theta_{0})(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0})(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}),$$

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

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

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

$$\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\eta_{i}\bar{\gamma}\chi_{t-1}\left(L;\xi_{0}\right)\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}-\dot{\tau}_{t}(\theta_{0})\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)+\chi_{t}\left(L;\xi_{0}\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)\right\} \times \left\{-\eta_{i}\bar{\gamma}\chi_{t-1}\left(L;\xi_{0}\right)\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}-\dot{\tau}_{t}(\theta_{0})\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)+\chi_{t}\left(L;\xi_{0}\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)\right\}' + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\eta_{i}\bar{\gamma}\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}-\tau_{t}(\theta_{0})\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)+\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right\} \times \left\{-\eta_{i}\bar{\gamma}b_{t}^{0}\left(L\right)\lambda_{t-1}\left(L;\theta_{0}\right)f_{t}-\ddot{\tau}_{t}(\theta_{0})\left(\varepsilon_{i0}-\hat{\phi}_{i}\bar{\varepsilon}_{0}\right)+b_{t}^{0}\left(L\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)\right\},$$

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 Nand 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. \Box

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 $\delta_{i0} \in \mathcal{D}$. \Box

8.3 Proofs for Section 6

Proofs of Theorems 4 and 5. For $\delta^* \geq 1$, write $\hat{\beta}_i(\delta^*) - \beta_{i0} = M_i + U_i$, where

$$M_{i} = (\mathcal{X}_{i} \bar{\mathcal{W}}_{T} \mathcal{X}_{i}')^{-1} \mathcal{X}_{i} \bar{\mathcal{W}}_{T} \mathcal{F}' \gamma_{i}$$

$$U_{i} = (\mathcal{X}_{i} \bar{\mathcal{W}}_{T} \mathcal{X}_{i}')^{-1} \mathcal{X}_{i} \bar{\mathcal{W}}_{T} (\Delta^{\delta^{*}-1} \Delta \lambda^{-1} (L; \theta_{i0}) \varepsilon_{i}))'$$

so that M_i is the projection component, and U_i is the usual regression-error component also carrying an initial condition term because

$$\Delta_{t-1}^{\delta^{*}-1}(\Delta\lambda_{t}^{-1}(L;\theta_{0})\varepsilon_{i}) = \lambda_{t}^{-1}(L;\delta_{i0}-\delta^{*},\xi_{0})\varepsilon_{i} - \pi_{t}(\delta^{*}-1)\varepsilon_{i0}$$

with $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT}).$

The asymptotic inference for $\hat{\beta}_i(\delta^*)$ is derived from $U_{1,i}$,

$$U_{1,i} = \left(\Delta^{\delta^* - \vartheta_i} \mathbf{E}_i \Delta^{\delta^* - \vartheta_i} \mathbf{E}'_i\right)^{-1} \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i \left(\lambda^{-1} \left(L; \delta_{i0} - \delta^*, \xi_0\right) \varepsilon_i - \pi_t (\delta^* - 1) \varepsilon_{i0}\right)'$$

where, noting that $\mathcal{W}_f \mathcal{X}_i = \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i$, we can write $U_i = U_{1,i} + U_{2,i}$ with $U_{2,i}$ being the error from approximating \mathcal{W}_f by $\overline{\mathcal{W}}_T$. We later show that both M_i and $U_{2,i}$, are negligible.

For the consistency proof of Theorem 4, we note that $\delta^* \geq 1$ implies $\vartheta_i + \delta_{i0} - 2\delta^* < 1$ and that under Assumption D,

$$T^{-1}\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i\Delta^{\delta^*-\vartheta_i}\mathbf{E}'_i \to_p \qquad \boldsymbol{\Sigma}_{i\mathbf{e}}(0) > 0$$
$$T^{-1}\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i\left(\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_0\right)\varepsilon_i - \pi_t(\delta^*_i-1)\varepsilon_{i0}\right)' \to_p \qquad 0,$$

as $(N,T)_j \to \infty$, exploiting the independence of \mathbf{E}_i and ε_i .

The asymptotic distributions in Theorem 5, correspond to those of $T^{1/2}U_{1,i}$, using a martingale CLT when $\delta^* = \delta_{i0}$ and $\psi(L, \xi_0) \equiv 1$, and using Theorem 1 in Robinson and Hidalgo (1997) when $\delta^* \neq \delta_{i0}$, whose conditions for the OLS estimate are implied by Assumption D.

We now show that M_i and $U_{2,i}$ are negligible. Write

$$ar{\mathbf{H}}' = \mathcal{F}'ar{\mathbf{C}} + \mathcal{X}_iar{\mathbf{V}}$$

where, $\Pi_T^* = (\pi_1 (\delta^* - 1), \dots, \pi_T (\delta^* - 1)),$

$$\bar{\mathbf{V}} = \begin{pmatrix} \overline{\Delta^{\delta^* \lambda^{-1}} (L; \theta_0) \varepsilon} - \mathbf{\Pi}_T^* \overline{\varepsilon_0} + \overline{\beta' \Delta^{\delta^* - \vartheta_0} \mathbf{e}} \\ \overline{\Delta^{\delta^* - \vartheta_0} \mathbf{e}}. \end{pmatrix}$$

Since

$$\mathcal{X}_i \left(\mathbf{I}_T - \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^- \bar{\mathbf{H}}' \right) \mathcal{F}' \gamma_i = \mathcal{X}_i \mathcal{F}' \gamma_i - \mathcal{X}_i \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^- \bar{\mathbf{H}}' \mathcal{F}' \gamma_i,$$

reasoning as in Pesaran (2006) we need to analyze the terms depending on $\bar{\mathbf{V}}$ in

$$\begin{aligned} \mathcal{X}_{i}\bar{\mathbf{H}} &= \frac{\mathcal{X}_{i}\mathcal{F}'\bar{\mathbf{C}}}{T} + \frac{\mathcal{X}_{i}\bar{\mathbf{V}}}{T}, \\ \bar{\mathbf{H}}'\bar{\mathbf{H}} &= \frac{\bar{\mathbf{C}}'\mathcal{F}\mathcal{F}'\bar{\mathbf{C}}}{T} + \frac{\bar{\mathbf{C}}'\mathcal{F}\bar{\mathbf{V}}}{T} + \frac{\bar{\mathbf{V}}'\mathcal{F}'\bar{\mathbf{C}}}{T} + \frac{\bar{\mathbf{V}}'\bar{\mathbf{V}}}{T}, \\ \bar{\mathbf{H}}'\mathcal{F}' &= \frac{\bar{\mathbf{C}}'\mathcal{F}\mathcal{F}'}{T} + \frac{\bar{\mathbf{V}}'\mathcal{F}'}{T}, \end{aligned}$$

where

$$\frac{\mathcal{F}\mathcal{F}'}{T} \to_p \Sigma_f > 0$$

as $T \to \infty$ with $\Sigma_f = \Sigma_f (\delta^* - \varrho) = \sum_{k=0}^{\infty} \Phi_k^f (\delta^* - \varrho) \, \Omega_f \Phi_k^f (\delta^* - \varrho)'$, where the weights $\Phi_k^f (\delta^* - \varrho)$ are square summable with $\delta^* \ge 1$ and incorporate also the fractional differencing effect, $\Phi_k^f (\delta^* - \varrho) = \sum_{j=0}^k \Phi_{k-j}^f \pi_j (\delta^* - \varrho)$, so that Σ_f is positive definite by Assumption D.2.

To show that all the error terms in the projection are negligible we first consider the case $\delta^* > 5/4$ so that $\vartheta_{max} - \delta^* < 1/4$ and $\varrho - \delta^* < 1/4$.

(a). Write $T^{-1}\bar{\mathbf{V}}'\bar{\mathbf{V}}$ as

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} \bar{v}_{t}' \bar{v}_{t} &= \frac{1}{T} \sum_{t=1}^{T} \left\{ \left(\overline{\Delta_{t}^{\delta^{*}} \lambda_{t}^{-1} \left(L; \theta_{0} \right) \varepsilon_{t}} \right)^{2} + \left(\overline{\pi_{t} (\delta^{*} - 1) \varepsilon_{0}} \right)^{2} + \left(\overline{\beta' \Delta_{t}^{\delta^{*} - \vartheta_{0}} e_{t}} \right)^{2} \\ &+ \left(\overline{\Delta_{t}^{\delta^{*} - \vartheta_{0}} e_{t}} \right)^{2} + 2 \overline{\Delta_{t}^{\delta^{*}} \lambda_{t}^{-1} \left(L; \theta_{0} \right) \varepsilon_{t}} \overline{\pi_{t} (\delta^{*} - 1) \varepsilon_{0}} \\ &+ 2 \overline{\Delta^{\delta^{*}} \lambda_{t}^{-1} \left(L; \theta_{0} \right) \varepsilon_{t}} \overline{\Delta_{t}^{\delta^{*} - \vartheta_{0}} e_{t}} + 2 \overline{\pi_{t} (\delta^{*} - 1) \varepsilon_{0}} \overline{\Delta_{t}^{\delta^{*} - \vartheta_{0}} e_{t}} \right\} \end{aligned}$$

whose expectation is $O(N^{-1})$, and its variance is proportional to $O((TN)^{-1})$. Thus,

$$\frac{1}{T}\sum_{t=1}^{T}\bar{v}_t'\bar{v}_t = O_p\left(\frac{1}{N} + \frac{1}{\sqrt{NT}}\right).$$

(b). The term $T^{-1}\bar{\mathbf{V}}'\mathcal{F}' = T^{-1}\sum_{t=1}^{T} \bar{v}_t f_t = O_p\left((NT)^{-1/2}\right)$ since it has zero expectation and using the independence of ε_{it} and f_t , its variance is

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{v}_{t}f_{t}\right) = \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(\bar{v}_{t}'\bar{v}_{t}\right)E\left(f_{t}'f_{t'}\right)$$

whose norm is $O(N^{-1})$ times

$$O\left(\begin{array}{c}T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left\{|t-t'|_{+}^{2(\max\{\delta_{max}-\delta^{*},\vartheta_{max}-\delta^{*}\})-1}+|t-t'|_{+}^{\max\{\delta_{max}-\delta^{*},\vartheta_{max}-\delta^{*}\}-1}\right\}\\ \times\left\{|t-t'|_{+}^{2(\varrho-\delta^{*})-1}+|t-t'|_{+}^{\varrho-\delta^{*}-1}\right\}\\ = O\left(T^{-1}\right).$$

(c). Lastly, $T^{-1} \sum_{t=1}^{T} \overline{\Delta_t^{\delta^* - \vartheta_0} e_t} \bar{\varepsilon}_t = O_p \left((NT)^{-1/2} \right)$ because it has zero expectation and using the independence of e_{it} and ε_{it} , its variance is proportional to $O(N^{-1})$ times

$$O\left(\begin{array}{c}T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left\{|t-t'|_{+}^{2(\max\{\delta_{max}-\delta^{*},\vartheta_{max}-\delta^{*}\})-1}+|t-t'|_{+}^{\max\{\delta_{max}-\delta^{*},\vartheta_{max}-\delta^{*}\}-1}\right\}\\\times\left\{|t-t'|_{+}^{2(\vartheta_{max}-\delta^{*})-1}+|t-t'|_{+}^{\vartheta_{max}-\delta^{*}-1}\right\}\end{array}\right),$$

which is $O(T^{-1})$.

Thus, for $\delta^* > 5/4$, the projection error is

$$M_i = O_p\left(\frac{1}{N} + \frac{1}{\sqrt{NT}}\right) = o_p\left(1\right)$$

as $(N,T)_j \to \infty$, and $T^{1/2}M_i = O_p\left(T^{1/2}N^{-1} + N^{-1/2}\right) = o_p(1)$ if $T^{1/2}N^{-1} \to 0$ as $(N,T)_j \to \infty$.

