Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Panel Data Models with Long-Range Dependence

Autor:

Yunus Emre Ergemen

Director/es:

Prof. Carlos Velasco

DEPARTAMENTO/INSTITUTO DE

ECONOMIA

Getafe, 2015

(a entregar en la Oficina de Posgrado, una vez nombrado el Tribunal evaluador , para preparar el documento para la defensa de la tesis)

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

PANEL DATA MODELS WITH LONG-RANGE DEPENDENCE

Autor: *Yunus Emre Ergemen*

Director/es: Prof. Carlos Velasco

Firma del Tribunal Calificador:

Firma

Presidente: Jesús Gonzalo Muñoz

Vocal: Javier Hualde

Secretario: Mª del Pilar Poncela Blanco

Calificación:

Getafe, de de 2015

Panel Data Models with Long-Range Dependence

Universidad Carlos III de Madrid

Yunus Emre Ergemen

08 May 2015

Abstract

This thesis comprises of three chapters that study panel data models with long-range dependence.

The first chapter is a coauthored paper with Prof. Carlos Velasco. 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-of-squares estimate is $\sqrt{NT}$ consistent but the normal asymptotic distribution might not be centered, requiring the time series dimension to grow faster than the cross-section size for correction. 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.

The second chapter extends the first chapter. In this paper, a general dynamic panel data model is considered that incorporates individual and interactive fixed effects and possibly correlated innovations. The model accommodates general stationary or nonstationary long-range dependence through interactive fixed effects and innovations, removing the necessity to perform a priori unit-root or stationarity testing. Moreover, persistence in innovations and interactive fixed effects allows for cointegration; innovations can also have vector-autoregressive dynamics; deterministic trends can be nested. Estimations are performed using conditional-sum-of-squares criteria based on projected series by which latent characteristics are proxied. Resulting estimates are consistent and asymptotically normal at parametric rates. A simulation study provides reliability on the estimation method. The method is then applied to the long-run relationship between debt and GDP.

The third and final chapter of the thesis is a coauthored paper with Prof. Abderrahim Taamouti. In this paper, a parametric portfolio policy function is considered that incorporates common stock volatility dynamics to optimally determine portfolio weights. Reducing dimension of the traditional portfolio selection problem significantly, only a number of policy parameters corresponding to first- and second-order characteristics are estimated based on a standard method-of-moments technique. The method, allowing for the calculation of portfolio weight and return statistics, is illustrated with an empirical application to 30 U.S. industries to study the economic activity before and after the recent financial crisis.

Acknowledgements

First and foremost, I would like to thank my family who always supported me, and this thesis is dedicated to them. Their support has always been incredible.

I wish to express my sincere gratitude to my supervisor Prof. Carlos Velasco, from whom I learned a great deal, for treating me as a colleague rather than just a student, continuously encouraging me to do better and always believing in me.

I am extremely grateful to Professors Jesús Gonzalo, Juan José Dolado and Abderrahim Taamouti, who have always been very kind to lend a hand when I needed, for being encouraging and supportive.

I want to place on record my sincere thanks to Professors Manuel Arellano, Yoosoon Chang, Miguel Delgado, Niels Haldrup, Javier Hualde, Serena Ng, Bent Nielsen, Peter M. Robinson, Enrique Sentana and the participants in CREATES Seminar 2015, RES Meeting 2015 and NBER-NSF Time Series Conference 2014, the 67^{th} Econometric Society European Meeting, CREATES Symposium on Long Memory 2013, Robust Econometric Methods for Modeling Economic and Financial Variables Conference 2012, UC3M Seminars, III_t, IV_t and V_t Workshop in Time Series Econometrics for helpful comments and discussions that prompted improvements in parts of this thesis.

I also would like to gratefully acknowledge financial support from the Spanish Plan Nacional de I+D+I (ECO2012-31748), Spanish Ministerio de Ciencia e Innovacion grant ECO2010-19357 and Consolider-2010 that made it possible for me to attend conferences and meetings all over the world.

Finally, I would like to thank (in no specific order) Anil Yildizparlak, Fabian Rinnen, Robert Kirkby, Eleonora Garlandi, Lian Allub, Albert Riera, Marta Sanz, Marta Rekas, Pedro H.C. Sant'anna, Nikolaos Tsakas, Xiaojun Song, Lovleen Kushwah, Victor Troster, Christos Mavridis, Sebastian Panthöfer, Mehdi Hamidisahneh, Federico Masera, Andres Garcia-Suaza and Marco Serena for being super fun to hang out with at work, and I am very grateful to Nazli Aktakke, Zeynep Ozkok, Tugba Taskiran, Ceren Genc, Giulia Bonnat, Matthias D'haene, Nicolas Garcia Sanchez, Antoine Mairal, Edu Nogales Corrales, Paula Toledo Piza, Firat Akcal, Georgette Rio Hewison, Andrzej Pioch, Anna Oldinger, Emily Wells, Christopher Kristiansen, Robbie Heim, Jean-Francois Mercier, Alex Barrachina, Egemen Eren and Rutkay Ardogan for constantly reminding me that there is a life outside academia. Life would not have been the same without you guys!

Contents

Chapter 1

Estimation of Fractionally Integrated Panel Data Models with Fixed Effects and Cross-Section Dependence (with Carlos Velasco)

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-of-squares estimate is $\sqrt{NT}$ consistent but the normal asymptotic distribution might not be centered, requiring the time series dimension to grow faster than the cross-section size for correction. 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.

1.1 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 cross-section 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, [29] 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 the unit-root case. For example, [24] extend the study by [29] to panels where observables and factors are integrated $I(1)$ processes while regression errors are $I(0)$. Furthermore, [5] and [3] propose unit-root testing procedures when idiosyncratic shocks and the common factor are both $I(1)$. Similarly, [27] 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. [20] propose a test for the memory parameter under a fractionally integrated panel setup with multiple time series. [39] 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 [20] and [39] 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 cross-section 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 [37] and [38] to panel data models and of [29] 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 time-series asymptotics, $\to_p$ to denote convergence in probability and $\to_d$ to denote convergence in distribution. All mathematical proofs and technical lemmas are collected in the appendix.

1.2 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

$$\lambda_t (L; \theta_0) (y_{it} - \alpha_i - \gamma_i f_t) = \varepsilon_{it}, \tag{1.1}$$

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

$$\lambda_t (L; \theta) = \sum_{j=0}^{t} \lambda_j (\theta) L^j \tag{1.2}$$

truncates $\lambda (L; \theta) = \lambda_\infty (L; \theta)$. We assume that $\lambda (L; \theta)$ has this particular structure,

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

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

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

the coefficients $\psi_j (\xi)$ satisfy

$$\psi_0 (\xi) = 1, \ \ |\psi_j (\xi)| = O (\exp (-c (\xi) j)), \tag{1.3}$$

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

$$\lambda_j(\theta) = \sum_{k=0}^{j} \pi_{j-k}(\delta)\, \psi_k(\xi), \quad j \geq 0, \tag{1.4}$$

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

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

see Robinson and Velasco [39], 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 ($\delta_0 \geq 1/2$) of $y_{it} - \alpha_i - \gamma_i f_t$ and $\psi(L;\xi)$ describes short memory dynamics.

The α_i are unobservable fixed effects, γ_i unobservable factor loadings and f_t is the unobservable common factor that is assumed to be an $I(\varrho)$ process, where we treat ϱ as a nuisance parameter. This way the model incorporates heterogeneity through α_i as well as γ_i and also introduces account cross-section dependence by means of the factor structure, $\gamma_i f_t$, which was not considered in [39]. When we write (1.1) as

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

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

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

which is then easily comparable to a standard dynamic AR(1) panel data model with cross-section dependence, e.g. that of [19],

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

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 (1.1) 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, (1.1) 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 [39], who only require T to grow in the asymptotics, while N could be constant or diverging simultaneously with T. In the first part of the paper we assume a common vector parameter, including a common integration parameter δ, for all cross-section units $i = 1, \ldots, N$. While the fractional integration parameter may as well be allowed to be heterogeneous, our approach is geared towards getting a pooled estimate for the entire panel exploiting potential efficiency gains. Further, this pooling has to control for potential distortions due to common factor elimination, that, as well as fixed effects removal, lead to some bias in the asymptotic distribution of parameter estimates, cf. [39].

We use the following assumptions throughout the paper:

Assumption A.

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

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

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

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

A.5. For $\xi \in \Xi$, $\psi(x; \xi)$ is differentiable in ξ and, for all $\xi \neq \xi_0$, $|\psi(x; \xi)| \neq |\psi(x; \xi_0)|$ on a subset of $\{x : |x| = 1\}$ of positive Lebesgue measure, and (1.3) holds for all $\xi \in \Xi$ with $c(\xi)$ satisfying

$$\inf_{\Xi} c(\xi) = c^* > 0. \tag{1.5}$$

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. [23] and [28]. 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. [29] and [2]. In particular, the condition on $\bar{\gamma}$ is related to Assumption 5(b) of [29] and used to guarantee that our projection to remove factors works in finite samples.

Assumption A.5 ensures that $\psi\left(L;\xi\right)$ is smooth for $\xi \in \Xi$, and the weights ψ_j lead to short-memory dynamics as is also assumed by Robinson and Velasco [39], where the parameter space Ξ can depend on stationarity and invertibility restrictions on $\psi\left(L;\xi\right)$.

1.3 Parameter Estimation

[2] and [29], among many others, study the estimation of panel data models with cross-section dependence. [2] 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, [29] 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

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

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

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

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

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

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

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

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

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

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

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

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

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

because $\lambda_t\left(L; \theta^{(-1)}\right)\Delta = \lambda_t(L; \theta)$, $t = 0, 1, \ldots$. Even when $\theta = \theta_0$, all residuals involve ε_{i0}, i.e. the initial condition, which is reflected in a bias term of $\hat{\theta}$ as in [39].

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. [28] and [23], 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 [39], 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

$$\max\{\varrho, \delta_0\} - 1/2 < \underline{\delta} \leq \delta_0. \tag{1.6}$$

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 [23] and [28] for fractional time series models and in [39] 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 (1.6) holds. This contrasts with the results in [39], where only T was required to grow and N could be fixed or increasing in the asymptotics. An increasing T therein is required to control for the initial condition contribution due to first differencing for fixed effects elimination, as is needed here, but projection on cross-section averages for factor removal further requires that both N and T grow.

Next, we establish the asymptotic distribution of the parameter estimates, for which we assume that $\psi(L;\xi)$ is twice continuously differentiable for all $\xi \in \Xi$ with $\dot{\psi}_t(L;\xi) = (d/d\xi)\psi_t(L;\xi)$ where it is assumed that $\left|\dot{\psi}_t(L;\xi)\right| = O\left(\exp(-c(\xi)j)\right)$. In establishing the asymptotic behaviour, the most delicate part is formulating the asymptotic bias. The initial condition (IC) bias of $(NT)^{1/2}\left(\hat{\theta} - \theta_0\right)$ is proportional to $T^{-1}\nabla_T(\theta_0)$, where

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

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

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

The term $\nabla_T(\theta_0)$, depending only on the unknown θ_0 and T, also found in [39], 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. $\max\{1/4, \varrho - 1/2, \varrho/2 - 1/12\} < \delta_0 \leq \min\{5/4, 5/2 - \varrho\}$.

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

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

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

Corollary 1. *Under Assumptions of Theorem 2,*

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

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 [39] additionally using Assumption C to control for the projection errors and requiring N to grow with T to remove the cross-sectionally averaged error terms, while the range of allowed values of δ_0 is limited in the same way. Assumption C.1 basically requires that T grows faster than N, but slower than N^3, so that different projection errors are not dominating to achieve the $\sqrt{NT}$ rate of convergence. This last restriction is milder than the related conditions that impose $TN^{-2}\to 0$ for slope estimation, e.g. [29], but we also need T to grow faster than N to control the initial condition bias.

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

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

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

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

Corollary 2. *Under Assumptions of Theorem 2,*

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

Bias correction cannot relax the lower bound restriction on the true fractional integration parameter δ_0, but eliminates some further restrictions on N and T though still requires Assumption C.1 which implies the restrictions of Theorem 5.2 of [39] for a similar result in the absence of factors.

1.3.1 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 (1.1) 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 $\boldsymbol{\varepsilon}_i=\left(\varepsilon_{i1},\ldots,\varepsilon_{iT}\right)$, 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 $\bar{\delta}$ in Assumptions A.1, A.5, B and C.2 with δ_{i0}, $\underline{\delta}_i$ and $\bar{\delta}_i$, respectively. We denote these conditions as A_i, B_i and C_i, and assume them to hold for all i.

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

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

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

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

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.

1.4 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 $i = 1, \ldots, N$ and $t = 0, 1, \ldots, T$, the model that generate the observed series y_{it} and X_{it} is given by

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta_{i0}' X_{it} + \gamma_i' f_t + \lambda_t^{-1}(L; \theta_{i0}) \varepsilon_{it}, \\
X_{it} &= \mu_i + \Gamma_i' f_t + e_{it}
\end{aligned}
\tag{1.7}
$$

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 [29] and [2] the possible endogenous covariates are $I(0)$, so they can only address cases in which there is no long-range dependence in the panel. [24] 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 (1.7).

Assumption D

D.1. The idiosyncratic shocks, ε_{it}, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ 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 $\Phi_k^f(s) = \sum_{k=0}^{\infty} \Phi_k^f s^k$, $\sum_{k=0}^{\infty} k \left\| \Phi_k^f \right\| < \infty$, $\det\left(\Phi_k^f(s)\right) \neq 0$ for $|s| \leq 1$ and $v_t^f \sim iid(0, \mathbf{\Omega}_f)$, $\mathbf{\Omega}_f > 0$,

$E\left\|v_t^f\right\|^4 < \infty$, 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 $\Phi_{ik}^e(s) = \sum_{k=0}^{\infty} \Phi_{ik}^e s^k$, $\sup_i \sum_{k=0}^{\infty} k \|\Phi_{ik}^e\| < \infty$, $\det(\Phi_{ik}^e(s)) \neq 0$ for $|s| \leq 1$ and $v_{it}^e \sim iid(0, \boldsymbol{\Omega}_{ie})$, $\boldsymbol{\Omega}_{ie} > 0$, $\sup_{i,t} E\|v_{it}^e\|^4 < \infty$.

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

$$\overline{\mathbf{C}}_N = \begin{pmatrix} \overline{\beta_0'\mathbf{\Gamma'}}_N + \overline{\gamma'}_N \\ \overline{\mathbf{\Gamma'}}_N \end{pmatrix}$$

with $\overline{\gamma}_N = N^{-1}\sum_{i=1}^N \gamma_i$, $\overline{\mathbf{\Gamma}}_N = N^{-1}\sum_{i=1}^N \Gamma_i$, $\overline{\beta_0'\mathbf{\Gamma'}}_N = N^{-1}\sum_{i=1}^N \beta_{i0}'\Gamma_i'$.

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 $k+1 \geq m$. 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. [29] and [24]. 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 $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the first-differences model, including only asymptotically stationary variables, is

$$\begin{aligned} \Delta y_{it} &= \beta_{i0}'\Delta X_{it} + \gamma_i'\Delta f_t + \Delta\lambda_t^{-1}(L;\theta_0)\,\varepsilon_{it}, \\ \Delta X_{it} &= \Gamma_i'\Delta f_t + \Delta e_{it}. \end{aligned} \tag{1.8}$$

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 [5]'s method based on first differencing), because if we write

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

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 [29], 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 (1.8) 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 $\mathbf{y}_i = (y_{i1},\ldots,y_{iT})$, $\mathbf{X}_i = (X_{i1},\ldots,X_{iT})$, $\mathbf{F} = (f_1,\ldots,f_T)$, $\mathbf{E}_i = (e_{i1},\ldots,e_{iT})$ and $\varepsilon_i = (\varepsilon_{i1},\ldots,\varepsilon_{iT})$. We can write down the model in first differences as

$$\begin{aligned}\Delta\mathbf{y}_i &= \beta_{i0}'\Delta\mathbf{X}_i + \gamma_i'\Delta\mathbf{F} + \Delta\lambda_t^{-1}(L;\theta_0)\,\varepsilon_i\\ \Delta\mathbf{X}_i &= \Gamma_i'\Delta\mathbf{F} + \Delta\mathbf{E}_i.\end{aligned}$$

Then, the projection matrix can be denoted by

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

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

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

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 [29], under the rank conditions in Assumptions D.2 and D.4, as $(N,T)_j \to \infty$, we have that

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

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 [29] 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

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

where the following identification condition is satisfied.

Assumption D.5. $\mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{X}_i'$ 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 [5] 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 [5]'s method based on first differencing.

Theorem 4. *Under Assumption D, $\delta^* \geq 1$, 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 $\delta^* \geq 1$,

$$\mathbf{\Sigma}_{ie}(j) = \sum_{k=0}^{\infty} \Phi_{ik}^e (\delta^* - \vartheta_i) \mathbf{\Omega}_{ie} \Phi_{ij+k}^e (\delta^* - \vartheta_i)', \quad j = 0, 1, \ldots,$$

$\mathbf{\Sigma}_{ie}(j) = \mathbf{\Sigma}_{ie}(-j)'$, $j < 0$, 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

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

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

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

$$\vartheta_{max} = \max_i \vartheta_i, \quad \delta_{max} = \max_i \delta_{i0}.$$

Assumption E. $\delta^* > 5/4$, or $\delta^* \geq 1$ 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 $T/N^2 \to 0$ as used by [29] 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

$$\boldsymbol{\Upsilon}_{\beta_i} = \sigma_i^2 \boldsymbol{\Sigma}_{ie}^{-1}(0) \boldsymbol{\Sigma}_{i0} \boldsymbol{\Sigma}_{ie}^{-1}(0).$$

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

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

Note that when $\delta^* = \delta_{i0}$ and $\psi(L;\xi) = 1$, $\boldsymbol{\Upsilon}_{\beta_i} = \sigma_i^2 \boldsymbol{\Sigma}_{ie}^{-1}(0)$, 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 [37] and [36] in terms of projected observations to eliminate factors and an estimate of δ_{i0} or the residual series.

1.4.1 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, $\mathcal{D} = \left[\underline{\delta}, \overline{\delta}\right] \subset (0, 3/2)$, and

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

where

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

and the vectors of observations $\tilde{\mathbf{y}}_i = \mathcal{Y}_i \bar{\mathcal{W}}_T$ and $\tilde{\mathbf{X}}_i = \mathcal{X}_i \bar{\mathcal{W}}_T$ 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 $(N,T)_j \to \infty$,*

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

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,

$$\begin{aligned} \widehat{\mathcal{Y}}_j = \widehat{\mathcal{Y}}_j(\hat{\theta}_i) &= \lambda_{t-1}\left(L; \hat{\theta}_i^{(-1)}\right) \Delta \mathbf{y}_j \\ \widehat{\mathcal{X}}_j = \widehat{\mathcal{X}}_j(\hat{\theta}_i) &= \lambda_{t-1}\left(L; \hat{\theta}_i^{(-1)}\right) \Delta \mathbf{X}_j \end{aligned}$$

for $j = 1, \ldots, N$, 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

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

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

Let

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

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 $\bar{\Phi}_{ik}^e = \sum_{j=0}^{k} \Phi_{ik-j}^e \lambda_j \left(\delta_{i0} - \vartheta_i, \xi_{i0}\right)$ incorporate the prewhitening effect.

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

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

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

1.4.2 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

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

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 $\left(T^{-1}\mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{X}_i'\right)^{-1}$ having finite second order moments for all i=1, …, 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 [29] and [24] for $I\left(0\right)$ and $I\left(1\right)$ 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 [29], 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

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

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

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

1.5 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.

1.5.1 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 $\varrho = 0.4, 1$. As is clear in Table 2.1, when the factors are less persistent ($\varrho = 0.4$), the estimate is heavily biased for the stationary case of $\delta_0 = 0.3$ while 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 $\varrho = 0.4$ although benefits are limited for $\varrho = 1$. While there is a monotonically decreasing pattern for increasing δ_0 in terms of bias both for the total bias and bias-corrected cases, magnitudes of biases increase when δ_0 leaves the neighbourhood of unity.

Table 2.1 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 2.2 shows the empirical coverage of 95% confidence intervals of δ_0 based on the asymptotics of our estimate. For $\varrho = 0.4, 1$, the true fractional parameter is poorly covered when $\delta_0 \leq 0.6$. Bias correction in these cases improves the results reasonably. For near-unit-root cases, the estimate achieves the most accurate coverage, especially by comparison with intervals based on estimates of $\delta_0 = 1.4$ and $\delta_0 \leq 0.6$.

1.5.2 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 2.3 and 2.4 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 $\varrho = 1$. 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 non-cointegration 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.

1.6 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-value-weighted 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

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

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

$$RVI_{i,t} = \left(\sum_{s\in t}^{N_t} e_{i,s}^2\right)^{1/2}, \qquad t = 1, 2, \ldots, T,$$

where $e_{i,s} = r_{i,s} - r_{m,s}$, cf. [10]. 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: book-to-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, [35], 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 1.5 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 $\varrho = 0.71, 0.66$ 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 (β_{i0}^0 and β_{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 1.6 presents the fractional order of residuals, from which the cointegrating relationships are confirmed based on the results presented in Table 1.5.

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 1.7, 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.

1.7 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 [39]; (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 [15] 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.

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

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

1.8.1 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

$$\hat{\phi}_i = \frac{\sum_{t=1}^{T} \Delta \bar{y}_t \Delta y_{it}}{\sum_{t=1}^{T} (\Delta \bar{y}_t)^2} = \frac{\gamma_i}{\bar{\gamma}} + \eta_i \tag{1.9}$$

where

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

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

$$L_{N,T}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\lambda_t^0 (L; \theta) \left(\varepsilon_{it} - \hat{\phi}_i \bar{\varepsilon}_t \right) - \tau_t(\theta)(\varepsilon_{i0} - \hat{\phi}_i \bar{\varepsilon}_0) - \eta_i \bar{\gamma} \lambda_{t-1} (L; \theta) f_t \right)^2 \tag{1.10}$$

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 (1.10) 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 [23] 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 [23]. For $\epsilon > 0$, define $Q_\epsilon = \{\theta : |\delta - \delta_0| < \epsilon\}$, $\overline{Q}_\epsilon = \{\theta : \theta \notin Q_\epsilon, \delta \in \mathcal{D}\}$. For small enough ϵ,

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

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

$$Pr\left(\inf_{\theta \in \overline{Q}_\epsilon \cap \Theta_k} S_{N,T}(\delta) \leq 0 \right) \to 0 \text{ as } (N,T)_j \to \infty, \; k = 1, \ldots, 4. \tag{1.11}$$

We write $L_{N,T}(\theta)$ in (1.10) as

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

The projection error component in the conditional sum of squares,

$$\sup_{\theta\in\Theta}\left|\bar{\gamma}^2\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{1}{T}\sum_{t=1}^{T}\left(\lambda_{t-1}\left(L;\theta\right)f_t\right)^2\right|, \tag{1.12}$$

is $O_p(T^{2\varrho+2\delta_0-6}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2})+O_p(T^{4\varrho+2(\delta_0-\underline{\delta})-7}+T^{2(\varrho-\underline{\delta}-1)}\log T$ $+N^{-1}T^{2(\varrho-\underline{\delta})+4\delta_0-7}+T^{2(\varrho-\underline{\delta})-1}N^{-2})=o_p\left(1\right)$ uniformly in $\theta\in\Theta$ by $\bar{\gamma}^2\to_p E\left[\gamma_i\right]^2$, 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,

$$\sup_{\theta\in\Theta}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t^2(\theta)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0)^2\right|=o_p(1), \tag{1.13}$$

because

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

uniformly in $\theta\in\Theta$ with $\underline{\delta}>0$, using $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}^2+\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2=O_p\left(1\right),\ \bar{\varepsilon}_0=O_p\left(N^{-1/2}\right)$ and Cauchy-Schwarz inequality, see Lemma 1, and therefore we find for the cross term corresponding to the sum of squares in (1.12) and (1.13)

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

uniformly in δ by (1.12), (1.13) and Cauchy-Schwarz inequality.

The other cross terms involving usual fractional residuals $\lambda_t^0\left(L;\theta\right)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right)$ are also uniformly $o_p\left(1\right)$ 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 $o_p\left(1\right)$ for $\theta\in\Theta_1\cup\Theta_2\cup\Theta_3$ under the assumptions of the theorem. Then to show (1.11) we only need to analyze the terms in $(\lambda_t^0\left(L;\theta\right)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right))^2$ for Θ_k, $k=1,\ldots,4$ as in [23].

Proof for $k = 4$. We show that

$$\sup_{\theta\in\Theta_4}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(\lambda_t^0(L;\theta)(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t))^2-\sigma^2\sum_{j=0}^{\infty}\lambda_j^0(\theta)^2\right]\right| = o_p(1), \tag{1.14}$$

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 [23],

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

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 $\theta\neq\theta_0$, (1.11) holds for $k=4$ if (1.14) holds and the contribution of cross-section averaged term, $\hat{\phi}_i\bar{\varepsilon}_t$, is negligible.

To check (1.14) 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

$$\begin{aligned}&\frac{1}{T}\sum_{j=0}^{T}\lambda_j^0(\theta)^2\frac{1}{N}\sum_{i=1}^{N}\sum_{l=0}^{T-j}(\varepsilon_{il}^2-\sigma^2)\\ &+\frac{2}{T}\sum_{j=0}^{T-1}\lambda_j^0(\theta)\lambda_k^0(\theta)\frac{1}{N}\sum_{i=1}^{N}\sum_{l=k-j+1}^{T-j}\varepsilon_{il}\varepsilon_{il-(k-j)} = (a)+(b).\end{aligned} \tag{1.15}$$

Then,

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

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

$$\begin{aligned}
& \frac{2\lambda_{T-1}^{0}(\theta)}{T}\sum_{j=0}^{T-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=j+1}^{T}\sum_{l=k-j+1}^{T-j}\lambda_{j}^{0}(\theta)\,\varepsilon_{il}\varepsilon_{il-(k-j)} \\
& -\frac{2}{T}\sum_{j=0}^{T-1}\lambda_{j}^{0}(\theta)\sum_{k=j+1}^{T}\left[\lambda_{k+1}^{0}(\theta)-\lambda_{k}^{0}(\theta)\right]\frac{1}{N}\sum_{i=1}^{N}\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{T-j}\varepsilon_{il}\varepsilon_{il-(r-j)} \\
= \;& (b_1)+(b_2).
\end{aligned}$$

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

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

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 (1.14). We write

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\phi}_i^2\left(\lambda_t^0(L;\theta)\,\bar{\varepsilon}_t\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2\frac{1}{T}\sum_{t=1}^{T}\left(\lambda_t^0(L;\theta)\,\bar{\varepsilon}_t\right)^2. \tag{1.16}$$

The average in i is $O_p(1)$ by Lemma 1, while the sum in t in the lhs (1.16) satisfies for θ^* with first component $\theta^*_{(1)} = \zeta - \frac{1}{2}$,

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

as $N\to\infty$, uniformly in $\theta\in\Theta_4$ as $T\to\infty$, and (1.16) 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 (1.14) is asymptotically negligible by Cauchy-Schwarz inequality. So we have proved (1.14), and therefore we have proved (1.11) for $k=4$.

Proof for $k = 3, 2$. The uniform convergence for the idiosyncratic component for the proof of (1.11) follows as in [23], since the average in $i = 1, \ldots, N$ adds no additional complication as in the case $k = 4$. The treatment for the cross-section averaged term and the cross-product term follows from the same steps as the idiosyncratic term as well as the results we derived for $k = 4$ using $\frac{1}{N}\sum_{i=1}^{N} \hat{\phi}_i^2 = O_p(1)$ and that $\bar{\varepsilon}_t$ has variance σ^2/N.

Proof for $k = 1$. Noting that

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

we write

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

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

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

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

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 [25], for each $i = 1, \ldots, N$,

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

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 [21] 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 double-index processes in [32] and $\frac{1}{N}\sum_{i=1}^{N}\phi_i^2 =$

$O_p(1)$,

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

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

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

so that

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

and (1.11) follows for $i=1$ as ϵ is arbitrarily small. □

1.8.2 Other Proofs in Section 3

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

Assumption C*.

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

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

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

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

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 $\theta=\theta_0$,

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

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

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

$$
\begin{aligned}
\sqrt{NT}\frac{\partial}{\partial\theta}L_{N,T}(\theta)|_{\theta=\theta_0} &= \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i^2\bar{\gamma}^2\lambda_{t-1}(L;\theta_0)f_t * \chi_{t-1}(L;\xi_0)\lambda_{t-1}(L;\theta_0)f_t && (1.17)\\
&+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)\dot{\tau}_t(\theta_0)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0)^2 && (1.18)\\
&+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\bar{\gamma}\lambda_{t-1}(L;\theta_0)f_t * \dot{\tau}_t(\theta_0)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0) && (1.19)\\
&-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\bar{\gamma}\lambda_{t-1}(L;\theta_0)f_t * \chi_t(L;\xi_0)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right) && (1.20)\\
&+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\bar{\gamma}\chi_{t-1}(L;\xi_0)\lambda_{t-1}(L;\theta_0)f_t * \tau_t(\theta_0)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0) && (1.21)\\
&-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_t(\theta_0)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0) * \chi_t(L;\xi_0)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right) && (1.22)\\
&-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\bar{\gamma}\chi_{t-1}(L;\xi_0)\lambda_{t-1}(L;\theta_0)f_t * (\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t) && (1.23)\\
&-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tau}_t(\theta_0)(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0)(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t) && (1.24)\\
&+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t) * \chi_t(L;\xi_0)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right). && (1.25)
\end{aligned}
$$

The term (1.17) is asymptotically negligible, since with Lemmas 1 and 2 and $\varrho-\delta_0<\frac{1}{2}$, we find that

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

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

In (1.18), 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

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

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

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

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

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

if $\delta_0 > 1/4$.

We show that (1.19) is $o_p(1)$ considering the contribution of

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

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

$$\begin{aligned}&2\sqrt{NT}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}^2\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\right)^{1/2}\left|\frac{1}{T}\sum_{t=1i}^{T}\lambda_{t-1}(L;\theta_0)\,f_t\dot{\tau}_t(\theta_0)\right|\\ =\ &O_p\left((NT)^{1/2}\left(T^{2(\varrho+\delta_0-3)}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2}\right)^{1/2}T^{-1}\right)\\ &+O_p\left((NT)^{1/2}\left(T^{2(\varrho+\delta_0-3)}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2}\right)^{1/2}\left\{T^{\varrho-2\delta_0-1/2}+T^{-\delta_0/2-1/2}\right\}\log T\right)\\ =\ &O_p\left(N^{1/2}\left(T^{2(\varrho+\delta_0-3)}+T^{-1}\log T+N^{-2}\right)^{1/2}T^{\varrho-2\delta_0}\log T\right)\\ &+O_p\left(N^{1/2}T^{\varrho+\delta_0-3}T^{-\delta_0/2}\log T\right)+o_p(1)\end{aligned}$$

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

For (1.20), we consider the contribution of

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

whose absolute value is bounded by

$$
\begin{aligned}
& 2\sqrt{NT}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}\left(L;\theta_0\right)f_t*\chi_t\left(L;\xi_0\right)\varepsilon_{it}\right)^2\right)^{1/2} \\
= \; & O_p\left(\left(NT\right)\left(T^{2\varrho+2\delta_0-6}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2}\right)T^{-1}\right)^{1/2} \\
= \; & O_p\left(N\left(T^{2\varrho+2\delta_0-6}+T^{-1}\log T+N^{-1}T^{4\varrho-6}\log T+N^{-2}\right)\right)^{1/2}=o_p\left(1\right)
\end{aligned}
$$

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

$$
\begin{aligned}
& E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}\left(L;\theta_0\right)f_t*\chi_t\left(L;\xi_0\right)\varepsilon_{it}\right)^2\right] \\
= \; & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\lambda_{t-1}\left(L;\theta_0\right)f_t*\chi_t\left(L;\xi_0\right)\varepsilon_{it}*\lambda_{r-1}\left(L;\theta_0\right)f_r*\chi_r\left(L;\xi_0\right)\varepsilon_{ir}\right] \\
= \; & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\lambda_{t-1}\left(L;\theta_0\right)f_t*\lambda_{r-1}\left(L;\theta_0\right)f_r\right]E\left[\chi_t\left(L;\xi_0\right)\varepsilon_{it}*\chi_r\left(L;\xi_0\right)\varepsilon_{ir}\right] \\
= \; & O\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{t}\left|t-r\right|^{2(\varrho-\delta_0)-2}\log t\right) \\
= \; & O\left(T^{-1}+T^{2(\varrho-\delta_0-1)}\log T\right)=O\left(T^{-1}\right).
\end{aligned}
$$

Then the term (1.20) is $o_p\left(1\right)$ because the factor depending on $\hat{\phi}_i\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t$ could be dealt with similarly using Cauchy-Schwarz inequality and Lemma 1.

The proof that the term (1.21) is $o_p\left(1\right)$ could be dealt with exactly as when bounding (1.19), while the proof that the term (1.23) is $o_p\left(1\right)$ could be dealt with in a similar but easier way than (1.20).

