System Estimation of Panel Data Models under Long-Range Dependence

Yunus Emre Ergemen

CREATES Research Paper 2016-2

System Estimation of Panel Data Models under Long-Range Dependence[*]

Yunus Emre Ergemen[†]

CREATES and Aarhus University

January 13, 2016

Abstract

A general dynamic panel data model is considered that incorporates individual and interactive fixed effects allowing for contemporaneous correlation in model 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 featured. 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 standard 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

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 aggregation of individual characteristics. Robinson (1978), Granger (1980) and Chambers (1998) show that aggregating autoregressive models can lead to fractionally integrated models that have dramatically different correlation structures for both dependent and independent individual series. On the empirical side, Gil-Alaña and Robinson (1997) 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. (2013). Furthermore, Michelacci and Zaffaroni (2000) 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 (2014) 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.

Based on the evidence in the literature 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. (2011) propose a test for memory in fractionally integrated panels. Robinson and Velasco (2015) 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 (2015) 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.

In this paper, we consider a fractionally integrated panel data model with individual and interactive fixed effects that allows for contemporaneous correlations between the innovations of the panel dependent variable and covariates. The model nests stationary $I(0)$ and nonstationary $I(1)$ autoregressive panel data models that are extensively used in economic modelling,

but unlike autoregressive ones, it has smoothness in the unit-root case (and elsewhere), thus the parameter estimates and related test statistics have standard asymptotic distributions, see also Robinson and Velasco (2015) for a discussion. The 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 studies dealing with possibly nonstationary variables. Furthermore, parameter heterogeneity is allowed in the model so that unit-specific inference can be obtained 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 in that a nonstationary variable is assumed to be $I(1)$ for all cross-section units merely based on unit-root testing.

This paper contributes to the literature in many ways. First, unlike in Hassler et al. (2011) and Robinson and Velasco (2015), 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 (2003) and Hualde and Robinson (2007). Second, unlike in Ergemen and Velasco (2015), 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, particularly in cases where endogeneity induced by the unobserved common factor may not be the only source of contemporaneous correlation, e.g. in the analysis of debt and GDP relationship in an economic union where feedback effects between these indicators depend also on country-specific innovations, such as a change in government spending. 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 factors is greater than those of the idiosyncratic innovations. Thus a pure time-series approach, which would automatically incorporate a latent factor structure in the error term, may fail to detect possible cointegrating relationships. In this case, possible cointegrating relationships can only be disclosed after the 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 analysis 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 (2007). 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. We then apply the method to the long-run relationship between log-debt and log-GDP for 20 high-income OECD countries.

The remainder of the paper proceeds as follows. Next section presents the model and the conditions imposed to study it. Section 3 details the estimation procedures for linear and fractional integration parameters and contains the main results. Section 4 briefly discusses the inclusion of deterministic trends. Section 5 presents a finite-sample study based on Monte Carlo experiments, and Section 6 presents the empirical application. Section 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; "$\rightarrow_p$" denotes convergence in probability; "$\rightarrow_d$" denotes convergence in distribution; and $\|A\| = (trace(AA'))^{1/2}$. All mathematical proofs and intermediate technical results are collected in an appendix at the end of the paper.

2 Model

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

$$y_{it} = \alpha_i + \beta'_{i0} x_{it} + \lambda'_i f_t + \Delta_t^{-d_{i0}} \epsilon_{1it}, \tag{1}$$

$$x_{it} = \mu_i + \gamma'_i f_t + \Delta_t^{-\vartheta_{i0}} \epsilon_{2it},$$

where the scalar y_{it} and $k \times 1$ vector of covariates x_{it} are observable with idiosyncratic innovations that have unknown true integration orders d_{i0} and ϑ_{i0}, respectively, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. All idiosyncratic components of x_{it} have the same integration order ϑ_{i0}. In this model, imposing $\beta_{i0} \neq 0$ for identification, y_{it} is treated as the dependent variable while x_{it} consists of possibly endogenous control, or explanatory, variables. In practice the choice of which variable to call the dependent variable is generally made based on the specific empirical target pursued, and in this regression-based framework a dependent variable is selected that ensures $\beta_{i0} \neq 0$. Furthermore in (1), f_t is the $m \times 1$ vector of unobserved common factors that may be integrated of a common unknown order δ, i.e. $f_t \sim I(\delta)$, and $m \times 1$ vector λ_i and $m \times k$ matrix γ_i contain the corresponding factor loadings. This multi-factor structure is in line with the factor structure used by Pesaran (2006) and generalizes the model for use in practice. 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

$$\Delta_t^{-d} \epsilon_{it} = \Delta^{-d} \epsilon_{it} 1(t > 0) = \sum_{j=0}^{t-1} \pi_j(-d) \epsilon_{it-j}, \quad \pi_j(-d) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \tag{2}$$