Alternatively, if we just take $\delta^* = 1$:

(a). Write

$$\frac{1}{T}\sum_{t=1}^{T}\bar{v}_{t}'\bar{v}_{t} = \frac{1}{T}\sum_{t=1}^{T}\left\{\left(\overline{\Delta\lambda_{t}^{-1}\left(L;\theta_{0}\right)\varepsilon_{t}}\right)^{2} + \left(\overline{\beta'\Delta_{t}^{1-\vartheta_{0}}e_{t}}\right)^{2} + \left(\overline{\Delta_{t}^{1-\vartheta_{0}}e_{t}}\right)^{2} + 2\overline{\Delta\lambda_{t}^{-1}\left(L;\theta_{0}\right)\varepsilon_{t}}\overline{\Delta_{t}^{1-\vartheta_{0}}e_{t}}\right\}$$

whose expectation is $O(N^{-1})$ times

$$O\left(1 + T^{2(\delta_{max}-1)-1} + T^{2(\vartheta_{max}-1)-1} + T^{\delta_{max}-3}\right) = O(1)$$

and its variance is proportional to $O(N^{-2})$ times

$$O\left(T^{-1} + T^{4(\delta_{max}-1)-2} + T^{2(\vartheta_{max}+\delta_{max}-2)-2} + T^{4(\vartheta_{max}-1)-2}\right).$$

Then

$$\frac{1}{T}\sum_{t=1}^{T}\bar{v}_{t}'\bar{v}_{t} = O_{p}\left(\frac{1}{N} + \frac{1}{N}\left\{T^{-1/2} + T^{2\delta_{max}-3} + T^{2\vartheta_{max}-3} + T^{\vartheta_{max}+\delta_{max}-3}\right\}\right) = O_{p}\left(N^{-1}\right).$$

(b). The term $T^{-1}\mathcal{F}\bar{\mathbf{V}} = T^{-1}\sum_{t=1}^{T} \bar{v}_t f_t$ has zero expectation and

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{v}_{t}f_{t}\right) = O\left(N^{-1}T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}|t-t'|_{+}^{2(\max\{\delta_{max}-1,\vartheta_{max}-1\})-1}|t-t'|_{+}^{2(\varrho-1)-1}\right)$$

that $T^{-1}\sum_{t=1}^{T}\bar{v}_{t}f_{t} = O_{p}\left((NT)^{-1/2} + N^{-1/2}\left\{T^{\delta_{max}+\varrho-3} + T^{\vartheta_{max}+\varrho-3}\right\}\right).$

(c). Lastly, $T^{-1} \sum_{t=1}^{T} \overline{\Delta^{1-\vartheta_0} e_t} \bar{v}_t$ has zero expectation and using the independence of e_{it} and ε_{it} , variance is proportional to $O(N^{-1})$ times

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T \sum_{t'=1}^T \left\{ \left| t - t' \right|_+^{2(\max\{\delta_{max} - 1, \vartheta_{max} - 1\}) - 1} + \left| t - t' \right|_+^{\max\{\delta_{max} - 2, \vartheta_{max} - 2\}} \right\} \left\{ \left| t - t' \right|_+^{2(\vartheta_{max} - 1) - 1} + \left| t - t' \right|_+^{\vartheta_{max} - 2} \right\} \\ &= O\left(\frac{1}{T^2} \sum_{t=1}^T \sum_{t'=1}^T \left\{ \left| t - t' \right|_+^{2(\delta_{max} + \vartheta_{max} - 2) - 2} + \left| t - t' \right|_+^{4(\vartheta_{max} - 1) - 2} \right\} + T^{-1} \right) \\ &= O\left(T^{-1} + T^{2(\delta_{max} + \vartheta_{max} - 3)} + T^{4(\vartheta_{max} - 1) - 2} \right) \end{aligned}$$

so that

 \mathbf{SO}

$$\frac{1}{T}\sum_{t=1}^{T}\overline{\Delta^{1-\vartheta_0}e_t}\varepsilon_t = O_p\left(N^{-1/2}\left\{T^{-1/2} + T^{\delta_{max}+\vartheta_{max}-3} + T^{2\vartheta_{max}-3}\right\}\right).$$

Thus the entire projection error is

$$\begin{split} M_{i} &= O_{p} \left(N^{-1} + N^{-1/2} \left\{ T^{-1/2} + T^{\delta_{max} + \vartheta_{max} - 3} + T^{2\vartheta_{max} - 3} + T^{\varrho + \delta_{max} - 3} + T^{\varrho + \vartheta_{max} - 3} \right\} \right) = o_{p} \left(1 \right) \\ \text{as} \ (N, T)_{j} \to \infty, \text{ and} \\ T^{1/2}M_{i} &= O_{p} \left(T^{1/2}N^{-1} + N^{-1/2} \left\{ 1 + T^{\delta_{max} + \vartheta_{max} - 5/2} + T^{2\vartheta_{max} - 5/2} + T^{\varrho + \vartheta_{max} - 5/2} + T^{\varrho + \vartheta_{max} - 5/2} \right\} \right). \\ \text{Therefore, if} \ \vartheta_{max} < 11/8 \ \text{and} \ \varrho + \delta_{max}, \varrho + \vartheta_{max}, \delta_{max} + \vartheta_{max} < 11/4, \ T^{1/2}M_{i} = o_{p} \left(1 \right) \text{ as} \\ (N, T)_{j} \to \infty \ \text{when} \ \delta^{*} = 1 \ \text{since} \ T^{1/2}N^{-1} = o \left(1 \right) \ \text{and} \ N^{1/2} = o \left(T^{-1/4} \right). \end{split}$$

The proof that the approximation term $U_{2,i}$ is negligible is similar and is omitted. \Box

Proof of Theorem 6. We first show the consistency of the parameter estimates. We can rewrite the projected variables entering in the concentrated log-likelihood as

$$\tilde{\mathbf{y}}_{i}\left(\delta^{*}\right) = \Delta^{\delta^{*}-1} \Delta \mathbf{y}_{i} - \hat{\Upsilon}'_{iy} \bar{\mathbf{H}} = \Delta^{\delta^{*}-1} \Delta \mathbf{y}_{i} - \Delta^{\delta^{*}-1} \Delta \mathbf{y}_{i} \bar{\mathbf{H}}' \left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-} \bar{\mathbf{H}}$$

which, after filtering each component of $\tilde{\mathbf{y}}_{i}(\delta^{*})$ by $\lambda_{t-1}(L;\theta) \Delta^{-\delta^{*}} = \lambda_{t-1}(L;\delta-\delta^{*},\xi)$ adapted to the prefiltering by $\Delta^{\delta^{*}}$ implicit in $\bar{\mathbf{H}}$ yields,

$$\lambda\left(L;\delta-\delta^{*},\xi\right)\tilde{\mathbf{y}}_{i}\left(\delta^{*}\right)=\psi\left(L;\xi\right)\Delta^{\delta-1}\Delta\mathbf{y}_{i}-\hat{\Upsilon}_{ij}^{\prime}\bar{\mathbf{H}}(\theta)$$

where $\hat{\Upsilon}_{iy} = \left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-1} \bar{\mathbf{H}}\Delta^{\delta^*-1}\Delta \mathbf{y}'_i$ and $\bar{\mathbf{H}}(\theta) = \lambda \left(L; \delta - \delta^*, \xi\right) \bar{\mathbf{H}}(\delta^*) = \psi \left(L; \xi\right) \Delta^{\delta - \delta^*} \bar{\mathbf{H}}(\delta^*)$, and likewise,

$$\lambda \left(L; \delta - \delta^*, \xi \right) \tilde{\mathbf{X}}_i \left(\delta^* \right) = \psi \left(L; \xi \right) \Delta^{\delta - 1} \Delta \mathbf{X}_i - \hat{\Upsilon}'_{ix} \bar{\mathbf{H}}(\theta).$$

Next, write for the components of the residuals

$$\lambda \left(L; \delta - \delta^*, \xi \right) \tilde{\mathbf{y}}_i \left(\delta^* \right) = P_{y,i} \left(\theta \right) + R_{y,i} \left(\theta \right)$$

where

$$P_{y,i}(\theta) = \lambda (L; \delta - 1, \xi) \Delta \mathbf{y}_i - \Delta^{\delta^* - 1} \Delta \mathbf{y}_i \mathcal{F}' (\mathcal{F}\mathcal{F}')^{-1} \mathbf{F}(\theta)$$

$$R_{y,i}(\theta) = \Delta^{\delta^* - 1} \Delta \mathbf{y}_i \left\{ \mathcal{F}' (\mathcal{F}\mathcal{F}')^{-1} \mathbf{F}(\theta) - \bar{\mathbf{H}}' (\bar{\mathbf{H}}\bar{\mathbf{H}}')^{-} \bar{\mathbf{H}}(\theta) \right\}$$

with $\mathbf{F}(\theta) = \lambda (L; \delta - \delta^*, \xi) \mathcal{F} = \psi (L; \xi) \Delta^{\delta} \mathbf{F}$, and similarly $\lambda (L; \delta - \delta^*, \xi) \tilde{\mathbf{X}}_i (\delta^*) = P_{x,i}(\theta) + R_{x,i}(\theta)$ for $P_{x,i}$ and $R_{x,i}$ defined replacing \mathbf{y}_i by \mathbf{x}_i .

Then, truncating the filters appropriately for each element and $\lambda^{0}(L;\theta) = \lambda(L;\theta) \lambda^{-1}(L;\theta_{i0})$,

$$P_{y,i}(\theta) = \lambda^{0}(L;\theta)\varepsilon_{i} + \beta_{i0}^{\prime}\psi(L;\xi)\Delta^{\delta-\vartheta_{i}}\mathbf{E}_{i} - \varsigma_{T}(\theta)\varepsilon_{i0} - \left[\lambda^{-1}(L;\delta_{i0}-\delta^{*},\xi_{0})\varepsilon_{i} + \beta_{i0}^{\prime}\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i} - \mathbf{\Pi}_{T}^{*}\varepsilon_{i0}\right]\mathcal{F}^{\prime}(\mathcal{FF}^{\prime})^{-}\mathbf{F}(\theta),$$

with $\varsigma_T(\theta) = (\tau_1(\theta), \dots, \tau_T(\theta))$ and

$$P_{x,i}(\theta) = \psi(L;\xi) \,\Delta^{\delta - \vartheta_i} \mathbf{E}_i - \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i \mathcal{F}'(\mathcal{F}\mathcal{F}')^- \mathbf{F}(\theta).$$

Also,

$$R_{y,i}(\theta) = \left[\lambda^{-1} (L; \delta_{i0} - \delta^*) \varepsilon_i + \beta'_{i0} \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i + (\beta'_{i0} \Gamma'_i + \gamma'_i) \mathcal{F} - \mathbf{\Pi}_T^* \varepsilon_{i0}\right] \\ \times \left[\mathcal{F}' (\mathcal{F}\mathcal{F}')^{-1} \mathbf{F}(\theta) - \bar{\mathbf{H}}' (\bar{\mathbf{H}}\bar{\mathbf{H}}')^{-} \bar{\mathbf{H}}(\theta)\right],$$

and $R_{x,i}$ can be written similarly.