The leading term of (1.24), 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 (1.22) is given in Lemma 5 and that of (1.25) in Lemma 6 and, combining the plims of (1.18) and (1.22), we obtain the definition of $\nabla_T(\delta)$.

Then collecting the results for all terms (1.17) to (1.25) 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 $\theta=\theta_0$, $\left(\partial^2/\partial\theta\partial\theta'\right)L_{N,T}(\theta)|_{\theta=\theta_0}$, which equals

$$\begin{aligned}
&\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\eta_i\bar{\gamma}\chi_{t-1}\left(L;\xi_0\right)\lambda_{t-1}\left(L;\theta_0\right)f_t-\dot{\tau}_t(\theta_0)\left(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0\right)+\chi_t\left(L;\xi_0\right)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right)\right\}\\
&\quad\times\left\{-\eta_i\bar{\gamma}\chi_{t-1}\left(L;\xi_0\right)\lambda_{t-1}\left(L;\theta_0\right)f_t-\dot{\tau}_t(\theta_0)\left(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0\right)+\chi_t\left(L;\xi_0\right)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right)\right\}'\\
&+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\eta_i\bar{\gamma}\lambda_{t-1}\left(L;\theta_0\right)f_t-\tau_t(\theta_0)\left(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0\right)+\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right\}\\
&\quad\times\left\{-\eta_i\bar{\gamma}b_t^0\left(L\right)\lambda_{t-1}\left(L;\theta_0\right)f_t-\ddot{\tau}_t(\theta_0)\left(\varepsilon_{i0}-\hat{\phi}_i\bar{\varepsilon}_0\right)+b_t^0\left(L\right)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right)\right\},
\end{aligned}$$

where $b_t^0(L)=\dot{\chi}_t\left(L;\xi_0\right)+\chi_t\left(L;\xi_0\right)\chi_t\left(L;\xi_0\right)'$, $\dot{\chi}_t\left(L;\xi\right)=\left(\partial/\partial\theta'\right)\chi_t\left(L;\xi\right)$ and $\ddot{\tau}_t(\theta)=\left(\partial^2/\partial\theta\partial\theta'\right)\tau_t\left(\theta\right)$. Using the same techniques as in the proof of Theorem 1, as N and T get larger, only the term on $\chi_t\left(L;\xi_0\right)\varepsilon_{it}\chi_t\left(L;\xi_0\right)'\varepsilon_{it}$ in the first element of the rhs contributes to the probability limit, see the proof of Theorem 5.2 in [39]. 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. □

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 [39].
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 $\delta_{i0}\in\mathcal{D}$. □

1.8.3 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

$$\begin{aligned} M_i &= \left(\mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{X}_i'\right)^{-1} \mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{F}' \gamma_i \\ U_i &= \left(\mathcal{X}_i \bar{\mathcal{W}}_T \mathcal{X}_i'\right)^{-1} \mathcal{X}_i \bar{\mathcal{W}}_T (\Delta^{\delta^*-1} \Delta \lambda^{-1}(L; \theta_{i0})\, \varepsilon_i))' \end{aligned}$$

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 $U_{1,i}$,

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

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 $\bar{\mathcal{W}}_T$. 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,

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

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 [37] 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

$$\bar{\mathbf{H}}' = \mathcal{F}' \bar{\mathbf{C}} + \mathcal{X}_i \bar{\mathbf{V}}$$

where, $\mathbf{\Pi}_T^* = (\pi_1(\delta^* - 1), \ldots, \pi_T(\delta^* - 1))$,

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

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 [29] 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 $\Sigma_f=\Sigma_f(\delta^*-\varrho)=\sum_{k=0}^{\infty}\Phi_k^f(\delta^*-\varrho)\,\mathbf{\Omega}_f\Phi_k^f(\delta^*-\varrho)'$, where the weights $\Phi_k^f(\delta^*-\varrho)$ are square summable with $\delta^*\geq 1$ 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

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\bar{v}_t'\bar{v}_t &= \frac{1}{T}\sum_{t=1}^{T}\Big\{\left(\overline{\Delta_t^{\delta^*}\lambda_t^{-1}(L;\theta_0)\,\varepsilon_t}\right)^2+\left(\overline{\pi_t(\delta^*-1)\varepsilon_0}\right)^2+\left(\overline{\beta'\Delta_t^{\delta^*-\vartheta_0}e_t}\right)^2\\
&+\left(\overline{\Delta_t^{\delta^*-\vartheta_0}e_t}\right)^2+2\overline{\Delta_t^{\delta^*}\lambda_t^{-1}(L;\theta_0)\,\varepsilon_t}\,\overline{\pi_t(\delta^*-1)\varepsilon_0}\\
&+2\overline{\Delta^{\delta^*}\lambda_t^{-1}(L;\theta_0)\,\varepsilon_t}\,\overline{\Delta_t^{\delta^*-\vartheta_0}e_t}+2\overline{\pi_t(\delta^*-1)\varepsilon_0}\,\overline{\Delta_t^{\delta^*-\vartheta_0}e_t}\Big\}
\end{aligned}
$$

whose expectation is $O(N^{-1})$, and its variance is proportional to $O\left((TN)^{-1}\right)$. 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 $T^{-1}\bar{\mathbf{V}}'\mathcal{F}'=T^{-1}\sum_{t=1}^{T}\bar{v}_tf_t=O_p\left((NT)^{-1/2}\right)$ since it has zero expectation and using the independence of ε_{it} and f_t, its variance is

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

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

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

(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 $O\left(N^{-1}\right)$ 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 $O\left(T^{-1}\right)$.

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

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

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

$$\frac{1}{T}\sum_{t=1}^{T}\bar{v}_t'\bar{v}_t = \frac{1}{T}\sum_{t=1}^{T}\left\{\left(\overline{\Delta\lambda_t^{-1}(L;\theta_0)\,\varepsilon_t}\right)^2+\left(\overline{\beta'\Delta_t^{1-\vartheta_0}e_t}\right)^2+\left(\overline{\Delta_t^{1-\vartheta_0}e_t}\right)^2+2\overline{\Delta\lambda_t^{-1}(L;\theta_0)\,\varepsilon_t}\,\overline{\Delta_t^{1-\vartheta_0}e_t}\right\}$$

whose expectation is $O\left(N^{-1}\right)$ 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 $O\left(N^{-2}\right)$ 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

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

so 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 $O\left(N^{-1}\right)$ times

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

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

Thus the entire projection error is

$$M_i=O_p\left(N^{-1}+N^{-1/2}\left\{T^{-1/2}+T^{\delta_{max}+\vartheta_{max}-3}+T^{2\vartheta_{max}-3}+T^{\varrho+\delta_{max}-3}+T^{\varrho+\vartheta_{max}-3}\right\}\right)=o_p(1)$$

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

$$T^{1/2}M_i=O_p\left(T^{1/2}N^{-1}+N^{-1/2}\left\{1+T^{\delta_{max}+\vartheta_{max}-5/2}+T^{2\vartheta_{max}-5/2}+T^{\varrho+\delta_{max}-5/2}+T^{\varrho+\vartheta_{max}-5/2}\right\}\right).$$

Therefore, if $\vartheta_{max}<11/8$ and $\varrho+\delta_{max},\varrho+\vartheta_{max},\delta_{max}+\vartheta_{max}<11/4$, $T^{1/2}M_i=o_p(1)$ as $(N,T)_j\to\infty$ when $\delta^*=1$ since $T^{1/2}N^{-1}=o(1)$ and $N^{1/2}=o\left(T^{-1/4}\right)$.

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

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

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

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 Δ^{δ^*} implicit in $\bar{\mathbf{H}}$ yields,

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

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

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

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

$$\begin{aligned}P_{y,i}\left(\theta\right) &= \lambda\left(L;\delta-1,\xi\right)\Delta\mathbf{y}_{i}-\Delta^{\delta^{*}-1}\Delta\mathbf{y}_{i}\mathcal{F}'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathbf{F}(\theta)\\ R_{y,i}\left(\theta\right) &= \Delta^{\delta^{*}-1}\Delta\mathbf{y}_{i}\left\{\mathcal{F}'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathbf{F}(\theta)-\bar{\mathbf{H}}'\left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-}\bar{\mathbf{H}}(\theta)\right\}\end{aligned}$$

with $\mathbf{F}(\theta)=\lambda\left(L;\delta-\delta^{*},\xi\right)\mathcal{F}=\psi\left(L;\xi\right)\Delta^{\delta}\mathbf{F}$, and similarly $\lambda\left(L;\delta-\delta^{*},\xi\right)\tilde{\mathbf{X}}_{i}\left(\delta^{*}\right)=P_{x,i}\left(\theta\right)+R_{x,i}\left(\theta\right)$ 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 $\lambda^{0}\left(L;\theta\right)=\lambda\left(L;\theta\right)\lambda^{-1}\left(L;\theta_{i0}\right)$,

$$\begin{aligned}P_{y,i}\left(\theta\right) &= \lambda^{0}\left(L;\theta\right)\varepsilon_{i}+\beta_{i0}'\psi\left(L;\xi\right)\Delta^{\delta-\vartheta_{i}}\mathbf{E}_{i}-\varsigma_{T}(\theta)\varepsilon_{i0}\\ &\quad-\left[\lambda^{-1}\left(L;\delta_{i0}-\delta^{*},\xi_{0}\right)\varepsilon_{i}+\beta_{i0}'\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}-\mathbf{\Pi}_{T}^{*}\varepsilon_{i0}\right]\mathcal{F}'\left(\mathcal{F}\mathcal{F}'\right)^{-}\mathbf{F}(\theta),\end{aligned}$$

with $\varsigma_{T}(\theta)=\left(\tau_{1}\left(\theta\right),\ldots,\tau_{T}\left(\theta\right)\right)$ and

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

Also,

$$\begin{aligned}R_{y,i}\left(\theta\right) &= \left[\lambda^{-1}\left(L;\delta_{i0}-\delta^{*}\right)\varepsilon_{i}+\beta_{i0}'\Delta^{\delta^{*}-\vartheta_{i}}\mathbf{E}_{i}+\left(\beta_{i0}'\mathbf{\Gamma}_{i}'+\gamma_{i}'\right)\mathcal{F}-\mathbf{\Pi}_{T}^{*}\varepsilon_{i0}\right]\\ &\quad\times\left[\mathcal{F}'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathbf{F}(\theta)-\bar{\mathbf{H}}'\left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-}\bar{\mathbf{H}}(\theta)\right],\end{aligned}$$

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

Therefore

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

where

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

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

$$\begin{aligned} \varepsilon_i^{(1)}(\theta) &= \lambda^0 (L;\theta)\, \varepsilon_i - \varsigma_T(\theta)\varepsilon_{i0} - \lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i W_f(\theta) - \mathbf{\Pi}^*_T \varepsilon_{i0} W_f(\theta) \\ \varepsilon_i^{(2)}(\theta) &= -\left(\beta_{i0} - \hat{\beta}_i(\delta^*)\right)' \left[\psi(L;\xi)\,\Delta^{\delta-\vartheta_i}\mathbf{E}_i - \Delta^{\delta^*-\vartheta_i}\mathbf{E}_i W_f(\theta)\right] \\ \varepsilon_i^{(3)}(\theta) &= \left[\left(\left(\beta_{i0} - \hat{\beta}_i(\delta^*)\right)' \mathbf{\Gamma}_i' + \gamma_i'\right)\mathcal{F} + \left(\beta_{i0} - \hat{\beta}_i(\delta^*)\right)' \Delta^{\delta^*-\vartheta_i}\mathbf{E}_i + \lambda^{-1}(L;\delta_{i0}-\delta^*)\,\varepsilon_i - \mathbf{\Pi}^*_T\varepsilon_{i0}\right] \\ &\quad \times \left(W_f(\theta) - W_h(\theta)\right). \end{aligned}$$

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

$$\begin{aligned} &\frac{1}{T}\left(\lambda^0(L;\theta)\,\varepsilon_i - \varsigma_T(\theta)\varepsilon_{i0}\right)\left(\lambda^0(L;\theta)\,\varepsilon_i - \varsigma_T(\theta)\varepsilon_{i0}\right)' \\ &+\frac{1}{T}\left(\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i W_f(\theta) - \mathbf{\Pi}^*_T\varepsilon_{i0}W_f(\theta)\right)\left(\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i W_f(\theta) - \mathbf{\Pi}^*_T\varepsilon_{i0}W_f(\theta)\right)' \\ &-\frac{2}{T}\left(\lambda^0(L;\theta)\,\varepsilon_i - \varsigma_T(\theta)\varepsilon_{i0}\right)\left(\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i W_f(\theta) - \mathbf{\Pi}^*_T\varepsilon_{i0}W_f(\theta)\right)'. \end{aligned}$$

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

$$\begin{aligned} &\frac{1}{T}\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i W_f(\theta)W_f(\theta)'\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i' \qquad (1.26) \\ =\ &\frac{1}{T}\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i\mathcal{F}'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathbf{F}(\theta)\mathbf{F}(\theta)'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathcal{F}\lambda^{-1}(L;\delta_{i0}-\delta^*,\xi_0)\,\varepsilon_i' \end{aligned}$$

where

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

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

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

we obtain that (1.26) 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

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

First,

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

because $\beta_{i0}-\hat{\beta}_i(\delta^*) = O_p\left(T^{-1/2}\right)$ by Theorem 5 and with $\vartheta_i - \underline{\delta} < 1$,

$$\sup_{\theta\in\Theta}\left|\frac{1}{T^2}\psi(L;\xi)\Delta^{\delta-\vartheta_i}\mathbf{E}_i\psi(L;\xi)\Delta^{\delta-\vartheta_i}\mathbf{E}_i'\right| = O\left(T^{-1}+T^{2(\vartheta_i-\underline{\delta}-1)}\right) = o_p(1).$$

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 $T^{-1}\varepsilon_i^{(3)}(\theta)\varepsilon_i^{(3)}(\theta)'$

$$\begin{aligned}\frac{1}{T}\quad&\left[\left(\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i+\lambda^{-1}(L;\delta_{i0}-\delta^*)\varepsilon_i-\mathbf{\Pi}_T^*\varepsilon_{i0}\right]\\ &\times\left(W_f(\theta)-W_h(\theta)\right)\left(W_f(\theta)-W_h(\theta)\right)'\\ &\times\left[\left(\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i+\lambda^{-1}(L;\delta_{i0}-\delta^*)\varepsilon_i-\mathbf{\Pi}_T^*\varepsilon_{i0}\right]'.\end{aligned}$$

First,

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\left(\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)\mathcal{F}\left(W_f(\theta)-(\theta)W_h\right)\left(W_f(\theta)-W_h(\theta)\right)'\mathcal{F}'\left(\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)'\right|$$

is $o_p(1)$ because

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

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,

$$\begin{aligned}
&\sqrt{T}\frac{\partial}{\partial\theta}L^*_{i,T}(\theta)\Big|_{\theta=\theta_{i0}} \\
= &\frac{2}{\sqrt{T}}\Big\{\left(\varepsilon_i-\varsigma_T(\theta_{i0})\varepsilon_{i0}-\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_iW_f\left(\theta_{i0}\right)+\mathbf{\Pi}^*_T\varepsilon_{i0}W_f\left(\theta_{i0}\right)\right) \\
&-\left(\beta_{i0}-\hat{\beta}_i\left(\delta^*\right)\right)'\left[\psi\left(L;\xi_{i0}\right)\Delta^{\delta_{i0}-\vartheta_i}\mathbf{E}_i-\Delta^{\delta^*-\vartheta_i}\mathbf{E}_iW_f\left(\theta_{i0}\right)\right] \\
&+\left[\left(\left(\beta_{i0}-\hat{\beta}_i\left(\delta^*\right)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_i\left(\delta^*\right)\right)'\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i+\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i-\mathbf{\Pi}^*_T\varepsilon_{i0}\right] \\
&\times\left(W_f(\theta_{i0})-W_h(\theta_{i0})\right)\Big\} \\
&\times\Big\{\left(\chi\left(L;\xi_{i0}\right)\varepsilon_i-\dot{\varsigma}_T(\theta_{i0})\varepsilon_{i0}-\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i\dot{W}_f\left(\theta_{i0}\right)+\mathbf{\Pi}^*_T\varepsilon_{i0}\dot{W}_f\left(\theta_{i0}\right)\right) \\
&-\left(\beta_{i0}-\hat{\beta}_i\left(\delta^*\right)\right)'\left[\chi\left(L;\xi_{i0}\right)\psi\left(L;\xi_{i0}\right)\Delta^{\delta_{i0}-\vartheta_i}\mathbf{E}_i-\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i\dot{W}_f\left(\theta_{i0}\right)\right] \\
&+\left[\left(\left(\beta_{i0}-\hat{\beta}_i\left(\delta^*\right)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_i\left(\delta^*\right)\right)\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i+\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i-\mathbf{\Pi}^*_T\varepsilon_{i0}\right] \\
&\times\left(\dot{W}_f\left(\theta_{i0}\right)-\dot{W}_h\left(\theta_{i0}\right)\right)\Big\}',
\end{aligned}$$

where

$$\begin{aligned}
\dot{W}_f\left(\theta_{i0}\right) &: = \mathcal{F}'\left(\mathcal{F}\mathcal{F}'\right)^{-}\dot{\mathbf{F}}(\theta_{i0}), \\
\dot{W}_h\left(\theta_{i0}\right) &: = \bar{\mathbf{H}}'\left(\bar{\mathbf{H}}\bar{\mathbf{H}}'\right)^{-}\dot{\bar{\mathbf{H}}}(\theta_{i0})
\end{aligned}$$

and $\dot{\mathbf{F}}(\theta) = (\partial/\partial\theta)\,\mathbf{F}(\theta)$, $\dot{\bar{\mathbf{H}}}(\theta) = (\partial/\partial\theta)\,\bar{\mathbf{H}}(\theta)$. 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

$$\frac{2}{\sqrt{T}}\left[\varepsilon_i - \varsigma_T(\theta_{i0})\varepsilon_{i0}\right]\left[\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i\dot{W}_f\left(\theta_{i0}\right) - \mathbf{\Pi}_T^*\varepsilon_{i0}\dot{W}_f\left(\theta_{i0}\right)\right]'. \tag{1.27}$$

Then,

$$\frac{1}{\sqrt{T}}\varepsilon_i\dot{W}_f\left(\theta_{i0}\right)'\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i' = \frac{1}{\sqrt{T}}\varepsilon_i\dot{\mathbf{F}}(\theta_{i0})'\left(\mathcal{F}\mathcal{F}'\right)^{-1}\mathcal{F}\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i' = o_p\left(1\right)$$

because $\rho - \delta_{i0} < 1/2$ so that $T^{-1}\mathcal{F}\mathcal{F}' \to_p \Sigma_f > 0$,

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

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

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

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

$$\frac{2}{\sqrt{T}}\varsigma_T(\theta_{i0})\dot{W}_f\left(\theta_{i0}\right)\mathbf{\Pi}_T^{*\prime}\varepsilon_{i0}^2 = O_p\left(T^{-1/2}\left(T^{-1}+T^{-1/2-\delta_{i0}/2}\right)\log T\right) = o_p(1),$$

and we can conclude that (1.27) is $o_p\left(1\right)$.

(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,

$$\frac{2}{\sqrt{T}}\left[\varepsilon_i - \varsigma_T(\theta_{i0})\varepsilon_{i0}\right]\left\{\left[\left(\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\mathbf{\Gamma}_i'+\gamma_i'\right)\mathcal{F}+\left(\beta_{i0}-\hat{\beta}_i(\delta^*)\right)'\Delta^{\delta^*-\vartheta_i}\mathbf{E}_i+\lambda^{-1}\left(L;\delta_{i0}-\delta^*,\xi_{i0}\right)\varepsilon_i-\mathbf{\Pi}_T^*\varepsilon_{i0}\right]\right.$$
$$\left.\times\left(\dot{W}_f\left(\theta_{i0}\right)-\dot{W}_h\left(\theta_{i0}\right)\right)\right\}',$$

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 [38] 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 $(N,T)_j \to \infty$. □

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. □

1.9 Lemmas

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

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\eta_i^2 &= O_p(T^{2\varrho+2\delta_0-6}+T^{-1}\log T+N^{-1}T^{4\delta_0-6}+N^{-2})\\ \frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2 &= O_p(1).\end{aligned}$$

Proof of Lemma 1. We only prove the first statement, since the second one is an easy consequence of the first one, (1.9) and $\bar{\gamma}^2 \to_p (E[\gamma_i])^2 > 0$ and $E[\gamma_i^2] < \infty$. Write

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

The denominator converges to a positive constant term because

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

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

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

since $\varrho < 2/3$ and the second 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,

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta\bar{y}_t\Delta\bar{y}_{t'}\sum_{i=1}^{N}\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'} \qquad (1.28)\\
=\;&\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta f_t\Delta f_{t'}\sum_{i=1}^{N}\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\\
&+\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\lambda_t\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_t\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_{t'}\sum_{i=1}^{N}\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\\
&+\frac{2}{NT^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Delta f_t\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_{t'}\sum_{i=1}^{N}\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}.
\end{aligned}$$

The expectation of the first term in (1.28), 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,

$$\begin{aligned}
E\left(\Delta f_t\Delta f_{t'}\right) &= O\left(|t-t'|_+^{2(\varrho-1)-1} + |t-t'|_+^{\varrho-2}\right)\\
E\left(\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right) &= O\left(|t-t'|_+^{2(\delta_0-1)-1} + |t-t'|_+^{\delta_0-2}\right)
\end{aligned}$$

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 (1.28), which is also positive, is

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left[\lambda_t\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_t\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_{t'}\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right]\\
=\;&\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}E\left[\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{jt}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{kt'}\lambda_t\left(L;\theta_0^{(-1)}\right)\varepsilon_{it}\lambda_{t'}\left(L;\theta_0^{(-1)}\right)\varepsilon_{it'}\right]\\
=\;&\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{a=1}^{t}\sum_{b=1}^{t'}\sum_{c=1}^{t}\sum_{d=1}^{t'}\tau_a^0\tau_b^0\tau_c^0\tau_\delta^0 E\left[\varepsilon_{jt-a}\varepsilon_{kt'-b}\varepsilon_{it-c}\varepsilon_{it'-d}\right],
\end{aligned}$$

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

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

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

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

proceeding as in Lemma 8. The second type is

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

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

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

The final fourth type involving fourth order cumulants is

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

The third term in (1.28) can be bounded using Cauchy-Schwarz inequality and the Lemma follows. $\square$

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

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

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

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

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

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}(\lambda_{t-1}(L;\theta))^2\right| \le \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{(\lambda_{t-1}(L;\theta)f_t)^2 - E\left[(\lambda_{t-1}(L;\theta)f_t)^2\right]\right\}\right| + \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}E\left[(\lambda_{t-1}(L;\theta)f_t)^2\right]\right|. \qquad (1.29)$$

Under Assumption 2, we have

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

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 $\theta\in\Theta$

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

while the first term on the lhs of (1.29) 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

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

Uniformly in j,

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

so,

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

where K is some arbitrarily large positive constant. Similarly,

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

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

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}\dot{\tau}_t(\theta_0)\dot{\tau}_r(\theta_0)E\left[\lambda_{t-1}(L;\theta_0) f_t \lambda_{r-1}(L;\theta_0) f_r\right]. \tag{1.30}$$

When $0 \leq \varrho - \delta_0 \leq 1$, $|E[\lambda_{t-1}(L;\theta_0) f_t * \lambda_{r-1}(L;\theta_0) f_r]| \leq K|t-r|_+^{2(\varrho-\delta_0)-1}$ and using that $|\dot{\tau}_t(\theta_0)| \leq K t^{-\delta_0} \log t$, (1.30) is

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

When $\varrho - \delta_0 < 0$, $|E[\lambda_{t-1}(L;\theta_0) f_t * \lambda_{r-1}(L;\theta_0) f_r]| \leq K|t-r|_+^{\varrho-\delta_0-1} r^{\varrho-\delta_0}$, $t > r$, see Lemma 8, so (1.30) is

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

and the result follows. □

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

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

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

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta) f_t * \lambda_t^0(L;\theta)\left(\varepsilon_{it}-\hat{\phi}_i\bar{\varepsilon}_t\right) \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta) f_t * \lambda_t^0(L;\theta)\,\varepsilon_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}(L;\theta) f_t * \lambda_t^0(L;\theta)\,\hat{\phi}_i\bar{\varepsilon}_t,
\end{aligned}
$$

where the first term is equal to

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\lambda_{t-1}\left(L;\theta\right)f_t*\lambda_t^0\left(L;\theta\right)\varepsilon_{it} \\
= & \frac{1}{T^{-1}\sum_t\left(\Delta\bar{y}_t\right)^2}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\Delta\bar{y}_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\left(\varepsilon_{ir}-\frac{\gamma_i}{\bar{\gamma}}\bar{\varepsilon}_r\right)*\lambda_{t-1}\left(L;\theta\right)f_t*\lambda_t^0\left(L;\theta\right)\varepsilon_{it} \\
= & \frac{1}{T^{-1}\sum_t\left(\Delta\bar{y}_t\right)^2}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\left(\bar{\gamma}\Delta f_r+\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_r\right)\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\left(\varepsilon_{ir}-\frac{\gamma_i}{\bar{\gamma}}\bar{\varepsilon}_r\right)*\lambda_{t-1}\left(L;\theta\right)f_t*\lambda_t^0\left(L;\theta\right)\varepsilon_{it}.
\end{aligned}
$$

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

$$
\begin{aligned}
& \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\left(\bar{\gamma}\Delta f_r+\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_r\right)\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\left(\varepsilon_{ir}-\frac{\gamma_i}{\bar{\gamma}}\bar{\varepsilon}_r\right)\lambda_{t-1}\left(L;\theta\right)f_t\lambda_t^0\left(L;\theta\right)\varepsilon_{it} \qquad (1.31) \\
& =\frac{\bar{\gamma}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\Delta f_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{t-1}\left(L;\theta\right)f_t\lambda_t^0\left(L;\theta\right)\varepsilon_{it} \\
& +\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{t-1}\left(L;\theta\right)f_t\lambda_t^0\left(L;\theta\right)\varepsilon_{it} \\
& -\frac{1}{NT^2\bar{\gamma}}\sum_{i=1}^{N}\gamma_i\sum_{t=1}^{T}\sum_{r=1}^{T}\bar{\gamma}\Delta f_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_r\lambda_{t-1}\left(L;\theta\right)f_t\lambda_t^0\left(L;\theta\right)\varepsilon_{it} \\
& -\frac{1}{NT^2\bar{\gamma}}\sum_{i=1}^{N}\gamma_i\sum_{t=1}^{T}\sum_{r=1}^{T}\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_r\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\bar{\varepsilon}_r\lambda_{t-1}\left(L;\theta\right)f_t\lambda_t^0\left(L;\theta\right)\varepsilon_{it}.
\end{aligned}
$$

The first term on the rhs of (1.31) can be written as $\bar{\gamma}$ times

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

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

$$
\begin{aligned}
& \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}E\left[\Delta f_r\lambda_{t-1}\left(L;\theta\right)f_t\right]E\left[\lambda_r^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_t^0\left(L;\theta\right)\varepsilon_{it}\right] \\
= \; & O\left(\begin{array}{c}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{r=1}^{T}\left(|t-r|_+^{2(\varrho-1)-\underline{\delta}}+|t-r|_+^{\varrho-1-\underline{\delta}}+|t-r|_+^{\varrho-2}\right) \\ \times\left(|t-r|_+^{2(\delta_0-1)-\underline{\delta}}+|t-r|_+^{\delta_0-1-\underline{\delta}}+|t-r|_+^{\delta_0-2}\right)\end{array}\right) \\
= \; & o\left(1\right)
\end{aligned}
$$

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

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

$=(a)+(b)$, 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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The bounds for the other terms on the rhs of (1.31) follow in a similar form, noting that the presence of cross section averages introduce a further $N^{-1/2}$ factor in the probability bounds. □

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

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

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

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0(L;\theta)\,\varepsilon_{it}\tau_t(\theta)\varepsilon_{i0} \tag{1.32}
$$

is

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

uniformly in δ, so we can show that the term (1.32) 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

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\lambda_t^0\left(\theta\right)\tau_t(\theta)\frac{1}{N}\sum_{i=1}^{N}\left\{\varepsilon_{i0}^2-\sigma^2\right\} \\ & +\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\lambda_j^0\left(\theta\right)\tau_t(\theta)\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it-j}\varepsilon_{i0} \end{aligned}$$

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

$$\begin{aligned} & \frac{1}{T}\sum_{j=0}^{T}\sum_{k=j+1}^{T}\lambda_j^0\left(\theta\right)\tau_k(\delta)\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{ik-j}\varepsilon_{i0} \\ = & \frac{\tau_T(\delta)}{T}\sum_{j=0}^{T}\lambda_j^0\left(\theta\right)\frac{1}{N}\sum_{i=1}^{N}\sum_{k=j+1}^{T}\varepsilon_{ik-j}\varepsilon_{i0} \\ & -\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=0}^{T}\lambda_j^0\left(\theta\right)\sum_{k=j+1}^{T}\left\{\tau_{k+1}(\delta)-\tau_k(\delta)\right\}\frac{1}{N}\sum_{i=1}^{N}\sum_{r=j+1}^{k}\varepsilon_{ir-j}\varepsilon_{i0} \\ = & \left(b_1\right)+\left(b_2\right). \end{aligned}$$

Then,

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

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,

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

The second term is

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

because we can show that

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

using the same method as for bounding (1.32), while

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

by Lemma 1 and Cauchy-Schwarz inequality.