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 $\Gamma(0)/\Gamma(0) = 1$. With the prime denoting transposition, $\epsilon_{it} = (\epsilon_{1it}, \epsilon'_{2it})'$ is a covariance stationary process, allowing for $Cov(\epsilon_{1it}, \epsilon_{2it}) \neq 0$, whose short-memory vector-autoregressive (VAR) dynamics are described by

$$B(L; \theta_i) \epsilon_{it} \equiv \left(I_{k+1} - \sum_{j=1}^{p} B_j(\theta_i) L^j \right) \epsilon_{it} = v_{it}, \tag{3}$$

where L is the lag operator, θ_i the short-memory parameters, I_{k+1} the $(k+1) \times (k+1)$ identity matrix, B_j the $(k+1) \times (k+1)$ upper-triangular matrices, and v_{it} is a $(k+1) \times 1$ sequence

that is identically and independently distributed across i and t with zero mean and variance-covariance matrix $\Omega_i > 0$. The arrays $\{\alpha_i, i \geq 1\}$ and $\{\mu_i, i \geq 1\}$ are unobserved individual fixed effects; $\{f_t, t > 0\}$ is the $I(\delta)$ vector of unobserved common factors that induces cross-section dependence and possibly further endogeneity in the system; $\{\lambda_i, i \geq 1\}$ and $\{\gamma_i, i \geq 1\}$ are vectors and matrices of unobserved factor loadings indicating how much each cross-section unit is affected by f_t.

Set

$$\vartheta_{max} = \max_i \vartheta_{i0} \quad \text{and} \quad d_{max} = \max_i d_{i0},$$

and let d^* be a prewhitening parameter that is chosen by the econometrician. Also denote the lower bounds of the allowed range of ϑ_{i0} and d_{i0} by $\underline{\vartheta}$ and $\underline{d}$, respectively. Then we introduce the following conditions to analyze the system in (1).

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 $rank(\overline{C}) = m \leq k+1$ where*

$$\overline{C} = \left(\overline{\gamma\beta} + \bar{\lambda} \quad \bar{\gamma} \right)$$

with $\bar{\gamma} = n^{-1} \sum_{i=1}^n \gamma_i$, $\bar{\lambda} = n^{-1} \sum_{i=1}^n \lambda_i$ and $\overline{\gamma\beta} = n^{-1} \sum_{i=1}^n \gamma_i\beta_i$.

Assumption 1.1 is a fairly general version of the assumptions used by e.g. Hualde and Robinson (2011) and Nielsen (2014), 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 singling out a fixed value although for

most applications $d^* = 1$ would suffice anticipating $\vartheta_{i0}, \delta, d_{i0} < 5/4$. The requirement on the lower bounds of the allowed range of memory values is necessary to ensure consistency of the parameter estimates while not being too restrictive since they can move to contain the true memory values in a radius of $1/2$. This further implies that $\vartheta_{i0} - d_{i0} < 1/2$, i.e. at most weak fractional cointegration. That $\vartheta_{i0} \neq d_{i0}$ is required to circumvent perfect collinearity in the regression equation is not restrictive because the methodology still allows for the observable series to have the same integration orders when $\delta > \vartheta_{i0}, d_{i0}$.

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 can be explained as follows: in general, if there are k covariates, the maximum number of factors that can be featured is $k + 1$ so that the factor space can be spanned when the number of factors $m \leq k + 1$. The non-zero mean possibility in common factors, i.e. when $\alpha^f \neq 0$, allows for a drift.

Assumptions 1.3 and 1.4 are standard in the factor models literature and have been used by e.g. Pesaran (2006) and Bai (2009). 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.