Therefore

$$\begin{split} \lambda\left(L;\delta-\delta^{*},\xi\right)\left\{\tilde{\mathbf{y}}_{i}\left(\delta^{*}\right)-\hat{\beta}_{i}\left(\delta^{*}\right)'\tilde{\mathbf{X}}_{i}\left(\delta^{*}\right)\right\}\\ &=P_{y,i}\left(\theta\right)+R_{y,i}\left(\theta\right)-\hat{\beta}_{i}\left(\delta^{*}\right)'\left(P_{x,i}\left(\theta\right)+R_{x,i}\left(\theta\right)\right)\\ &=\lambda^{0}\left(L;\theta\right)\varepsilon_{i}-\varsigma_{T}\left(\theta\right)\varepsilon_{i0}-\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{0}\right)\varepsilon_{i}W_{f}\left(\theta\right)-\mathbf{\Pi}_{T}^{*}\varepsilon_{i0}W_{f}\left(\theta\right)\\ &-\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\left[\psi\left(L;\xi\right)\Delta^{\delta-\vartheta_{i}}\mathbf{E}_{i}-\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}W_{f}\left(\theta\right)\right]\\ &+\left[\left(\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\mathbf{\Gamma}_{i}'+\gamma_{i}'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}+\lambda^{-1}\left(L;\delta_{i0}-\delta^{*}\right)\varepsilon_{i}-\mathbf{\Pi}_{T}^{*}\varepsilon_{i0}\right]\\ &\times\left(W_{f}\left(\theta\right)-W_{h}\left(\theta\right)\right)\end{split}$$

where

$$W_f(\theta) := \mathcal{F}' \left(\mathcal{F} \mathcal{F}' \right)^- \mathbf{F}(\theta)$$
$$W_h(\theta) := \mathbf{\bar{H}}' \left(\mathbf{\bar{H}} \mathbf{\bar{H}}' \right)^- \mathbf{\bar{H}}(\theta),$$

and the residuals $\varepsilon_i(\theta)$ in the CSS $L^*_{i,T}(\theta) = T^{-1}\varepsilon_i(\theta)\varepsilon_i(\theta)'$ can be written as

$$\varepsilon_i(\theta) = \varepsilon_i^{(1)}(\theta) + \varepsilon_i^{(2)}(\theta) + \varepsilon_i^{(3)}(\theta),$$

with

$$\begin{split} \varepsilon_{i}^{(1)}(\theta) &= \lambda^{0} \left(L;\theta\right) \varepsilon_{i} - \varsigma_{T}(\theta) \varepsilon_{i0} - \lambda^{-1} \left(L;\delta_{i0} - \delta^{*},\xi_{0}\right) \varepsilon_{i} W_{f}(\theta) - \mathbf{\Pi}_{T}^{*} \varepsilon_{i0} W_{f}(\theta) \\ \varepsilon_{i}^{(2)}(\theta) &= -\left(\beta_{i0} - \hat{\beta}_{i} \left(\delta^{*}\right)\right)' \left[\psi\left(L;\xi\right) \Delta^{\delta-\vartheta_{i}} \mathbf{E}_{i} - \Delta^{\delta^{*}-\vartheta_{i}} \mathbf{E}_{i} W_{f}\left(\theta\right)\right] \\ \varepsilon_{i}^{(3)}(\theta) &= \left[\left(\left(\beta_{i0} - \hat{\beta}_{i} \left(\delta^{*}\right)\right)' \mathbf{\Gamma}_{i}' + \gamma_{i}'\right) \mathcal{F} + \left(\beta_{i0} - \hat{\beta}_{i} \left(\delta^{*}\right)\right)' \Delta^{\delta^{*}-\vartheta_{i}} \mathbf{E}_{i} + \lambda^{-1} \left(L;\delta_{i0} - \delta^{*}\right) \varepsilon_{i} - \mathbf{\Pi}_{T}^{*} \varepsilon_{i0}\right] \\ \times \left(W_{f}(\theta) - W_{h}(\theta)\right). \end{split}$$

Now we study the contribution of each (cross-) product $\varepsilon_i^{(j)}(\theta)\varepsilon_i^{(k)}(\theta)'$, j, k = 1, 2, 3, to $L_{i,T}^*$.

(a). Write can write the term
$$T^{-1}\varepsilon_{i}^{(1)}(\theta)\varepsilon_{i}^{(1)}(\theta)'$$
 as

$$\frac{1}{T} \left(\lambda^{0}(L;\theta)\varepsilon_{i} - \varsigma_{T}(\theta)\varepsilon_{i0}\right) \left(\lambda^{0}(L;\theta)\varepsilon_{i} - \varsigma_{T}(\theta)\varepsilon_{i0}\right)'$$

$$+\frac{1}{T} \left(\lambda^{-1}(L;\delta_{i0} - \delta^{*},\xi_{0})\varepsilon_{i}W_{f}(\theta) - \mathbf{\Pi}_{T}^{*}\varepsilon_{i0}W_{f}(\theta)\right) \left(\lambda^{-1}(L;\delta_{i0} - \delta^{*},\xi_{0})\varepsilon_{i}W_{f}(\theta) - \mathbf{\Pi}_{T}^{*}\varepsilon_{i0}W_{f}(\theta)\right)'$$

$$-\frac{2}{T} \left(\lambda^{0}(L;\theta)\varepsilon_{i} - \varsigma_{T}(\theta)\varepsilon_{i0}\right) \left(\lambda^{-1}(L;\delta_{i0} - \delta^{*},\xi_{0})\varepsilon_{i}W_{f}(\theta) - \mathbf{\Pi}_{T}^{*}\varepsilon_{i0}W_{f}(\theta)\right)'.$$

The first term converges uniformly in θ and is minimized for $\theta = \theta_{i0}$ as in the proof of Theorem 1. To show that the second term is negligible, it suffices to check the squared terms only. First, take

$$\frac{1}{T}\lambda^{-1} (L; \delta_{i0} - \delta^*, \xi_0) \varepsilon_i W_f(\theta) W_f(\theta)' \lambda^{-1} (L; \delta_{i0} - \delta^*, \xi_0) \varepsilon'_i$$

$$= \frac{1}{T}\lambda^{-1} (L; \delta_{i0} - \delta^*, \xi_0) \varepsilon_i \mathcal{F}' (\mathcal{F}\mathcal{F}')^{-1} \mathbf{F}(\theta) \mathbf{F}(\theta)' (\mathcal{F}\mathcal{F}')^{-1} \mathcal{F}\lambda^{-1} (L; \delta_{i0} - \delta^*, \xi_0) \varepsilon'_i$$
(26)

where

$$\frac{\mathcal{F}\mathcal{F}'}{T} \to_p \qquad \Sigma_f > 0,$$
$$\sup_{\theta \in \Theta} \left| \frac{\mathbf{F}(\theta)\mathbf{F}(\theta)'}{T} \right| = O_p \left(1 + T^{2(\varrho - \underline{\delta}) - 1} \right) = O_p \left(1 \right)$$

since $\rho - \underline{\delta} \leq 1/2$. Then, because

$$\frac{\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i\mathcal{F}'}{T} = O_p\left(T^{-1/2}+T^{\delta_0+\varrho-2\delta^*-1}\right) = o_p(1),$$

we obtain that (26) is $o_p(1)$ uniformly for $\theta \in \Theta$.

Next,

$$\frac{\mathbf{\Pi}_T^* \mathcal{F}'}{T} = O_p \left(T^{-1/2} \right) = o_p(1)$$

implies that

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \mathbf{\Pi}_T^* W_f(\theta) W_f(\theta)' \mathbf{\Pi}_T^{*\prime} \varepsilon_{i0}^2 \right| = o_p(1),$$

and all the other cross terms can be bounded uniformly in θ by the Cauchy-Schwarz inequality.

(b). Next, write
$$T^{-1}\varepsilon_i^{(2)}(\theta)\varepsilon_i^{(2)}(\theta)'$$
 as

$$\frac{1}{T} \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \right)' \left[\psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}_i - \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i W_f\left(\theta\right) \right] \left[\psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}_i - \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i W_f\left(\theta\right) \right]' \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \mathbf{E}_i \right) \right] \left[\psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}_i - \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i W_f\left(\theta\right) \right]' \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \mathbf{E}_i \right) \right] \left[\psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}_i \right]$$

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \right)' \psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}_i \psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}'_i \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \right) \right| = o_p(1)$$

because $\beta_{i0} - \hat{\beta}_i(\delta^*) = O_p\left(T^{-1/2}\right)$ by Theorem 5 and with $\vartheta_i - \underline{\delta} < 1$, $\sup_{\theta \in \Theta} \left| \frac{1}{T^2} \psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}_i \psi\left(L;\xi\right) \Delta^{\delta - \vartheta_i} \mathbf{E}'_i \right| = O\left(T^{-1} + T^{2(\vartheta_i - \underline{\delta} - 1)}\right) = o_p(1).$

Next,

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \right)' \Delta^{\delta^* - \vartheta_i} \mathbf{E}_i W_f(\theta) W_f(\theta)' \Delta^{\delta^* - \vartheta_i} \mathbf{E}'_i \left(\beta_{i0} - \hat{\beta}_i(\delta^*) \right) \right| = o_p(1)$$

since

$$\frac{\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i\mathcal{F}'}{T} = O_p\left(T^{-1/2} + T^{\vartheta_i+\varrho-2\delta^*-1}\right) = o_p(1),$$

and the cross-term is negligible by Cauchy-Schwarz inequality under the same conditions.

(c). Finally, write
$$T^{-1}\varepsilon_{i}^{(3)}(\theta)\varepsilon_{i}^{(3)}(\theta)'$$

$$\frac{1}{T} \qquad \left[\left(\left(\beta_{i0} - \hat{\beta}_{i}(\delta^{*}) \right)' \mathbf{\Gamma}_{i}' + \gamma_{i}' \right) \mathcal{F} + \left(\beta_{i0} - \hat{\beta}_{i}(\delta^{*}) \right)' \Delta^{\delta^{*}-\vartheta_{i}} \mathbf{E}_{i} + \lambda^{-1} \left(L; \delta_{i0} - \delta^{*} \right) \varepsilon_{i} - \mathbf{\Pi}_{T}^{*} \varepsilon_{i0} \right] \\
\times \left(W_{f}(\theta) - W_{h}(\theta) \right) \left(W_{f}(\theta) - W_{h}(\theta) \right)' \\
\times \left[\left(\left(\beta_{i0} - \hat{\beta}_{i}(\delta^{*}) \right)' \mathbf{\Gamma}_{i}' + \gamma_{i}' \right) \mathcal{F} + \left(\beta_{i0} - \hat{\beta}_{i}(\delta^{*}) \right)' \Delta^{\delta^{*}-\vartheta_{i}} \mathbf{E}_{i} + \lambda^{-1} \left(L; \delta_{i0} - \delta^{*} \right) \varepsilon_{i} - \mathbf{\Pi}_{T}^{*} \varepsilon_{i0} \right]'.$$

First,

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \left(\left(\beta_{i0} - \hat{\beta}_{i} \left(\delta^{*} \right) \right)' \mathbf{\Gamma}_{i}' + \gamma_{i}' \right) \mathcal{F} \left(W_{f} \left(\theta \right) - \left(\theta \right) W_{h} \right) \left(W_{f} \left(\theta \right) - W_{h} \left(\theta \right) \right)' \mathcal{F}' \left(\left(\beta_{i0} - \hat{\beta}_{i} \left(\delta^{*} \right) \right)' \mathbf{\Gamma}_{i}' + \gamma_{i}' \right)' \right) \right|$$

is $o_p(1)$ because

$\mathcal{F}W_{h}W_{h}'\mathcal{F}' = \mathcal{F}\bar{\mathbf{H}}'\left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-}\bar{\mathbf{H}}(\theta)\bar{\mathbf{H}}(\theta)'\left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-}\bar{\mathbf{H}}\mathcal{F}'$

for which it can be easily shown following the projection details above that

$$\begin{aligned} \mathcal{F}\bar{\mathbf{H}}' &= \frac{\mathcal{F}\mathcal{F}'}{T}\bar{\mathbf{C}}' + O_p\left(\frac{1}{N} + \frac{1}{\sqrt{NT}}\right) \\ \frac{\bar{\mathbf{H}}\bar{\mathbf{H}}'}{T} &= \bar{\mathbf{C}}\frac{\mathcal{F}\mathcal{F}'}{T}\bar{\mathbf{C}}' + O_p\left(\frac{1}{N} + \frac{1}{\sqrt{NT}}\right) \\ \sup_{\theta\in\Theta} \left|\frac{\bar{\mathbf{H}}(\theta)\bar{\mathbf{H}}(\theta)'}{T}\right| &= \bar{\mathbf{C}}\frac{\mathbf{F}(\theta)\mathbf{F}(\theta)'}{T}\bar{\mathbf{C}}' + O_p\left(\frac{1}{N} + \frac{1}{\sqrt{NT}} + \frac{T^{2(\vartheta_{max}-\underline{\delta})-1}}{\sqrt{N}} + \frac{T^{\vartheta_{max}+\varrho-2\underline{d}-1}}{\sqrt{N}}\right) \end{aligned}$$

where the projection errors are $o_p(1)$ if $\vartheta_{max} - \underline{\delta} < 1/2$, and $\vartheta_{max} + \varrho - 2\underline{\delta} - 1 < 0$ which is implied by $\vartheta_{max} - \underline{\delta} < 1/2$ and $\varrho - \underline{\delta} < 1/2$.