The third term,

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

is negligible because, on the one hand

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

because $\bar{\varepsilon}_0 = O_p\left(N^{-1/2}\right)$, $\bar{\gamma}^{-1} = O_p\left(1\right)$ and the average can be bounded as (1.32) since γ_i is independent of ε_{it}, which is zero mean, and on the other hand under Assumption B*,

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

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

Finally, the last term,

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\hat{\phi}_i^2\bar{\varepsilon}_t\tau_t(\theta)\bar{\varepsilon}_0 &= \bar{\varepsilon}_0\frac{1}{T}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta\right)\bar{\varepsilon}_t\tau_t(\theta)\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_i^2\\ &= O_p\left(N^{-1/2}\right)o_p\left(1\right)O_p\left(1\right) = o_p\left(1\right),\end{aligned}$$

is also negligible, proceeding as before. □

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

$$\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\tau_t(\theta_0)\tau_r(\theta_0)Cov\left[\varepsilon_{i0}*\chi_t\left(L;\xi_0\right)\varepsilon_{it},\varepsilon_{i0}*\chi_r\left(L;\xi_0\right)\varepsilon_{ir}\right]\\
= & \frac{1}{T}\sum_{t=1}^{T}\tau_t(\theta_0)^2\left[\sigma^4\left(\sum_{j=0}^{t}j^{-2}+t^{-2}\right)+\{\kappa_4\}\right]\\
& +\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{t}\tau_t(\theta_0)\tau_r(\theta_0)\left[\sigma^4\left(\sum_{j=0}^{t}j^{-1}\left(t-r+j\right)^{-1}+t^{-1}r^{-1}\right)+\kappa_4t^{-2}1\left\{t=r\right\}\right]\\
= & O\left(T^{-1}+T^{-2\delta_0}\right)+O\left(T^{-1}\sum_{t=1}^{T}\sum_{r=1}^{t}\left(rt\right)^{-\delta_0}\left(\left|t-r\right|^{-1}\log t+\left(tr\right)^{-1}\right)\right)\\
= & O\left(T^{-1}+T^{-2\delta_0}\right)+O\left(T^{-1}\sum_{t=1}^{T}t^{-\delta_0}\left(t^{-\delta_0}\log^2 t+t^{-1}\log t\right)\right)=O\left(T^{-1}\log^4 T+T^{-2\delta_0}\log^2 T\right)=o\left(1\right)
\end{aligned}$$

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

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

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

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

while

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

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

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

and the proof is completed. □

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

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

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

$$\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{it}*\chi_t\left(L;\xi_0\right)\varepsilon_{it}-\varepsilon_{it}\hat{\phi}_i\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t-\hat{\phi}_i\bar{\varepsilon}_t\chi_t\left(L;\xi_0\right)\varepsilon_{it}+\hat{\phi}_i^2\bar{\varepsilon}_t*\chi_t\left(L;\xi_0\right)\bar{\varepsilon}_t\right\} \quad (1.33)$$

where Proposition 2 in [39] shows the asymptotic $\mathcal{N}\left(0,4B\left(\xi_0\right)\right)$ distribution of the first term as $\left(N,T\right)_j\to\infty$, and we now show that the remainder terms are negligible. Then the second term on (1.33) 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 $2\left(NT\right)^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\gamma_i\chi_t\left(L;\xi_0\right)\varepsilon_{jt}$ has zero expectation and variance proportional to

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

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

$$
\begin{aligned}
\left|\frac{2}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\eta_i\chi_t\left(L;\xi_0\right)\varepsilon_{jt}\right| \leq & \frac{2}{\sqrt{NT}}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\chi_t\left(L;\xi_0\right)\varepsilon_{jt}\right)^2\right)^{1/2} \\
= & O_p\left(\left(NT\right)^{-1/2}\left(T^{2\varrho+2\delta_0-6}+T^{-1}\right)^{1/2}\left(NT\right)^{1/2}\right) \\
= & O_p\left(\left(T^{2\varrho+2\delta_0-6}+T^{-1}\right)^{1/2}\right) = o_p\left(1\right)
\end{aligned}
$$

because

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

The third term in (1.33) 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 $2\left(NT\right)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_i\chi_t\left(L;\xi_0\right)\varepsilon_{it}\bar{\varepsilon}_t$ has zero expectation and variance

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

while

$$\left|\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_i\chi_t(L;\xi_0)\varepsilon_{it}\bar{\varepsilon}_t\right| \le \frac{2N}{\sqrt{NT}}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\chi_t(L;\xi_0)\varepsilon_{it}\bar{\varepsilon}_t\right)^2\right)^{1/2}$$
$$= O_p\left(N^{1/2}T^{-1/2}(T^{2\varrho+2\delta_0-6}+T^{-1})^{1/2}\left(N^{-1}T\right)^{1/2}\right)$$
$$= O_p\left((T^{2\varrho+2\delta_0-6}+T^{-1})^{1/2}\right) = o_p(1)$$

because

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

Finally, the fourth term in (1.33) is also negligible, since

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

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

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

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

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

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

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

$$\begin{aligned}
E\left[\Delta f_r \lambda_{t-1}(L;\theta) f_t\right] &= O\begin{pmatrix} |t-r|_+^{2(\varrho-1)-\delta} \\ +|t-r|_+^{\varrho-1-\delta} r^{\varrho-1} 1\{\varrho-1<0\} 1\{r<t\} \\ +|t-r|_+^{\varrho-2} t^{\varrho-\delta} 1\{\varrho-\delta<0\} 1\{t<r\} \end{pmatrix} \\
&= O\left(|t-r|_+^{2(\varrho-1)-\delta} + |t-r|_+^{\varrho-1-\delta} + |t-r|_+^{\varrho-2}\right) \\
E\left[\lambda_{t-1}^{-1}\left(L;\theta_0^{(-1)}\right)\varepsilon_{ir}\lambda_{t-1}^{0}(L;\theta)\varepsilon_{it}\right] &= O\begin{pmatrix} |t-r|_+^{2(\delta_0-1)-\delta} \\ +|t-r|_+^{\delta_0-1-\delta} r^{\delta_0-1} 1\{\delta_0-1<0\} 1\{r<t\} \\ +|t-r|_+^{\delta_0-2} t^{\delta_0-\delta} 1\{\delta_0-\delta<0\} 1\{t<r\} \end{pmatrix} \\
&= O\left(|t-r|_+^{2(\delta_0-1)-\delta} + |t-r|_+^{\delta_0-1-\delta} + |t-r|_+^{\delta_0-2}\right),
\end{aligned}$$

where $|a|_+ = \max\{|a|,1\}$ *and*

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

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

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

where $d_j(a) = \sum_{k=0}^{j} \varphi_k^f \pi_{j-k}(a) \sim c j^{-a-1}$ and $c_j(a) = c_j(a,\xi) = \sum_{k=0}^{j} \varphi_k^f \lambda_{j-k}(a,\xi) \sim c j^{-a-1}$ 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, $\varrho \geq 1$,

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

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

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

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

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

If $r > t$

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

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

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

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

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

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

Similarly, if $r = t$

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

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

Bibliography

[1] Baglan, D. and E. Yoldas (2013): "Government Debt and Macroeconomic Activity: A Predictive Analysis for Advanced Economies," *Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.*

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

[3] ——— (2010): "Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation," *Econometric Theory*, 26, 1088–1114.

[4] Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

[5] ——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127–1177.

[6] ——— (2013): "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176, 18–29.

[7] Blanchard, O. J. and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79(4), 655–73.

[8] Bollerslev, T., D. Osterrieder, N. Sizova, and G. Tauchen (2013): "Risk and Return: Long-Run Relationships, Fractional Cointegration, and Return Predictability," *Journal of Financial Economics*, 108(2), 409–424.

[9] Chambers, M. J. (1998): "Long Memory and Aggregation in Macroeconomic Time Series," *International Economic Review*, 39(4), 1053–1072.

[10] 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.).*

[11] Chudik, A., K. Mohaddes, H. Pesaran, and M. Raissi (2013): "Debt, Inflation and Growth: Robust Estimation of Long-Run Effects in Dynamic Panel Data Models," *Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper No. 162.*

[12] Chudik, A., H. Pesaran, and E. Tosetti (2011): “Weak and Strong Cross-Section Dependence and Estimation of Large Panels,” *The Econometrics Journal*, 14(1), C45–C90.

[13] DeLong, J. B. and L. H. Summers (2012): “Fiscal Policy in a Depressed Economy,” *Brookings Papers on Economic Activity*, 233–297.

[14] Elmendorf, D. W. and G. N. Mankiw (1999): *Government Debt*, Elsevier, chap. Volume 1, Part C, 1615–1669.

[15] Ergemen, Y. E. (2015): “Fractionally Integrated Panel Data Systems,” *Preprint UC3M*, http://dx.doi.org/10.2139/ssrn.2521050.

[16] Ergemen, Y. E. and C. Velasco (2015): “Estimation of Fractionally Integrated Panels with Fixed-Effects and Cross-Section Dependence,” *Preprint UC3M*.

[17] Gil-Alaña, L. and P. Robinson (1997): “Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series,” *Journal of Econometrics*, 80(2), 241–268.

[18] Granger, C. (1980): “Long Memory Relationships and the Aggregation of Dynamic Models,” *Journal of Econometrics*, 14, 227–238.

[19] 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.

[20] 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.

[21] Hosoya, Y. (2005): “Fractional Invariance Principle,” *Journal of Time Series Analysis*, 26, 463–486.

[22] Hualde, J. and P. M. Robinson (2007): “Root-N-Consistent Estimation of Weak Fractional Cointegration,” *Journal of Econometrics*, 140, 450–484.

[23] ——— (2011): “Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models,” *The Annals of Statistics*, 39(6), 3152–3181.

[24] Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): “Panels with Non-Stationary Multifactor Error Structures,” *Journal of Econometrics*, 160(2), 326–348.

[25] Marinucci, D. and P. Robinson (2000): “Weak Convergence of Multivariate Fractional Processes,” *Stochastic Processes and their Applications*, 86, 103–120.

[26] Michelacci, C. and P. Zaffaroni (2000): “(Fractional) Beta Convergence,” *Journal of Monetary Economics*, 45, 129–153.

[27] MOON, H. R. AND B. PERRON (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122(1), 81–126.

[28] 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.

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

[30] PESARAN, H. AND E. TOSETTI (2011): "Large Panels with Common Factors and Spatial Correlation," *Journal of Econometrics*, 161(2), 182–202.

[31] PESARAN, M. H. AND A. CHUDIK (2014): "Aggregation in Large Dynamic Panels," *Journal of Econometrics*, 178, 273–285.

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

[33] REINHART, C. AND K. S. ROGOFF (2010): "Growth in a Time of Debt," *American Economic Review*, 100(2), 573–578.

[34] ROBINSON, P. M. (1978): *Comments on "Some consequences of temporal aggregation in seasonal time analysis models" by W. W. S. Wei*, United States Department of Commerce, Bureau of the Census, Washington, DC, 445–447.

[35] ——— (1995): "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23(5), 1630–1661.

[36] ——— (2005): "Robust Covariance Matrix Estimation : 'HAC' Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21(1), 171–180.

[37] ROBINSON, P. M. AND J. HIDALGO (1997): "Time Series Regression with Long-Range Dependence," *The Annals of Statistics*, 25(1), 77–104.

[38] ROBINSON, P. M. AND J. HUALDE (2003): "Cointegration in Fractional Systems with Unknown Integration Orders," *Econometrica*, 71(6), 1727–1766.

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

[40] SIMS, C. A. (1987): "A Rational Expectations Framework for Short Run Policy Analysis," doi=10.1.1.211.9699.

Table 1.1: Empirical bias and RMSE of $\hat{\delta}$ and $\tilde{\delta}$

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

Table 1.2: Empirical coverage of 95% CI based on $\hat{\delta}$ and $\tilde{\delta}$

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

Table 1.3: Preliminary and Joint Estimation Bias and RMSE's with $N = 10$ and $T = 50$ ($\delta^* = 1$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$
$\varrho = 0.4$:										
Bias of $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	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
	$\hat{\beta}_{MG}(\hat{\delta})$	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 $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	0.0621	0.0567	0.0529	0.0611	0.0573	0.0552	0.0538	0.0536	0.0555
	$\hat{\beta}_{CC}(\delta^*)$	0.0621	0.0569	0.0531	0.0609	0.0571	0.0551	0.0518	0.0501	0.0518
	$\hat{\beta}_{MG}(\hat{\delta})$	0.0621	0.0567	0.0529	0.0611	0.0570	0.0550	0.0531	0.0512	0.0525
	$\hat{\beta}_{CC}(\hat{\delta})$	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
$\varrho = 1$:										
Bias of $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	-0.0029	-0.0019	0.0017	-0.0039	-0.0052	-0.0024	-0.0070	-0.0131	-0.0140
	$\hat{\beta}_{CC}(\delta^*)$	0.0006	0.0006	0.0008	0.0006	0.0007	0.0011	0.0001	0.0002	0.0007
	$\hat{\beta}_{MG}(\hat{\delta})$	0.0001	0.0001	0.0001	0.0002	0.0002	0.0005	0.0001	0.0002	0.0006
	$\hat{\beta}_{CC}(\hat{\delta})$	0.0436	0.0145	0.0012	0.0327	0.0127	0.0015	0.0146	0.0067	0.0012
RMSE of $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	0.0624	0.0573	0.0537	0.0617	0.0580	0.0559	0.0545	0.0539	0.0555
	$\hat{\beta}_{CC}(\delta^*)$	0.0626	0.0577	0.0541	0.0618	0.0581	0.0563	0.0533	0.0517	0.0534
	$\hat{\beta}_{MG}(\hat{\delta})$	0.0624	0.0573	0.0537	0.0616	0.0577	0.0559	0.0540	0.0523	0.0537
	$\hat{\beta}_{CC}(\hat{\delta})$	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 1.4: Preliminary and Joint Estimation Bias and RMSE's with $N = 20$ and $T = 100$ ($\delta^* = 1$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$	$\delta_0 = 0.5$	$\delta_0 = 0.75$	$\delta_0 = 1$
$\varrho = 0.4$:										
Bias of $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	-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
	$\hat{\beta}_{MG}(\hat{\delta})$	-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 $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	0.0295	0.0270	0.0254	0.0290	0.0271	0.0265	0.0256	0.0251	0.0262
	$\hat{\beta}_{CC}(\delta^*)$	0.0299	0.0274	0.0258	0.0296	0.0276	0.0269	0.0251	0.0241	0.0252
	$\hat{\beta}_{MG}(\hat{\delta})$	0.0294	0.0270	0.0254	0.0290	0.0271	0.0265	0.0250	0.0240	0.0250
	$\hat{\beta}_{CC}(\hat{\delta})$	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
$\varrho = 1$:										
Bias of $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	-0.0031	-0.0026	-0.0027	0.0001	0.0008	0.0003	0.0068	0.0082	0.0082
	$\hat{\beta}_{CC}(\delta^*)$	-0.0013	-0.0015	-0.0015	-0.0013	-0.0016	-0.0019	-0.0009	-0.0013	-0.0018
	$\hat{\beta}_{MG}(\hat{\delta})$	-0.0013	-0.0015	-0.0016	-0.0012	-0.0015	-0.0018	-0.0008	-0.0012	-0.0018
	$\hat{\beta}_{CC}(\hat{\delta})$	0.0588	0.0155	-0.0015	0.0423	0.0130	-0.0018	0.0159	0.0062	-0.0017
RMSE of $\hat{\beta}$	$\hat{\beta}_{MG}(\delta^*)$	0.0297	0.0273	0.0258	0.0293	0.0274	0.0267	0.0263	0.0258	0.0267
	$\hat{\beta}_{CC}(\delta^*)$	0.0302	0.0277	0.0261	0.0300	0.0280	0.0273	0.0258	0.0248	0.0259
	$\hat{\beta}_{MG}(\hat{\delta})$	0.0296	0.0272	0.0257	0.0293	0.0274	0.0268	0.0255	0.0245	0.0255
	$\hat{\beta}_{CC}(\hat{\delta})$	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

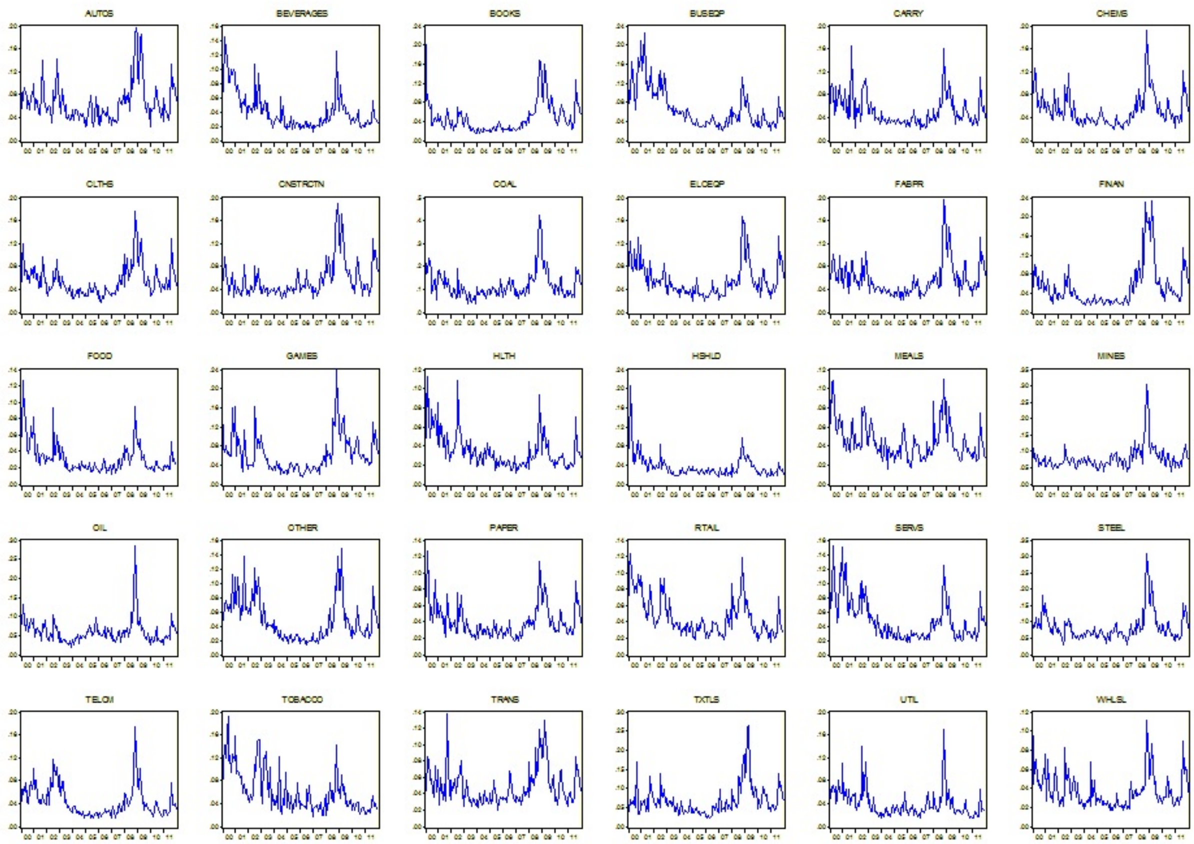


Figure 1.1: Monthly Realized Volatilities across Industries

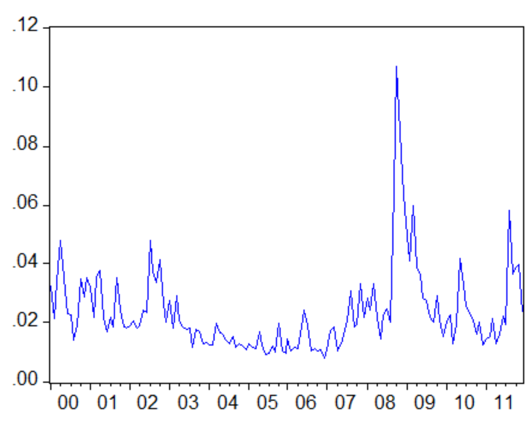


Figure 1.2: Monthly Realized Volatility in the Composite Market

Table 1.5: Estimated Integration Orders of Industry Realized Volatilities

<u>$m = 20$:</u>

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

<u>$m = 32$:</u>

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 1.6: 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.*

Table 1.7: Estimated Slope Parameters across Industry Realized Volatilities

	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)
$\hat{\beta}_i$	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
$\hat{\beta}_i^0$	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)
$\hat{\beta}_i$	-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
	(0.0961)	(0.1271)	(0.1530)	(0.0768)	(0.0751)	(0.0603)	(0.1308)	(0.1020)
$\hat{\beta}_i$	0.1399	-0.3669	0.0311	-1.0433	-0.1778	-2.4097	2.6804	-0.6838
	(0.0628)	(0.1329)	(0.1718)	(0.0686)	(0.1065)	(0.1122)	(0.0832)	(0.0820)
		Chems	Txtls	Oil	Util	Finan	Other	
$\hat{\beta}_i^0$		0.7898	0.4888	0.7927	0.6498	0.5316	0.1067	
		(0.0516)	(0.0981)	(0.0852)	(0.0925)	(0.0986)	(0.0632)	
$\hat{\beta}_i$		-0.0546	-0.1731	-0.1238	-0.4930	-0.8456	-0.1933	
		(0.0419)	(0.1665)	(0.0982)	(0.0828)	(0.1838)	(0.0881)	

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.*

Chapter 2

System Estimation of Panel Data Models under Long-Range Dependence

Abstract

A general dynamic panel data model is considered that incorporates individual and interactive fixed effects and possibly correlated innovations. The model accommodates general stationary or nonstationary long-range dependence through interactive fixed effects and innovations, removing the necessity to perform a priori unit-root or stationarity testing. Moreover, persistence in innovations and interactive fixed effects allows for cointegration; innovations can also have vector-autoregressive dynamics; deterministic trends can be nested. Estimations are performed using conditional-sum-of-squares criteria based on projected series by which latent characteristics are proxied. Resulting estimates are consistent and asymptotically normal at parametric rates. A simulation study provides reliability on the estimation method. The method is then applied to the long-run relationship between debt and GDP.
KEYWORDS: Long memory, factor models, panel data, endogeneity, fixed effects, debt and GDP.
JEL CLASSIFICATION: C32, C33

2.1 Introduction

In economics, long-range dependence can arise due to aggregation. It is common practice to assume that laws of motion of capital, consumption and borrowing rates follow an autoregressive process in economic modelling under a heterogeneous-agents setting. However, economic theories are described for a representative agent whose behaviour reflects the average behaviour, which requires aggregation of individual characteristics. This in turn leads to the necessity of aggregating laws of motions in a given economic model so that conclusions can be drawn for the representative agent. Robinson [34] and Granger [18] prove that aggregating autoregressive models can lead to fractionally integrated models that have dramatically different correlation structures for both dependent and independent individual series as is the case when aggregating micro variables such as total personal income, unemployment, consumption of non-durable goods, inventories, and profits. Chambers [9] shows that U.K. macroeconomic series exhibit fractional long-range dependence when the dynamic models describing the series are cross-sectionally or temporally aggregated. In a pure time-series context, Gil-Alaña and Robinson [17] show that unemployment rate, CPI, industrial production and money stock (M2) exhibit non-integer values of integration, and similar conclusions arise for many financial series such as real exchange rates, equity and stock market realized volatility, see e.g. Bollerslev et al. [8]. Furthermore, Michelacci and Zaffaroni [26] find that aggregate GDP shocks exhibit long memory and show that output convergence to steady state is intertwined with this property. Recently, Pesaran and Chudik [31] show that aggregation of linear dynamic panel data models can lead to long memory and use this property to investigate the source of persistence in aggregate inflation.

In order to get a solid empirical perspective, several indicators are frequently organized in a panel data structure to incorporate the characteristics of different units, such as countries or assets, while describing their time-series dynamics. The examples of macroeconomic panel data indicators include GDP, interest, inflation and unemployment rates, and in finance, it is standard to use a panel data structure in portfolio performance evaluations and risk management. Analysis of such panel indicators has been carried out using both static and dynamic models. To be more realistic, recent research in panel data theory focuses on developing inference when unobserved heterogeneity and interactions between cross-section units are present based on stationary $I(0)$ variables; see e.g. Pesaran [29]. The research on nonstationary panel data models, on the other hand, has typically developed in an autoregressive framework with $I(1)$ variables. For instance, Phillips and Moon [32] develop limit theory for heterogeneous panel data models with $I(1)$ series. Different nonstationary settings have also been considered to account for individual cross-section characteristics and interactions between cross-section units. For example, Bai and Ng [5] and Bai [3] propose unit-root testing procedures when idiosyncratic innovations and the common factor are both $I(1)$, and Moon and Perron [27] propose the use of dynamic factors to test for unit roots in cross-sectionally dependent panels.

Since several studies have repeatedly shown that many economic and financial time series exhibit fractional long-range dependence (possibly due to aggregation) and many macroeconomic and financial indicators are presented in the form of panels, panel data models should also account for such characteristics. To the best of our knowledge, only few papers study fractional long-range dependence in panel data models. Hassler et al. [20] propose a test for memory in fractionally integrated panels. Robinson and Velasco [39] employ different estimation techniques to obtain efficient inference on the memory parameter in a fractional panel setting with fixed effects. Extending the latter, Ergemen and Velasco [16] incorporate cross-section dependence and exogenous covariates to estimate slope and memory parameters in a single-equation setting, which enables disclosing possible cointegrating relationships between the unobserved independent idiosyncratic components.

This paper contributes to the literature in many ways. First, unlike in Hassler et al. [20] and Robinson and Velasco [39], we explicitly model cross-section dependence and allow for cointegrating relationships in the unobserved components. However, under our setup, there is no cointegration requirement for obtaining valid inference, which removes the necessity of a priori cointegration testing as required by Robinson and Hualde [38] and Hualde and Robinson [22]. Second, unlike in Ergemen and Velasco [16], we allow for contemporaneous correlations in the idiosyncratic innovations, which calls for system estimation on the defactored observed series. Allowing for endogeneity via the idiosyncratic innovations leads the model to achieve wider empirical applicability, especially in cases where endogeneity induced by the unobserved common factor is not the only source of contemporaneous correlation. For example, empirical analyses of endogenous growth theories and the purchasing power parity hypothesis generally require that the idiosyncratic errors be correlated even after the factor structure is removed due to prevailing two-way endogeneity in data. Third, our model can successfully address the cases in which a time series cointegration approach would lead to invalid results. The observable series can display the same memory level when the integration order of the common factor is greater than those of the idiosyncratic innovations. Thus a pure time-series approach may fail to detect possible cointegrating relationships. In this case, possible cointegrating relationships can only be disclosed after the common factor structure is projected out, implying that accounting for individual unit characteristics and cross-section interactions is essential in obtaining valid inference, as is the case under our setup.

The methodology that we develop in this paper can be used, for instance, as a country-specific inference tool for analyses of economic unions. In our econometric framework, country-specific characteristics are captured by individual and interactive fixed effects. To get heterogeneous inference in an economic union, we allow for long-range dependence in both idiosyncratic innovations and the common factor structure capturing possible interactions between countries, while letting the country-specific innovations be also contemporaneously correlated. These properties in turn introduce the possibility of cointegrated system estimation in the classical sense, by which an equilibrium analysis can be carried out in macroeconomic terms.

In the estimation of the slope and long-range dependence parameters, we use an equation-by-equation conditional-sum-of-squares (CSS) approach, in a similar way to Hualde and Robinson [22]. The estimation procedure is based on the defactored variables obtained after projections on the sample means of fractionally differenced data, leading to GLS-type estimates for slope parameters. The resulting individual slope and long-range dependence estimates are $\sqrt{T}$ consistent with a centered asymptotic normal distribution, and the mean-group slope estimate is $\sqrt{n}$ consistent and asymptotically normally distributed, irrespective of cointegrating relationships, where n is the number of cross-section units and T is the length of time series. We explore the small-sample behaviour of our estimates by means of Monte Carlo experiments both when autocorrelations and/or endogeneity are absent and present, and find that the estimates behave well even in relatively small panels.

In the empirical application, we investigate the long-run relationship between real GDP and debt/GDP growth rates as well as debt and real GDP in log-levels for 20 high-income OECD countries for the time period 1955-2008. We find that GDP growth does not respond to a growth in debt/GDP for most of the countries at the 5% level. On the other hand, real GDP and debt in log-levels have a significant relationship for all countries but New Zealand and the United States, and this relationship is cointegrating for several countries, which we can find using our panel approach but not using a pure time series cointegration methodology as we show comparing our results to those that would be obtained by Hualde and Robinson [22]'s method. The empirical application stresses that our panel data approach provides correct inference particularly when the main source of persistence in the indicators is cross-country dependence.

The remainder of the paper proceeds as follows. Next section contains estimation details of slope and fractional integration parameters. Section 2.3 lists all the conditions needed and contains the main results. Section 2.4 briefly discusses the inclusion of deterministic trends. Section 2.5 presents a finite-sample study based on Monte Carlo experiments, and Section 2.6 presents the empirical application. Section 2.7 contains the final comments.

Throughout the paper, "$(n,T)_j$" denotes joint asymptotics in which both the cross-section size and time-series length are growing; " $\to_p$ " denotes convergence in probability; and " $\to_d$ " denotes convergence in distribution. All mathematical proofs and intermediate technical results are collected in an appendix at the end of the paper.

2.2 Model, Discussion and Parameter Estimation

We consider the following triangular array describing a type-II fractionally integrated panel data model of the observed series (y_{it}, x_{it}) :

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}\beta_{i0} + f_t\lambda_i + \Delta_t^{-d_{i0}}\epsilon_{1it}, \\ x_{it} &= \mu_i + f_t\gamma_i + \Delta_t^{-\vartheta_{i0}}\epsilon_{2it}, \end{aligned} \tag{2.1}$$

where y_{it} and x_{it} are scalars whose idiosyncratic innovations have unknown true integration orders d_{i0} and ϑ_{i0} for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and f_t is an unobserved common factor that may be integrated of an unknown order δ. While vector x_{it} may also be analyzed allowing for a multiple regression setting, we consider the simplest case to focus on the main ideas. Throughout the paper, the subscript at the fractional differencing operator attached to a vector or scalar ϵ_{it} (i.e. a type-II process) has the meaning

$$\Delta_t^{-d}\epsilon_{it} = \Delta^{-d}\epsilon_{it}1(t > 0) = \sum_{j=0}^{t-1} \pi_j(-d)\epsilon_{it-j}, \tag{2.2}$$
$$\pi_j(-d) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)},$$

where $1(\cdot)$ is the indicator function, and $\Gamma(\cdot)$ denotes the gamma function such that $\Gamma(d) = \infty$ for $d = 0, -1, -2, \ldots$, and $\Gamma(0)/\Gamma(0) = 1$ by convention. With the prime denoting transposition, $\epsilon_{it} = (\epsilon_{1it}, \epsilon_{2it})'$ is a bivariate covariance stationary process, allowing for $Cov(\epsilon_{1it}, \epsilon_{2it}) \neq 0$, whose short-memory vector-autoregressive (VAR) dynamics are described by

$$B(L;\theta_i)\epsilon_{it} \equiv \left(I_2 - \sum_{j=1}^{p} B_j(\theta_i)L^j\right)\epsilon_{it} = v_{it}, \tag{2.3}$$

where L is the lag operator, θ_i the short-memory parameters, I_2 the 2×2 identity matrix, B_j the 2×2 upper-triangular matrices, and v_{it} is a bivariate sequence that is identically and independently distributed across i and t with zero mean and covariance matrix $\Omega_i > 0$. The upper-triangularity assumption on the short-memory matrices, B_j, provides a great deal of parsimony in the asymptotics as it further develops the triangular structure of the system, and it is in line with the long-run VAR restriction of Blanchard and Quah [7] and the short-run VAR restriction of Sims [40]. The arrays $\{\alpha_i, i \geq 1\}$ and $\{\mu_i, i \geq 1\}$ are unobserved individual fixed effects; $\{f_t, t > 0\}$ is the $I(\delta)$ unobserved common factor that induces cross-section dependence and possibly further endogeneity in the system; $\{\lambda_i, i \geq 1\}$ and $\{\gamma_i, i \geq 1\}$ are unobserved factor loadings indicating how much each cross-section unit is affected by f_t. In addition to these general dynamics, autoregressive conditional heteroskedasticity can also be featured in the common factor so that the model can be suitable also for applications in finance.

After explaining the technical details of the model, it is also important to show the usefulness of it in economic analysis. First, the panel data model in (2.1) nests stationary $I(0)$ and nonstationary $I(1)$ autoregressive panel data models that are extensively used in economic modelling, but unlike in the $I(1)$ autoregressive case, (2.1) has smoothness everywhere, thus the test statistics for the parameter estimates obtained under (2.1) are χ^2 distributed. Second, allowance for general long-range dependence through model innovations and the common factor structure is mainly motivated by a desire to avoid a priori unit-root or stationarity testing as is currently carried out in empirical analyses dealing with possibly nonstationary variables. Third, parameter heterogeneity in (2.1) allows for obtaining unit-specific inference in an economy while latent individual characteristics and possible interactions of the units are also taken into account through fixed effects and common factor structures. Heterogeneity in the memory parameters allows for each unit to exhibit different persistence characteristics. This contrasts with the standard approach in the literature when a nonstationary variable is assumed to be $I(1)$ for all cross-section units merely based on unit-root testing.

2.2.1 Prewhitening and Projection of the Common Factor Structure

In a standard way, we first-difference (2.1) to remove the fixed effects,

$$\begin{aligned} \Delta y_{it} &= \Delta x_{it}\beta_{i0} + \Delta f_t \lambda_i + \Delta_t^{1-d_{i0}} \epsilon_{1it}, \\ \Delta x_{it} &= \Delta f_t \gamma_i + \Delta_t^{1-\vartheta_{i0}} \epsilon_{2it}, \end{aligned} \tag{2.4}$$

for $i = 1, \ldots, n$ and $t = 2, \ldots, T$. After this transformation, it becomes clear that there is a mismatch between the sample available and the lengths of the fractional filters $\Delta_t^{1-d_{i0}}$ and $\Delta_t^{1-\vartheta_{i0}}$, which involve ϵ_{1i1} and ϵ_{2i1}, i.e. the initial conditions, while in practice only the filter Δ_{t-1} can be used. We argue that initial conditions in the idiosyncratic innovations are negligible since the second-order bias caused by initial conditions asymptotically vanishes in time-series length under a heterogeneous setup; see Ergemen and Velasco [16].

Setting

$$\vartheta_{max} = \max_i \vartheta_i \quad \text{and} \quad d_{max} = \max_i d_i,$$

(2.4) can be prewhitened from idiosyncratic long-range dependence for some fixed exogenous differencing choice, d^*, using which all variables become asymptotically stationary with their sample means converging to population limits.

Let us introduce the notation $a_{it}(\tau) = \Delta_{t-1}^{\tau-1} \Delta a_{it}$ for any τ. Then the prewhitened model is given by

$$\begin{aligned} y_{it}(d^*) &= x_{it}(d^*)\beta_{i0} + f_t(d^*)\lambda_i + \epsilon_{1it}(d^* - d_{i0}), \\ x_{it}(d^*) &= f_t(d^*)\gamma_i + \epsilon_{2it}(d^* - \vartheta_{i0}). \end{aligned} \tag{2.5}$$

Thus, using the notation $z_{it}(\tau_1, \tau_2) = (y_{it}(\tau_1), x_{it}(\tau_2))'$, (2.5) can be written in the vectorized form as

$$z_{it}(d^*, d^*) = \zeta x_{it}(d^*)\beta_{i0} + F_t(d^*)L_i + \epsilon_{it}\left(d^* - d_{i0}, d^* - \vartheta_{i0}\right), \tag{2.6}$$

where $\zeta = (1, 0)'$, $F_t(d^*) = f_t(d^*) \otimes I_2$, $L_i = (\lambda_i, \gamma_i)'$, and f_t, λ_i and γ_i are scalars.

The structure $F_t(d^*)L_i$ in (2.6) induces cross-section correlation between units i through $F_t(d^*)$.