The other squared terms contain the initial memory value $\delta^* \geq 1$ which make them stationary. Thus it can be shown in a similar way to the analysis above that they are $o_p(1)$, and the proof of consistency is then complete.

Proof of asymptotic normality. The \sqrt{T} -normalized score evaluated at the true value,

$$\begin{split} &\sqrt{T}\frac{\partial}{\partial\theta}L_{i,T}^{*}(\theta)\Big|_{\theta=\theta_{i0}} \\ &= \frac{2}{\sqrt{T}}\left\{\left(\varepsilon_{i}-\varsigma_{T}(\theta_{i0})\varepsilon_{i0}-\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{i0}\right)\varepsilon_{i}W_{f}\left(\theta_{i0}\right)+\Pi_{T}^{*}\varepsilon_{i0}W_{f}\left(\theta_{i0}\right)\right)\right. \\ &-\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\left[\psi\left(L;\xi_{i0}\right)\Delta^{\delta_{i0}-\vartheta_{i}}\mathbf{E}_{i}-\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}W_{f}\left(\theta_{i0}\right)\right] \\ &+\left[\left(\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\Gamma_{i}'+\gamma_{i}'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}+\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{i0}\right)\varepsilon_{i}-\Pi_{T}^{*}\varepsilon_{i0}\right)\right. \\ &\times\left(W_{f}(\theta_{i0})-W_{h}(\theta_{i0})\right)\right\} \\ &\times\left\{\left(\chi\left(L;\xi_{i0}\right)\varepsilon_{i}-\dot{\varsigma}_{T}(\theta_{i0})\varepsilon_{i0}-\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{i0}\right)\varepsilon_{i}\dot{W}_{f}\left(\theta_{i0}\right)+\Pi_{T}^{*}\varepsilon_{i0}\dot{W}_{f}\left(\theta_{i0}\right)\right)\right. \\ &-\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\left[\chi\left(L;\xi_{i0}\right)\psi\left(L;\xi_{i0}\right)\Delta^{\delta_{i0}-\vartheta_{i}}\mathbf{E}_{i}-\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}\dot{W}_{f}\left(\theta_{i0}\right)\right] \\ &+\left[\left(\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)'\Gamma_{i}'+\gamma_{i}'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_{i}\left(\delta^{*}\right)\right)\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}+\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{i0}\right)\varepsilon_{i}-\Pi_{T}^{*}\varepsilon_{i0}\right] \\ &\times\left(\dot{W}_{f}\left(\theta_{i0}\right)-\dot{W}_{h}\left(\theta_{i0}\right)\right)\right\}', \end{split}$$

where

$$\dot{W}_{f}(\theta_{i0}) := \mathcal{F}'(\mathcal{F}\mathcal{F}')^{-} \dot{\mathbf{F}}(\theta_{i0}),
\dot{W}_{h}(\theta_{i0}) := \bar{\mathbf{H}}' (\bar{\mathbf{H}}\bar{\mathbf{H}}')^{-} \dot{\mathbf{H}}(\theta_{i0})$$

and $\dot{\mathbf{F}}(\theta) = (\partial/\partial\theta) \mathbf{F}(\theta)$, $\dot{\mathbf{\bar{H}}}(\theta) = (\partial/\partial\theta) \mathbf{\bar{H}}(\theta)$. Taking N = 1, the treatment for $\frac{2}{\sqrt{T}} \left[\varepsilon_i - \varsigma_T(\theta_{i0}) \varepsilon_{i0} \right] \left[\chi\left(L; \xi_{i0}\right) \varepsilon_i - \dot{\varsigma}_T(\theta_{i0}) \varepsilon_{i0} \right]$

has been shown in the proof of Theorem 2, where the term leads to the asymptotic normal distribution with an initial condition bias, that does not appear now because normalization is only by $T^{1/2}$. In what follows, we only check that the dominating terms are negligible since terms containing the estimation effect and/or δ^* have smaller sizes.

(a) First consider

$$\frac{2}{\sqrt{T}} \left[\varepsilon_i - \varsigma_T(\theta_{i0}) \varepsilon_{i0} \right] \left[\lambda^{-1} \left(L; \delta_{i0} - \delta^*, \xi_{i0} \right) \varepsilon_i \dot{W}_f(\theta_{i0}) - \Pi_T^* \varepsilon_{i0} \dot{W}_f(\theta_{i0}) \right]'.$$
(27)

Then,

$$\frac{1}{\sqrt{T}}\varepsilon_{i}\dot{W}_{f}\left(\theta_{i0}\right)'\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{i0}\right)\varepsilon_{i}'=\frac{1}{\sqrt{T}}\varepsilon_{i}\dot{\mathbf{F}}\left(\theta_{i0}\right)'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathcal{F}\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{i0}\right)\varepsilon_{i}'=o_{p}\left(1\right)$$

because $\rho-\delta_{i0}<1/2$ so that $T^{-1}\mathcal{F}\mathcal{F}'\to_{p}\Sigma_{f}>0$,

$$\frac{\varepsilon_i \dot{\mathbf{F}}(\theta_{i0})'}{T} = O_p \left(T^{-1/2} + T^{\varrho - \delta_{i0} - 1} \log T \right)$$
$$\frac{\mathcal{F}\lambda^{-1} \left(L; \delta_{i0} - \delta^*, \xi_{i0} \right) \varepsilon'_i}{T} = O_p \left(T^{-1/2} + T^{\varrho + \delta_{i0} - 2\delta^* - 1} \right).$$

Using the methods of the proof of Lemma 2(c), it can be shown that, using $\rho - \delta_{i0} < 1/2$,

$$\frac{1}{T} \sum_{t=1}^{T} \pi_t \left(\delta^* - 1 \right) \chi_t \left(L; \xi_{i0} \right) \lambda_t \left(L; \theta_{i0} \right) f_t = O_p \left(T^{-1} \log T \right) \frac{1}{T} \sum_{t=1}^{T} \tau_t \left(\theta_{i0} \right) \Delta^{\delta^*} f_t = O_p \left(T^{-1} + T^{-1/2 - \delta_{i0}/2} \right)$$

because $\delta^* \geq 1$ and Assumption E, and therefore following the same steps,

$$\frac{2}{\sqrt{T}}\varsigma_T(\theta_{i0})\dot{W}_f(\theta_{i0})\,\mathbf{\Pi}_T^{*'}\varepsilon_{i0}^2 = O_p\left(T^{-1/2}\left(T^{-1} + T^{-1/2-\delta_{i0}/2}\right)\log T\right) = o_p(1),$$

and we can conclude that (27) is $o_p(1)$.

(b) To show that

$$\frac{2}{\sqrt{T}} \left[\varepsilon_{i} - \varsigma_{T}(\theta_{i0})\varepsilon_{i0}\right] \left(\left(\beta_{i0} - \hat{\beta}_{i}\left(\delta^{*}\right)\right)' \left[\chi\left(L;\xi_{i0}\right)\psi\left(L;\xi_{i0}\right)\Delta^{\delta_{i0}-\vartheta_{i}}\mathbf{E}_{i} - \Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}\dot{W}_{f}\left(\theta_{i0}\right)\right] \right)' = o_{p}\left(1\right)$$

if $\vartheta_i - \delta_{i0} < 1/2$ it suffices to check that

$$\frac{2}{\sqrt{T}}\varepsilon_i\chi\left(L;\xi_{i0}\right)\psi\left(L;\xi_{i0}\right)\Delta^{\delta_{i0}-\vartheta_i}\mathbf{E}'_i\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right) = O_p\left(T^{-1/2}+T^{\vartheta_i-\delta_{i0}-1}\log T\right)$$

which is $o_p(1)$ because $\vartheta_i - \delta_{i0} < 1/2$ and the remaining terms have smaller orders.

(c) The term dealing with the projection approximation,

$$\frac{2}{\sqrt{T}} \left[\varepsilon_{i} - \varsigma_{T}(\theta_{i0})\varepsilon_{i0}\right] \left\{ \left[\left(\left(\beta_{i0} - \hat{\beta}_{i}(\delta^{*})\right)' \mathbf{\Gamma}_{i}' + \gamma_{i}' \right) \mathcal{F} + \left(\beta_{i0} - \hat{\beta}_{i}(\delta^{*})\right)' \Delta^{\delta^{*} - \vartheta_{i}} \mathbf{E}_{i} + \lambda^{-1} \left(L; \delta_{i0} - \delta^{*}, \xi_{i0}\right) \varepsilon_{i} - \mathbf{\Pi}_{T}^{*} \varepsilon_{i0} \right] \times \left(\dot{W}_{f}\left(\theta_{i0}\right) - \dot{W}_{h}\left(\theta_{i0}\right) \right) \right\}',$$

can be shown to be $o_p(1)$ following the same steps described earlier since, for instance,

$$\frac{1}{\sqrt{T}}\varepsilon_{i}\left(\dot{W}_{f}\left(\theta_{i0}\right)-\dot{W}_{h}\left(\theta_{i0}\right)\right)'\mathcal{F}'=o_{p}\left(1\right).$$

All other cross terms have a similar structure, and showing their orders to be $o_p(1)$ is analogous to what has been discussed so far, so the result follows. Then the convergence of the Hessian can be studied as in Theorem 2 but in a simpler way and the proof is complete.

Proof of Theorem 7. Using the result obtained in Corollary 2, and noting that this result satisfies the requirement, $\theta_i - \theta_{i0} = O_p(T^{-\kappa})$, $\kappa > 0$, for Theorem 1 of Robinson and Hualde (2003) along with the other conditions therein, it also holds that

$$\sqrt{T}\left(\hat{\beta}_{i}(\theta_{i})-\beta_{i0}\right)=\left(T^{-1}\mathcal{X}_{i}\bar{\mathcal{W}}_{T}\mathcal{X}_{i}'\right)^{-1}T^{-1/2}\mathcal{X}_{i}\bar{\mathcal{W}}_{T}\varepsilon_{i}'+o_{p}\left(1\right)+O_{p}\left(N^{-1}\sqrt{T}\right),$$

where the latter $O_p(\cdot)$ term stems from the projection and is removed if $\sqrt{T}/N \to 0$ as $(N,T)_j \to \infty$. \Box

Proof of Theorem 8. The properties of the mean group estimate follow as in Pesaran (2006) under the rank condition and the random coefficients model, we omit the details. \Box

9. 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 $\bar{\gamma}^2 \rightarrow_p (E[\gamma_i])^2 > 0$ and $E[\gamma_i^2] < \infty$. Write

$$\frac{1}{N}\sum_{i=1}^{N}\eta_{i}^{2} = \frac{\frac{1}{NT^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta\bar{y}_{t}\Delta\bar{y}_{t'}\sum_{i=1}^{N}\lambda_{t}\left(L;\theta_{0}^{(-1)}\right)\left(\varepsilon_{it}-\frac{\gamma_{i}}{\bar{\gamma}}\bar{\varepsilon}_{t}\right)\lambda_{t'}\left(L;\theta_{0}^{(-1)}\right)\left(\varepsilon_{it'}-\frac{\gamma_{i}}{\bar{\gamma}}\bar{\varepsilon}_{t'}\right)}{\left(\frac{1}{T}\sum_{t=1}^{T}(\Delta\bar{y}_{t})^{2}\right)^{2}}$$

The denominator converges to a positive constant term because

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

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 $\rho < 2/3$ and the second and third term are negligible due to cross section averaging.

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,

$$\frac{1}{NT^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \Delta \bar{y}_{t} \Delta \bar{y}_{t'} \sum_{i=1}^{N} \lambda_{t} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it} \lambda_{t'} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it'} \tag{28}$$

$$= \frac{1}{NT^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \Delta f_{t} \Delta f_{t'} \sum_{i=1}^{N} \lambda_{t} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it} \lambda_{t'} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it'} + \frac{1}{NT^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \lambda_{t} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{t} \lambda_{t'} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{t'} \sum_{i=1}^{N} \lambda_{t} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it'} + \frac{2}{NT^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \Delta f_{t} \lambda_{t'} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{t'} \sum_{i=1}^{N} \lambda_{t} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it} \lambda_{t'} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{it'}.$$

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,

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

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

$$\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} E\left[\lambda_{t}\left(L;\theta_{0}^{(-1)}\right) \bar{\varepsilon}_{t}\lambda_{t'}\left(L;\theta_{0}^{(-1)}\right) \bar{\varepsilon}_{t'}\lambda_{t}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{it}\lambda_{t'}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{it'}\right] \\
= \frac{1}{N^{2}T^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{j=1}^{N} \sum_{k=1}^{N} E\left[\lambda_{t}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{jt}\lambda_{t'}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{kt'}\lambda_{t}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{it}\lambda_{t'}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{it}\lambda_{t'}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{it'}\right] \\
= \frac{1}{N^{2}T^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{a=1}^{L} \sum_{b=1}^{t'} \sum_{c=1}^{t} \sum_{d=1}^{t'} \tau_{a}^{0} \tau_{b}^{0} \tau_{c}^{0} \tau_{\delta}^{0} E\left[\varepsilon_{jt-a} \varepsilon_{kt'-b} \varepsilon_{it-c} \varepsilon_{it'-d}\right],$$

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

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

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

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

proceeding as in Lemma 8. The second type is

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

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

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

The final fourth type involving fourth order cumulants is

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

The third term in (28) can be bounded using Cauchy-Schwarz inequality and the Lemma follows. \Box

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

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

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

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

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

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} (\lambda_{t-1} (L; \theta))^2 \right| \leq \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \left\{ (\lambda_{t-1} (L; \theta) f_t)^2 - E \left[(\lambda_{t-1} (L; \theta) f_t)^2 \right] \right\} \right|$$
(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 $c_j = c_j(\delta - \varrho, \xi) = \sum_{k=0}^j \varphi_k^f \lambda_{j-k}(\delta - \varrho, \xi) \sim cj^{\varrho - \delta - 1}$ as $j \to \infty$ under Assumption A.2.

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

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

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

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

say. Then, with $\gamma_{v}(j) = E[v_{0}v_{j}]$,

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

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

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

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

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

Uniformly in j,

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

 $\mathrm{so},$

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

where K is some arbitrarily large positive constant. Similarly,

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

since

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

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

To prove (c) 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

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

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

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

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

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

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

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

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

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

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

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

and the result follows. $\hfill \square$

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

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

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

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\lambda_{t-1}\left(L;\theta\right)f_{t}*\lambda_{t}^{0}\left(L;\theta\right)\left(\varepsilon_{it}-\hat{\phi}_{i}\bar{\varepsilon}_{t}\right)$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\lambda_{t-1}\left(L;\theta\right)f_{t}*\lambda_{t}^{0}\left(L;\theta\right)\varepsilon_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i}\lambda_{t-1}\left(L;\theta\right)f_{t}*\lambda_{t}^{0}\left(L;\theta\right)\hat{\phi}_{i}\bar{\varepsilon}_{t},$$

where the first term is equal to

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

$$= \frac{1}{T^{-1} \sum_{t} (\Delta \bar{y}_{t})^{2}} \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} \Delta \bar{y}_{r} \lambda_{r}^{-1} (L; \theta_{0}^{(-1)}) \left(\varepsilon_{ir} - \frac{\gamma_{i}}{\bar{\gamma}} \bar{\varepsilon}_{r}\right) * \lambda_{t-1} (L; \theta) f_{t} * \lambda_{t}^{0} (L; \theta) \varepsilon_{it}$$

$$= \frac{1}{T^{-1} \sum_{t} (\Delta \bar{y}_{t})^{2}} \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} \left(\bar{\gamma} \Delta f_{r} + \lambda_{r}^{-1} (L; \theta_{0}^{(-1)}) \bar{\varepsilon}_{r}\right) \lambda_{r}^{-1} (L; \theta_{0}^{(-1)}) \left(\varepsilon_{ir} - \frac{\gamma_{i}}{\bar{\gamma}} \bar{\varepsilon}_{r}\right) * \lambda_{t-1} (L; \theta) f_{t} * \lambda_{t}^{0} (L; \theta) \varepsilon_{it}$$

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

$$\frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} \left(\bar{\gamma} \Delta f_{r} + \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \right) \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \left(\varepsilon_{ir} - \frac{\gamma_{i}}{\bar{\gamma}} \bar{\varepsilon}_{r} \right) \lambda_{t-1} \left(L; \theta \right) f_{t} \lambda_{t}^{0} \left(L; \theta \right) \varepsilon_{it} \tag{31}$$

$$= \frac{\bar{\gamma}}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} \Delta f_{r} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{ir} \lambda_{t-1} \left(L; \theta \right) f_{t} \lambda_{t}^{0} \left(L; \theta \right) \varepsilon_{it} + \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \varepsilon_{ir} \lambda_{t-1} \left(L; \theta \right) f_{t} \lambda_{t}^{0} \left(L; \theta \right) \varepsilon_{it} + \frac{1}{NT^{2} \bar{\gamma}} \sum_{i=1}^{N} \gamma_{i} \sum_{t=1}^{T} \sum_{r=1}^{T} \bar{\gamma} \Delta f_{r} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \lambda_{t-1} \left(L; \theta \right) f_{t} \lambda_{t}^{0} \left(L; \theta \right) \varepsilon_{it} - \frac{1}{NT^{2} \bar{\gamma}} \sum_{i=1}^{N} \gamma_{i} \sum_{t=1}^{T} \sum_{r=1}^{T} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \lambda_{t-1} \left(L; \theta \right) f_{t} \lambda_{t}^{0} \left(L; \theta \right) \varepsilon_{it} - \frac{1}{NT^{2} \bar{\gamma}} \sum_{i=1}^{N} \gamma_{i} \sum_{t=1}^{T} \sum_{r=1}^{T} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \lambda_{r}^{-1} \left(L; \theta_{0}^{(-1)} \right) \bar{\varepsilon}_{r} \lambda_{t-1} \left(L; \theta \right) f_{t} \lambda_{t}^{0} \left(L; \theta \right) \varepsilon_{it}.$$

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

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

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

$$\frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} E\left[\Delta f_{r} \lambda_{t-1}\left(L;\theta\right) f_{t}\right] E\left[\lambda_{r}^{-1}\left(L;\theta_{0}^{(-1)}\right) \varepsilon_{ir} \lambda_{t}^{0}\left(L;\theta\right) \varepsilon_{it}\right] \\
= O\left(\frac{\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{r=1}^{T} \left(|t-r|_{+}^{2(\varrho-1)-\underline{\delta}} + |t-r|_{+}^{\varrho-1-\underline{\delta}} + |t-r|_{+}^{\varrho-2}\right) \\
\times \left(|t-r|_{+}^{2(\delta_{0}-1)-\underline{\delta}} + |t-r|_{+}^{\delta_{0}-1-\underline{\delta}} + |t-r|_{+}^{\delta_{0}-2}\right)\right) \\
= o\left(1\right)$$

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

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

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

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

= (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

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

where

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

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

$$\begin{split} &\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 \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 \Delta_{r'}^{1-\varrho} z_{r'} \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] \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_r^{-1} \left(L; \theta_0^{(-1)}\right) \varepsilon_{ir'} + 2E\left[z_\ell \Delta^{1-\varrho} z_r\right] E\left[z_\ell \Delta_{r'}^{1-\varrho} z_{r'}\right]\right) \\ &\times \left(E\left[\varepsilon_{i\ell}^2\right] E\left[\Delta_r^{-1} \left(L; \theta_0^{(-1)}\right) \varepsilon_{ir'} + 2E\left[\varepsilon_{i\ell} \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] \right\} \\ \end{array}$$

and $\sum_{\ell=1}^{T-j} Var[A_{i,\ell,\ell}]$ is, using Lemma 8,

$$O\left(\frac{1}{T^2}\sum_{\ell=1}^{T-j}\sum_{r=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c}\left(|r-r'|_{+}^{2(\varrho-1)-1}+|r-r'|_{+}^{\varrho-2}+|r-\ell|^{\varrho-2}|r'-\ell|^{\varrho-2}\right)\\\times\left(|r-r'|_{+}^{2(\delta_{0}-1)-1}+|r-r'|_{+}^{\delta_{0}-2}+|r-\ell|^{\delta_{0}-2}|r'-\ell|^{\delta_{0}-2}\right)\end{array}\right\}\right)$$
$$=O\left(\log T+T^{2(\varrho-1)+2(\delta_{0}-1)-1}\right),$$

while using a similar argument

$$\begin{split} &Co\varepsilon \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] \right\} \\ = & \frac{1}{T^2} \sum_{r=1}^T \sum_{r'=1}^T \left\{ \begin{array}{c} 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] \right\} \\ = & \frac{1}{T^2} \sum_{r=1}^T \sum_{r'=1}^T \left\{ \begin{array}{c} E \left[z_{\ell} \varepsilon_{\ell'} \Delta_r^{1-\varrho} z_r \lambda_r^{-1} \left(L; \theta_0^{-1}\right) \varepsilon_{ir'} \right] E \left[\varepsilon_{i\ell} \varepsilon_{i\ell'} \Delta_{r'}^{1-\varrho} \varepsilon_{ir'} \Delta_{r+1}^{1-\varepsilon_{ir'}} \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_1^{1-\varrho} z_r \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[\varepsilon_{i\ell} \varepsilon_{\ell'} \Delta_{r'}^{1-\varrho} z_{r'} \right] + E \left[z_{\ell} \Delta^{1-\varrho} z_r \right] E \left[\varepsilon_{\ell'} \Delta_{r'}^{1-\varrho} z_{r'} \right] E \left[z_{\ell'} \Delta_{r'}^{1-\varrho} z_{r'} \right] \\ + E \left[\varepsilon_{i\ell} \varepsilon_{i\ell'} \sum_{r'=1}^{1-\varrho} E \left[\varepsilon_{i\ell'} \lambda_{r'}^{-1} \left(L; \theta_0^{(-1)}\right) \varepsilon_{ir'} \right] \right] \\ + E \left[\varepsilon_{i\ell'} \delta_{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] E \left[z_{\ell'} \Delta^{1-\varrho} z_{r'} \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[\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[\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[\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[\varepsilon_{\ell'} \Delta^{1-\varrho} z_{r''} \right] E \left[\varepsilon_{\ell'} \Delta^{1-\varrho} z_{r''} \right] E \left[\varepsilon_{\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[\varepsilon_{\ell'} \Delta^{1-\varrho} z_{r''} \right] E \left[\varepsilon_{\ell'} \Delta^{1-\varrho} z_{r'''} \right] E \left[\varepsilon_{\ell'}$$

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

$$O\left(\frac{1}{T^2}\sum_{\ell=1}^{T-j}\sum_{r'=1}^{T-j}\sum_{r'=1}^{T}\sum_{r'=1}^{T}\left\{\begin{array}{c} \left(\frac{|\ell-\ell'|^{-2}\left(|r-r'|^{2(\varrho-1)-1}_{+}+|r-r'|^{\varrho-2}_{+}\right)\right)}{+|r-\ell|^{\varrho-2}|r'-\ell'|^{\varrho-2}+|r'-\ell|^{\varrho-2}|r-\ell'|^{\varrho-2}}\right)\\ \times \left(\begin{array}{c} |\ell-\ell'|^{-2}\left(|r-r'|^{2(\delta_{0}-1)-1}+|r-r'|^{\delta_{0}-2}_{+}\right)\\ +|r-\ell|^{\delta_{0}-2}|r'-\ell'|^{\delta_{0}-2}+|r-\ell'|^{\delta_{0}-2}|r'-\ell|^{\delta_{0}-2}\end{array}\right)\end{array}\right\}\right)$$
$$= O\left(\log T + T^{2(\varrho-1)+2(\delta_{0}-1)-1}\right).$$

Then, using $\left|\lambda_{j}\left(\delta-\varrho,\xi\right)\lambda_{j}^{0}\left(\theta\right)\right| \leq C j^{\varrho+\delta_{0}-2\delta-2}$,

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

= $O\left(N^{-1} \left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1} \right) \left(T^{-2} + \sup_{\delta} T^{2(\varrho-1)+2(\delta_0-1)-4\delta} \right) \right)^{1/2}$
= $O(1) + O\left(N^{-1} T^{4(\varrho-1)+4(\delta_0-1)-1-4\delta} \right)^{1/2} = O(1)$