The common factor may also be allowed to feature breaks both at levels and in persistence under higher order assumptions, which we do not explore in this paper. Several techniques for eliminating or estimating $I(0)$ common-factor structures have been proposed in the literature. Pesaran [29] suggests using cross-section averages of the observed series as proxies to asymptotically replace the common factor structure. A different version of this procedure has been recently adopted in case of persistent common factors by Ergemen and Velasco [16]. There has also been some focus on estimating the factor loadings and common factors up to a rotation, in $I(0)$ or $I(1)$ cases, which enables their use as plug-in estimates. The well-known principal components approach (PCA) has been greatly extended in factor analysis by e.g. Bai and Ng [4] and Bai and Ng [6]. While factor structure estimates, obtained by principal components analysis, can be used as plug-in estimates thus allowing for the exploitation of more information in forecasting studies, they cause size distortions leading to lower finite-sample performance in testing as pointed out by Pesaran [29]. Moreover, PCA estimation of factors with fractional long-range dependence has not been explored in the literature yet. Bearing in mind this fact, we project out the common factor structure using the cross-section averages of prewhitened data, by which the projection errors vanish asymptotically in cross-section size.

The estimation methodology is primarily based on proxying the latent common factor structure using projections. To give the details about projection, let us denote $\bar{z}_t(d^*, d^*) = n^{-1} \sum_{i=1}^n z_{it}(d^*, d^*)$ to write (2.6) in cross-section averages as

$$\bar{z}_t(d^*, d^*) = \overline{\zeta x_t(d^*)\beta_0} + F_t(d^*)\bar{L} + \bar{\epsilon}_t \left(d^* - d_0, d^* - \vartheta_0\right), \tag{2.7}$$

where $\bar{\epsilon}_t \left(d^* - d_0, d^* - \vartheta_0\right)$ is $O_p(n^{-1/2})$ for large enough d^*. Thus, $\bar{z}_t(d^*, d^*)$ and $\overline{\zeta x_t(d^*)\beta_0}$ asymptotically capture all the information provided by the common factor provided that $\bar{L}$ is full rank. Note that $\bar{x}_t(d^*)$ is readily contained in $\bar{z}_t(d^*, d^*)$ and β_{i0} do not have any contribution in terms of dynamics in $\overline{\zeta x_t(d^*)\beta_0}$ since they are fixed for each i. This is why, $\bar{z}_t(d^*, d^*)$ alone can span the factor space.

Let us write the time-stacked observed series as $\mathbf{x}_i(d^*) = (x_{i2}(d^*), \ldots, x_{iT}(d^*))'$ and $\mathbf{z}_i(d^*, d^*) = (z_{i2}(d^*, d^*), \ldots, z_{iT}(d^*, d^*))'$ for $i = 1, \ldots, n$. Then, for each $i = 1, \ldots, n$,

$$\mathbf{z}_i(d^*, d^*) = \mathbf{x}_i(d^*)\beta_{i0}\zeta' + \mathbf{F}(d^*)\mathbf{L}_i + \mathbf{E}_i \left(d^* - d_{i0}, d^* - \vartheta_{i0}\right), \tag{2.8}$$

where $\mathbf{E}_i \left(d^* - d_{i0}, d^* - \vartheta_{i0}\right) = \left(\epsilon_{i2} \left(d^* - d_{i0}, d^* - \vartheta_{i0},\right), \ldots, \epsilon_{iT} \left(d^* - d_{i0}, d^* - \vartheta_{i0}\right)\right)'$ and $\mathbf{F}(d^*) = \left(\mathbf{vec}\left[F_2(d^*)\right], \ldots, \mathbf{vec}\left[F_T(d^*)\right]\right)'$.

The common factor structure, for $T_1 = T - 1$, can asymptotically be removed by the $T_1 \times T_1$ projection matrix

$$\bar{\mathbf{M}}_{T_1}(d^*) = \mathbf{I}_{T_1} - \bar{\mathbf{z}}(d^*, d^*)(\bar{\mathbf{z}}'(d^*, d^*)\bar{\mathbf{z}}(d^*, d^*))^{-}\bar{\mathbf{z}}'(d^*, d^*), \tag{2.9}$$

where $\bar{\mathbf{z}}(d^*, d^*) = n^{-1} \sum_{i=1}^n \mathbf{z}_i(d^*, d^*)$, and P^- denotes the generalized inverse of a matrix P. When the projection matrix is built with the original (possibly nonstationary) series, it is impossible to ensure the asymptotic replacement of the latent factor structure by cross-section averages because the noise in (2.6) may be too persistent when $d^* = 0$. On the other hand, using some $d^* > \max\{\vartheta_{max}, d_{max}, \delta\} - 1/4$ for prewhitening guarantees that the projection errors vanish asymptotically.

Based on (2.8), the defactored observed bivariate series for each $i = 1, \ldots, n$,

$$\tilde{\mathbf{z}}_i(d^*, d^*) = \tilde{\mathbf{x}}_i(d^*)\beta_{i0}\zeta' + \tilde{\mathbf{E}}_i\left(d^* - d_{i0}, d^* - \vartheta_{i0}\right), \tag{2.10}$$

where $\tilde{\mathbf{z}}_i(d^*, d^*) = \bar{\mathbf{M}}_{T_1}(d^*)\mathbf{z}_i(d^*, d^*)$, $\tilde{\mathbf{x}}_i(d^*) = \bar{\mathbf{M}}_{T_1}(d^*)\mathbf{x}_i(d^*)$ and $\tilde{\mathbf{E}}_i(d^*) = \bar{\mathbf{M}}_{T_1}(d^*)\mathbf{E}_i(d^*)$. The projection error, $\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*)$, is of order $O_p\left(n^{-1} + (nT)^{-1/2}\right)$ as shown in Appendix A.1.

2.2.2 Estimation of Linear Model Parameters

Writing (2.10) for $i = 1, \ldots, n$, and $t = 2, \ldots, T$ we now integrate the defactored series back by d^* to their original integration orders, to perform estimations, as

$$\tilde{z}^*_{it}(d_i, \vartheta_i) = \zeta\tilde{x}^*_{it}(d_i)\beta_{i0} + \tilde{\epsilon}^*_{it}\left(d_i - d_{i0}, \vartheta_i - \vartheta_{i0}\right), \tag{2.11}$$

where the first and second equations of (2.11) are obtained, respectively, by

$$\tilde{y}^*_{it}(d_i) = \Delta_{t-1}^{d_i - d^*}\tilde{y}_{it}(d^*) \quad \text{and} \quad \tilde{x}^*_{it}(\vartheta_i) = \Delta_{t-1}^{\vartheta_i - d^*}\tilde{x}_{it}(d^*),$$

where we omit the dependence on d^* in the notation and assume away the initial conditions.

To explicitly show the short-memory dynamics in the model based on (2.3), (2.11) can be written as

$$\begin{aligned}\tilde{z}^*_{it}(d_i, \vartheta_i) - &\sum_{j=1}^{p} B_j(\theta_i)\tilde{z}^*_{it-j}(d_i, \vartheta_i) \\ &= \left\{\zeta\tilde{x}^*_{it}(d_i) - \sum_{j=1}^{p} B_j(\theta_i)\zeta\tilde{x}^*_{it-j}(d_i)\right\}\beta_{i0} + \tilde{v}^*_{it}\left(d_i - d_{i0}, \vartheta_i - \vartheta_{i0}\right),\end{aligned} \tag{2.12}$$

whose second equation, noting that $\tilde{z}^*_{it}(d_i, \vartheta_i) = (\tilde{y}^*_{it}(d_i), \tilde{x}^*_{it}(\vartheta_i))'$, is

$$\tilde{x}^*_{it}(\vartheta_i) - \sum_{j=1}^{p} B_{2j}(\theta_i)\tilde{z}^*_{it-j}(d_i, \vartheta_i) = \left(-\sum_{j=1}^{p} B_{2j}(\theta_i)\zeta\tilde{x}^*_{it-j}(d_i)\right)\beta_{i0} + \tilde{v}^*_{2it}\left(\vartheta_i - \vartheta_{i0}\right) \tag{2.13}$$

and the first equation can be organized to account for the contemporaneous correlation if we write

$\tilde{y}^*_{it}(d_i) - \rho_i \tilde{x}^*_{it}(\vartheta_i)$ as

$$\begin{aligned}\tilde{y}^*_{it}(d_i) =& \tilde{x}^*_{it}(d_i)\beta_{i0} + \tilde{x}^*_{it}(\vartheta_i)\rho_i + \sum_{j=1}^{p} \left(B_{1j}(\theta_i) - \rho_i B_{2j}(\theta_i)\right) \tilde{z}^*_{it-j}(d_i, \vartheta_i) \\ & - \left(\sum_{j=1}^{p} \left(B_{1j}(\theta_i) - \rho_i B_{2j}(\theta_i)\right) \zeta \tilde{x}^*_{it-j}(d_i) \right) \beta_{i0} + \tilde{v}^*_{1it}\left(d_i - d_{i0}\right) - \rho_i \tilde{v}^*_{2it}\left(\vartheta_i - \vartheta_{i0}\right)\end{aligned} \tag{2.14}$$

with B_{kj} denoting the k-th row of B_j, and $\rho_i = E[\tilde{v}^*_{1it}\tilde{v}^*_{2it}]/E[{\tilde{v}^*_{2it}}^2]$.

Under (2.14), cointegration (i.e. $\vartheta_{i0} > d_{i0}$) is useful in the estimation of β_{i0} since the signal that can be extracted from $\tilde{x}^*_{it}(d_i)$ is stronger than that from $\tilde{x}^*_{it}(\vartheta_i)$. However, identification of β_{i0} is still possible in a spurious regression where $d_{i0} > \vartheta_{i0}$ since the error term in (2.14) is orthogonal to $\tilde{v}^*_{2it}(\cdot)$ given that v_{it} are identically and independently distributed so that $\tilde{v}^*_{1it}(\cdot) - \rho_i \tilde{v}^*_{2it}(\cdot)$ is uncorrelated with $\tilde{v}^*_{2it}(\cdot)$. The only exclusion we have under a spurious setting is the case in which $\vartheta_{i0} = d_{i0}$, which leads to collinearity in (2.14) thus rendering the identification of β_{i0} and ρ_i impossible. The spurious estimation case in which $d_{i0} > \vartheta_{i0}$ is evidently more relevant when the interest is in the estimation of contemporaneous correlations between series more than in the estimation of slope parameters. While the triangular array structure of the system readily leads to the identification of β_{i0} and ρ_i so long as $\vartheta_{i0} \neq d_{i0}$, some B_{kj} may still be left unidentified. In that case, imposing an upper-triangular structure in $B_j(\cdot)$ to further develop the triangular structure of the system leads to identification of B_{kj}.

The case in which $\rho_i = 0$, corresponding to exogenous regressors, has been developed by Ergemen and Velasco [16], where estimation is carried out for the parameters only in the first equation and ϑ_i are treated as nuisance parameters. In the present paper, while the main parameter of interest is still β_{i0}, we can also obtain the estimates of d_{i0}, ϑ_{i0}, ρ_i and $B_j(\theta_i)$.

In this paper, short-memory dynamics are not our main concern so we treat $B_j(\cdot)$ as nuisance parameters. First, we use a $q \times (3p+2)$ restriction matrix Q that is I_{3p+2} when there are no prior zero restrictions on B_j, and a $q < 3p+2$ matrix with prior zero restrictions that is obtained by dropping rows of Q corresponding to restrictions, which may improve efficiency by eliminating some lagged values of the series. Then, write (2.14) as

$$\tilde{y}^*_{it}(d_i) = \omega_i' Q \tilde{Z}^*_{it}(d_i, \vartheta_i) + \tilde{v}^*_{1it}\left(d_i - d_{i0}\right) - \rho_i \tilde{v}^*_{2it}\left(\vartheta_i - \vartheta_{i0}\right) \tag{2.15}$$

with

$$\begin{aligned}\tilde{Z}^*_{it}(d_i, \vartheta_i) &= \left(\tilde{x}^*_{it}(d_i), \tilde{x}^*_{it}(\vartheta_i), \tilde{u}^{*\prime}_{it-1}(d_i, \vartheta_i), \ldots, \tilde{u}^{*\prime}_{it-p}(d_i, \vartheta_i)\right)', \\ \tilde{u}^*_{it-k}(d_i, \vartheta_i) &= \left(\tilde{x}^*_{it-k}(d_i), \tilde{x}^*_{it-k}(\vartheta_i), \tilde{y}^*_{it-k}(d_i)\right)', \quad k = 1, \ldots, p,\end{aligned}$$

and ω_i being the vector of coefficients that are functions of β_i, ρ_i and $B_{kj}(\theta_i)$ whose least-squares estimate is given by

$$\hat{\omega}_i(\tau_1,\tau_2) := M_i(\tau_1,\tau_2)^{-1} m_i(\tau_1,\tau_2) \tag{2.16}$$

with

$$M_i(\tau_1,\tau_2) = Q\frac{1}{T}\sum_{t=p+1}^{T} \tilde{Z}^*_{it}(\tau_1,\tau_2)\tilde{Z}^{*\prime}_{it}(\tau_1,\tau_2)Q' \quad \text{and} \quad m_i(\tau_1,\tau_2) = Q\frac{1}{T}\sum_{t=p+1}^{T} \tilde{Z}^*_{it}(\tau_1,\tau_2)\tilde{y}^*_{it}(\tau_1)$$

where (τ_1,τ_2) denotes the infeasible cases of (d_{i0},ϑ_{i0}), $(\hat{d}_i,\vartheta_{i0})$, $(d_{i0},\hat{\vartheta}_i)$ and the feasible case of $(\hat{d}_i,\hat{\vartheta}_i)$.

In most empirical work, the main parameter of interest is β_{i0}, for which the estimate can simply be obtained from (2.16) as

$$\hat{\beta}_i(\tau_1,\tau_2) = \psi_\beta' \hat{\omega}_i(\tau_1,\tau_2), \quad \psi_\beta = (1,0,\ldots,0)'. \tag{2.17}$$

While $\hat{\beta}_i$ in (2.17) is less efficient than the Gaussian maximum likelihood estimate in the VAR ϵ_{it} case, it is computationally much simpler in practice. Ergemen and Velasco [16] discuss the case in which $\hat{\beta}_i$ is efficient when $Cov(\epsilon_{1it},\epsilon_{2it}) = 0$.

When the interest is in the estimation of contemporaneous correlation between the idiosyncratic innovations, the vector ψ can be adjusted accordingly so that

$$\hat{\rho}_i(\tau_1,\tau_2) = \psi_\rho' \hat{\omega}_i(\tau_1,\tau_2), \quad \psi_\rho = (0,1,\ldots,0)'.$$

Short-memory matrices $B_j(\theta_i)$ and, in case of knowledge on the mappings $B_j(\cdot)$, thereof short-memory parameters can be estimated similarly taking e.g. $\psi_\theta = (0,0,1,\ldots,1)'$.

2.2.3 Estimation of Long-Range Dependence Parameters

For the estimation of long memory or fractional integration parameters, we only consider the empirically relevant case of unknown d_i and ϑ_i. Estimation of long-range dependence parameters in the panel data context is a relatively new topic. Robinson and Velasco [39] propose several techniques for estimating a pooled fractional integration parameter under a fractional panel setting with no covariates or cross-section dependence. Extending their study, Ergemen and Velasco [16] propose fractional panel data models with fixed effects and cross-section dependence in which the long-range dependence parameter is estimated, also when their general model features exogenous covariates, in first differences.

In order to estimate both long-range dependence parameters under our setup, we use an equation-by-equation CSS approach. First, we estimate the second equation of (2.12). Assuming an upper-triangular structure for $B_j(\theta_i)$ in (2.3) for parsimony, we write (2.13) as

$$\tilde{x}_{it}^*(\vartheta_i) - \phi_i' R \tilde{X}_{it}^*(\vartheta_i) = \tilde{v}_{2it}^*(\vartheta_i - \vartheta_{i0})$$

with

$$\tilde{X}_{it}^*(\vartheta_i) = \left(\tilde{x}_{it-1}^*(\vartheta_i), \ldots, \tilde{x}_{it-p}^*(\vartheta_i)\right)',$$

the $r \times p$ matrix $R = I_p$ for $r = p$, but for $r < p$, R is obtained by dropping rows from I_p, and ϕ_i collecting the B_{22j} that are nonzero a priori. Then an estimate of ϕ_i,

$$\hat{\phi}_i(\vartheta) := G_i(\vartheta)^{-1} g_i(\vartheta) \tag{2.18}$$

where

$$G_i(\cdot) = R\frac{1}{T}\sum_{t=p+1}^{T} \tilde{X}_{it}^*(\cdot)\tilde{X}_{it}^{*\prime}(\cdot)R' \quad \text{and} \quad g_i(\cdot) = R\frac{1}{T}\sum_{t=p+1}^{T} \tilde{X}_{it}^*(\cdot)\tilde{x}_{it}^*(\cdot).$$

Having obtained (2.18), ϑ_{i0} can be estimated by

$$\hat{\vartheta}_i = \arg\min_{\vartheta \in \mathcal{V}} \sum_{t=p+1}^{T} \left\{\tilde{x}_{it}^*(\vartheta) - \hat{\phi}_i(\vartheta)' R \tilde{X}_{it}^*(\vartheta)\right\}^2,$$

with $\mathcal{V} = [\underline{\vartheta}, \bar{\vartheta}] \subset \left(0, \frac{3}{2}\right)$.

Then d_{i0} can be estimated from (2.15) by

$$\hat{d}_i = \arg\min_{d \in \mathcal{D}} \sum_{t=p+1}^{T} \left\{\tilde{y}_{it}^*(d) - \hat{\omega}_i(d, \hat{\vartheta}_i)' Q \tilde{Z}_{it}^*(d, \hat{\vartheta}_i)\right\}^2,$$

with $\mathcal{D} = [\underline{d}, \bar{d}] \subset \left(0, \frac{3}{2}\right)$.

The lower-bound restrictions on the sets $\mathcal{V}$ and $\mathcal{D}$, i.e. $\underline{d}, \underline{\vartheta} > 0$, ensure that the initial-condition terms are asymptotically negligible because they are of size $O_p(T^{-d_i})$ and $O_p(T^{-\vartheta_i})$. The upper-bound restrictions are a consequence of the first-differencing transformation, which is mirrored by working with $d^* \geq 1$.

The estimates $\hat{\vartheta}_i$ and $\hat{d}_i$ are not efficient since they are not jointly estimated. To update the estimates to efficiency, a single Newton step may be taken from these initial estimates, $\hat{\tau}_i = (\hat{d}_i, \hat{\vartheta}_i)$, whose $\sqrt{T}-$consistency we establish in Section 3, as

$$\tau_i = \hat{\tau}_i - \mathbf{H}_T^{-1}(\hat{\tau}_i)\mathbf{h}_T(\hat{\tau}_i), \tag{2.19}$$

where

$$\mathbf{H}_T(\tau) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial\hat{\tilde{v}}_{it}^*(\tau)}{\partial\tau'}\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\tilde{v}}_{it}^*(\tau)\hat{\tilde{v}}_{it}^*(\tau)'\right)^{-1}\frac{\partial\hat{\tilde{v}}_{it}^*(\tau)}{\partial\tau'},$$

and

$$\mathbf{h}_T(\tau) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial\hat{\tilde{v}}_{it}^*(\tau)}{\partial\tau'}\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\tilde{v}}_{it}^*(\tau)\hat{\tilde{v}}_{it}^*(\tau)'\right)^{-1}\hat{\tilde{v}}_{it}^*(\tau)$$

with

$$\hat{\tilde{v}}_{it}^*(\hat{d}_i,\hat{\vartheta}_i) = \tilde{z}_{it}^*(\hat{d}_i,\hat{\vartheta}_i) - \sum_{j=1}^{p}\hat{B}_j(\theta_i)\tilde{z}_{it-j}^*(\hat{d}_i,\hat{\vartheta}_i) - \left\{\zeta\tilde{x}_{it}^*(\hat{d}_i) - \sum_{j=1}^{p}\hat{B}_j(\theta_i)\zeta\tilde{x}_{it-j}^*(\hat{d}_i)\right\}\hat{\beta}_i(\hat{d}_i,\hat{\vartheta}_i).$$

2.2.4 Common Correlated Mean-Group Slope Estimate

In many empirical applications, there is also an interest in obtaining inference on the panel rather than individual series alone. Given the linearity of the model in β_i, we consider the common-correlation mean-group estimate,

$$\hat{\beta}_{CCMG}\left(\hat{d},\hat{\vartheta}\right) := \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i\left(\hat{d}_i,\hat{\vartheta}_i\right). \tag{2.20}$$

This estimate is essentially a GLS mean-group estimate based on the average of individual feasible slope estimates. For the asymptotic analysis of the mean-group estimate, it is standard to use a random coefficients model as in

$$\beta_i = \beta_0 + w_i, \quad w_i \sim iid\left(0,\Omega_w\right),$$

with w_i independent of all other model variables.

2.3 Assumptions and Main Results

We impose and discuss a set of regularity conditions that allow us to derive our asymptotic results.

Assumption 1 (**Long-range dependence and common-factor structure**). *Persistence and cross-section dependence are introduced according to the following:*

1. *The fractional integration parameters, with true values $\vartheta_{i0} \neq d_{i0}$, satisfy $max\{\vartheta_{max}, d_{max}, \delta\} - min\{\underline{\vartheta}, \underline{d}\} < 1/2$, and either $max\{\vartheta_{max}, d_{max}, \delta\} < 5/4$ with $d^* = 1$, or $d^* > max\{\vartheta_{max}, d_{max}, \delta\} - 1/4$.*

2. *The common factor vector satisfies* $f_t = \alpha^f + \Delta_t^{-\delta} z_t^f$, *where* $z_t^f = \sum_{k=0}^{\infty} \Psi_k^f \varepsilon_{t-k}^f$ *with* $\sum_{k=0}^{\infty} k \left|\Psi_k^f\right| < \infty$, *and* $\varepsilon_t^f \sim iid(0, \sigma_f)$, $E\left|\varepsilon_t^f\right|^4 < \infty$.

3. f_t *and* $\epsilon_{\cdot it}$ *are independent, and independent of factor loadings* λ_i *and* γ_i *for all* i *and* t.

4. *Factor loadings* λ_i *and* γ_i *are independent across* i, *and the matrix*

$$\begin{pmatrix} \overline{\gamma\beta} + \bar{\lambda} \\ \bar{\gamma} \end{pmatrix}$$

is full rank.

Assumption 1.1 is a fairly general version of the assumptions used by e.g. Hualde and Robinson [23] and Nielsen [28], additionally ensuring that the projection errors asymptotically vanish with the prescribed choice of d^*. To simplify the presentation, we consider a large enough d^* prescribed in Assumption 1.1 without pointing out a fixed value although for most applications $d^* = 1$ would suffice anticipating $\vartheta_{i0}, \delta, d_{i0} < 5/4$. This condition also requires that the lower bounds of the sets $\mathcal{V}$ and $\mathcal{D}$ not be too apart from other memory parameters when $d_{i0} \in \mathcal{D}$ and $\vartheta_{i0} \in \mathcal{V}$, in which case it is further implied that $\vartheta_{i0} - d_{i0} < 1/2$, i.e. at most weak fractional cointegration.

Assumption 1.2 allows for long-range dependence in the common factors that may also have short-memory dynamics, where the $I(0)$ innovations of f_t are not collinear. The restriction on the number of factors may be relaxed when more covariates are introduced: in general, if there are r covariates, the maximum number of factors that can be featured is $1 + r$ so that the factor space can be spanned. The non-zero mean possibility in the common factor, i.e. when $\alpha^f \neq 0$, allows for a drift in the common factor.

Assumptions 1.3 and 1.4 are standard in the factor models literature and have been used by e.g. Pesaran [29] and Bai [2]. The full rank condition on the factor loadings matrix simplifies the identification of factors with no loss of generality requiring that there be sufficiently many covariates whose sample averages can span the factor space. This is straightforwardly satisfied in case of one common factor.

Assumption 2 (System errors). *The process* ϵ_{it} *has the representation*

$$\epsilon_{it} = \Psi(L; \theta_i) v_{it}$$

where

$$\Psi(s; \theta_i) = I_2 + \sum_{j=1}^{\infty} \Psi_j(\theta_i) s^j$$

and the 2×2 *matrices* Ψ_j *satisfy that*

1. $\sum_{j=1}^{\infty} j \left\|\Psi_j\right\| < \infty$, *det* $\{\Psi(s; \theta_i)\} \neq 0, \quad |s| = 1$ *for* $\theta_i \in \Theta$;

2. *$\Psi(L;\theta_i)$ is twice continuously differentiable in θ_i on a closed neighborhood $\mathcal{N}_r(\theta_{i0})$ of radius $0 < r < 1/2$ about θ_{i0};*

3. *the v_{it} are identically and independently distributed vectors across i and t with zero mean and positive-definite covariance matrix Ω_i, and have bounded fourth-order moments.*

Assumptions 2.1-2.3 are quite standard in the analysis of stationary VAR processes, as were also used by Robinson and Hualde [38], constituting the counterpart conditions for B_j. The first condition rules out possible collinearity in the innovations imposing a standard summability requirement and ensures well-defined functional behaviour at zero frequency, allowing for invertibility. The second condition is needed for the uniform convergence of the Hessian in the asymptotic distribution, and finally the moment requirement in the third condition is in general easily satisfied under Gaussianity. The iid requirement in the last condition may be relaxed to martingale difference innovations whose conditional and unconditional third and fourth order moments are equal, which indicates iid behaviour up to fourth moments.

Assumption 3 (Rank condition). *Based on the time-stacked version of the vector of observables $\tilde{Z}^*_{it}$, $\tilde{\boldsymbol{Z}}^*_i$, the following conditions are satisfied:*

1. *$T^{-1}\tilde{\boldsymbol{Z}}^*_i\tilde{\boldsymbol{Z}}^{*\prime}_i$ is full rank;*

2. *$\left(T^{-1}\tilde{\boldsymbol{Z}}^*_i\tilde{\boldsymbol{Z}}^{*\prime}_i\right)^{-1}$ has finite second order moments.*

Assumption 3.1 is a regularity condition ensuring the existence of the least-square estimate in (2.16) and thus of the slope estimate in (2.17) while Assumption 3.2 is used in the derivation of asymptotic results of the common-correlation mean group estimate described in (2.20). These conditions are used by Pesaran [29] based on stationary $I(0)$ variables.

Under our setup, the common-factor structure that accounts for cross-sectional dependence is projected out, and this adds the extra complexity of dealing with projection errors. In a pure time-series context, Hualde and Robinson [22] derive joint asymptotics for memory and slope parameters without accounting for individual or interactive characteristics of the series. Although the results by Hualde and Robinson [22] are similar to ours, showing our results relies heavily on the projection algebra due to the allowance of cross-section dependence.

The next theorem presents the consistency of slope and long-range dependence parameter estimates that are mainly of interest in structural estimation.

Theorem 1. *Under Assumptions 1-3, as $(n,T)_j \to \infty$,*

$$\begin{Bmatrix} \hat{\beta}_i(\hat{d}_i,\hat{\vartheta}_i) - \beta_{i0} \\ \hat{d}_i - d_{i0} \\ \hat{\vartheta}_i - \vartheta_{i0} \end{Bmatrix} \to_p 0.$$

This result does not require a rate condition on n and T so long as they jointly grow in the asymptotics, and it can be readily extended to include also the other model parameters. This contrasts with the results derived by Robinson and Velasco [39], where only T is required to grow and n can be fixed or increasing in the asymptotics. An increasing T is needed therein since it yields the asymptotics, as is needed here, but projection on cross-section averages for factor structure removal further requires that n grow because the projection errors are of size $O_p\left(n^{-1} + (nT)^{-1/2}\right)$ as shown in Appendix A.1.

Next, we show the joint asymptotic distribution of the parameters, where a rate condition on n and T is imposed to remove the projection error.

Theorem 2. *Under Assumptions 1-3, and if* $\sqrt{T}/n \to 0$ *as* $(n,T)_j \to \infty$,

$$\sqrt{T}\begin{Bmatrix} \hat{\beta}_i(\hat{d}_i, \hat{\vartheta}_i) - \beta_{i0} \\ \hat{d}_i - d_{i0} \\ \hat{\vartheta}_i - \vartheta_{i0} \end{Bmatrix} \to_d N\left(0, A_i B_i A_i'\right).$$

The variance-covariance matrix $A_i B_i A_i'$ has a highly involved analytic expression, but definitions of the estimates $\hat{A}_i$ and $\hat{B}_i$, thus forming the positive semi-definite covariance matrix estimate $\hat{A}_i \hat{B}_i \hat{A}_i'$, are provided in Appendix 2.8.4.

This joint estimation result differs from the one by Robinson and Hualde [38] but is similar to that by Hualde and Robinson [22] in that there can at most be weak cointegration under our setup. Removal of common factors that allow for cross-section dependence brings the extra condition that $Tn^{-2} \to 0$ along with more involved derivations, leading to substantially different proofs from those only outlined in Hualde and Robinson [22]. Under lack of autocorrelation and endogeneity induced by the idiosyncratic innovations, Ergemen and Velasco [16] establish the $\sqrt{T}$-convergence rate in the joint estimation of both slope and fractional integration parameters under weak cointegration, with which our results are also parallel.

We finally consider the asymptotic behaviour of the common correlated mean-group slope estimate.

Theorem 3. *Under Assumptions 1-3, as* $(n,T)_j \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{CCMG}\left(\hat{d}, \hat{\vartheta}\right) - \beta_0\right) \to_d N\left(0, \Omega_w\right).$$

This theorem extends the results by Pesaran [29] and Kapetanios et al. [24] on $I(0)$ and $I(1)$ variables, where this GLS-type estimate now converges at the $\sqrt{n}$ rate without requiring any conditions on the relative growth of n to T. The asymptotic variance-covariance matrix, Ω_w, can be estimated nonparametrically based on the GLS slope estimates by

$$\hat{\Omega}_w\left(\hat{\boldsymbol{d}},\hat{\boldsymbol{\vartheta}}\right)=\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\beta}_i\left(\hat{d}_i,\hat{\vartheta}_i\right)-\hat{\beta}_{CCMG}\left(\hat{\boldsymbol{d}},\hat{\boldsymbol{\vartheta}}\right)\right)\left(\hat{\beta}_i\left(\hat{d}_i,\hat{\vartheta}_i\right)-\hat{\beta}_{CCMG}\left(\hat{\boldsymbol{d}},\hat{\boldsymbol{\vartheta}}\right)\right)'$$

since variability only depends on the heterogeneity of the β_i, and bold indicates parameter vectors.

2.4 Deterministic Trends

While our model in (2.1) can accommodate both deterministic and stochastic unobserved trends via the common factor f_t, this imposes that the trending behaviour be shared by some cross-section units, in particular by those with nonzero factor loadings. This then indicates that among those cross-section units sharing the same trend, the difference is only up to a constant, based on λ_i and γ_i. To relax such a restriction and allow for separate time trends, we extend the model in (2.1) as

$$\begin{aligned}y_{it}&=\alpha_i+\alpha_i^1 q(t)+x_{it}\beta_{i0}+f_t\lambda_i+\Delta_t^{-d_{i0}}\epsilon_{1it},\\x_{it}&=\mu_i+\mu_i^1 r(t)+f_t\gamma_i+\Delta_t^{-\vartheta_{i0}}\epsilon_{2it},\end{aligned}\tag{2.21}$$

where now $q(t)$ and $r(t)$ are known time trends.

The case in which $q(t)$ and $r(t)$ in (2.21) are linear, possibly with drifts, can be straightforwardly analyzed in second differences, at whose first and second differences the time trends are reduced to constants and removed, respectively. Alternatively, projections can be carried out in first differences using an augmented version of the projection matrix described in (2.9) to include ones at its first column, which then mirrors fixed-effects estimation in first differences. In both of these approaches, asymptotics remain the same under the conditions prescribed in Section 2.3: although the series may be overdifferenced in the beginning, they are integrated back by the order of their initial differencing orders after projections to their original integration orders, e.g. for double differencing, as in

$$\Delta_{t-1}^{d-2}\Delta^2 y_{it}\approx\Delta_{t-1}^{d}y_{it}\quad\text{and}\quad\Delta_{t-1}^{\vartheta-2}\Delta^2 x_{it}\approx\Delta_{t-1}^{\vartheta}x_{it}.$$

In cases of (possibly fractional) nonlinearity in $q(t)$ and $r(t)$, such as $t^2, t^3, \log t$ and $\Delta^{-\varphi}1$ with $\varphi>1/2$, removal or estimation of trends may become more complicated as opposed to the linear case. When the orders of trend polynomials are known, the first column of the projection matrix in (2.9) can be augmented accordingly to remove the trending behaviour. Even when $q(t)$ and $r(t)$ are functional trends of known orders, such projection matrix augmentation may prove useful. However, when the orders of trend polynomials are unknown, removal of trends based on projection is not straightforward, though some nonparametric GLS detrending approach might be used. This case is beyond the scope of the present paper and is not further explored.

2.5 Simulations

In this section, we investigate the finite-sample behaviour of our estimates, $\hat{\beta}_i(d_{i0}, \vartheta_{i0})$, $\hat{d}_i$, $\hat{\vartheta}_i$ and $\hat{\beta}_i(\hat{d}_i, \hat{\vartheta}_i)$, by means of Monte Carlo experiments. While we estimate the parameters for each i separately, we can only report the average characteristics. We draw the mean zero Gaussian idiosyncratic innovations vector v_{it} with covariance matrix

$$\Omega = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

where we allow for variations in the signal-to-noise ratio, $\tau = a_{22}/a_{11}$, and the correlation $\rho = a_{12}/(a_{11}a_{22})^{1/2}$. We take $a_{11} = 1$ with no loss of generality, and introduce the short-memory dynamics taking $B_j(\theta_i) = diag\{\theta_{1i}, \theta_{2i}\}$ to generate ϵ_{it}.