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

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

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

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

$$KT^{-1}T^{\delta_0-\underline{\delta}-1}\sum_{j=1}^{T} j^{\varrho-\underline{d}-1}TN^{-1/2} \left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2}$$

$$\leq KN^{-1/2}T^{\delta_0-\underline{\delta}-1} \left(1+T^{\varrho-\underline{\delta}}\right) \left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right)^{1/2}$$

$$\leq K \left\{ N^{-1} \left(T^{2(\delta_0-1)-2\underline{\delta}} + T^{2(\delta_0+\varrho-1)-4\underline{\delta}}\right) \left(\log T + T^{2(\varrho-1)+2(\delta_0-1)-1}\right) \right\}^{1/2}$$

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

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

which is o(1) 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. \Box

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

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

is

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

uniformly in δ , so we can show that the term (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 $O\left(N^{-1/2}\left(T^{-1}+T^{-2\delta}\right)\right) = o_p(1)$, uniformly in δ , while the second can be written using summation by parts as

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

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

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

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

Then,

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

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

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

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

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

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

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

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

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

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

because $\bar{\varepsilon}_0 = O_p(N^{-1/2})$, $\bar{\gamma}^{-1} = O_p(1)$ 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^{*},

$$\left|\frac{\bar{\varepsilon}_{0}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{t}^{0}\left(L;\theta\right)\varepsilon_{it}\tau_{t}(\theta)\eta_{i}\right| \leq \left|\bar{\varepsilon}_{0}\right|\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_{t}^{0}\left(L;\theta\right)\varepsilon_{it}\right)^{2}\tau_{t}^{2}(\theta)\right|^{1/2}\left|\frac{1}{N}\sum_{i=1}^{N}\eta_{i}^{2}\right|^{1/2} = o_{p}\left(1\right)$$

because we can show that

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

using again the same methods, $|\bar{\varepsilon}_0| = O_p(N^{-1/2})$ 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. \Box

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

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r=1}^{T} \tau_{t}(\theta_{0}) \tau_{r}(\theta_{0}) Cov \left[\varepsilon_{i0} * \chi_{t} \left(L; \xi_{0}\right) \varepsilon_{it}, \varepsilon_{i0} * \chi_{r} \left(L; \xi_{0}\right) \varepsilon_{ir}\right] \\
= \frac{1}{T} \sum_{t=1}^{T} \tau_{t}(\theta_{0})^{2} \left[\sigma^{4} \left(\sum_{j=0}^{t} j^{-2} + t^{-2} \right) + \left\{ \kappa_{4} \right\} \right] \\
+ \frac{1}{T} \sum_{t=1}^{T} \sum_{r=1}^{T} \tau_{t}(\theta_{0}) \tau_{r}(\theta_{0}) \left[\sigma^{4} \left(\sum_{j=0}^{t} j^{-1} \left(t - r + j\right)^{-1} + t^{-1}r^{-1} \right) + \kappa_{4}t^{-2}1\left\{ t = r \right\} \right] \\
= O\left(T^{-1} + T^{-2\delta_{0}}\right) + O\left(T^{-1} \sum_{t=1}^{T} \sum_{r=1}^{t} \left(rt\right)^{-\delta_{0}} \left(|t - r|^{-1}\log t + (tr)^{-1}\right)\right) \\
= O\left(T^{-1} + T^{-2\delta_{0}}\right) + O\left(T^{-1} \sum_{t=1}^{T} t^{-\delta_{0}} \left(t^{-\delta_{0}}\log^{2} t + t^{-1}\log t\right)\right) = O\left(T^{-1}\log^{4} T + T^{-2\delta_{0}}\log^{2} T\right) = o\left(1\right)$$

while for the other three terms, we can check in turn that

$$-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{t}(\theta_{0})\varepsilon_{i0}\hat{\phi}_{i}\chi_{t}(L;\xi_{0})\bar{\varepsilon}_{t} = O_{p}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i0}\hat{\phi}_{i}\sum_{t=1}^{T}\tau_{t}(\theta_{0})\chi_{t}(L;\xi_{0})\bar{\varepsilon}_{t}\right)$$
$$= O_{p}\left((T/N)^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}\hat{\phi}_{i}\sum_{t=1}^{T}\tau_{t}(\theta_{0})\chi_{t}(L;\xi_{0})\bar{\varepsilon}_{t}\right)$$
$$= O_{p}\left((T/N)^{-1/2}N^{-1/2}\left\{1+T^{1/2-\delta_{0}}\log^{1/2}T\right\}\right)$$

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

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

while

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

using $\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{i}=O_{p}(1)$ and the same argument as for N=1, and finally

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

and the proof is completed. \Box

Lemma 6. Under the conditions of Theorem 2,

$$\frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \left(\varepsilon_{it} - \hat{\phi}_{i} \bar{\varepsilon}_{t} \right) \left[\chi_{t} \left(L; \xi_{0} \right) \varepsilon_{it} - \hat{\phi}_{i} \chi_{t} \left(L; \xi_{0} \right) \bar{\varepsilon}_{t} \right] \right\} \rightarrow_{d} \mathcal{N} \left(0, 4B \left(\xi_{0} \right) \right)$$

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

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{it} * \chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it} - \varepsilon_{it}\hat{\phi}_{i}\chi_{t}\left(L;\xi_{0}\right)\bar{\varepsilon}_{t} - \hat{\phi}_{i}\bar{\varepsilon}_{t}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it} + \hat{\phi}_{i}^{2}\bar{\varepsilon}_{t} * \chi_{t}\left(L;\xi_{0}\right)\bar{\varepsilon}_{t}\right\}$$
(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

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

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

$$\frac{1}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\sum_{t'=1}^{N}\sum_{j'=1}^{T}E\left[\varepsilon_{it}\gamma_{i}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{jt}\varepsilon_{i't'}\gamma_{i'}\chi_{t'}\left(L;\xi_{0}\right)\varepsilon_{j't'}\right]$$

$$=\frac{1}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\sum_{t'=1}^{T}E\left[\gamma_{i}\gamma_{i'}\right]E\left[\varepsilon_{it}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{jt}\varepsilon_{i't'}\chi_{t'}\left(L;\xi_{0}\right)\varepsilon_{j't'}\right]$$

$$=\frac{1}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E\left[\gamma_{i}^{2}\right]E\left[\varepsilon_{it}^{2}\right]E\left[\left\{\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{jt}\right\}^{2}\right]=O\left(N^{-1}\right)=o\left(1\right)$$

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

$$\begin{aligned} \left| \frac{2}{\sqrt{NT}} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \eta_{i} \chi_{t} \left(L; \xi_{0} \right) \varepsilon_{jt} \right| &\leq \frac{2}{\sqrt{NT}} \left(\frac{1}{N} \sum_{i=1}^{N} \eta_{i}^{2} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \chi_{t} \left(L; \xi_{0} \right) \varepsilon_{jt} \right)^{2} \right)^{1/2} \\ &= O_{p} \left((NT)^{-1/2} \left(T^{2\varrho+2\delta_{0}-6} + T^{-1} \right)^{1/2} \left(NT \right)^{1/2} \right) \\ &= O_{p} \left(\left(T^{2\varrho+2\delta_{0}-6} + T^{-1} \right)^{1/2} \right) = o_{p} \left(1 \right) \end{aligned}$$

because

$$E\left[\left(\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{jt}\right)^{2}\right] = \sum_{j=1}^{N}\sum_{j'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left[\varepsilon_{it}\varepsilon_{it'}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{jt}\chi_{t'}\left(L;\xi_{0}\right)\varepsilon_{j't'}\right]\right]$$
$$= \sum_{j=1}^{N}\sum_{t=1}^{T}E\left[\varepsilon_{it}^{2}\right]E\left[\left\{\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{jt}\right\}^{2}\right] = O\left(NT\right).$$

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

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\hat{\phi}_{i}\bar{\varepsilon}_{t} = \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{\gamma_{i}}{\bar{\gamma}} + \eta_{i}\right\}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\bar{\varepsilon}_{t}$$

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

$$\frac{2}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\sum_{t'=1}^{T}E\left[\gamma_{i}\gamma_{i'}\right]E\left[\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\varepsilon_{jt}\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{i't'}\varepsilon_{j't'}\right]$$
$$=\frac{2}{NT}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E\left[\gamma_{i}^{2}\right]E\left[\varepsilon_{jt}^{2}\right]E\left[\left\{\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\right\}^{2}\right]=O\left(N^{-1}\right)$$

while

$$\begin{aligned} \left| \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_{i} \chi_{t} \left(L; \xi_{0} \right) \varepsilon_{it} \bar{\varepsilon}_{t} \right| &\leq \frac{2N}{\sqrt{NT}} \left(\frac{1}{N} \sum_{i=1}^{N} \eta_{i}^{2} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \chi_{t} \left(L; \xi_{0} \right) \varepsilon_{it} \bar{\varepsilon}_{t} \right)^{2} \right)^{1/2} \\ &= O_{p} \left(N^{1/2} T^{-1/2} (T^{2\varrho+2\delta_{0}-6} + T^{-1})^{1/2} \left(N^{-1} T \right)^{1/2} \right) \\ &= O_{p} \left((T^{2\varrho+2\delta_{0}-6} + T^{-1})^{1/2} \right) = O_{p} \left(1 \right) \end{aligned}$$

because

$$E\left[\left(\sum_{t=1}^{T}\varepsilon_{it}\bar{\varepsilon}_{t}\right)^{2}\right] = \frac{1}{N^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{j=1}^{N}\sum_{j'=1}^{N}E\left[\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\varepsilon_{jt}\chi_{t'}\left(L;\xi_{0}\right)\varepsilon_{it'}\varepsilon_{j't'}\right]$$
$$= \frac{1}{N^{2}}\sum_{t=1}^{T}\sum_{j=1}^{N}E\left[\varepsilon_{jt}^{2}\right]E\left[\left\{\chi_{t}\left(L;\xi_{0}\right)\varepsilon_{it}\right\}^{2}\right] = O\left(TN^{-1}\right).$$

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 $O_{p}(T^{1/2})$ because it has zero expectation and variance

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

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

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

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 . \Box

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(\begin{array}{c} |t-r|_{+}^{2(\varrho-1)-\delta} \\ +|t-r|_{+}^{\varrho-1-\delta}r^{\varrho-1}1\left\{\varrho-1<0\right\}1\left\{r

$$= O\left(|t-r|_{+}^{2(\varrho-1)-\delta} + |t-r|_{+}^{\varrho-1-\delta} + |t-r|_{+}^{\varrho-2}\right)$$

$$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(\begin{array}{c} |t-r|_{+}^{\delta_{0}-1-\delta}r^{\delta_{0}-1}1\left\{\delta_{0}-1<0\right\}1\left\{r

$$= O\left(|t-r|_{+}^{2(\delta_{0}-1)-\delta} + |t-r|_{+}^{\delta_{0}-1-\delta} + |t-r|_{+}^{\delta_{0}-2}\right),$$$$$$

where $|a|_{+} = \max\{|a|, 1\}$ 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 $E[\Delta f_r \lambda_{t-1}(L;\theta) f_t]$, 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 $d_j(a) = \sum_{k=0}^j \varphi_k^f \pi_{j-k}(a) \sim cj^{-a-1}$ and $c_j(a) = c_j(a,\xi) = \sum_{k=0}^j \varphi_k^f \lambda_{j-k}(a,\xi) \sim cj^{-a-1}$ as $j \to \infty$, $d_j(0) = \varphi_j^f$ and $\sum_{j=0}^\infty d_j(a) = \sum_{j=0}^\infty c_j(a) = 0$ if $a > 0, \xi \in \Xi$, so that the absolute value of the last expression is bounded by, $\varrho \ge 1$,

$$K\sum_{j=0}^{r} |d_{j}(1-\varrho)| (j+t-r)^{\varrho-\delta-1} \leq K(t-r)^{\varrho-\delta-1} \sum_{j=0}^{t-r} |d_{j}(1-\varrho)| + K\sum_{j=t-r+1}^{r} j^{2\varrho-\delta-3}$$

$$\leq K(t-r)^{\varrho-\delta-1} (t-r)^{\varrho-1} + K(t-r)^{2(\varrho-1)-\delta}$$

$$= O\left((t-r)^{2(\varrho-1)-\delta}\right)$$

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

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

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

using that $c_{j+t-r+1} \left(\delta - \varrho\right) - c_{j+t-r} \left(\delta - \varrho\right) = c_{j+t-r+1} \left(\delta - \varrho + 1\right)$.