We draw the factor loadings as $U(-0.5, 1)$, and then generate serially correlated common factors based on iid innovations drawn as standard normal. The fixed effects are left unspecified since projections and estimations are carried out in first differences. Fixing the cross-section size and time-series length to $n = 10$ and $T = 50$, respectively, we consider the parameter values $\vartheta = 0.75, 1, 1.25$, $d = 0.5, 0.75, 1$, covering both cointegration and noncointegration cases, and $\theta_1 = \theta_2 = 0, 0.5$ with $\rho = 0, 0.5$ for $\delta = 0.4, 1$. For this study, we fix $\beta_{i0}, \tau, d^* = 1$. Simulations are carried out via 1,000 replications.

Tables 2.1 and 2.2 present the bias and RMSE profiles of our estimates for $\theta_1 = \theta_2 = \rho = 0$ and $\theta_1 = \theta_2 = \rho = 0.5$, respectively. Both the feasible and infeasible versions of $\hat{\beta}_{MG}$ have considerably small biases under absence of autocorrelation and endogeneity, with the biases further decreasing in ϑ although their magnitudes increase in δ. In the second setup, where both endogeneity and autocorrelation are present, biases of all parameter estimates show an increase in magnitude due to the simultaneous equation bias stemming from prevalent contemporaneous correlations. Biases of slope estimates are decreasing in the order of cointegration, i.e. $\vartheta - d$. The fractional parameter estimate $\hat{\vartheta}$ remains robust in terms of bias for a given ϑ, and the estimate $\hat{d}$ has a bias generally decreasing in d.

In terms of performance, slope estimates behave well both under absence and presence of autocorrelation and endogeneity, in most cases standard deviations dominating biases in terms of contribution to root mean square errors (RMSE). The fractional parameter estimates $\hat{\vartheta}$ and $\hat{d}$ also perform well.

In order to investigate the contributions of endogeneity and short-memory dynamics separately, we next consider $\theta_1 = \theta_2 = 0$ with $\rho = 0.5$ as well as $\theta_1 = \theta_2 = 0.5$ with $\rho = 0$. Table 2.3 presents the case of endogeneity without short-memory dynamics. Compared to the results in Table 2.1, slope estimates mainly suffer from the simultaneous equation bias caused by $\rho \neq 0$ while the performance of fractional integration parameters are slightly ameliorated. When autocorrelation is introduced instead of endogeneity in Table 2.4, slope estimates perform similarly to the results in Table 2.1. The performance of fractional parameter estimates $\hat{\vartheta}$ and $\hat{d}$, however, are slightly worsened compared to the results in Table 2.1. A further comparison between Tables 2.2 and 2.3 reveals that under endogeneity, short-memory dynamics help both the feasible and infeasible slope estimates in terms of performance in some cases. Introducing endogeneity when short-memory dynamics are already present improves the performance of fractional integration parameter estimates to some extent as can be concluded from the comparison of Tables 2.2 and 2.4.

We also explore the finite-sample behaviour of our estimates under (2.21) taking $r(t), q(t) = t$. As before, estimations are performed in first differences, but the projection matrix in (2.9) is now augmented to include ones in its first column. This way, the estimation method mimics fixed-effects estimation in first differences, and the corresponding bias and RMSE profiles are shown in Tables 2.5 and 2.6. The results in Tables 2.5 and 2.1 are comparable as are the results in Tables 2.6 and 2.2. With the inclusion of linear trends, while both the infeasible and feasible slope estimates have positive and small biases, the fractional integration parameter estimates appear to have been underestimated in general.

Finally, we replicate the results in Table 2.2 taking $n = 5$ and $T = 25$ to explore the small-sample behaviour of the estimates in the most difficult case since the projection errors have a larger role. These results are reported in Table 2.7. In terms of performance, the standard errors roughly double (although for individual estimates the convergence rate is $\sqrt{T}$, through averaging the rate becomes approximately $\sqrt{nT}$) while the bias profiles of slope estimates remain more or less the same. However, fractional integration parameter estimates generally suffer from larger biases compared to the results in Table 2.2.

2.6 An Analysis of the Long-Run Debt and GDP Relationship

2.6.1 Related Literature and Empirical Strategy

The relationship between debt and economic growth has been extensively analyzed based on several different approaches leading to mixed results. Among many others, Elmendorf and Mankiw [14] argue for the negative effect of public debt on growth. Reinhart and Rogoff [33] use a debt-bracketing approach coupled with threshold estimation to conclude that high debt hinders economic growth in developed countries. Baglan and Yoldas [1] show that nonlinearities caused by a common debt-level threshold is insignificant and suggest grouping the countries according to their debt-to-GDP ratios to conclude a common negative relationship between GDP growth and debt for countries with chronically high debt. In line with these findings, Chudik et al. [11] show that debt has a negative and significant effect on growth in the long-run and that debt-level thresholds have no significant effects thus refuting the nonlinearity arguments based on thresholding in debt dynamics. Contrary to these views, DeLong and Summers [13] find a positive effect of debt on

GDP growth arguing that recession periods can lead to a situation in which expansionary fiscal policies may have positive effect on long-run GDP growth.

Overall the existing literature has provided ambiguous conclusions as to whether the relationship between debt and GDP growth is negative or positive due to large differences in their estimation methodologies. Except for the econometric specification by Chudik et al. [11], which constitutes the AR alternative of ours, all others rely on homogeneous slope estimation methods, completely disregarding country characteristics and possible interactions between countries. Such homogeneity assumption on the slope parameter implies that different countries converge to their equilibrium at the same rate and that there is no debt overhang from one country to another, which is implausible given the increasing interdependencies between economies. Although Chudik et al. [11] can address these issues in their cross-sectionally augmented autoregressive distributed lag estimation strategy, they restrict their analyses to $I(0)$ and $I(1)$ assumptions. Just like in the other references, their decision on the stationarity of the dynamics of debt-to-GDP ratio and GDP growth is merely based on unit-root testing. However, as is well known by now, rejecting the null of a unit root does not imply $I(0)$ stationarity in the series since stationarity may also be rejected. Our methodology does not require a priori unit-root or stationarity testing because these and in-between cases are flexibly nested.

We analyze the relationship between real GDP and debt-to-GDP growth rates and the relationship between real GDP and debt in log-levels separately in the following subsections. The former application is aimed at contrasting our findings to those in the literature, and the latter is included for the sake of simplicity in interpretations. In the first part, we use post-war yearly data on debt-to-GDP ratios from Reinhart's database and real GDP data from Angus Maddison's website spanning the time period 1955-2008 for 20 high-income OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, United Kingdom and United States. Real GDP growth rates and debt-to-GDP ratios for each country are plotted in Figures 1 and 2, respectively. In the second part using the same datasets, we use the PPP-based GDP data and construct debt data based on that and debt/GDP data for the time period 1955-2008. Since using only level data might invalidate the results if residuals obtained from regressions are trending, we perform the analysis in logs to ensure this is not the case.

2.6.2 Empirical Analysis of the GDP Growth and Debt-to-GDP Ratio Relationship

We examine the effect of debt[1] on economic growth using our fractionally integrated panel data estimation methodology. Using our approach, we incorporate country-specific characteristics and the interactions between countries while also allowing for endogeneity without having to restrict

[1] We use data on central government debt since this is the only available data for all the high-income OECD countries that we consider.

our analysis to $I(0)$ and $I(1)$ cases, by which we are able to detect stationarity and nonstationarity of fractional orders.

From Figures 1 and 2, it is evident that real GDP growth rates show more oscillations, which is a typical behaviour of stationary series, than debt-to-GDP ratios for all countries. The average growth rate for all countries over time is 3.37% while the average debt-to-GDP ratio is 53.21%. In line with the literature, the correlation coefficient between these averaged series is -0.0983 implying an inverse relationship between debt and growth. Furthermore, we account for cross-section mean and variance characteristics of the series so that we can get accurate inference on the long-run relationship between growth rates and debt-to-GDP ratios, if any.

First, we estimate the fractional integration orders of real GDP growth rates and debt-to-GDP ratios using local Whittle estimation based on Robinson [35] with bandwidth choices of $m = 10, 14$. Given that the sample contains 54 time-series data points, choosing higher Fourier frequencies will lead to short-memory contamination in the estimates. The estimation results are collected in Table 8.

The results in Table 8 suggest that real GDP growth rates may in fact be integrated of fractional orders and even be mildly nonstationary[2] although they are always considered to be $I(0)$ variables in the literature. While the null of $I(0)$ stationarity in GDP growth rates cannot be rejected for several countries given the standard errors of their memory estimates, there are also other countries in our sample whose growth rates are significantly fractionally integrated of different orders, thus justifying our approach.

The integration order estimates of debt-to-GDP ratios presented in Table 8 are all significant and around unity, indicating high persistence but of varying orders. Chudik et al. [11] use debt-to-GDP growth rates in their analysis, for which we present the integration orders also in Table 8. These fractional integration or memory estimates suggest that debt-to-GDP growth can still be persistent for some countries with varying magnitudes.

We also estimate the fractional integration order of the common factor based on the cross-section average of both of the series together, which proxy the factor structure well as is evident from (2.7), using local Whittle estimation based on Robinson [35]. The common factor is integrated of orders 0.7577 and 0.7067 for $m = 10, 14$, respectively, providing evidence that the cross-section dependence is persistent itself, which has not been considered in this literature so far.

Having obtained the integration order estimates for GDP growth and debt-to-GDP ratio as well as debt-to-GDP ratio growth, we analyze the relationship between real GDP growth rates and debt-to-GDP growth rates, as is the case in Chudik et al. [11], for two reasons: first, regressing GDP growth, which is stationary for most countries, on debt-to-GDP ratio, which is highly nonstationary, is completely uninformative whereas a regression based on the change in the debt-to-GDP ratio, which has almost the same persistence characteristics as GDP growth, can prove insightful; second, interpretation of the results is more useful since our primary focus is on

[2] Chudik et al. [11] also point out that growth rates may be mildly nonstationary and use this information to select sufficiently many lags in their ARDL specification.

determining how economic growth responds to a change in the debt-to-GDP ratio.

We therefore estimate (2.1) taking y_{it} as the real GDP growth and x_{it} as the debt-to-GDP ratio growth of country i, based on our methodology in which we account for country-specific characteristics, such as institutions and geographical location, as well as characteristics that are common for all countries – OECD membership, high income, etc. Our estimation methodology also allows for the two-way endogeneity between the debt-to-GDP ratio and real GDP growth since the idiosyncratic innovations are allowed to be correlated in the model, which is called for in this analysis as has been discussed by Baglan and Yoldas [1] and Chudik et al. [11]. The estimation results, taking $d^* = 1.25$ and assuming a VAR(1) structure in the idiosyncratic innovations, are reported in Table 9. For all countries but Italy, slope coefficient estimates are insignificant at the 5% level, indicating that debt-to-GDP growth and GDP growth do not have a relationship. For Italy, the slope estimate is positive and significant, but the long-range dependence parameter estimates are both insignificant, implying that the relationship between debt-to-GDP growth and GDP growth only has a short-term nature.

Moreover, there is no statistically significant evidence for a cointegrating relationship between economic growth and debt growth for any of the countries, which can be simply checked by means of a $t-$test constructed as $t = (\hat{\vartheta}_i - \hat{d}_i)/s.e.(\hat{\vartheta}_i - \hat{d}_i)$ in the direction $\hat{\vartheta}_i > \hat{d}_i$. This leads to the conclusion that there is no long-lasting equilibrium relationship between GDP growth and debt growth. Along with most of the claims in the literature, this could be due to the net direction of the causality between these variables being undetermined in the longer run: while high debt burden may have an adverse impact on economic growth, low GDP growth (by reducing tax revenues and increasing public expenditures) could also lead to high debt-to-GDP ratios.

2.6.3 Empirical Analysis of the Relationship between GDP and Debt in Log-Levels

In structural estimation, using comparable level data, such as GDP and debt, leads to easy-to-interpret results. With this in mind, we repeat the analysis in the previous subsection using real GDP and debt in log-levels, whose persistence characteristics we expect to be similar, so that we can identify possible long-run relationships. This way, we can guarantee that the results have clear interpretations.

We find that both real GDP and debt levels exhibit different cross-section mean and volatility characteristics, which we take into account so that valid comparisons can be made. We plot real GDP and debt at levels after normalizations in Figures 3 and 4, respectively.

For both series, there is a clear trending behaviour, leading us to think that they are both nonstationary series. To verify this, we carry out local Whittle estimations on logs of the level series using $m = 10, 14$ Fourier frequencies. The results are collected in Table 10.

The estimation results show that real GDP and debt in logs are integrated of an order around unity, which is in line with the literature where they are treated as $I(1)$ variables. The common factor of real GDP and debt is estimated based on the cross-section averages of the stacked series and is integrated of orders 1.0042 and 0.9272 for $m = 10, 14$, respectively, indicating that removing the common factor is essential for disclosing possible cointegrating relationships. To verify this statement, we provide benchmark estimation results based on the pure time-series estimation approach by Hualde and Robinson [22] assuming a VAR(1) structure. Along this line, to understand the long-run relationships, we are interested in identifying cointegrating relationships. Nontrivial cointegrating relationships between real GDP and debt exist if a) the slope coefficients are significantly different from zero; b) the estimated integration orders of debt in log-levels are significantly larger than those of the estimation residuals, i.e. $\hat{\vartheta}_i > \hat{d}_i$. These benchmark estimation results are collected in Table 11.

According to the results in Table 11, all the estimates are significant for all countries except Australia and Canada with mixed signs. From these results, it is further indicated that real GDP and debt in logs do not have a cointegrating relationship for any of the countries, which can be simply checked by means of a $t-$test constructed as $t = (\hat{\vartheta}_i - \hat{d}_i)/s.e.(\hat{\vartheta}_i - \hat{d}_i)$ in the direction $\hat{\vartheta}_i > \hat{d}_i$. This result can be explained as follows. A time-series regression conceptually omits the common-factor structure accounting for cross-section dependence and when the common factor is the main source of persistence, the resulting regression residuals turn out to be persistent thus hindering the identification of a possible cointegrating relationship.

Now, using our model, we check the long-run relationship between real GDP and debt in logs, again assuming a VAR(1) structure. These estimation results are reported in Table 12.

A positive (or negative) slope estimate indicates that a unit-percent change in debt leads to an increase (decrease) in real GDP by $\hat{\beta}_i$%. According to the estimation results in Table 12, we find that debt and real GDP in logs have a significant relationship for all countries except New Zealand and the United States. The significant effect of debt on GDP is positive for Belgium, Canada, Finland, France, Germany, Ireland, Spain and Sweden, and it is negative and significant for the remaining countries. While a negative and significant effect of debt on real GDP is generally reported in the literature, a positive effect can be, for example, due to the debt increasing because of government spending while also fuelling real GDP; also see DeLong and Summers [13].

The relationship between real GDP and debt does not have a cointegration nature for Australia, Belgium, Canada, Finland, Netherlands, Norway and the United Kingdom, which suggests that the significant interplay between the variables has a short-term nature. On the other hand, we find a cointegrating relationship between real GDP and debt for Austria, Denmark, France, Germany, Greece, Ireland, Italy, Japan, Portugal, Spain and Sweden. While it cannot exactly be claimed that real GDP and debt have a long-term equilibrium relationship in the strict macroeconomic terms when $\vartheta_{i0} - d_{i0} > 1/2$, there still is a clear co-movement between these indicators.

To conclude, using our methodology we find that real GDP and debt have a cointegrating relationship for several high-income OECD countries while the impact can be positive or negative across countries. These cointegration findings contrast well to the benchmark estimation results in Table 11 where we could not find any cointegration due to the negligence of individual country characteristics and cross-country dependence. That is to say, if heterogeneity and interdependencies across countries are not taken into account in analyses of economic unions, as in a pure time-series estimation, identifying the true nature of the relationships between these variables will not be possible.

2.7 Final Comments

We have considered a fractionally integrated panel data system with individual stochastic components and cross-section dependence, which allows for a cointegrated system analysis in the defactored observed series. Although the present paper is quite general in that it incorporates long-range dependence and short-memory dynamics with the allowance of deterministic time trends, it nevertheless can be extended nontrivially in the following directions. The parametric factor structure inducing cross-section dependence in our model may be assumed to have been approximated by weak factors thus capturing spatial dependence in the idiosyncratic innovations; see Chudik et al. [12]. While this is a theoretical possibility in (2.1) with additional conditions on the common factor, f_t, we do not analyze spatial dependence explicitly. Parametric modelling of spatial dependence, see e.g. Pesaran and Tosetti [30], may provide further insights. Moreover, a multiple regression framework can be considered through the allowance of vector x_{it} whose elements display different degrees of persistence. While the extension is trivial when the entire vector displays the same persistence characteristics, the treatment of unit-varying persistence is likely to complicate the uniformity arguments shown in this paper. This extension, however, may allow for the identification of multiple cointegrating relationships. Finally, the fractionally integrated latent factor structure may be estimated and those estimates may be used as plug-in estimates in drawing inference on other model parameters, thus allowing the model to be used in forecasting studies. PCA estimation of fractionally integrated factor models are yet to be explored in the literature.

2.8 Technical Appendix

2.8.1 Proof of Theorem 1

Projections are carried out based on (2.9). Denoting $\bar{\mathbf{z}}(d^*, d^*) \equiv \bar{\mathbf{z}}(d^*)$, let us write

$$\mathbf{x}_i'(d^*)\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) = \mathbf{x}_i'(d^*)\mathbf{I}_{T_1}\mathbf{F}(d^*) - \mathbf{x}_i'(d^*)\bar{\mathbf{z}}(d^*)(\bar{\mathbf{z}}'(d^*)\bar{\mathbf{z}}(d^*))^{-}\bar{\mathbf{z}}'(d^*)\mathbf{F}(d^*), \tag{2.22}$$

with

$$\bar{\mathbf{z}}(d^*) = \mathbf{F}(d^*)\bar{\mathbf{C}} + \bar{\mathcal{E}}\left(\mathbf{d}^* - \mathbf{d}, \mathbf{d}^* - \boldsymbol{\vartheta}\right) \tag{2.23}$$

where bold indicates the vector of parameters with the critical parameter values being d_{max} and ϑ_{max}, and

$$\bar{\mathbf{C}} = \begin{pmatrix} \overline{\gamma\beta} + \bar{\lambda} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \bar{\gamma} \end{pmatrix} \quad \text{and} \quad \bar{\mathcal{E}}\left(\mathbf{d}^* - \mathbf{d}, \mathbf{d}^* - \boldsymbol{\vartheta}\right) = \bar{\boldsymbol{\epsilon}}\left(\mathbf{d}^* - \mathbf{d}, \mathbf{d}^* - \boldsymbol{\vartheta}\right) + \bar{\boldsymbol{\epsilon}}_2\left(\mathbf{d}^* - \boldsymbol{\vartheta}\right)\bar{\beta}\zeta'.$$

Suppressing the notation as $\bar{\mathcal{E}}\left(\mathbf{d}^* - \mathbf{d}, \mathbf{d}^* - \boldsymbol{\vartheta}\right) \equiv \bar{\mathcal{E}}$, the elements of the second term on the RHS of (2.22) can be expressed as

$$\begin{aligned}
T_1^{-1}\mathbf{x}_i'(d^*)\bar{\mathbf{z}}(d^*) &= T_1^{-1}\mathbf{x}_i'(d^*)\mathbf{F}(d^*)\bar{\mathbf{C}} + T_1^{-1}\mathbf{x}_i'(d^*)\bar{\mathcal{E}} \\
T_1^{-1}\bar{\mathbf{z}}'(d^*)\bar{\mathbf{z}}(d^*) &= T_1^{-1}\bar{\mathbf{C}}'\mathbf{F}'(d^*)\mathbf{F}(d^*)\bar{\mathbf{C}} + T_1^{-1}\bar{\mathbf{C}}'\mathbf{F}'(d^*)\bar{\mathcal{E}} + T_1^{-1}\bar{\mathcal{E}}'\mathbf{F}(d^*)\bar{\mathbf{C}} + T_1^{-1}\bar{\mathcal{E}}'\bar{\mathcal{E}} \\
T_1^{-1}\bar{\mathbf{z}}'(d^*)\mathbf{F}(d^*) &= T_1^{-1}\bar{\mathbf{C}}'\mathbf{F}'(d^*)\mathbf{F}(d^*) + T_1^{-1}\bar{\mathcal{E}}'\mathbf{F}(d^*).
\end{aligned}$$

By Assumption 2,

$$\bar{\boldsymbol{\epsilon}}_t = \Psi(L; \boldsymbol{\theta})\bar{\mathbf{v}}_t, \quad \boldsymbol{\theta} \in \Theta,$$

with $\sum_{j=1}^{\infty} j \left\|\Psi_j\right\| < K$, where K is a positive constant. Thus, projections based on $\bar{\mathbf{v}}_t$ and $\bar{\boldsymbol{\epsilon}}_t$ incur errors of the same asymptotic size, and we will show the results in this simpler case to motivate the main ideas.

Then, by Lemma 1, as $n \to \infty$, the projection error, which is the sum of the terms containing $\bar{\mathcal{E}}$, is of size

$$O_p\left(\frac{1}{n} + \frac{1}{\sqrt{nT}}\right) = o_p(1).$$

Denote the projection matrix containing the true factors $\mathbf{M}_F$. By the idempotence of the projection matrix, this result implies that

$$\mathbf{x}_i'(d^*)\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) = \mathbf{x}_i'(d^*)\mathbf{M}_F\mathbf{F}(d^*) + O_p\left(\frac{1}{n} + \frac{1}{\sqrt{nT}}\right), \tag{2.24}$$

indicating that $\bar{\mathbf{M}}_{T_1}$ can replace $\mathbf{M}_F$ as $n \to \infty$, which is useful for the asymptotic analysis. Furthermore,

$$T_1^{1/2}\mathbf{x}_i'(d^*)\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) = T_1^{1/2}\mathbf{x}_i'(d^*)\mathbf{M}_F\mathbf{F}(d^*) + O_p\left(\frac{\sqrt{T}}{n}\right). \tag{2.25}$$

Using the projection arguments above, we first show the consistency of $\hat{\beta}_i(d_{i0}, \vartheta_{i0})$, taking for simplicity $p = 0$ together with the notation $d = d_{i0}$ and $\vartheta = \vartheta_{i0}$, corresponding to the unfeasible

LS estimate with no short-memory dynamics. Then in (2.14), denoting $\sum_t = \sum_{t=2}^{T}$,

$$\hat{\beta}_i(d,\vartheta) = \frac{\sum_t \tilde{x}_{it}^{*\,2}(\vartheta)\sum_t \tilde{x}_{it}^{*}(d)\tilde{y}_{it}^{*}(d) - \sum_t \tilde{x}_{it}^{*}(d)\tilde{x}_{it}^{*}(\vartheta)\sum_t \tilde{x}_{it}^{*}(\vartheta)\tilde{y}_{it}^{*}(d)}{\sum_t \tilde{x}_{it}^{*\,2}(d)\sum_t \tilde{x}_{it}^{*\,2}(\vartheta) - \left(\sum_t \tilde{x}_{it}^{*}(d)\tilde{x}_{it}^{*}(\vartheta)\right)^2},$$

from which we can write

$$\hat{\beta}_i(d,\vartheta) - \beta_{i0} = \frac{\sum_t \tilde{x}_{it}^{*\,2}(\vartheta)\sum_t \tilde{x}_{it}^{*}(d)\tilde{v}_{1.2it}^{*} - \sum_t \tilde{x}_{it}^{*}(d)\tilde{x}_{it}^{*}(\vartheta)\sum_t \tilde{x}_{it}^{*}(\vartheta)\tilde{v}_{1.2it}^{*}}{\sum_t \tilde{x}_{it}^{*\,2}(d)\sum_t \tilde{x}_{it}^{*\,2}(\vartheta) - \left(\sum_t \tilde{x}_{it}^{*}(d)\tilde{x}_{it}^{*}(\vartheta)\right)^2}, \tag{2.26}$$

where $\tilde{v}_{1.2it}^{*} = \tilde{v}_{1it}^{*} - \rho_i \tilde{v}_{2it}^{*}$. Now noting that $Cov\left(\tilde{v}_{2it}^{*}, \tilde{v}_{1.2it}^{*}\right) = 0$, and using the projection arguments above,

$$\begin{aligned}\hat{\beta}_i(d,\vartheta) - \beta_{i0} &= O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{n}\right) \\ &= o_p(1).\end{aligned}$$

We then show the consistency of $\hat{\vartheta}_i$ taking $p = 0$ because the proof follows exactly the same steps for other p values. Write the time-stacked CSS as

$$L_{i,T}(\vartheta) = \frac{1}{T}\tilde{\mathbf{x}}_i^{*}(\vartheta)\tilde{\mathbf{x}}_i^{*\prime}(\vartheta), \tag{2.27}$$

for $\vartheta \in \mathcal{V} = [\underline{\vartheta}, \bar{\vartheta}] \subset \left(0, \frac{3}{2}\right)$. Now,

$$\tilde{\mathbf{x}}_i^{*}(\vartheta) = \Delta^{\vartheta - d^*}\Delta^{d^*-1}\Delta\tilde{\mathbf{x}}_i,$$

where

$$\begin{aligned}\Delta^{d^*-1}\Delta\tilde{\mathbf{x}}_i &= \Delta^{d^*-1}\Delta\mathbf{x}_i - \hat{\varsigma}_x\bar{\mathbf{z}}(d^*) \\ &= \Delta^{d^*-1}\Delta\mathbf{x}_i - \Delta^{d^*-1}\Delta\mathbf{x}_i\bar{\mathbf{z}}'(d^*)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\bar{\mathbf{z}}(d^*)\end{aligned}$$

so that

$$\Delta^{\vartheta - d^*}\Delta^{d^*-1}\Delta\tilde{\mathbf{x}}_i = \Delta^{\vartheta-1}\Delta\mathbf{x}_i - \hat{\varsigma}_x\bar{\mathbf{z}}(\vartheta).$$

Next, to be able to make use of (2.24), let us write

$$\Delta^{\vartheta-1}\Delta\tilde{\mathbf{x}}_i = I_x + J_x$$

with

$$\begin{aligned}I_x &= \Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i} - \Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\mathbf{F}'(d^*)\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\mathbf{F}(\vartheta), \\ J_x &= \Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\left\{\mathbf{F}'(d^*)\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\mathbf{F}(\vartheta) - \bar{\mathbf{z}}'(d^*)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\bar{\mathbf{z}}(\vartheta)\right\}\end{aligned}$$

where $\mathbf{F}(d^*) = (f_2(d^*), \ldots, f_T(d^*))'$. Then using the notation

$$
\begin{aligned}
M_f &:= M_f(\vartheta) = \mathbf{F}'(d^*) \left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1} \mathbf{F}(\vartheta), \\
M_z &:= M_z(\vartheta) = \bar{\mathbf{z}}'(d^*)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\bar{\mathbf{z}}(\vartheta),
\end{aligned}
$$

we can write (2.27) as

$$
\begin{aligned}
&\frac{1}{T}\left\{\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i} - \Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}M_f + \Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\left(M_f - M_z\right)\right\} \\
&\quad \times \left\{\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i} - \Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}M_f + \Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\left(M_f - M_z\right)\right\}',
\end{aligned}
$$

where it suffices to check only the squared terms since the cross terms are bounded from above by the Cauchy-Schwarz inequality. The first squared term,

$$
\frac{1}{T}\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i}\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i}',
$$

converges uniformly in ϑ to the variance of $\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i}$ and is minimized for $\vartheta = \vartheta_{i0}$ as in the proof of Theorem 3.3 of Robinson and Velasco [39] and Theorem 1 of Ergemen and Velasco [16]. To show that the second squared term is negligible, write

$$
\frac{1}{T}\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}M_f M_f'\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}'
$$

where

$$
M_f M_f' = \mathbf{F}'(d^*)\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\mathbf{F}(\vartheta)\mathbf{F}(\vartheta)'\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\mathbf{F}(d^*) \tag{2.28}
$$

satisfying under Assumption 1 that

$$
\begin{aligned}
\frac{\mathbf{F}(d^*)\mathbf{F}'(d^*)}{T} &\to_p \Sigma_f > 0 \\
\sup_{\vartheta\in\mathcal{V}}\left|\frac{\mathbf{F}(\vartheta)\mathbf{F}(\vartheta)'}{T}\right| &= O_p\left(1 + T^{2(\delta-\underline{\vartheta})-1}\right) = O_p(1)
\end{aligned}
$$

which is shown by Lemma 2. Now since, by Lemma 3,

$$
\frac{\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\mathbf{F}'(d^*)}{T} = O_p\left(T^{-1/2} + T^{\vartheta_{i0}+\delta-2d^*-1}\right) = o_p(1),
$$

and applying (2.28), we have that

$$
\sup_{\vartheta\in\mathcal{V}}\left|\frac{1}{T}\Delta^{d^*-\vartheta}\mathbf{v}_{2i}M_f M_f'\Delta^{d^*-\vartheta}\mathbf{v}_{2i}'\right| = o_p(1).
$$

The third squared term

$$\sup_{\vartheta\in\mathcal{V}}\left|\frac{1}{T}\Delta^{d^*-\vartheta}\mathbf{v}_{2i}\left(M_f-M_z\right)\left(M_f-M_z\right)'\Delta^{d^*-\vartheta}\mathbf{v}_{2i}'\right|=o_p(1)$$

because

$$\mathbf{F}(d^*)M_zM_z'\mathbf{F}'(d^*)=\mathbf{F}(d^*)\bar{\mathbf{z}}'(d^*)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\bar{\mathbf{z}}(\vartheta)\bar{\mathbf{z}}'(\vartheta)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\bar{\mathbf{z}}(\vartheta)\mathbf{F}'(d^*)$$

for which it is shown in Lemma 4 that

$$\sup_{\vartheta\in\mathcal{V}}\left|\frac{\mathbf{F}(d^*)M_zM_z'\mathbf{F}'(d^*)}{T}\right|=O_p\left(\frac{1}{n}+\frac{1}{\sqrt{nT}}+\frac{T^{2(\vartheta_{max}-\underline{\vartheta})-1}}{\sqrt{n}}+\frac{T^{\vartheta_{max}+\delta-2\underline{\vartheta}-1}}{\sqrt{n}}\right)=o_p(1).$$

The proof of consistency for $\hat{\vartheta}_i$ is then complete.

The consistency of $\hat{d}_i$ in the time-stacked CSS

$$\hat{d}_i=\arg\min_{d\in\mathcal{D}}\frac{1}{T}\left(\tilde{\mathbf{y}}_i^*(d)-\hat{\omega}_i(d,\hat{\vartheta}_i)'Q\tilde{\mathbf{Z}}_i^*(d,\hat{\vartheta}_i)\right)\left(\tilde{\mathbf{y}}_i^*(d)-\hat{\omega}_i(d,\hat{\vartheta}_i)'Q\tilde{\mathbf{Z}}_i^*(d,\hat{\vartheta}_i)\right)'$$

with $\mathcal{D}=[\underline{d},\bar{d}]\subset\left(0,\frac{3}{2}\right)$ can be shown using exactly the same line of reasoning as above additionally incorporating the estimation effects of $\hat{\omega}_i$ that are uniformly $O_p(T^{-1/2})$ in d based on the arguments in Hualde and Robinson [22], and thus the proof is omitted.

Finally, establishing

$$\hat{\beta}_i(\hat{d}_i,\hat{\vartheta}_i)-\beta_{i0}=o_p(1)$$

follows from the Mean Value Theorem writing

$$\hat{\beta}_i(\hat{\tau})-\beta_{i0}=\hat{\beta}_i(\hat{\tau})-\hat{\beta}_i(\tau)+\hat{\beta}_i(\tau)-\beta_{i0}\quad\text{with}\quad\tau=(d_{i0},\vartheta_{i0}),\tag{2.29}$$

where

$$\hat{\beta}_i(\hat{\tau})-\hat{\beta}_i(\tau)=\dot{\hat{\beta}}_i(\tau^{\ddagger})\left(\hat{\tau}-\tau\right)$$

with $\dot{\hat{\beta}}_i(\tau^{\ddagger})=O_p(1)$ for some intermediate-value vector $\tau^{\ddagger}$, as is the case in Robinson and Hualde [38], and using that $\hat{\tau}-\tau=O_p\left(T^{-1/2}\right)$. □

2.8.2 Proof of Theorem 2

Asymptotic normality of the slope estimates can readily be established based on (2.29), (2.26) and (2.25)

$$\sqrt{T}\left(\hat{\beta}_i(\hat{d}_i,\hat{\vartheta}_i)-\beta_{i0}\right)=N(0,\Sigma_\beta)+O_p\left(\frac{\sqrt{T}}{n}\right)$$

where Σ_β is the variance-covariance matrix obtained from (2.26) in the usual way, and the O_p term on the RHS appears due to projection error, which is removed if $\sqrt{T}/n\to 0$ as $n\to\infty$.