If
$$r > t$$

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

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

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

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

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

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

Similarly, if r = t

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

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. \Box

References

- BAI, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77(4), 1229–1279.
- (2010): "Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation," *Econometric Theory*, 26, 1088–1114.
- BAI, J. AND S. NG (2004): "A PANIC Attack on Unit Roots and Cointegration," Econometrica, 72(4), 1127–1177.
- CHAUVET, M., Z. SENYUZ, AND E. YOLDAS (2012): "What Does Realized Volatility Tell Us About Macroeconomic Fluctuations?" *Finance and Economics Discussion Series*, *Board of Governors of the Federal Reserve System (U.S.).*
- ERGEMEN, Y. E. (2015): "Fractionally Integrated Panel Data Systems," *Preprint UC3M*, http://dx.doi.org/10.2139/ssrn.2521050.
- 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.
- KAPETANIOS, G., M. H. PESARAN, AND T. YAMAGATA (2011): "Panels with Non-Stationary Multifactor Error Structures," *Journal of Econometrics*, 160(2), 326–348.
- MARINUCCI, D. AND P. ROBINSON (2000): "Weak Convergence of Multivariate Fractional Processes," *Stochastic Processes and their Applications*, 86, 103–120.
- MOON, H. R. AND B. PERRON (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122(1), 81–126.
- 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.
- PHILLIPS, P. AND H. R. MOON (1999): "Linear Regression Limit Theory For Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

- ROBINSON, P. M. (1995): "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23(5), 1630–1661.
 - (2005): "Robust Covariance Matrix Estimation : 'HAC' Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21(1), 171–180.
- ROBINSON, P. M. AND J. HIDALGO (1997): "Time Series Regression with Long-Range Dependence," *The Annals of Statistics*, 25(1), 77–104.
- ROBINSON, P. M. AND J. HUALDE (2003): "Cointegration in Fractional Systems with Unknown Integration Orders," *Econometrica*, 71(6), 1727–1766.
- ROBINSON, P. M. AND C. VELASCO (2015): "Efficient Estimation of Fractionally Integrated Panel Data Models with Fixed Effects," *Journal of Econometrics*, 185, 435–452.

		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	0, 50)	(10, 100)	(20, 50)	(20, 100)			
$\varrho = 0.4:$													
$\overline{\delta_0 = 0.3}$	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			
$\rho = 1:$													
$\overline{\delta_0} = 0.3$	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 1: Empirical bias and RMSE of $\hat{\delta}$ and $\tilde{\delta}$

		J	Incorrected	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)			
$\varrho = 0.4:$												
$\overline{\delta_0 = 0.3}$		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			
$\varrho = 1$:												
$\overline{\delta_0 = 0.3}$		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 2: Empirical coverage of 95% CI based on $\hat{\delta}$ and $\tilde{\delta}$

			$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$	5
		$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$
$\rho = 0.4$:										
Bias of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	0.0016	0.0005	0.0023	-0.0026	-0.0058	-0.0046	-0.0086	-0.0159	-0.0179
	$\hat{\beta}_{CC}(\delta^*)$	0.0012	0.0012	0.0012	0.0012	0.0012	0.0114	0.0005	0.0006	0.0009
	$\hat{eta}_{MG}(\hat{\delta})$	0.0007	0.0006	0.0006	0.0008	0.0008	0.0009	0.0005	0.0006	0.0010
	$\hat{\beta}_{CC}(\hat{\delta})$	0.0149	0.0054	0.0014	0.0089	0.0044	0.0016	0.0028	0.0018	0.0011
RMSE of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	0.0621	0.0567	0.0529	0.0611	0.0573	0.0552	0.0538	0.0536	0.0555
	$\hat{eta}_{CC}(\delta^*)$	0.0621	0.0569	0.0531	0.0609	0.0571	0.0551	0.0518	0.0501	0.0518
	$\hat{eta}_{MG}(\hat{\delta})$	0.0621	0.0567	0.0529	0.0611	0.0570	0.0550	0.0531	0.0512	0.0525
	$\hat{eta}_{CC}(\hat{\delta})$	0.0589	0.0559	0.0531	0.0454	0.0520	0.0550	0.0293	0.0403	0.0517
Bias of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.0854	0.0218	-0.0089	0.1133	0.0302	-0.0083	0.1635	0.0488	-0.0082
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.0840	0.0211	-0.0089	0.1100	0.0288	-0.0083	0.1573	0.0462	-0.0083
RMSE of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.0968	0.0458	0.0402	0.1245	0.0512	0.0399	0.1762	0.0673	0.0406
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.0956	0.0456	0.0403	0.1217	0.0506	0.0401	0.1711	0.0660	0.0410
$\underline{\varrho = 1:}$										
Bias of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	-0.0029	-0.0019	0.0017	-0.0039	-0.0052	-0.0024	-0.0070	-0.0131	-0.0140
	$\hat{eta}_{CC}(\delta^*)$	0.0006	0.0006	0.0008	0.0006	0.0007	0.0011	0.0001	0.0002	0.0007
	$\hat{eta}_{MG}(\hat{\delta})$	0.0001	0.0001	0.0001	0.0002	0.0002	0.0005	0.0001	0.0002	0.0006
	$\hat{eta}_{CC}(\hat{\delta})$	0.0436	0.0145	0.0012	0.0327	0.0127	0.0015	0.0146	0.0067	0.0012
RMSE of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	0.0624	0.0573	0.0537	0.0617	0.0580	0.0559	0.0545	0.0539	0.0555
	$\hat{eta}_{CC}(\delta^*)$	0.0626	0.0577	0.0541	0.0618	0.0581	0.0563	0.0533	0.0517	0.0534
	$\hat{eta}_{MG}(\hat{\delta})$	0.0624	0.0573	0.0537	0.0616	0.0577	0.0559	0.0540	0.0523	0.0537
	$\hat{eta}_{CC}(\hat{\delta})$	0.1033	0.0678	0.0539	0.0873	0.0648	0.0562	0.0577	0.0516	0.0533
Bias of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.1735	0.0609	0.0030	0.1870	0.0661	0.0033	0.2196	0.0816	0.0049
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.1724	0.0600	0.0031	0.1868	0.0651	0.0033	0.2179	0.0800	0.0049
RMSE of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.1903	0.0821	0.0427	0.2017	0.0862	0.0430	0.2327	0.1003	0.0451
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.1891	0.0816	0.0429	0.2010	0.0855	0.0433	0.2309	0.0991	0.0454

Table 3: Preliminary and Joint Estimation Bias and RMSE's with N = 10 and T = 50 ($\delta^* = 1$)

		$\vartheta = 0.75$				$\vartheta = 1$				
		$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	 $\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$
$\rho = 0.4$:										
$\frac{1}{\text{Bias of }\hat{\beta}}$	$\hat{eta}_{MG}(\delta^*)$	-0.0022	-0.0013	-0.0009	0.0004	0.0015	0.0016	0.0058	0.0074	0.0080
	$\hat{\beta}_{CC}(\delta^*)$	-0.0011	-0.0013	-0.0014	-0.0011	-0.0014	-0.0017	-0.0006	-0.0011	-0.0017
	$\hat{eta}_{MG}(\hat{\delta})$	-0.0011	-0.0013	-0.0014	-0.0010	-0.0013	-0.0017	-0.0006	-0.0010	-0.0016
	$\hat{\beta}_{CC}(\hat{\delta})$	0.0136	0.0026	-0.0013	0.0076	0.0018	-0.0017	0.0022	0.0005	-0.0016
RMSE of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	0.0295	0.0270	0.0254	0.0290	0.0271	0.0265	0.0256	0.0251	0.0262
	$\hat{eta}_{CC}(\delta^*)$	0.0299	0.0274	0.0258	0.0296	0.0276	0.0269	0.0251	0.0241	0.0252
	$\hat{eta}_{MG}(\hat{\delta})$	0.0294	0.0270	0.0254	0.0290	0.0271	0.0265	0.0250	0.0240	0.0250
	$\hat{eta}_{CC}(\hat{\delta})$	0.0341	0.0279	0.0258	0.0239	0.0258	0.0269	0.0131	0.0189	0.0252
Bias of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.0681	0.0174	-0.0028	0.0984	0.0257	-0.0012	0.1640	0.0490	0.0019
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.0679	0.0173	-0.0028	0.0975	0.0253	-0.0012	0.1616	0.0482	0.0019
RMSE of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.0723	0.0259	0.0189	0.1046	0.0329	0.0187	0.1739	0.0573	0.0195
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.0721	0.0259	0.0189	0.1038	0.0327	0.0187	0.1720	0.0568	0.0195
$\underline{\varrho} = 1:$										
Bias of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	-0.0031	-0.0026	-0.0027	0.0001	0.0008	0.0003	0.0068	0.0082	0.0082
	$\hat{eta}_{CC}(\delta^*)$	-0.0013	-0.0015	-0.0015	-0.0013	-0.0016	-0.0019	-0.0009	-0.0013	-0.0018
	$\hat{eta}_{MG}(\hat{\delta})$	-0.0013	-0.0015	-0.0016	-0.0012	-0.0015	-0.0018	-0.0008	-0.0012	-0.0018
	$\hat{eta}_{CC}(\hat{\delta})$	0.0588	0.0155	-0.0015	0.0423	0.0130	-0.0018	0.0159	0.0062	-0.0017
RMSE of $\hat{\beta}$	$\hat{eta}_{MG}(\delta^*)$	0.0297	0.0273	0.0258	0.0293	0.0274	0.0267	0.0263	0.0258	0.0267
	$\hat{eta}_{CC}(\delta^*)$	0.0302	0.0277	0.0261	0.0300	0.0280	0.0273	0.0258	0.0248	0.0259
	$\hat{eta}_{MG}(\hat{\delta})$	0.0296	0.0272	0.0257	0.0293	0.0274	0.0268	0.0255	0.0245	0.0255
	$\hat{eta}_{CC}(\hat{\delta})$	0.0927	0.0403	0.0260	0.0713	0.0371	0.0272	0.0362	0.0264	0.0258
Bias of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.1383	0.0406	0.0017	0.1545	0.0468	0.0032	0.2019	0.0680	0.0074
	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.1390	0.0404	0.0017	0.1570	0.0466	0.0032	0.2028	0.0676	0.0074
RMSE of $\hat{\delta}$	$\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$	0.1479	0.0494	0.0194	0.1628	0.0548	0.0198	0.2103	0.0765	0.0224
- •	$\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$	0.1482	0.0491	0.0195	0.1646	0.0546	0.0198	0.2107	0.0761	0.0224

Table 4: Preliminary and Joint Estimation Bias and RMSE's with N = 20 and T = 100 ($\delta^* = 1$)

70



Figure 1: Monthly Realized Volatilities across Industries



Figure 2: Monthly Realized Volatility in the Composite Market
m = 20:

Food	Bvrgs	Tobac	Games	Books	Hshld	Clths	Hlth	Chems	Txtls	Market
0.51	0.77	0.71	0.75	0.84	0.51	0.70	0.72	0.68	0.69	0.59
Cnstr	Steel	FabPr	ElcEq	Autos	Carry	Mines	Coal	Oil	Util	
0.73	0.71	0.73	0.86	0.74	0.70	0.47	0.71	0.56	0.52	
Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Finan	Other	
0.83	0.66	0.85	0.78	0.61	0.52	0.67	0.56	0.98	0.77	

m = 32:

Food	Bvrgs	Tobac	Games	Books	Hshld	Clths	Hlth	Chems	Txtls	Market
0.66	0.78	0.63	0.57	0.63	0.46	0.60	0.71	0.67	0.59	0.64
Cnstr	Steel	FabPr	ElcEq	Autos	Carry	Mines	Coal	Oil	Util	
0.74	0.72	0.64	0.69	0.56	0.55	0.54	0.63	0.58	0.58	
Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Finan	Other	
0.79	0.75	0.78	0.60	0.57	0.62	0.77	0.57	0.90	0.78	

Note: This table reports the local Whittle estimation results of the individual integration orders of industry and market realized volatilities with bandwidth choices of m = 20, 32. Estimates are rounded to two digits after zero. Standard errors of the estimates are 0.112 and 0.088 respectively for m = 20, 32.

Table 6: Residual Integration Order Estimates $(\hat{\delta}_i)$ of Industry Realized Volatilities

Food	Bvrgs	Tobac	Games	Books	Hshld	Clths	Hlth	Chems	Txtls
0.50	0.54	0.49	0.48	0.59	0.54	0.30	0.50	0.42	0.40
Cnstr	Steel	FabPr	ElcEq	Autos	Carry	Mines	Coal	Oil	Util
0.48	0.50	0.30	0.50	0.30	0.29	0.45	0.48	0.50	0.37
Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Finan	Other
0.51	0.58	0.65	0.43	0.42	0.28	0.65	0.54	0.53	0.43

Note: This table reports the estimation results of the integration order of individual industry realized volatility residuals. Estimations are performed based on our general model where the projections are carried out with $\delta^* = 1$. Values are rounded to two digits after zero. Standard error of these estimates is 0.065.

	Food	Bvrgs	Tobac	Games	Books	Hshld	Clths	Hlth
$\hat{\beta}_i^0$	0.5422	0.4002	0.3376	0.6896	0.6503	0.2707	0.7446	0.4289
	(0.1097)	(0.1379)	(0.1452)	(0.0762)	(0.0769)	(0.1234)	(0.0607)	(0.1199)
\hat{eta}_{i}	1.8145	1.4060	-0.1814	0.1361	0.4119	-0.2088	2.4219	-0.6377
	(0.0856)	(0.1006)	(0.1328)	(0.0559)	(0.1144)	(0.0864)	(0.0602)	(0.0830)
	Cnstr	Steel	FabPr	ElcEq	Autos	Carry	Mines	Coal
\hat{eta}_i^0	0.7346	0.8571	0.9094	0.6970	0.8332	0.6176	0.8373	0.7691
	(0.0821)	(0.0633)	(0.0413)	(0.0758)	(0.0523)	(0.0814)	(0.0854)	(0.0807)
\hat{eta}_{i}	-0.4109	0.1789	-0.4298	-0.3442	-0.3635	1.7414	-0.5087	0.3626
	(0.1266)	(0.0782)	(0.0537)	(0.0768)	(0.0765)	(0.0772)	(0.1335)	(0.1219)
	Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals
-								
$\hat{\beta}_i^0$	0.7190	0.6178	0.5250	0.6223	0.6183	0.8722	0.4078	0.5382
$\hat{\beta}_i^0$	$\begin{array}{c} 0.7190 \\ (0.0961) \end{array}$	$0.6178 \\ (0.1271)$	$0.5250 \\ (0.1530)$	$0.6223 \\ (0.0768)$	$0.6183 \\ (0.0751)$	$0.8722 \\ (0.0603)$	$0.4078 \\ (0.1308)$	(0.5382) (0.1020)
$\hat{\beta}_i^0$ $\hat{\beta}_i$	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \end{array}$	0.6178 (0.1271) -0.3669	$\begin{array}{c} 0.5250 \ (0.1530) \ 0.0311 \end{array}$	0.6223 (0.0768) -1.0433	0.6183 (0.0751) -0.1778	0.8722 (0.0603) -2.4097	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \end{array}$	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \end{array}$
$ \hat{\beta}_i^0 \\ \hat{\beta}_i $	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \\ (0.0628) \end{array}$	$\begin{array}{c} 0.6178 \\ (0.1271) \\ -0.3669 \\ (0.1329) \end{array}$	$\begin{array}{c} 0.5250 \\ (0.1530) \\ 0.0311 \\ (0.1718) \end{array}$	$\begin{array}{c} 0.6223 \\ (0.0768) \\ -1.0433 \\ (0.0686) \end{array}$	$\begin{array}{c} 0.6183 \\ (0.0751) \\ -0.1778 \\ (0.1065) \end{array}$	$\begin{array}{c} 0.8722 \\ (0.0603) \\ -2.4097 \\ (0.1122) \end{array}$	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \\ (0.0832) \end{array}$	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \\ (0.0820) \end{array}$
$\hat{\beta}_i^0$ $\hat{\beta}_i$	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \\ (0.0628) \end{array}$	$\begin{array}{c} 0.6178 \\ (0.1271) \\ -0.3669 \\ (0.1329) \end{array}$	$\begin{array}{c} 0.5250 \\ (0.1530) \\ 0.0311 \\ (0.1718) \end{array}$	$\begin{array}{c} 0.6223 \\ (0.0768) \\ -1.0433 \\ (0.0686) \end{array}$	$\begin{array}{c} 0.6183 \\ (0.0751) \\ -0.1778 \\ (0.1065) \end{array}$	$\begin{array}{c} 0.8722 \\ (0.0603) \\ -2.4097 \\ (0.1122) \end{array}$	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \\ (0.0832) \end{array}$	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \\ (0.0820) \end{array}$
$\hat{\beta}_i^0$ $\hat{\beta}_i$	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \\ (0.0628) \end{array}$	0.6178 (0.1271) -0.3669 (0.1329) Chems	0.5250 (0.1530) 0.0311 (0.1718) Txtls	0.6223 (0.0768) -1.0433 (0.0686) Oil	$0.6183 \\ (0.0751) \\ -0.1778 \\ (0.1065) \\ Util$	0.8722 (0.0603) -2.4097 (0.1122) Finan	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \\ (0.0832) \end{array}$ Other	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \\ (0.0820) \end{array}$
$\frac{\hat{\beta}_i^0}{\hat{\beta}_i}$	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \\ (0.0628) \end{array}$	0.6178 (0.1271) -0.3669 (0.1329) Chems 0.7898	$\begin{array}{c} 0.5250 \\ (0.1530) \\ 0.0311 \\ (0.1718) \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$	0.6223 (0.0768) -1.0433 (0.0686) Oil 0.7927	$\begin{array}{c} 0.6183 \\ (0.0751) \\ -0.1778 \\ (0.1065) \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$	0.8722 (0.0603) -2.4097 (0.1122) Finan 0.5316	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \\ (0.0832) \end{array}$ Other 0.1067	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \\ (0.0820) \end{array}$
$\frac{\hat{\beta}_i^0}{\hat{\beta}_i}$	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \\ (0.0628) \end{array}$	$\begin{array}{c} 0.6178 \\ (0.1271) \\ -0.3669 \\ (0.1329) \end{array}$ Chems $\begin{array}{c} 0.7898 \\ (0.0516) \end{array}$	$\begin{array}{c} 0.5250\\ (0.1530)\\ 0.0311\\ (0.1718)\\ \hline \\ Txtls\\ 0.4888\\ (0.0981)\\ \end{array}$	$\begin{array}{c} 0.6223 \\ (0.0768) \\ -1.0433 \\ (0.0686) \end{array}$	$\begin{array}{c} 0.6183 \\ (0.0751) \\ -0.1778 \\ (0.1065) \\ \hline \\ Util \\ 0.6498 \\ (0.0925) \end{array}$	$\begin{array}{c} 0.8722 \\ (0.0603) \\ -2.4097 \\ (0.1122) \\ \hline \\ Finan \\ 0.5316 \\ (0.0986) \end{array}$	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \\ (0.0832) \end{array}$ $\begin{array}{c} \text{Other} \\ 0.1067 \\ (0.0632) \end{array}$	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \\ (0.0820) \end{array}$
$ \begin{array}{c} \hat{\beta}_i^0 \\ \hat{\beta}_i \\ \hline \\ \hat{\beta}_i \\ \hat{\beta}_i \\ \hat{\beta}_i \end{array} $	$\begin{array}{c} 0.7190 \\ (0.0961) \\ 0.1399 \\ (0.0628) \end{array}$	$\begin{array}{c} 0.6178 \\ (0.1271) \\ -0.3669 \\ (0.1329) \end{array}$ Chems $\begin{array}{c} 0.7898 \\ (0.0516) \\ -0.0546 \end{array}$	$\begin{array}{c} 0.5250 \\ (0.1530) \\ 0.0311 \\ (0.1718) \end{array}$ Txtls $\begin{array}{c} 0.4888 \\ (0.0981) \\ -0.1731 \end{array}$	$\begin{array}{c} 0.6223 \\ (0.0768) \\ -1.0433 \\ (0.0686) \\ \hline \\ Oil \\ 0.7927 \\ (0.0852) \\ -0.1238 \end{array}$	$\begin{array}{c} 0.6183 \\ (0.0751) \\ -0.1778 \\ (0.1065) \\ \hline \\ Util \\ 0.6498 \\ (0.0925) \\ -0.4930 \\ \end{array}$	$\begin{array}{c} 0.8722 \\ (0.0603) \\ -2.4097 \\ (0.1122) \\ \\ \hline \\ Finan \\ 0.5316 \\ (0.0986) \\ -0.8456 \end{array}$	$\begin{array}{c} 0.4078 \\ (0.1308) \\ 2.6804 \\ (0.0832) \\ \hline \\ Other \\ 0.1067 \\ (0.0632) \\ -0.1933 \\ \end{array}$	$\begin{array}{c} 0.5382 \\ (0.1020) \\ -0.6838 \\ (0.0820) \end{array}$

Table 7: Estimated Slope Parameters across Industry Realized Volatilities

Note: This table reports the estimation results of the individual slope parameters across industry realized volatilities, where $\hat{\beta}_i^0$ is the coefficient of market realized volatility, and $\hat{\beta}_i$ is the coefficient of the average effect of Fama-French factors. Estimations are performed based on our general model where the projections are carried out with $\delta^* = 1$. Robust standard errors are reported in parentheses.

Research Papers 2013



- 2015-18: Peter Reinhard Hansen: A Martingale Decomposition of Discrete Markov Chains
- 2015-19: Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov: A Markov Chain Estimator of Multivariate Volatility from High Frequency Data
- 2015-20: Henri Nyberg and Harri Pönkä: International Sign Predictability of Stock Returns: The Role of the United States
- 2015-21: Ulrich Hounyo and Bezirgen Veliyev: Validity of Edgeworth expansions for realized volatility estimators
- 2015-22: Girum D. Abate and Niels Haldrup: Space-time modeling of electricity spot prices
- 2015-23: Eric Hillebrand, Søren Johansen and Torben Schmith: Data revisions and the statistical relation of global mean sea-level and temperature
- 2015-24: Tommaso Proietti and Alessandra Luati: Generalised partial autocorrelations and the mutual information between past and future
- 2015-25: Bent Jesper Christensen and Rasmus T. Varneskov: Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination
- 2015-26: Ulrich Hounyo and Rasmus T. Varneskov: A Local Stable Bootstrap for Power Variations of Pure-Jump Semimartingales and Activity Index Estimation
- 2015-27: Rasmus Søndergaard Pedersen and Anders Rahbek: Nonstationary ARCH and GARCH with t-distributed Innovations
- 2015-28: Tommaso Proietti and Eric Hillebrand: Seasonal Changes in Central England Temperatures
- 2015-29: Laurent Callot and Johannes Tang Kristensen: Regularized Estimation of Structural Instability in Factor Models: The US Macroeconomy and the Great Moderation
- 2015-30: Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Testing for Level Shifts in Fractionally Integrated Processes: a State Space Approach
- 2015-31: Matias D. Cattaneo, Michael Jansson and Whitney K. Newey: Treatment Effects with Many Covariates and Heteroskedasticity
- 2015-32: Jean-Guy Simonato and Lars Stentoft: Which pricing approach for options under GARCH with non-normal innovations?
- 2015-33: Nina Munkholt Jakobsen and Michael Sørensen: Efficient Estimation for Diffusions Sampled at High Frequency Over a Fixed Time Interval
- 2015-34: Wei Wei and Denis Pelletier: A Jump-Diffusion Model with Stochastic Volatility and Durations
- 2015-35: Yunus Emre Ergemen and Carlos Velasco: Estimation of Fractionally Integrated Panels with Fixed Effects and Cross-Section Dependence