Showing the asymptotic normality of $\hat{\vartheta}_i$ and $\hat{d}_i$ follows the same steps, which is why we only prove the result for $\hat{\vartheta}_i$ to focus on the main ideas. The $\sqrt{T}-$normalized score evaluated at the true value, ϑ_{i0}, is given by

$$\begin{aligned}\sqrt{T}\left.\frac{\partial L_{i,T}(\vartheta)}{\partial\vartheta}\right|_{\vartheta=\vartheta_{i0}} &= \frac{2}{\sqrt{T}}\left\{\mathbf{v}_{2i}-\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}M_{f,0}+\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\left(M_{f,0}-M_{z,0}\right)\right\}\\ &\quad\times\left\{\left(\log\Delta_t\right)\mathbf{v}_{2i}-\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\dot{M}_{f,0}+\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\left(\dot{M}_{f,0}-\dot{M}_{z,0}\right)\right\}'\end{aligned}$$

where

$$\begin{aligned}M_{f,0} &:= M_f(\vartheta_{i0})=\mathbf{F}'(d^*)\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\mathbf{F}(\vartheta_{i0}),\\ M_{z,0} &:= M_z(\vartheta_{i0})=\bar{\mathbf{z}}'(d^*)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\bar{\mathbf{z}}(\vartheta_{i0}),\\ \dot{M}_{f,0} &:= \dot{M}_f(\vartheta_{i0})=\mathbf{F}'(d^*)\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\dot{\mathbf{F}}(\vartheta_{i0}),\\ \dot{M}_{z,0} &:= \dot{M}_z(\vartheta_{i0})=\bar{\mathbf{z}}'(d^*)(\bar{\mathbf{z}}(d^*)\bar{\mathbf{z}}'(d^*))^{-1}\dot{\bar{\mathbf{z}}}(\vartheta_{i0}),\end{aligned}$$

and $\dot{\mathbf{F}}(\vartheta)=(\partial/\partial\vartheta)\,\mathbf{F}(\vartheta)$. Taking $n=1$, as $T\to\infty$, the term

$$\frac{2}{\sqrt{T}}\mathbf{v}_{2i}\left[\left(\log\Delta_t\right)\mathbf{v}_{2i}\right]'\to_d N(0,4\sigma_{v_2})$$

applying a central limit theorem for martingale difference sequences as shown by Robinson and Velasco [39].

Next, we show that the remaining terms are negligible. To do so, we only check the dominating terms since the other terms containing d^* have smaller sizes. The expression

$$\frac{2}{\sqrt{T}}\mathbf{v}_{2i}\dot{M}_{f,0}\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}'=\frac{2}{\sqrt{T}}\mathbf{v}_{2i}\mathbf{F}'(d^*)\left(\mathbf{F}(d^*)\mathbf{F}'(d^*)\right)^{-1}\dot{\mathbf{F}}(\vartheta_{i0})\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}'=o_p(1)$$

based on the results in Lemma 5.

The term dealing with the projection approximation,

$$\frac{2}{\sqrt{T}}\mathbf{v}_{2i}\left(\dot{M}_{f,0}-\dot{M}_{z,0}\right)\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}'$$

can easily be shown as in Ergemen and Velasco [16] to be $o_p(1)$ following the same steps described earlier. All other cross terms are negligible using similar arguments so the result follows.

Finally, uniform convergence of the Hessian can be shown following the arguments in Theorem 2 of Hualde and Robinson [23], and the proof is then complete. □

2.8.3 Proof of Theorem 3

The asymptotic behaviour of the mean-group slope estimate is readily shown in Pesaran [29] under the rank condition and the random coefficients model we described. The long-range dependence parameter estimation effects are $O_p(T^{-1/2})$, for which we need that $T \to \infty$ (as well as $n \to \infty$ that yields the asymptotics), but no further condition on the relative growth of n or T is needed. □

2.8.4 Covariance Matrix Estimate $\hat{A}_i\hat{B}_i\hat{A}_i'$

Definitions of the variance-covariance matrix components are comparable to those obtained by Hualde and Robinson [22]. The main exception under our setup is that these matrices must be constructed based on the projected series, which is clearly not a concern in the pure time series setup of Hualde and Robinson [22].

Denote $\hat{M}_i \equiv M_i(\hat{d}_i, \hat{\vartheta}_i)$, $\hat{\omega}_i \equiv \hat{\omega}_i(\hat{d}_i, \hat{\vartheta}_i)$, $\hat{G}_i \equiv G_i(\hat{\vartheta}_i)$, and $\hat{\phi}_i \equiv \hat{\phi}_i(\hat{\vartheta}_i)$. Then,

$$\hat{A}_i = \begin{pmatrix} \hat{a}_{i1}' & \hat{a}_{i2} & \hat{a}_{i3} \\ (0,\ldots,0)' & \hat{a}_{i4} & \hat{a}_{i5} \\ (0,\ldots,0)' & 0 & \hat{a}_{i6} \end{pmatrix},$$

with

$$\hat{a}_{i1}' = (1,0,\ldots,0)'\hat{M}_i^{-1}, \quad \hat{a}_{i2} = -(1,0,\ldots,0)'\hat{\omega}_{i\tau_1}\hat{s}_{i\tau_1\tau_1}^{-1},$$
$$\hat{a}_{i3} = (1,0,\ldots,0)'\hat{\omega}_{i\tau_1}\hat{s}_{i\tau_1\tau_1}^{-1}\hat{s}_{i\tau_1\tau_2}\hat{s}_{i\tau_2\tau_2}^{-1} - (1,0,\ldots,0)'\hat{\omega}_{i\tau_2}\hat{s}_{i\tau_2\tau_2}^{-1},$$
$$\hat{a}_{i4} = -\hat{s}_{i\tau_1\tau_1}^{-1}, \quad \hat{a}_{i5} = \hat{s}_{i\tau_1\tau_1}^{-1}\hat{s}_{i\tau_1\tau_2}\hat{s}_{i\tau_2\tau_2}^{-1}, \quad \hat{a}_{i6} = -\hat{s}_{i\tau_2\tau_2}^{-1},$$

where

$$
\begin{aligned}
\hat{\omega}_{i\tau_1} &= \hat{M}_i^{-1}\left(\hat{m}_{i\tau_1} - \hat{M}_{i\tau_1}^{-1}\hat{\omega}_i\right), \quad \hat{\omega}_{i\tau_2} = \hat{M}_i^{-1}\left(\hat{m}_{i\tau_2} - \hat{M}_{i\tau_2}^{-1}\hat{\omega}_i\right), \\
\hat{m}_{i\tau_1} &= Q\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{Z}^*_{it\tau_1}(\hat{d}_i)\tilde{y}^*_{it}(\hat{d}_i) + \tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i)\tilde{y}^*_{it\tau_1}(\hat{d}_i)\right\}, \\
\hat{M}_{i\tau_1} &= Q\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{Z}^*_{it\tau_1}(\hat{d}_i)\tilde{Z}^{*\prime}_{it}(\hat{d}_i,\hat{\vartheta}_i) + \tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i)\tilde{Z}^{*\prime}_{it\tau_1}(\hat{d}_i)\right\}Q', \\
\hat{m}_{i\tau_2} &= Q\frac{1}{T}\sum_{t=p+1}^{T}\tilde{Z}^*_{it\tau_2}(\hat{\vartheta}_i)\tilde{y}^*_{it}(\hat{d}_i), \\
\hat{M}_{i\tau_2} &= Q\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{Z}^*_{it\tau_2}(\hat{\vartheta}_i)\tilde{Z}^{*\prime}_{it}(\hat{d}_i,\hat{\vartheta}_i) + \tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i)\tilde{Z}^{*\prime}_{it\tau_2}(\hat{\vartheta}_i)\right\}Q'
\end{aligned}
$$

with the parameter subscripts denoting the first partial derivative as in

$$
\begin{aligned}
\tilde{y}^*_{it\tau_1}(\hat{d}_i) &= (\log\Delta)\tilde{y}^*_{it}(\hat{d}_i), \\
\tilde{Z}^*_{it\tau_1}(\hat{d}_i) &= (\log\Delta)\left\{\tilde{x}^*_{it}(\hat{d}_i), 0, \tilde{x}^*_{it-1}(\hat{d}_i), 0, \tilde{y}^*_{it-1}(\hat{d}_i), \ldots, \tilde{x}^*_{it-p}(\hat{d}_i), 0, \tilde{y}^*_{it-p}(\hat{d}_i)\right\}', \\
\tilde{Z}^*_{it\tau_2}(\hat{\vartheta}_i) &= (\log\Delta)\left\{0, \tilde{x}^*_{it}(\hat{\vartheta}_i), 0, \tilde{x}^*_{it-1}(\hat{\vartheta}_i), 0, \ldots, 0, \tilde{x}^*_{it-p}(\hat{\vartheta}_i), 0\right\}'
\end{aligned}
$$

and also

$$
\begin{aligned}
\hat{s}_{i\tau_1\tau_1} &= \frac{1}{T}\sum_{t=p+1}^{T}{\hat{v}^*_{it\tau_1}}^2, \quad \hat{s}_{i\tau_1\tau_2} = \frac{1}{T}\sum_{t=p+1}^{T}\hat{v}^*_{it\tau_1}\hat{v}^*_{it\tau_2}, \quad \hat{s}_{i\tau_2\tau_2} = \frac{1}{T}\sum_{t=p+1}^{T}{\hat{w}^*_{it\tau_2}}^2, \\
\hat{v}^*_{it\tau_1} &= \tilde{y}^*_{it\tau_1}(\hat{d}_i) - \hat{\omega}'_{i\tau_1}Q\tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i) - \hat{\omega}'_iQ\tilde{Z}^*_{it\tau_1}(\hat{d}_i), \\
\hat{v}^*_{it\tau_2} &= -\hat{\omega}'_{i\tau_2}Q\tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i) - \hat{\omega}'_iQ\tilde{Z}^*_{it\tau_2}(\hat{\vartheta}_i), \\
\hat{w}^*_{it\tau_2} &= \tilde{x}^*_{it\tau_2}(\hat{\vartheta}_i) - \hat{\phi}'_{i\tau_2}R\tilde{X}^*_{it}(\hat{\vartheta}_i) - \hat{\phi}'_iR\tilde{X}^*_{it\tau_2}(\hat{\vartheta}_i), \\
\tilde{x}^*_{it\tau_2}(\hat{\vartheta}_i) &= (\log\Delta)\tilde{x}^*_{it}(\hat{\vartheta}_i), \quad \tilde{X}^*_{it\tau_2}(\hat{\vartheta}_i) = (\log\Delta)\tilde{X}^*_{it}(\hat{\vartheta}_i), \\
\hat{\phi}_{i\tau_2} &= \hat{G}_i^{-1}\left(\hat{g}_{i\tau_2} - \hat{G}_{i\tau_2}\hat{\phi}_i\right), \\
\hat{g}_{i\tau_2} &= R\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{X}^*_{it\tau_2}(\hat{\vartheta}_i)\tilde{x}^*_{it}(\hat{\vartheta}_i) + \tilde{X}^*_{it}(\hat{\vartheta}_i)\tilde{x}^*_{it\tau_2}(\hat{\vartheta}_i)\right\}, \\
\hat{G}_{i\tau_2} &= R\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{X}^*_{it\tau_2}(\hat{\vartheta}_i)\tilde{X}^*_{it}(\hat{\vartheta}_i)' + \tilde{X}^*_{it}(\hat{\vartheta}_i)\tilde{X}^*_{it\tau_2}(\hat{\vartheta}_i)'\right\}R'.
\end{aligned}
$$

Finally,

$$\hat{B}_i = \frac{1}{T}\sum_{t=p+1}^{T}\begin{bmatrix}\hat{v}^*_{1.2,it}(\hat{d}_i,\hat{\vartheta}_i)Q\tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i)\\ \hat{v}^*_{1.2,it}(\hat{d}_i,\hat{\vartheta}_i)\hat{v}^*_{it\tau_1}\\ \hat{v}^*_{1.2,it}(\hat{d}_i,\hat{\vartheta}_i)\hat{w}^*_{it\tau_2}\end{bmatrix}\begin{bmatrix}\hat{v}^*_{1.2,it}(\hat{d}_i,\hat{\vartheta}_i)Q\tilde{Z}^*_{it}(\hat{d}_i,\hat{\vartheta}_i)\\ \hat{v}^*_{1.2,it}(\hat{d}_i,\hat{\vartheta}_i)\hat{v}^*_{it\tau_1}\\ \hat{v}^*_{2,it}(\hat{\vartheta}_i)\hat{w}^*_{it\tau_2}\end{bmatrix}',$$

where

$$\begin{aligned}\hat{v}^*_{1.2,it}(\hat{d}_i,\hat{\vartheta}_i) &= \hat{v}^*_{1it}(\hat{d}_i) - \rho_i\hat{v}^*_{2it}(\hat{\vartheta}_i),\\ \hat{v}^*_{2it}(\hat{\vartheta}_i) &= \tilde{x}^*_{it}(\hat{\vartheta}_i) - \hat{\phi}_i' R\tilde{X}^*_{it}(\hat{\vartheta}_i).\end{aligned}$$

2.9 Lemmas

Lemma 1. *For some $d^* > \max\{\vartheta_{max}, d_{max}, \delta\} - 1/4$, following are the stochastic orders of the projection components:*

a.

$$T_1^{-1}\bar{\mathcal{E}}'\bar{\mathcal{E}} = O_p\left(\frac{1}{n} + \frac{1}{\sqrt{nT}}\right),$$

b.

$$T_1^{-1}\bar{\mathcal{E}}'\boldsymbol{F}(d^*) = O_p\left(\frac{1}{\sqrt{nT}}\right),$$

c.

$$T_1^{-1}\bar{\boldsymbol{\epsilon}}_2'(d^* - \vartheta_{max})\bar{\mathcal{E}} = O_p\left(\frac{1}{n} + \frac{1}{\sqrt{nT}}\right),$$

where $\bar{\mathcal{E}} = (\bar{\boldsymbol{\varepsilon}}_2, \ldots, \bar{\boldsymbol{\varepsilon}}_T)'$.

Proof of Lemma 1.a. Let us write

$$\bar{\boldsymbol{\varepsilon}}_t = \begin{pmatrix}\overline{\Delta_t^{d^*-d_{max}}\epsilon_{1t}} + \overline{\Delta_t^{d^*-\vartheta_{max}}\epsilon_{2t}}\\ \overline{\Delta_t^{d^*-\vartheta_{max}}\epsilon_{2t}}\end{pmatrix}.$$

Then,

$$\begin{aligned}T_1^{-1}\left(\sum_{t=2}^{T}\bar{\boldsymbol{\varepsilon}}_t'\bar{\boldsymbol{\varepsilon}}_t\right) &= T_1^{-1}\sum_{t=2}^{T}\left(\overline{\Delta_t^{d^*-d_{max}}\epsilon_{1t}}\right)^2 + T_1^{-1}\sum_{t=2}^{T}\left(\overline{\Delta_t^{d^*-\vartheta_{max}}\epsilon_{2t}}\right)^2 + T_1^{-1}\sum_{t=2}^{T}\left(\overline{\Delta_t^{d^*-\vartheta_{max}}\epsilon_{2t}}\right)^2\\ &\quad + 2T_1^{-1}\sum_{t=2}^{T}\overline{\Delta_t^{d^*-d_{max}}\epsilon_{1t}\Delta_t^{d^*-\vartheta_{max}}\epsilon_{2t}},\end{aligned}$$

whose expectation is $O\left(n^{-1}\right)$ and variance is $O\left((nT)^{-1}\right)$, using Cauchy-Schwarz inequality. Thus,

$$T_1^{-1}\left(\sum_{t=2}^{T}\bar{\boldsymbol{\varepsilon}}_t'\bar{\boldsymbol{\varepsilon}}_t\right) = O_p\left(\frac{1}{n}+\frac{1}{\sqrt{nT}}\right).$$

b. The expression has zero expectation. Using the independence of $\mathbf{f}_t$ and $\bar{\boldsymbol{\varepsilon}}_t$,

$$Var\left(\frac{\sum_{t=2}^{T}\bar{\boldsymbol{\varepsilon}}_t'\mathbf{f}_t}{T_1}\right) = \frac{\sum_{t=2}^{T}\sum_{t'=2}^{T}E(\mathbf{f}_t\mathbf{f}_{t'}')E(\bar{\boldsymbol{\varepsilon}}_t\bar{\boldsymbol{\varepsilon}}_{t'}')}{T_1^2}.$$

which is $O\left(n^{-1}\right)$ times

$$\frac{1}{T_1^2}\sum_{t=2}^{T}\sum_{t'=2}^{T}|t-t'|^{2(\max\{d_{max}-d^*,\vartheta_{max}-d^*\})-1}|t-t'|^{2(\delta-d^*)-1}. \tag{2.30}$$

Take with no loss of generality, $\vartheta_{max} > d_{max}$. Then (2.30) becomes

$$\frac{1}{T_1^2}\sum_{t=2}^{T}\sum_{t'=2}^{T}|t-t'|^{2(\delta+\vartheta_{max}-2d^*-1)} = O\left(T^{-1}\right).$$

Thus, $\frac{\sum_{t=2}^{T}\bar{\boldsymbol{\varepsilon}}_t'\mathbf{f}_t}{T_1} = O_p\left((nT)^{-1/2}\right)$.

c. The expectation of $T_1^{-1}\left(\sum_{t=2}^{T}\bar{\boldsymbol{\varepsilon}}_t\bar{\epsilon}_{2t}\right)$ is $O\left(n^{-1}\right)$ and its variance is $O\left((nT)^{-1/2}\right)$, which can be shown as in Lemma 1.a. Thus, $T_1^{-1}\left(\sum_{t=2}^{T}\bar{\boldsymbol{\varepsilon}}_t\bar{\epsilon}_{2t}\right) = O_p\left(n^{-1}+(nT)^{-1/2}\right)$. □

Lemma 2. *Under Assumption 1,*

$$\begin{aligned}\sup_{\vartheta\in\mathcal{V}}\left|\frac{\boldsymbol{F}(\vartheta)\boldsymbol{F}(\vartheta)'}{T}\right| &= O_p\left(1+T^{2(\delta-\underline{\vartheta})-1}\right)\\ &= O_p(1)\end{aligned}$$

Proof of Lemma 2. The result follows from the arguments in the proofs of Theorems 4-6 of Ergemen and Velasco [16].

Lemma 3. *Under Assumption 1,*

$$\frac{\Delta^{d^*-\vartheta_{i0}}\boldsymbol{v}_{2i}\boldsymbol{F}'(d^*)}{T} = O_p\left(T^{-1/2}+T^{\vartheta_{i0}+\delta-2d^*-1}\right) = o_p(1),$$

Proof of Lemma 3. The result follows from the arguments in the proofs of Theorems 4-6 of Ergemen and Velasco [16].

Lemma 4. *Under Assumption 1,*

$$\sup_{\vartheta \in \mathcal{V}} \left| \frac{\boldsymbol{F}(d^*) M_z M_z' \boldsymbol{F}'(d^*)}{T} \right| = O_p \left(\frac{1}{n} + \frac{1}{\sqrt{nT}} + \frac{T^{2(\vartheta_{max} - \underline{\vartheta}) - 1}}{\sqrt{n}} + \frac{T^{\vartheta_{max} + \delta - 2\underline{\vartheta} - 1}}{\sqrt{n}} \right) = o_p(1).$$

Proof of Lemma 4. The result follows from the arguments in the proofs of Theorems 4-6 of Ergemen and Velasco [16].

Lemma 5. *Under Assumption 1,*

$$\frac{\boldsymbol{v}_{2i} \boldsymbol{F}'(d^*)}{T} = O_p \left(T^{-1/2} + T^{\delta - d^* - 1/2} \right)$$
$$\frac{\dot{\boldsymbol{F}}(\vartheta_{i0}) \Delta^{d^* - \vartheta_{i0}} \boldsymbol{v}_{2i}'}{T} = O_p \left(T^{-1/2} + T^{\delta - d^* - 1} \log T \right).$$

Proof of Lemma 5. The result follows from the arguments in in the proof of Theorem 7 of Ergemen and Velasco [16].

Bibliography

[1] Baglan, D. and E. Yoldas (2013): "Government Debt and Macroeconomic Activity: A Predictive Analysis for Advanced Economies," *Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.*

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

[3] ——— (2010): "Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation," *Econometric Theory*, 26, 1088–1114.

[4] Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

[5] ——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127–1177.

[6] ——— (2013): "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176, 18–29.

[7] Blanchard, O. J. and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79(4), 655–73.

[8] Bollerslev, T., D. Osterrieder, N. Sizova, and G. Tauchen (2013): "Risk and Return: Long-Run Relationships, Fractional Cointegration, and Return Predictability," *Journal of Financial Economics*, 108(2), 409–424.

[9] Chambers, M. J. (1998): "Long Memory and Aggregation in Macroeconomic Time Series," *International Economic Review*, 39(4), 1053–1072.

[10] 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.).*

[11] Chudik, A., K. Mohaddes, H. Pesaran, and M. Raissi (2013): "Debt, Inflation and Growth: Robust Estimation of Long-Run Effects in Dynamic Panel Data Models," *Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper No. 162.*

[12] Chudik, A., H. Pesaran, and E. Tosetti (2011): "Weak and Strong Cross-Section Dependence and Estimation of Large Panels," *The Econometrics Journal*, 14(1), C45–C90.

[13] DeLong, J. B. and L. H. Summers (2012): "Fiscal Policy in a Depressed Economy," *Brookings Papers on Economic Activity*, 233–297.

[14] Elmendorf, D. W. and G. N. Mankiw (1999): *Government Debt*, Elsevier, chap. Volume 1, Part C, 1615–1669.

[15] Ergemen, Y. E. (2015): "Fractionally Integrated Panel Data Systems," *Preprint UC3M*, http://dx.doi.org/10.2139/ssrn.2521050.

[16] Ergemen, Y. E. and C. Velasco (2015): "Estimation of Fractionally Integrated Panels with Fixed-Effects and Cross-Section Dependence," *Preprint UC3M*.

[17] Gil-Alaña, L. and P. Robinson (1997): "Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series," *Journal of Econometrics*, 80(2), 241–268.

[18] Granger, C. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

[19] 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.

[20] 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.

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

[22] Hualde, J. and P. M. Robinson (2007): "Root-N-Consistent Estimation of Weak Fractional Cointegration," *Journal of Econometrics*, 140, 450–484.

[23] ——— (2011): "Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models," *The Annals of Statistics*, 39(6), 3152–3181.

[24] Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): "Panels with Non-Stationary Multifactor Error Structures," *Journal of Econometrics*, 160(2), 326–348.

[25] Marinucci, D. and P. Robinson (2000): "Weak Convergence of Multivariate Fractional Processes," *Stochastic Processes and their Applications*, 86, 103–120.

[26] Michelacci, C. and P. Zaffaroni (2000): "(Fractional) Beta Convergence," *Journal of Monetary Economics*, 45, 129–153.

[27] Moon, H. R. and B. Perron (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122(1), 81–126.

[28] 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.

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

[30] Pesaran, H. and E. Tosetti (2011): "Large Panels with Common Factors and Spatial Correlation," *Journal of Econometrics*, 161(2), 182–202.

[31] Pesaran, M. H. and A. Chudik (2014): "Aggregation in Large Dynamic Panels," *Journal of Econometrics*, 178, 273–285.

[32] Phillips, P. and H. R. Moon (1999): "Linear Regression Limit Theory For Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

[33] Reinhart, C. and K. S. Rogoff (2010): "Growth in a Time of Debt," *American Economic Review*, 100(2), 573–578.

[34] Robinson, P. M. (1978): *Comments on "Some consequences of temporal aggregation in seasonal time analysis models" by W. W. S. Wei*, United States Department of Commerce, Bureau of the Census, Washington, DC, 445–447.

[35] ——— (1995): "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23(5), 1630–1661.

[36] ——— (2005): "Robust Covariance Matrix Estimation : 'HAC' Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21(1), 171–180.

[37] Robinson, P. M. and J. Hidalgo (1997): "Time Series Regression with Long-Range Dependence," *The Annals of Statistics*, 25(1), 77–104.

[38] Robinson, P. M. and J. Hualde (2003): "Cointegration in Fractional Systems with Unknown Integration Orders," *Econometrica*, 71(6), 1727–1766.

[39] Robinson, P. M. and C. Velasco (2015): "Efficient Inference on Fractionally Integrated Panel Data Models with Fixed Effects," *Journal of Econometrics*, 185, 435–452.

[40] Sims, C. A. (1987): "A Rational Expectations Framework for Short Run Policy Analysis," doi=10.1.1.211.9699.

Table 2.1: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0$ and $\rho = 0$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0015	-0.0016	-0.0015	-0.0007	-0.0011	-0.0015	0.0001	-0.0002	-0.0009
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0017	-0.0018	-0.0016	-0.0007	-0.0012	-0.0016	0.0001	-0.0001	-0.0007
	$\hat{\vartheta}$	0.0194	0.0187	0.0160	-0.0072	-0.0070	-0.0075	-0.0056	-0.0055	-0.0056
	$\hat{d}$	0.0052	-0.0092	-0.0201	0.0107	-0.0131	-0.0259	0.0222	-0.0188	-0.0375
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0497	0.0526	0.0510	0.0421	0.0495	0.0527	0.0364	0.0408	0.0497
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0496	0.0526	0.0511	0.0419	0.0493	0.0527	0.0350	0.0408	0.0495
	$\hat{\vartheta}$	0.0320	0.0316	0.0303	0.0256	0.0255	0.0257	0.0133	0.0131	0.0132
	$\hat{d}$	0.0435	0.0435	0.0466	0.0489	0.0445	0.0495	0.0605	0.0483	0.0567
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0018	-0.0018	-0.0016	-0.0015	-0.0016	-0.0018	-0.0008	-0.0009	-0.0014
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0020	-0.0019	-0.0017	-0.0018	-0.0018	-0.0019	-0.0008	-0.0009	-0.0014
	$\hat{\vartheta}$	0.0526	0.0519	0.0495	-0.0025	-0.0027	-0.0032	-0.0047	-0.0047	-0.0049
	$\hat{d}$	0.0704	0.0184	-0.0118	0.0708	0.0133	-0.0177	0.0745	0.0062	-0.0285
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0629	0.0547	0.0514	0.0536	0.0514	0.0530	0.0448	0.0427	0.0498
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0570	0.0542	0.0515	0.0489	0.0510	0.0530	0.0400	0.0425	0.0496
	$\hat{\vartheta}$	0.0644	0.0638	0.0620	0.0249	0.0250	0.0253	0.0120	0.0120	0.0123
	$\hat{d}$	0.0906	0.0487	0.0431	0.0921	0.0479	0.0455	0.0969	0.0485	0.0517

Table 2.2: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0150	-0.0171	-0.0132	-0.0122	-0.0216	-0.0198	-0.0097	-0.0286	-0.0414
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0088	-0.0168	-0.0239	-0.0071	-0.0137	-0.0193	-0.0086	-0.0215	-0.0320
	$\hat{\vartheta}$	0.0368	0.0364	0.0336	0.0234	0.0250	0.0252	-0.0004	-0.0003	-0.0002
	$\hat{d}$	-0.0016	-0.0189	-0.0407	-0.0009	-0.0203	-0.0430	-0.0077	-0.0243	-0.0464
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0450	0.0486	0.0468	0.0379	0.0481	0.0505	0.0301	0.0462	0.0608
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0440	0.0485	0.0513	0.0374	0.0455	0.0502	0.0308	0.0432	0.0550
	$\hat{\vartheta}$	0.0423	0.0420	0.0397	0.0290	0.0303	0.0307	0.0123	0.0124	0.0120
	$\hat{d}$	0.0357	0.0408	0.0551	0.0349	0.0405	0.0564	0.0378	0.0414	0.0589
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0162	-0.0168	-0.0106	-0.0107	-0.0189	-0.0150	-0.0088	-0.0256	-0.0349
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0138	-0.0166	-0.0215	-0.0122	-0.0131	-0.0149	-0.0132	-0.0218	-0.0273
	$\hat{\vartheta}$	0.0437	0.0432	0.0403	0.0246	0.0254	0.0248	-0.0003	-0.0003	-0.0003
	$\hat{d}$	0.0277	-0.0072	-0.0336	0.0244	-0.0097	-0.0369	0.0149	-0.0143	-0.0405
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0486	0.0482	0.0449	0.0414	0.0467	0.0474	0.0331	0.0445	0.0555
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0473	0.0482	0.0492	0.0417	0.0452	0.0475	0.0353	0.0437	0.0514
	$\hat{\vartheta}$	0.0497	0.0493	0.0468	0.0300	0.0306	0.0303	0.0122	0.0121	0.0120
	$\hat{d}$	0.0493	0.0373	0.0498	0.0465	0.0373	0.0520	0.0435	0.0374	0.0544

Table 2.3: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0$ and $\rho = 0.5$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0109	-0.0158	-0.0155	-0.0033	-0.0125	-0.0162	0.0008	-0.0092	-0.0187
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0115	-0.0155	-0.0200	-0.0130	-0.0133	-0.0156	-0.0106	-0.0116	-0.0156
	$\hat{\vartheta}$	0.0202	0.0197	0.0165	-0.0072	-0.0070	-0.0073	-0.0061	-0.0058	-0.0056
	$\hat{d}$	0.0211	0.0007	-0.0153	0.0267	-0.0019	-0.0195	0.0412	0.0016	-0.0202
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0443	0.0477	0.0463	0.0381	0.0443	0.0481	0.0345	0.0369	0.0466
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0449	0.0477	0.0485	0.0403	0.0450	0.0480	0.0345	0.0385	0.0458
	$\hat{\vartheta}$	0.0334	0.0332	0.0317	0.0248	0.0248	0.0251	0.0132	0.0129	0.0127
	$\hat{d}$	0.0432	0.0369	0.0400	0.0479	0.0358	0.0410	0.0619	0.0350	0.0402
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0230	-0.0276	-0.0215	-0.0053	-0.0165	-0.0188	0.0006	-0.0098	-0.0189
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0261	-0.0247	-0.0274	-0.0284	-0.0210	-0.0184	-0.0255	-0.0190	-0.0180
	$\hat{\vartheta}$	0.0540	0.0534	0.0505	-0.0021	-0.0021	-0.0025	-0.0052	-0.0051	-0.0050
	$\hat{d}$	0.0917	0.0352	0.0014	0.0867	0.0267	-0.0085	0.0925	0.0275	-0.0093
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0664	0.0567	0.0494	0.0541	0.0490	0.0494	0.0456	0.0407	0.0471
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0593	0.0539	0.0526	0.0556	0.0505	0.0493	0.0468	0.0443	0.0472
	$\hat{\vartheta}$	0.0654	0.0649	0.0627	0.0240	0.0241	0.0243	0.0119	0.0119	0.0117
	$\hat{d}$	0.1048	0.0538	0.0369	0.1003	0.0478	0.0373	0.1069	0.0478	0.0370

Table 2.4: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0008	-0.0017	-0.0021	0.0001	-0.0004	-0.0014	0.0004	0.0003	-0.0002
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0006	-0.0018	-0.0023	0.0004	-0.0001	-0.0013	0.0006	0.0005	0.0002
	$\hat{\vartheta}$	0.0347	0.0345	0.0321	0.0232	0.0242	0.0238	-0.0002	-0.0002	-0.0002
	$\hat{d}$	-0.0487	-0.0585	-0.0716	-0.0523	-0.0712	-0.0855	-0.0565	-0.0861	-0.1053
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0586	0.0660	0.0641	0.0455	0.0585	0.0658	0.0333	0.0447	0.0587
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0612	0.0702	0.0693	0.0473	0.0623	0.0720	0.0344	0.0474	0.0642
	$\hat{\vartheta}$	0.0403	0.0402	0.0382	0.0290	0.0299	0.0297	0.0115	0.0114	0.0117
	$\hat{d}$	0.0659	0.0730	0.0838	0.0704	0.0840	0.0964	0.0757	0.0979	0.1152
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0010	-0.0018	-0.0023	-0.0003	-0.0009	-0.0018	0.0000	-0.0001	0.0007
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0009	-0.0018	-0.0024	-0.0003	-0.0007	-0.0017	0.0002	0.0001	-0.0003
	$\hat{\vartheta}$	0.0420	0.0416	0.0390	0.0239	0.0243	0.0233	-0.0002	-0.0001	-0.0002
	$\hat{d}$	-0.0208	-0.0496	-0.0684	-0.0255	-0.0609	-0.0806	-0.0316	-0.0746	-0.0985
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0657	0.0677	0.0651	0.0511	0.0596	0.0662	0.0373	0.0456	0.0585
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0667	0.0714	0.0700	0.0518	0.0630	0.0718	0.0378	0.0479	0.0635
	$\hat{\vartheta}$	0.0479	0.0476	0.0453	0.0297	0.0301	0.0293	0.0115	0.0114	0.0117
	$\hat{d}$	0.0523	0.0656	0.0807	0.0566	0.0756	0.0919	0.0618	0.0884	0.1089

Table 2.5: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0$ and $\rho = 0$ with linear trends)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	0.0011	0.0013	0.0013	0.0008	0.0011	0.0013	0.0004	0.0008	0.0012
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0011	0.0013	0.0014	0.0009	0.0011	0.0014	0.0005	0.0007	0.0011
	$\hat{\vartheta}$	0.0078	0.0068	0.0041	-0.0494	-0.0494	-0.0500	-0.0421	-0.0418	-0.0420
	$\hat{d}$	-0.0136	-0.0573	-0.0804	-0.0156	-0.0629	-0.0863	-0.0159	-0.0711	-0.0970
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0507	0.0511	0.0490	0.0464	0.0505	0.0513	0.0394	0.0460	0.0507
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0510	0.0516	0.0495	0.0466	0.0512	0.0520	0.0394	0.0464	0.0515
	$\hat{\vartheta}$	0.0311	0.0310	0.0309	0.0567	0.0568	0.0573	0.0498	0.0496	0.0498
	$\hat{d}$	0.0447	0.0728	0.0931	0.0465	0.0771	0.0981	0.0495	0.0843	0.1076
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	0.0002	0.0009	0.0012	-0.0002	0.0006	0.0012	-0.0005	0.0002	0.0009
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0003	0.0009	0.0012	-0.0001	0.0006	0.0011	-0.0003	0.0002	0.0008
	$\hat{\vartheta}$	0.0217	0.0208	0.0184	-0.0442	-0.0445	-0.0452	-0.0398	-0.0398	-0.0403
	$\hat{d}$	0.0281	-0.0350	-0.0708	0.0247	-0.0415	-0.0771	0.0220	-0.0501	-0.0874
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0563	0.0522	0.0489	0.0516	0.0517	0.0512	0.0447	0.0474	0.0507
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0553	0.0528	0.0495	0.0509	0.0524	0.0521	0.0438	0.0481	0.0518
	$\hat{\vartheta}$	0.0389	0.0387	0.0381	0.0522	0.0525	0.0532	0.0477	0.0478	0.0484
	$\hat{d}$	0.0582	0.0591	0.0853	0.0573	0.0630	0.0904	0.0576	0.0696	0.0993

Table 2.6: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$ with linear trends)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0146	-0.0173	-0.0131	-0.0140	-0.0231	-0.0201	-0.0159	-0.0329	-0.0376
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0067	-0.0171	-0.0284	-0.0043	-0.0112	-0.0198	-0.0065	-0.0153	-0.0228
	$\hat{\vartheta}$	0.0121	0.0116	0.0093	-0.0017	-0.0006	-0.0011	-0.0072	-0.0062	-0.0056
	$\hat{d}$	-0.0343	-0.0709	-0.0991	-0.0351	-0.0701	-0.1001	-0.0386	-0.0670	-0.0953
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0474	0.0493	0.0469	0.0436	0.0512	0.0513	0.0387	0.0534	0.0599
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0462	0.0495	0.0539	0.0430	0.0476	0.0513	0.0375	0.0459	0.0523
	$\hat{\vartheta}$	0.0257	0.0256	0.0250	0.0190	0.0192	0.0195	0.0137	0.0126	0.0123
	$\hat{d}$	0.0504	0.0814	0.1078	0.0506	0.0803	0.1085	0.0532	0.0771	0.1038
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0147	-0.0158	-0.0100	-0.0124	-0.0198	-0.0151	-0.0143	-0.0289	-0.0308
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0098	-0.0166	-0.0256	-0.0072	-0.0101	-0.0158	-0.0091	-0.0142	-0.0180
	$\hat{\vartheta}$	0.0145	0.0138	0.0117	-0.0006	-0.0002	-0.0013	-0.0062	-0.0057	-0.0059
	$\hat{d}$	-0.0175	-0.0618	-0.0919	-0.0193	-0.0615	-0.0933	-0.0237	-0.0590	-0.0891
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0480	0.0481	0.0448	0.0437	0.0491	0.0481	0.0389	0.0505	0.0546
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0471	0.0487	0.0514	0.0439	0.0467	0.0487	0.0389	0.0449	0.0492
	$\hat{\vartheta}$	0.0274	0.0273	0.0267	0.0191	0.0192	0.0196	0.0125	0.0122	0.0125
	$\hat{d}$	0.0433	0.0739	0.1011	0.0435	0.0732	0.1022	0.0452	0.0706	0.0982

Table 2.7: Bias and RMSE Profiles with $n = 5$ and $T = 25$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\underline{\delta = 0.4:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0149	-0.0175	-0.0112	-0.0155	-0.0262	-0.0200	-0.0192	-0.0429	-0.0514
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0082	-0.0169	-0.0271	-0.0079	-0.0152	-0.0191	-0.0141	-0.0303	-0.0366
	$\hat{\vartheta}$	0.0405	0.0400	0.0361	0.0280	0.0296	0.0286	-0.0031	-0.0028	-0.0029
	$\hat{d}$	-0.0133	-0.0442	-0.0841	-0.0149	-0.0445	-0.0871	-0.0290	-0.0496	-0.0899
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0951	0.1007	0.0994	0.0851	0.0979	0.1033	0.0745	0.0948	0.1097
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0973	0.1023	0.1047	0.0879	0.0985	0.1047	0.0771	0.0931	0.1051
	$\hat{\vartheta}$	0.0604	0.0604	0.0584	0.0492	0.0504	0.0506	0.0210	0.0206	0.0212
	$\hat{d}$	0.0798	0.0919	0.1173	0.0776	0.0903	0.1193	0.0803	0.0901	0.1203
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0149	-0.0163	-0.0081	-0.0127	-0.0222	-0.0148	-0.0167	-0.0381	-0.0444
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0119	-0.0164	-0.0246	-0.0105	-0.0132	-0.0148	-0.0167	-0.0285	-0.0314
	$\hat{\vartheta}$	0.0452	0.0448	0.0414	0.0298	0.0306	0.0290	-0.0029	-0.0028	-0.0031
	$\hat{d}$	0.0115	-0.0340	-0.0789	0.0060	-0.0354	-0.0821	-0.0110	-0.0412	-0.0845
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.1001	0.1005	0.0972	0.0901	0.0971	0.1005	0.0784	0.0930	0.1044
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.1009	0.1019	0.1025	0.0923	0.0979	0.1020	0.0815	0.0922	0.1008
	$\hat{\vartheta}$	0.0641	0.0643	0.0628	0.0503	0.0511	0.0513	0.0213	0.0211	0.0219
	$\hat{d}$	0.0831	0.0887	0.1145	0.0805	0.0882	0.1171	0.0791	0.0873	0.1171

Table 2.8: Local Whittle Estimates of the Integration Orders, 1955-2008.

	Real GDP Growth		Debt-to-GDP Ratio		Debt-to-GDP Growth	
	$m = 10$	$m = 14$	$m = 10$	$m = 14$	$m = 10$	$m = 14$
Australia	0.4020	0.1109	0.8650	0.9464	-0.0771	0.5730
Austria	0.5601	0.3823	1.2679	1.0508	0.2740	0.1598
Belgium	0.5381	0.3680	1.1100	1.0690	1.0367	0.7376
Canada	0.1561	0.1935	0.7857	0.9584	0.2617	0.2098
Denmark	0.2710	0.2308	1.2061	1.3360	0.6254	0.7541
Finland	0.1762	0.1521	1.1762	1.4459	0.2082	0.3580
France	0.5129	0.4893	1.0009	1.0574	-0.0749	0.0674
Germany	0.7708	0.3244	0.9499	0.9817	0.1914	0.2627
Greece	0.4891	0.4299	1.4586	1.2520	0.2659	0.0700
Ireland	0.4383	0.4777	1.1871	1.2057	0.3821	0.2798
Italy	0.3190	0.4618	1.0425	1.0079	0.4096	0.5611
Japan	0.8071	0.6454	1.0626	1.0816	0.3307	0.4167
Netherlands	0.5373	0.2805	0.9796	1.1010	0.4785	0.5248
New Zealand	0.1095	0.1641	0.9079	0.9543	0.3042	0.4457
Norway	0.2428	0.1299	0.5582	0.8187	-0.2899	-0.1075
Portugal	0.3924	0.3498	0.9801	0.9790	0.2199	0.1075
Spain	0.3323	0.4371	0.8882	0.9566	0.3719	0.4193
Sweden	0.5035	0.3662	1.0963	1.3101	0.4868	0.8311
UK	-0.2749	-0.1820	1.0077	1.0214	0.0430	0.1795
US	-0.2500	-0.1440	0.9839	1.0336	0.4658	0.4645
s.e.	(0.1581)	(0.1336)	(0.1581)	(0.1336)	(0.1581)	(0.1336)

Note: *This table reports the local Whittle estimation results of the indicators across countries. Since the local Whittle estimates are inconsistent for values greater than one, we estimate the memory in the increments and add back one to ensure that we get valid estimates.*

Table 2.9: Estimation Results for the Slope and Long-Range Parameters

	Australia	Austria	Belgium	Canada	Denmark	Finland	France
$\hat{\beta}_i$	-0.1570	-0.1491	0.1338	0.0058	0.0014	-0.0469	-0.0330
s.e.($\hat{\beta}_i$)	(0.1222)	(0.1015)	(0.0728)	(0.1089)	(0.1062)	(0.0780)	(0.0593)
$\hat{\vartheta}_i$	0.6590	0.6310	0.6807	0.4485	0.6333	0.4936	0.3166
s.e.($\hat{\vartheta}_i$)	(0.7460)	(0.9062)	(0.4571)	(0.9749)	(0.6667)	(0.6742)	(0.9044)
$\hat{d}_i$	0.0680	0.8910	0.7840	0.7420	0.9140	0.5220	0.7780
s.e.($\hat{d}_i$)	(0.8862)	(0.8129)	(0.4733)	(0.7825)	(0.8320)	(0.6112)	(0.4992)

	Italy	Japan	Netherlands	New Zealand	Norway	Portugal	Spain
$\hat{\beta}_i$	**0.2130**	-0.0525	0.0320	-0.0854	0.0584	-0.0148	-0.0151
s.e.($\hat{\beta}_i$)	(0.0758)	(0.0723)	(0.0972)	(0.1241)	(0.1140)	(0.0966)	(0.1041)
$\hat{\vartheta}_i$	0.6628	0.8257	0.8856	0.7009	0.6508	0.5088	0.4492
s.e.($\hat{\vartheta}_i$)	(0.8827)	(0.7943)	(0.8311)	(0.7973)	(1.0853)	(1.1659)	(0.7980)
$\hat{d}_i$	0.2420	0.6170	0.4790	0.4250	0.7240	0.4310	0.8170
s.e.($\hat{d}_i$)	(0.5823)	(0.5097)	(0.7144)	(1.0181)	(0.9625)	(0.7491)	(0.8009)

	Germany	Sweden	Greece	Ireland	UK	US
$\hat{\beta}_i$	0.0451	-0.0342	-0.0130	-0.0676	0.0925	-0.1672
s.e.($\hat{\beta}_i$)	(0.0861)	(0.0544)	(0.1192)	(0.0879)	(0.0983)	(0.0922)
$\hat{\vartheta}_i$	0.5828	0.7782	0.5790	1.0122	0.7174	0.7290
s.e.($\hat{\vartheta}_i$)	(0.9618)	(0.5916)	(1.1156)	(0.9208)	(1.0774)	(0.6255)
$\hat{d}_i$	0.7700	0.0001	0.7690	0.8910	0.8080	0.8010
s.e.($\hat{d}_i$)	(0.6699)	(0.4798)	(0.8654)	(0.7612)	(0.8263)	(0.6867)

Note: *This table reports the estimation results of the individual slope and memory parameters across countries. Estimations are performed based on (2.1) where the projections are carried out with* $d^* = 1.25$. *Robust standard errors are reported in parentheses. Bold indicates significance up to the 5% level.*

Table 2.10: Local Whittle Estimates of the Integration Orders, 1955-2008.

	Real GDP (Log-level)		Debt (Log-level)	
	$m = 10$	$m = 14$	$m = 10$	$m = 14$
Australia	0.9716	0.9686	0.9785	0.9920
Austria	0.9536	0.9368	0.9954	0.9700
Belgium	0.9938	0.9794	0.9844	0.9864
Canada	0.9879	0.9667	0.9523	0.9874
Denmark	0.9355	0.9384	0.9082	0.9565
Finland	0.9420	0.9496	0.9248	0.9629
France	0.9778	0.9550	0.9820	0.9755
Germany	0.9149	0.9139	0.9817	0.9823
Greece	0.9591	0.9344	0.9660	0.9423
Ireland	0.9905	0.9869	0.9873	1.0014
Italy	0.9668	0.9564	0.9794	0.9828
Japan	0.9957	0.9812	0.9463	0.9493
Netherlands	0.9725	0.9764	0.9874	0.9990
New Zealand	0.9129	0.9236	0.9850	0.9992
Norway	0.9938	0.9937	0.9599	0.9799
Portugal	0.9921	0.9920	0.9890	0.9671
Spain	0.9956	0.9620	0.9491	0.9672
Sweden	0.9196	0.9392	0.9630	0.9704
UK	0.9784	0.9790	0.9164	1.0086
US	0.9964	0.9902	0.9884	0.9933
s.e.	(0.1581)	(0.1336)	(0.1581)	(0.1336)

Note: *This table reports the local Whittle estimation results of the indicators across countries. Since the local Whittle estimates are inconsistent for values greater than one, we estimate the memory in the increments and add back one to ensure that we get valid estimates.*

Figure 2.1: Real GDP Growth Rates, 1955-2008.

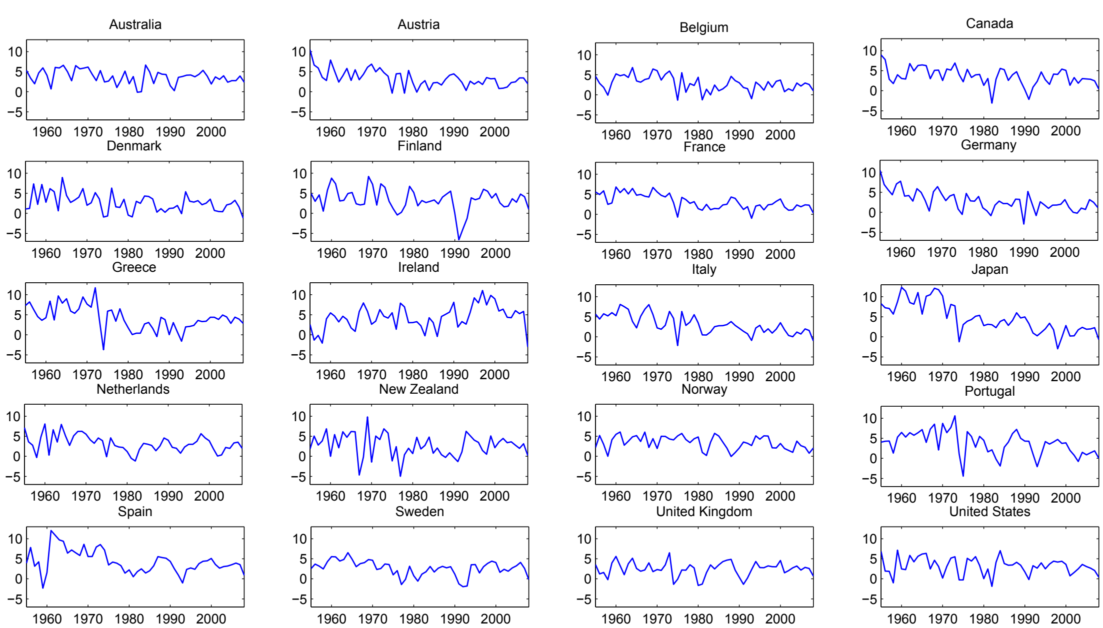


Table 2.11: Benchmark Estimation Results for the Slope and Long-Range Parameters based on Hualde and Robinson [22]

	Australia	Austria	Belgium	Canada	Denmark	Finland	France
$\hat{\beta}_i$	0.0070	**-0.0845**	**-0.1427**	0.0072	**0.0706**	**-0.2099**	**-0.0133**
s.e.($\hat{\beta}_i$)	(0.0075)	(0.0061)	(0.0061)	(0.0055)	(0.0088)	(0.0138)	(0.0054)
$\hat{\vartheta}_i$	**1.4900**	**1.3114**	**1.4900**	**1.1980**	**1.4899**	**1.4899**	**1.3220**
s.e.($\hat{\vartheta}_i$)	(0.3833)	(0.0834)	(0.0459)	(0.2112)	(0.1023)	(0.1310)	(0.1108)
$\hat{d}_i$	**1.4999**	**1.4999**	**1.4670**	**1.4999**	**1.4110**	**1.3830**	**1.4999**
s.e.($\hat{d}_i$)	(0.0495)	(0.0443)	(0.0440)	(0.0415)	(0.0606)	(0.0838)	(0.0350)

	Italy	Japan	Netherlands	New Zealand	Norway	Portugal	Spain
$\hat{\beta}_i$	**0.0596**	**0.0191**	**0.0519**	**0.0478**	**0.0140**	**0.0613**	**-0.0219**
s.e.($\hat{\beta}_i$)	(0.0062)	(0.0063)	(0.0066)	(0.0136)	(0.0043)	(0.0070)	(0.0060)
$\hat{\vartheta}_i$	**1.3982**	**1.4899**	**1.3458**	**1.3144**	**1.1701**	**1.1871**	**1.4512**
s.e.($\hat{\vartheta}_i$)	(0.0530)	(0.0546)	(0.1157)	(0.2474)	(0.2311)	(0.1329)	(0.1092)
$\hat{d}_i$	**1.4999**	**1.4999**	**1.4910**	**1.3130**	**1.4999**	**1.4610**	**1.4999**
s.e.($\hat{d}_i$)	(0.0436)	(0.0358)	(0.0452)	(0.0885)	(0.0381)	(0.0513)	(0.0385)

	Germany	Sweden	Greece	Ireland	UK	US
$\hat{\beta}_i$	**-0.1778**	**-0.0667**	**0.1017**	**-0.0917**	**0.0441**	**0.1131**
s.e.($\hat{\beta}_i$)	(0.0098)	(0.0069)	(0.0060)	(0.0079)	(0.0193)	(0.0056)
$\hat{\vartheta}_i$	**1.3256**	**1.4899**	**1.2705**	**1.3687**	**1.2739**	**1.4899**
s.e.($\hat{\vartheta}_i$)	(0.0950)	(0.0835)	(0.0850)	(0.1285)	(0.3629)	(0.0536)
$\hat{d}_i$	**1.4350**	**1.4999**	**1.4999**	**1.4999**	**1.3800**	**1.4720**
s.e.($\hat{d}_i$)	(0.0628)	(0.0447)	(0.0472)	(0.0575)	(0.1267)	(0.0443)

Note: *This table reports the estimation results of the individual slope and memory parameters across countries based on the pure time-series estimation technique by Hualde and Robinson [22] that disregards individual country characteristics and cross-country dependence. Robust standard errors are reported in parentheses. Bold indicates significance up to the 5% level.*

Table 2.12: Estimation Results for the Slope and Long-Range Parameters based on (2.21)

	Australia	Austria	Belgium	Canada	Denmark	Finland	France
$\hat{\beta}_i$	**-0.0532**	**-0.1252**†	**0.0203**	**0.0374**	**-0.0185**†	**0.3127**	**0.0159**†
s.e.($\hat{\beta}_i$)	(0.0027)	(0.0041)	(0.0034)	(0.0036)	(0.0048)	(0.0044)	(0.0023)
$\hat{\vartheta}_i$	**1.4096**	**1.0773**	**1.4900**	**1.1152**	**1.4899**	**1.2886**	**1.1490**
s.e.($\hat{\vartheta}_i$)	(0.2734)	(0.0722)	(0.0381)	(0.2036)	(0.0926)	(0.1025)	(0.1053)
$\hat{d}_i$	**1.3220**	**0.4420**	**1.4999**	**0.7800**	**1.0510**	**1.1780**	**0.8110**
s.e.($\hat{d}_i$)	(0.0302)	(0.0276)	(0.0213)	(0.0255)	(0.0378)	(0.0357)	(0.0175)

	Italy	Japan	Netherlands	New Zealand	Norway	Portugal	Spain
$\hat{\beta}_i$	**-0.1089**†	**-0.0882**†	**-0.2528**	-0.0189	**-0.1079**	**-0.0253**†	**0.0940**†
s.e.($\hat{\beta}_i$)	(0.0038)	(0.0035)	(0.0039)	(0.0099)	(0.0028)	(0.0045)	(0.0041)
$\hat{\vartheta}_i$	**1.1971**	**1.4899**	**1.1607**	**1.2143**	**1.1632**	**1.0529**	**1.2252**
s.e.($\hat{\vartheta}_i$)	(0.0453)	(0.0510)	(0.1087)	(0.2029)	(0.2047)	(0.1293)	(0.0995)
$\hat{d}_i$	**0.9360**	**0.6630**	**1.2130**	**0.9020**	**1.0590**	**0.2710**	**0.8460**
s.e.($\hat{d}_i$)	(0.0291)	(0.0255)	(0.0266)	(0.0677)	(0.0240)	(0.0338)	(0.0300)

	Germany	Sweden	Greece	Ireland	UK	US
$\hat{\beta}_i$	**0.1521**†	**0.1119**†	**-0.1535**†	**0.6534**†	**-0.2177**	0.0054
s.e.($\hat{\beta}_i$)	(0.0047)	(0.0032)	(0.0047)	(0.0044)	(0.0042)	(0.0037)
$\hat{\vartheta}_i$	**1.0892**	**1.4899**	**1.0464**	**1.0776**	**1.2887**	**1.4899**
s.e.($\hat{\vartheta}_i$)	(0.0897)	(0.0732)	(0.0788)	(0.1191)	(0.3276)	(0.0428)
$\hat{d}_i$	**0.5780**	**1.0120**	**0.3890**	**0.5120**	**1.2890**	**1.3480**
s.e.($\hat{d}_i$)	(0.0330)	(0.0244)	(0.0359)	(0.0325)	(0.0343)	(0.0311)

Note: *This table reports the estimation results of the individual slope and memory parameters across countries. Estimations are performed based on (2.21) where the projections are carried out with* $d^* = 1.25$*. Robust standard errors are reported in parentheses. Bold indicates significance up to the 5% level.* † *indicates a cointegrating relationship between real GDP and debt in logs at the 5% level.*

Figure 2.2: Debt-to-GDP Ratios, 1955-2008.

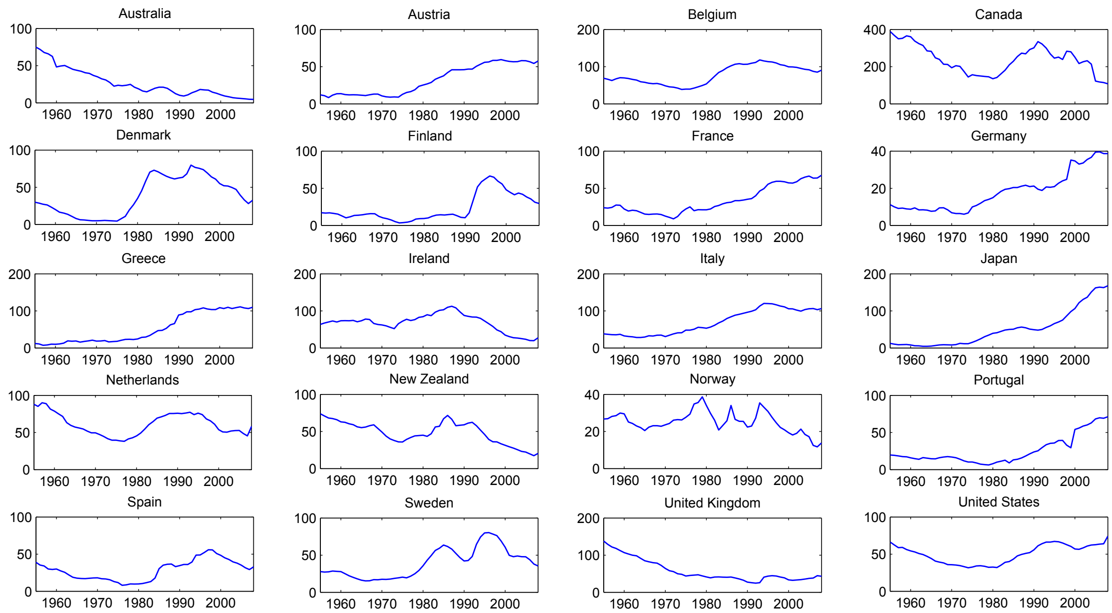


Figure 2.3: Real GDP in Logs, 1955-2008.

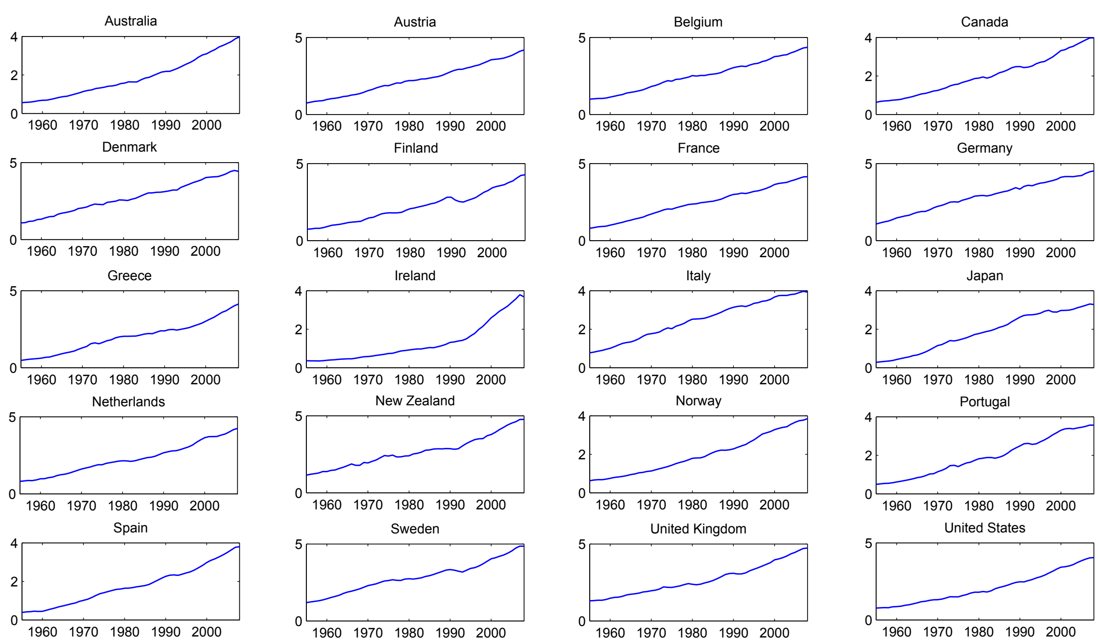
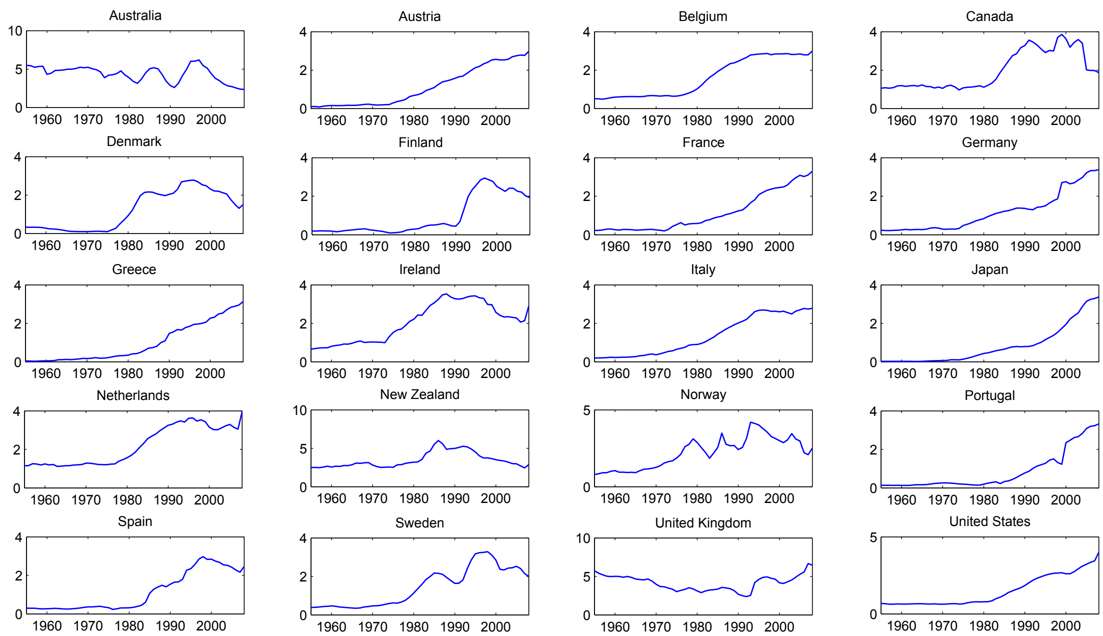


Figure 2.4: Debt in Logs, 1955-2008.



Chapter 3

Parametric Portfolio Policies with Common Volatility Dynamics (with Abderrahim Taamouti)

Abstract

A parametric portfolio policy function is considered that incorporates common stock volatility dynamics to optimally determine portfolio weights. Reducing dimension of the traditional portfolio selection problem significantly, only a number of policy parameters corresponding to first- and second-order characteristics are estimated based on a standard method-of-moments technique. The method, allowing for the calculation of portfolio weight and return statistics, is illustrated with an empirical application to 30 U.S. industries to study the economic activity before and after the recent financial crisis.

Keywords: Parametric portfolio policy, stock characteristics, volatility common factors.
JEL classification: C13, C21, C23, C58, G11, G15.

3.1 Introduction

Portfolio selection problems have been traditionally studied based on the portfolio theory by Markowitz (1952), which requires modeling the joint distribution of returns. Portfolios selected based on Markowitz approach, however, do not completely take into account the risk borne by the investor because only the mean and variance are known but not the entire distribution.

Brand et al. (2009) (BSCV (2009) hereafter) proposes a parametric portfolio policy in that weights of stocks depend on stock characteristics. Their approach removes the necessity of modeling the joint distribution of returns and only a small number of parameters are estimated to determine optimal portfolio weights. While this approach is much easier to use in practice compared to the traditional Markowitz approach, it also lacks the ability to explicitly account for the risk borne by the investor in the weights function.

This paper considers a parametric portfolio policy with common volatility dynamics to explicitly incorporate the impact of risk borne by the investor in portfolio selection decisions. Our portfolio policy function is based on stock characteristics as proposed by BSCV (2009), but unlike theirs, ours is augmented by the estimates of volatility common factors. This way, the portfolio policy not only accounts for the first-order (stock) characteristics but also the second-order (volatility) characteristics thus providing the investor with the ability to base his decision also on risk.

Our portfolio policy contains only a number of stock characteristics and nests long-short portfolios of Fama and French (1993), Carhart (1997) and Fama and French (2015), but it additionally accounts for common volatility dynamics of the stocks. Since only a number of common stock characteristics are considered instead of historical stock returns and their joint distribution, dimensionality is significantly reduced. Therefore our approach is easy to implement in practice and it avoids possible imprecision due to overfitting.

In the analysis, volatility common factors are estimated first. Stock realized volatilities (RV's hereafter), which we calculate based on the jump-robust realized bipower variation measure due to Barndorff-Nielsen and Shephard (2004), exhibit fractional long-range dependence as shown by Bollerslev et al. (2013). This requires that stock RV's be appropriately differenced with their corresponding integration orders so that a principal components (PC) estimation can be employed to obtain the estimates of volatility common factors. These estimates are then plugged in to the parametric portfolio policy function of BSCV (2009) to determine optimal portfolio weights.

In the estimation of portfolio policy parameters, a generalized method-of-moments estimation is employed that is shown to produce consistent, asymptotically normal and efficient estimates as shown by Hansen (1982) within the class of estimators that employ the same set of moment conditions as ours. Based on these estimates, portfolio weight and return statistics can be calculated.

To illustrate the effectiveness of our approach, we use montly return data on 30 U.S. industries spanning the time period January 1966 - December 2014, which we split to January 1966 - August 2008 in-sample and September 2008 - December 2014 out-of-sample periods with the purpose of studying the impact of the recent crisis. We compare the performance of the portfolio policy that incorporates the common volatility dynamics to that which only considers first-order (stock) characteristics. The findings indicate that accounting for common volatility dynamics leads the investor to select an optimal portfolio with higher returns, reduced risk, higher Sharpe ratios and positive skewness in sample and out of sample.

The remainder of the paper is organized as follows. Next section explains the estimation of volatility common factors. Section 3 gives details on the parametric policy function incorporating common volatility dynamics. Section 4 provides an empirical illustration with data, and finally Section 5 concludes the paper.

3.2 Common Dynamics in Realized Volatilities

It is intuitive and clear that risk associated with the volatility of a stock affects the investment decision taken by the investor. That said, volatility associated with each stock can be treated separately to make allocation decisions but when large number of assets are analyzed instead, volatility-return assessment becomes cumbersome from an empirical point of view. With this in mind, we suggest using a common-factor model to capture the information about realized volatilities to reduce the dimension of the problem significantly. Common factors in the treatment of high-dimensional data has been used in several different setups; see e.g. Pesaran (2006) and Bai and Ng (2013).

We first construct the realized volatility measures based on bipower variation that is robust to jumps, following Barndorff-Nielsen and Shephard (2004). Let us denote an excess return at time t corresponding to industry i, $r_{i,t}$. Then the monthly realized bipower variation (RBV) is given by

$$RBV_{i,t} = \sum_{j=1}^{M-1} |r_{i,j}| \, |r_{i,j+1}| \, , \tag{3.1}$$

where M is the number of trading days in a month. Barndorff-Nielsen and Shephard (2004) argue that RBV converges to realized variance in the limit assuming asset prices follow a stochastic-volatility process and the limiting RBV measure is robust to rare jumps. Therefore, a jump-robust realized volatility measure can be envisaged as the square-root of RBV in (3.1).

To investigate the common dynamics of RV's, a common factor model can be employed as follows:

$$RV_{i,t} = \lambda_i' f_t + \epsilon_{i,t} \tag{3.2}$$

where λ_i are unobserved factor loadings indicating how much each cross-section unit is affected by the unobserved common factors f_t, and $\epsilon_{i,t}$ are assumed to be identically and independently distributed volatility shocks with mean zero and variance σ_i^2. In the estimation of common factor models, the use of principal components (PC) analysis, see e.g. Bai and Ng (2002, 2004, 2013), is standard to get the estimates of factor loadings and common factors, $\hat{\lambda}_i$ and $\hat{f}_t$. Restricting the attention to (3.2), the estimates $\hat{f}_t$ constitute the common dynamics of RV's and are much easier to use in portfolio choice problems than individual RV's due to reduced dimensionality providing a portfolio policy rather than requiring a stock-specific treatment. Asymptotic theory for $\hat{\lambda}_i$ and $\hat{f}_t$ is derived by Bai and Ng (2002, 2004) in case of stationary $I(0)$ and nonstationary $I(1)$ dependent variables, respectively.

Among others, Bollerslev et al. (2013) show that RV's exhibit long memory properties. This requires that RV's be appropriately differenced to stationarity before attempting to estimate (3.2). Bai and Ng (2004) use a similar approach in that they first-difference $I(1)$ data to obtain stationary variables to get factor structure estimates. Let us denote the fractional integration order of $RV_{i,t}$ by δ_i so that $RV_{i,t}$ is $I(\delta_i)$, where δ_i is positive. Then, using that $\Delta = 1 - L$ with the lag operator L, the common-factor structure estimates are obtained from the equation,

$$\Delta_t^{\delta_i} RV_{i,t} = \lambda_i' f_t + \epsilon_{i,t}. \tag{3.3}$$

For some $\delta > 0$,

$$\Delta_t^{\delta} = \Delta^{\delta} 1(t > 0) = \sum_{j=0}^{t-1} \pi_j(\delta) L^j, \tag{3.4}$$

$$\pi_j(\delta) = \frac{\Gamma(j-\delta)}{\Gamma(j+1)\Gamma(-\delta)},$$

where $1(\cdot)$ is the indicator function, and $\Gamma(\cdot)$ denotes the gamma function such that $\Gamma(d) = \infty$ for $d = 0, -1, -2, \ldots$, but $\Gamma(0)/\Gamma(0) = 1$. The expression in (3.4) bestows long-memory dynamics, in which autocorrelations show an algebraic rather than exponential decay because $\pi_j(\mu) \sim C j^{-\mu-1}$ as $j \to \infty$ for $\mu > 0$. So, these weights are appropriate to control for inherent long memory in RV's as shown by Bollerslev et al. (2013) and $\Delta_t^{\delta_i} RV_{i,t}$ becomes $I(0)$.

When δ_i are known, this differencing can be directly carried out. However, in practice δ_i are unknown and must be estimated. For the estimation, a parametric approach or a semiparametric approach such as a local Whittle estimation, e.g. by Robinson (1995), can be used to obtain consistent estimates for δ_i. Then, we are simply interested in obtaining factor-structure estimates using a standard PC approach on the equation,

$$\Delta_t^{\hat{\delta}_i} RV_{i,t} = \lambda_i' f_t + \epsilon_{i,t}, \tag{3.5}$$

for which limiting theory is readily established in the literature, e.g. by Bai and Ng (2013). The number of common factors to be retained in the analysis can be determined based on the number of eigenvalues exceeding the mean eigenvalue. Denote $\hat{f}_t^*$ the vector of retained common factor estimates that is a subset of the factor estimates obtained from (3.5). Then, $\hat{f}_t^*$ can be used in different regression settings as plug-in estimates to serve, for example, as volatility common factor augmentation. The estimates $\hat{f}_t^*$ can also be used solely to capture the common volatility information, measuring whose impact on invesment decisions is generally of interest.

3.3 Optimal portfolio policy with common dynamics of volatility

In the setup, we consider that at time t, there are N_t number of stocks that are investable. Each stock i has a return of $r_{i,t+1}$ from time t to $t+1$ and is associated with a vector of firm characteristics $x_{i,t}$ and retained estimates of common volatility factors $\hat{f}_t^*$ observed at time t. The stock characteristics can contain, among others, the market capitalization of the stock and the book-to-market ratio of the stock. The investor's problem is then to maximize the conditional expected utility of the portfolio return $r_{p,t+1}$ by choosing the weights $w_{i,t}$ optimally, i.e.,

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} E_t[u(r_{p,t+1})] = E_t\left[u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right]. \tag{3.6}$$

Adopting BSCV (2009), we parameterize the portfolio weights as a function of stock characteristics as well as the common dynamics of stock volatilities,

$$w_{i,t} = g(x_{it}, \hat{f}_t^*; \theta, \gamma). \tag{3.7}$$

In particular, we focus on a linear specification of the portfolio weight function:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\left(\theta' \tilde{x}_{i,t} + \gamma' \hat{f}_t^*\right), \tag{3.8}$$

where $\bar{w}_{i,t}$ is the weight of the stock i at time t in a benchmark portfolio, e.g. the value-weighted market portfolio, θ and γ are coefficients to be estimated, $\hat{f}_t^*$ is the vector of common factors of volatilities, and $\tilde{x}_{i,t}$ are the characteristics of stock i, standardized cross-sectionally to have zero mean and unit standard deviation across all stocks at time t. The interest is in estimating weights as a single function of characteristics, as in BSCV (2009), and also common volatility drivers that applies to all stocks over time.

The parameterization in (3.8) brings in the possibility to deviate from the benchmark portfolio, whose weights are given by $\bar{w}_{i,t}$, based on $\tilde{x}_{i,t}$ and $\hat{f}_t^*$. In practice, standardization of characteristics and the normalization factor $1/N_t$ are necessary to ensure that weights are not mischosen; see BSCV (2009) for a discussion.

The coefficient vectors to be estimated, θ and γ, do not vary over time, which implies that portfolio weights depend only on firm and common volatility characteristics and not on historical returns. Time-invariant coefficients also imply that the coefficients that maximize the conditional expected utility of the investor also maximize his unconditional expected utility. Therefore, the maximization problem can be formulated using (3.7) as

$$\max_{\theta,\gamma} E\left[u\left(r_{p,t+1}\right)\right] = E\left[u\left(\sum_{i=1}^{N_t} g(x_{it}, \hat{f}_t^*; \theta, \gamma) r_{i,t+1}\right)\right]. \tag{3.9}$$

Since, under some regularity conditions, the empirical moment of the expected utility function converges to the theoretical one, in practice θ and γ will be estimated by maximizing the sample analogue of the unconditional expected utility,

$$\max_{\theta,\gamma}\left\{\frac{1}{T}\sum_{t=0}^{T-1} u(r_{p,t+1})\right\} = \max_{\theta,\gamma}\left\{\frac{1}{T}\sum_{t=0}^{T-1}\left[u\left(\sum_{i=1}^{n} g(x_{it}, \hat{f}_t^*; \theta, \gamma) r_{i,t+1}\right)\right]\right\}, \tag{3.10}$$

for some prespecified choice of $u(\cdot)$, e.g. log, quadratic or a general constant relative risk aversion (CRRA) function. While the specification of $u(\cdot)$ is a matter of choice, the power-utility function of the form

$$u(c) = \frac{(1+c)^{1-\zeta}}{1-\zeta} \tag{3.11}$$

helps realize the implicit assumption made by time-invariant coefficients in (3.7) that the stock characteristics fully capture all aspects of the joint distribution of returns that are relevant for forming optimal portfolios because (3.11) not only takes into account the mean and variance, but also higher-order moments such as skewness and kurtosis. Moreover, CRRA is directly imposed by this functional form which shows sensitivity to different risk aversion levels through the parameter ζ.

Using (3.8), (3.10) can be expressed as

$$\max_{\theta,\gamma}\left\{\frac{1}{T}\sum_{t=0}^{T-1} u(r_{p,t+1})\right\} = \max_{\theta,\gamma}\left\{\frac{1}{T}\sum_{t=0}^{T-1}\left[u\left(\sum_{i=1}^{n}\left(\bar{w}_{i,t} + \frac{1}{N_t}\left(\theta'\tilde{x}_{i,t} + \gamma'\hat{f}_t^*\right)\right) r_{i,t+1}\right)\right]\right\}. \tag{3.12}$$

It is important to note that (3.12) contains parameter vectors θ and γ that are of small dimensions because there are only a limited number of stock characteristics and very few (just one or two) common drivers of stock volatility, which makes their estimations computationally easy. Using this parametric portfolio policy also reduces the risk of imprecise estimation due to overfitting.[1]

A portfolio policy generated by (3.8) nests the long-short portfolios. Let us write the return of the portfolio policy in (3.8),

$$\begin{aligned} r_{p,t+1} &= \sum_{i=1}^{N_t} \bar{w}_{i,t+1} r_{i,t+1} + \sum_{i=1}^{N_t}\left(\frac{1}{N_t}\left(\theta'\tilde{x}_{i,t} + \gamma'\hat{f}_t^*\right)\right) r_{i,t+1} \\ &= r_{m,t+1} + r_{h,t+1}, \end{aligned} \tag{3.13}$$

[1]For an extensive discussion see BSCV (2009).

where m denotes the benchmark value-weighted market, and h denotes a long-short hedge fund with weights $\frac{1}{N_t}\left(\theta'\tilde{x}_{i,t} + \gamma'\hat{f}_t^*\right)$ summing up to zero. The linear portfolio policy weights in (3.8) therefore also nests the popular portfolios of Fama and French (1993, 2015) and Carhart (1997). For example, the return of the three-factor portfolio by Fama and French (1993) additionally incorporating volatility common factors can be expressed as

$$r_{p,t+1} = r_{m,t+1} + \theta_{smb} r_{smb,t+1} + \theta_{hml} r_{hml,t+1} + \gamma' \hat{f}_t^* \frac{1}{N_t} \sum_{i=1}^{N_t} r_{i,t+1} \tag{3.14}$$

where $r_{smb,t+1}$ and $r_{hml,t+1}$ are the returns to small-minus-big and high-minus-low portfolios, respectively.

Having formulated the optimal portfolio weights selection problem as an expected utility maximization problem, we can obtain the estimates $\hat{\theta}$ and $\hat{\gamma}$ resorting to methods of moments estimation. The estimates $\hat{\theta}$ and $\hat{\gamma}$, defined by the optimization problem in (3.12) satisfy the first-order conditions

$$\frac{1}{T}\sum_{t=0}^{T-1}\left\{u_\theta(r_{p,t+1})\left(\frac{1}{N_t}\hat{x}_t' r_{t+1}\right) + u_\gamma(r_{p,t+1})\left(\hat{f}_t^* \frac{1}{N_t}\sum_{i=1}^{N_t} r_{i,t+1}\right)\right\} = 0$$

where $u_\varsigma = (\partial/\partial\varsigma)\, u$. The asymptotic variance-covariance matrix and its estimate can be envisaged following Hansen (1982) who shows that GMM estimates such as the ones we have are consistent, asymptotically normal and efficient within the class of estimators employing the same set of moment conditions. In practice, estimation may be performed based on multi-step or continuous-updating GMM procedures to acquire a desired level of parameter convergence.

3.4 Empirical illustration with data

3.4.1 Data description and empirical strategy

To illustrate the impact of incorporating common volatility dynamics into the parametric portfolio policy function by BSCV (2009), we explore the performance of industry portfolios because they are more informative about economic activity rather than being of specific investment interest.

We use daily return data on 30 U.S. industries and the composite average index of NYSE, NASDAQ and AMEX for the time period January 1966 - December 2014 downloaded from Ken French's Data Library along with the risk-free rates to calculate monthly industry and market RV's employing (3.1). We otherwise use the monthly data readily available for the three Fama-French factors in French's Data library. In the application, the investor is restricted to invest only in stocks. As also discussed by BSCV (2009), the reason for not including the risk-free asset as an investment opportunity is that the varying leverage induced by the risk-free asset only corresponds to a change in the scale of the stock portfolio weights.

The raw data requires standardization so that the results become comparable. The stock characteristics, x_{it}, show varying cross-sectional means and standard deviations that we take into account. The risk aversion is taken to be five. The CRRA utility function in (3.11) is used in a two-step GMM setting to determine the optimal portfolio weights.

With the goal of studying the predictive ability of the portfolio using common volatility factors, we divide the study sample into two groups: the in-sample analysis uses equity return data from January 1966 to August 2008 (512 data points), and the out-of-sample analysis focuses on the period September 2008 - December 2014 (76 data points), including the recent financial crisis. There is no specific reason as to why we split the sample to these two periods apart from the interest in investigating whether there are huge differences in terms of portfolio performance between pre- and post-crisis periods. Clearly different out-of-sample periods can also be considered.

We first estimate the common factors of industry RV's to be able to use them as further characteristics in the portfolio weight function. We then estimate the parameters of the portfolio whose returns are given by (3.14). Based on these estimates, we calculate portfolio weight statistics alongside with the unconditional mean, standard deviation, skewness and Sharpe ratio of the optimal portfolio.

3.4.2 Estimation of the common factors in industry RV's

First, we estimate the fractional integration orders of industry and market RV's based on Robinson (1995)'s local Whittle method that requires specifying the number of Fourier frequencies to be used. It is well known that long memory should be investigated in lower frequencies since higher frequencies are susceptible to short-memory contamination. This is why, we focus on $m = 45, 71$ Fourier frequencies corresponding to $T^{.6}$ and $T^{.67}$ with $T = 588$ the time-series length in our dataset.

The nonstationarity bound for long-memory processes is $\delta_i = 0.5$, so an indicator exhibits nonstationary long memory for $\delta_i \geq 0.5$ and stationary long memory for $\delta_i < 0.5$ and $\delta_i \neq 0$. Based on the results in Table 1, industry RV's show some heterogeneity in terms of stationarity while the market RV is stationary. This stresses the importance of appropriately differencing the RV's before carrying out PC estimation to obtain factor structure estimates.

After differencing the industry RV's by their corresponding integration orders[2], we carry out a PC estimation on (3.5) to get the common factor estimates. The PC estimation indicates that there is only one common factor driving the industry RV's, as can also be seen from the screeplot in Figure 1. This common factor explains 69.64% of the total variation in the industry RV's.

It is also important to show that a common-factor model fits the industry RV's well. This can be checked by the uniqueness of variances that are not captured by the common factor: if uniqueness ratios are small, or equivalently if communality=1-uniqueness is large, then there is evidence that a common-factor model is well suited to the analysis of industry RV's. Table 2 below shows that the factor loadings estimates are positive and large while the uniqueness ratios

[2] $m = 45$ Fourier frequencies were used.

are small. So, a common-factor model indeed fits industry RV's well.

3.4.3 Portfolio performance incorporating the common factor of industry RV's

In Section 3, we have shown that the linear portfolio policy in (3.8) nests many widely analyzed portfolios, such as those of Fama and French (1993), Carhart (1997) and Fama and French (2015). To simply illustrate the impact of incorporating common volatility dynamics into the parametric policy function of BSCV (2009), we restrict our attention to the portfolio of Fama and French (1993) that we discussed in (3.14). That said, obviously other portfolios can also be analyzed but the impact of common volatility dynamics on portfolio selection can be determined more easily in this less complicated setting.

We first consider the optimization problem in (3.12) as is, and then restrict $\gamma = 0$ to be able to determine the impact of $\hat{f}_t^*$ on optimal portfolio selection. A generalized method of moments estimation for the portfolio policy incorporating volatility common factor in (3.8) based on (3.11) leads to the results in Table 3.

The first six rows of Table 3 present the estimated coefficients of parametric portfolio policy function with volatility common factor along with their standard errors. These coefficients indicate that the optimal portfolio is determined by choosing small firms, value stocks and less volatile stocks since the coefficients are positive and statistically significant for *smb* and *hml* while it is negative for $\hat{f}_t^*$. The finding that the deviation of the optimal weights from the benchmark weights increases with *smb* and *hml* and decreases with $\hat{f}_t^*$ is quite intuitive and mirrors the findings in the literature.

Rows seven to eleven of Table 3 describe the weights of the optimized portfolio. The average absolute weight of the optimal portfolio is equal to 0.3871% in sample and 1.6822% out of sample. The average (over time) maximum and minimum weights of the optimal portfolio are 1.0639% and -4.4701% for the in-sample period and 4.0439% and -3.6111% for the out-of-sample period, respectively. The average sum of negative weights in the optimal portfolio is -0.4930 in sample and -0.1308 out of sample. The average fraction of negative weights (shorted equities) in the optimal portfolio is 0.2047 for in sample and 0.0933 for out of sample. Therefore, the optimal portfolio using common RV factor does not reflect unreasonably extreme bets on individual equities and could well be implemented by a combination of an index fund that reflects the market and a long-short equity hedge fund as in (3.13).

The remaining rows of Table 3 characterize the performance of the optimal portfolio. The optimal portfolio has an average monthly return of 0.51% in sample and 1.87% out of sample. The standard deviation of the optimal portfolio returns is 0.0161 and 0.0359, respectively, for in sample and out of sample that translates into Sharpe ratios of 0.3158 and 0.5211, respectively. Skewness is positive and large for both split-sample periods indicating that there is a decreased likelihood of encountering a large negative return.

In order to show that accounting for common volatility dynamics leads to better portfolio performance, we consider the parametric portfolio policy restricting the attention to *smb* and *hml* only, i.e. $\gamma = 0$. The estimation results along with portfolio weight and return statistics are reported in Table 4.

The estimated coefficients are positive for both *smb* and *hml* in sample and out of sample. That is, small firms and value stocks are positively weighed in for the selection of the optimal portfolio, which is in line with the findings in the literature. In the out-of-sample period, *smb* does not have a significant role in the determination of optimal portfolio weights but the coefficient of *hml* remains significant, indicating that in the post-crisis period the investment decision is based on high value stocks regardless of firm size.

Rows seven to eleven of Table 4 describe the weights of the optimized portfolio that does not account for common volatility dynamics. The average absolute weight of this portfolio is equal to 0.1949% in sample and 1.3333% out of sample. The average (over time) maximum and minimum weights of this portfolio are 0.2113% and 0.1807% for the in-sample period and 1.3984% and 1.2521% for the out-of-sample period, respectively. The average fraction of negative weights (shorted equities) in the optimal portfolio is 0 for in sample and out of sample, indicating that this portfolio policy recommends not shorting any of the equities. These findings contrast with the portfolio weight statistics in Table 3 in that accounting for common volatility dynamics leads to the recommendation to short equities whose risk is high.

The remaining rows of Table 4 summarizes the optimal portfolio return statistics. The optimal portfolio has an average monthly return of 0.19% in sample and 1.63% out of sample. The standard deviation of the optimal portfolio returns is 0.0182 and 0.0762, respectively, for in sample and out of sample that translates into Sharpe ratios of 0.1044 and 0.2138, respectively. Skewness is negative for both split-sample periods indicating that there is a likelihood of encountering a large negative return. These results contrast poorly to the optimal portfolio return statistics in Table 3 in that the portfolio policy accounting for common volatility dynamics has higher average monthly returns, reduced portfolio risk, higher Sharpe ratios and positive skewness both in sample and out of sample.

3.4.4 The relationship between common factor of industry RV's and variance risk premium

When an analysis is carried out at the macroeconomic level based on industry portfolios, it may also be interesting to establish the ties between the factor-structure estimates obtained from (3.5) and a general measure such as variance risk premium (VRP) since an economic discussion can then be pursued.

Common volatility dynamics can be linked to variance risk premium that is defined as the difference between the ex-ante risk neutral expectation of the future stock market return variance and the expectation of the stock market return variance between time t and $t+1$:

$$VRP_t \equiv E_t^{\mathbb{Q}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right) - E_t^{\mathbb{P}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right),$$

where "$E_t^{\mathbb{P}}$" denotes the conditional expectation with respect to physical probability. VRP_t is unobservable and can be estimated by replacing $E_t^{\mathbb{Q}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right)$ and $E_t^{\mathbb{P}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right)$ by their estimates $\hat{E}_t^{\mathbb{Q}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right)$ and $\hat{E}_t^{\mathbb{P}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right)$, respectively,

$$\widehat{VRP}_t \equiv \hat{E}_t^{\mathbb{Q}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right) - \hat{E}_t^{\mathbb{P}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right),$$

where in practice $\hat{E}_t^{\mathbb{Q}}\left(Var_{t,t+1}\left(r_{t+1}\right)\right)$ and the true variance $Var_{t,t+1}\left(r_{t+1}\right)$ are replaced by the squared VIX and realized variance, respectively.

We then consider the regression for the time period January 1990 - December 2012 whose data we borrow from Zhou (2010):

$$\widehat{VRP}_t = \xi_0 + \xi_1' \hat{f}_t^* + \varepsilon_{i,t}. \tag{3.15}$$

The estimation results are summarized in Table 5. These results indicate that the common factor of industry RV's are positively linked to the estimate of variance risk premium. The common factor of industry RV's is a systematic risk measure while VRP is a measure of the degree of risk aversion in an economy rather than a market risk measure as argued by Bollerslev et al. (2009). The positive relationship between VRP and common factor of industry RV's can then be explained as follows: an increase (decrease) in systematic risk leads risk-averse agents to cut (increase) their consumption and investment expenditures and shift their portfolios from more (less) risky assets to less (more) risky ones, which is also a consequence of an increase (decrease) in the degree of risk aversion, as reflected by VRP.

3.5 Conclusion

We have proposed incorporating common volatility dynamics as a determinant of the optimal portfolio weights that contrasts well with both the traditional Markowitz approach and the approach by BSCV (2009) who did not account for volatility effects in their portfolio selection methods. We have empirically illustrated the positive impact of accounting for common volatility dynamics on portfolio performance in a parametric portfolio setting, and linked the common volatility factor to VRP, which is widely used in empirical analyses.

While we restricted our attention to industry portfolios in the empirical analysis to be able to understand general economic activity, further research can be undertaken considering other investment-purpose portfolios. It could be also interesting to develop forecasting methods using the parametric portfolio policy that incorporates common volatility dynamics. Finally, further work is warranted for additional portfolio statistics, such as turnover ratios and truncated weights, which we purposefully neglect in this paper to focus on the main ideas.

Bibliography

[1] Bai, J., and S. Ng. (2002). "Determining the Number of Factors in Approximate Factor Models," *Econometrica,* 77(4), pp. 1229–1279.

[2] Bai, J., and S. Ng. (2004). "A PANIC Attack on Unit Roots and Cointegration," *Econometrica,* 72(4), pp. 1127–1177.

[3] Bai, J., and S. Ng. (2013). "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics,* 176, pp. 18–29.

[4] Bakshi, G., and D. Madan. (2006). "A Theory of Volatility Spread," *Management Science,* 52, pp. 1945–56.

[5] Barndorff-Nielsen, O. E., and N. Shephard. (2004). "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics,* 2(1), pp. 1–37.

[6] Bollerslev, T., D. Osterreider, N. Sizova and G. Tauchen. (2013). "Risk and Return: Long-Run Relations, Fractional Cointegration, and Return Predictability," *Journal of Financial Economics*, 108, pp. 409–424.

[7] Bollerslev, T., G. Tauchen, and H. Zhou. (2009). "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22(11), pp. 4463–4492.

[8] Brandt, M. W., P. Santa-Clara, and R. Valkanov. (2009). "Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns," *The Review of Financial Studies*, 22(9), pp. 3411–3447.

[9] Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance," *The Journal of Finance,* 52(1), pp. 57–82.

[10] Fama, E. F. and K. R. French. (1993). "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics,* 33(1), pp. 3–56.

[11] Fama, E. F. and K. R. French. (2015). "A Five-Factor Asset Pricing Model," *Journal of Financial Economics,* 116(1), pp. 1–22.

[12] Hansen, L.P. (1982). "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, pp. 1029–1054.

[13] Markowitz, H. (1952). "Portfolio Selection," *The Journal of Finance*, 7(1), pp. 77–91.

[14] Pesaran, H. (2006). "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), pp. 967–1012.

[15] Robinson, P. M. (1995). "Gaussian Semiparametric Estimation of Long-Range Dependence," *The Annals of Statistics*, 23(5), pp. 1630–1661.

[16] Zhou, H. (2010). "Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty," Working paper Federal Reserve Board, Washington, D.C.

Table 3.1: Estimated Integration Orders of Industry Realized Volatilities

$m = 45$:

Food	Bvrgs	Tobac	Games	Books	Hshld	Clths	Hlth	Chems	Txtls	**Market**
0.36	0.47	0.61	0.45	0.50	0.37	0.48	0.35	0.49	0.53	0.41
Cnstr	Steel	FabPr	ElcEq	Autos	Carry	Mines	Coal	Oil	Util	
0.48	0.49	0.44	0.38	0.45	0.41	0.52	0.61	0.49	0.45	
Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Finan	Other	
0.46	0.45	0.56	0.42	0.39	0.34	0.49	0.48	0.52	0.49	

$m = 71$:

Food	Bvrgs	Tobac	Games	Books	Hshld	Clths	Hlth	Chems	Txtls	**Market**
0.35	0.45	0.49	0.41	0.51	0.33	0.47	0.33	0.45	0.57	0.40
Cnstr	Steel	FabPr	ElcEq	Autos	Carry	Mines	Coal	Oil	Util	
0.44	0.51	0.45	0.42	0.50	0.40	0.45	0.53	0.43	0.44	
Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Finan	Other	
0.48	0.42	0.54	0.42	0.40	0.34	0.42	0.45	0.65	0.47	

Note: *This table reports the local Whittle estimation results of the individual integration orders of industry and market realized volatilities with* $m = 45, 71$ *Fourier frequencies. Estimates are rounded to two digits after zero. Standard errors of the estimates are 0.0745 and 0.0593 respectively for* $m = 45, 71$.

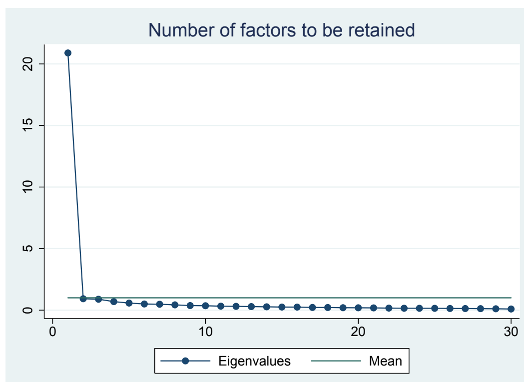


Figure 3.1: This screeplot draws the eigenvalues associated with factors and the mean eigenvalue which is equal to 1. Only eigenvalues greater than 1 are retained.

Table 3.2: Estimated Factor Loadings and Uniqueness of Variances

RV_i	Factor loadings	Ratio of variance unique to RV_i
food	0.8743	0.2357
beer	0.7593	0.4235
smoke	0.5088	0.7411
games	0.8544	0.2699
books	0.8530	0.2724
hshld	0.8622	0.2566
clths	0.8600	0.2605
hlth	0.8230	0.3227
chems	0.8934	0.2018
txtls	0.8017	0.3572
cnstr	0.9080	0.1755
steel	0.8537	0.2712
fabpr	0.9286	0.1377
elceq	0.8920	0.2044
autos	0.8528	0.2727
carry	0.8516	0.2748
mines	0.7192	0.4828
coal	0.6890	0.5252
oil	0.8161	0.3340
util	0.7699	0.4073
telcm	0.8178	0.3312
servs	0.8708	0.2418
buseq	0.8055	0.3512
paper	0.8852	0.2165
trans	0.8692	0.2444
whlsl	0.9059	0.1793
rtail	0.8707	0.2418
meals	0.8156	0.3348
fin	0.8397	0.2948
other	0.8696	0.2439

Note: *This table reports the PC estimation results for industry RV's. The uniqueness ratios are quite small indicating that the common factor explains much of the variance of each industry RV.*

Table 3.3: Portfolio performance with common volatility factor

Parameters	In-Sample	Out-of-Sample
$\hat{\theta}_{smb}$	0.0217***	0.0067***
	(0.0042)	(0.0015)
$\hat{\theta}_{hml}$	0.0084***	0.0033***
	(0.0022)	(0.0012)
$\hat{\gamma}$	-0.0756***	-0.0254***
	(0.0107)	(0.0058)
$\lvert w_i \rvert \times 100$	0.3871	1.6822
$\max w_i \times 100$	1.0639	4.0439
$\min w_i \times 100$	-4.4701	-3.6111
$\sum w_i I(w_i < 0)$	-0.4930	-0.1308
$\sum I(w_i \leq 0)/n$	0.2047	0.0933
$\bar{r}$	0.51%	1.87%
$\sigma(r)$	0.0161	0.0359
Skewness	5.4814	3.1426
Sharpe Ratio	0.3158	0.5211

Note: This table reports the estimation results of portfolio policy in (3.8). In-sample study covers the period from January 1966 to August 2008, and the out-of-sample study, carried out based on a rolling window of 12 months, covers the period from September 2008 to December 2014. Rows 7 to 11 show statistics of the portfolio weights averaged across time. These statistics include average absolute portfolio weight ($\lvert w_i \rvert \times 100$), the average maximum ($\max w_i \times 100$) and minimum ($\min w_i \times 100$) portfolio weights, the average sum of negative portfolio weights ($\sum w_i I(w_i < 0)$) and the fraction of the negative portfolio weights ($\sum I(w_i \leq 0)/n$), respectively. Rows 12 to 15 display the monthly portfolio statistics: average monthly return ($\bar{r}$), standard deviation ($\sigma(r)$), skewness and Sharpe ratio. Risk aversion is assumed to be equal to five. "***" indicates statistical significance at the 1% level.

Table 3.4: Portfolio performance without common volatility factor

Parameters	In-Sample	Out-of-Sample
$\hat{\theta}_{smb}$	0.00018**	0.00011
	(0.00008)	(0.00015)
$\hat{\theta}_{hml}$	0.00058***	0.00061***
	(0.00008)	(0.00011)
$\lvert w_i \rvert \times 100$	0.1949	1.3333
$\max w_i \times 100$	0.2113	1.3984
$\min w_i \times 100$	0.1807	1.2521
$\sum w_i I(w_i < 0)$	0	0
$\sum I(w_i \leq 0)/n$	0	0
$\bar{r}$	0.19%	1.63%
$\sigma(r)$	0.0182	0.0762
Skewness	-0.4519	-0.5559
Sharpe Ratio	0.1044	0.2138

Note: This table reports the estimation results of portfolio policy in (3.8) without the common factor of industry RV's, i.e. $\gamma = 0$. In-sample study covers the period from January 1966 to August 2008, and the out-of-sample study, carried out based on a rolling window of 12 months, covers the period from September 2008 to December 2014. Rows 7 to 11 show statistics of the portfolio weights averaged across time. These statistics include average absolute portfolio weight ($\lvert w_i \rvert \times 100$), the average maximum ($\max w_i \times 100$) and minimum ($\min w_i \times 100$) portfolio weights, the average sum of negative portfolio weights ($\sum w_i I(w_i < 0)$) and the fraction of the negative portfolio weights ($\sum I(w_i \leq 0)/n$), respectively. Rows 12 to 15 display the monthly portfolio statistics: average monthly return ($\bar{r}$), standard deviation ($\sigma(r)$), skewness and Sharpe ratio. Risk aversion is assumed to be equal to five. "***" and "**" indicate statistical significance at the 1% and 5% level, respectively.

Table 3.5: VRP and Common Factor of Industry RV's

Estimates	$\hat{\xi}_0$	$\hat{\xi}_1$
	0.0088	0.5459***
	(0.0479)	(0.0433)
	[0.8550]	[0.0000]

Note: *This table reports the regression results of the variance risk premium estimate on the common factor of industry RV's based on (3.15). Heteroskedasticity and autocorrelation robust standard errors are reported in parantheses and the corresponding p-values in square brackets. *** indicates significance at the 1% level.*