Assumption 2 (**System errors**). *In the representation*

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

1. *$B_j(\cdot)$ are upper-triangular matrices that satisfy $\sum_{j=1}^{\infty} j \, \|B_j\| < \infty$, $det \, \{B(s; \theta_i)\} \neq 0$, $|s| = 1$ for $\theta_i \in \Theta$;*

2. *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.*

Assumption 2.1 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 upper-triangularity assumption on the short-memory matrices, B_j, implies that y_{it} depends on $y_{it-1}, x_{it}, y_{it-2}, x_{it-1}, \dots$ while each component of x_{it} depends only on its past values. This condition permits the use of univariate nonlinear optimizations, which introduces a great deal of parsimony further developing the triangular structure of the

system, and it is in line with the long-run VAR restriction of Blanchard and Quah (1989) and the short-run VAR restriction of Sims (1987). In practice, the choice of a dependent variable should be made accordingly.

The moment requirement in Assumption 2.2 is very standard and in general easily satisfied under Gaussianity. The iid requirement therein 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.

3   Parameter Estimation

3.1   Prewhitening and Projection of the Common Factor Structure

In a standard way, we first-difference (1) to remove the fixed effects,

$$\Delta y_{it} = \beta_{i0}'\Delta x_{it} + \lambda_i'\Delta f_t + \Delta_t^{1-d_{i0}}\epsilon_{1it}, \tag{4}$$

$$\Delta x_{it} = \gamma_i'\Delta f_t + \Delta_t^{1-\vartheta_{i0}}\epsilon_{2it},$$

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 (2015).

The first-differenced model in (4) can be prewhitened from idiosyncratic long-range dependence for some fixed exogenous differencing choice, d^* as prescribed in Assumption 1.1, 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

$$y_{it}(d^*) = \beta_{i0}'x_{it}(d^*) + \lambda_i'f_t(d^*) + \epsilon_{1it}(d^* - d_{i0}), \tag{5}$$

$$x_{it}(d^*) = \gamma_i'f_t(d^*) + \epsilon_{2it}(d^* - \vartheta_{i0}).$$

Thus, using the notation $z_{it}(\tau_1, \tau_2) = (y_{it}(\tau_1), x'_{it}(\tau_2))'$, (5) can be written in the vectorized form as

$$z_{it}(d^*, d^*) = \zeta\beta'_{i0}x_{it}(d^*) + \Lambda'_i f_t(d^*) + \epsilon_{it}(d^* - d_{i0}, d^* - \vartheta_{i0}), \qquad (6)$$

where $\zeta = (1, 0, \ldots, 0)'$, and $\Lambda_i = (\lambda_i\ \gamma_i)$.

The structure $\Lambda'_i f_t(d^*)$ in (6) induces cross-section correlation between units i through $f_t(d^*)$. The common factors 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 (2006) 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 (2015). 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 (PC) approach has been greatly extended in factor analysis by e.g. Bai and Ng (2002) and Bai and Ng (2013). 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 (2006). Moreover, PC 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 (6) in cross-section averages as

$$\bar{z}_t(d^*, d^*) = \overline{\zeta\beta'_0 x_t(d^*)} + \bar{\Lambda}' f_t(d^*) + \bar{\epsilon}_t(d^* - d_0, d^* - \vartheta_0), \qquad (7)$$

where $\bar{\epsilon}_t(d^* - d_0, d^* - \vartheta_0)$ is $O_p(n^{-1/2})$ for large enough d^*. Thus, $\bar{z}_t(d^*, d^*)$ and $\overline{\zeta\beta'_0 x_t(d^*)}$ asymptotically capture all the information provided by the common factor provided that $\bar{\Lambda}$

9

is full rank. Note that $\bar{x}_t(d^*)$ is readily contained in $\bar{z}_t(d^*, d^*)$ and the β_{i0} do not have any contribution in terms of dynamics in $\overline{\zeta\beta_0' x_t(d^*)}$ 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 $i = 1, \ldots, n$,

$$\mathbf{z}_i(d^*, d^*) = \mathbf{x}_i(d^*)\beta_{i0}\zeta' + \mathbf{F}(d^*)\mathbf{\Lambda}_i + \mathbf{E}_i\left(d^* - d_{i0}, d^* - \vartheta_{i0}\right), \qquad (8)$$

where $\mathbf{E}_i\left(d^* - d_{i0}, d^* - \vartheta_{i0}\right) = (\epsilon_{i2}\left(d^* - d_{i0}, d^* - \vartheta_{i0}, \right), \ldots, \epsilon_{iT}\left(d^* - d_{i0}, d^* - \vartheta_{i0}\right))'$ and $\mathbf{F}(d^*) = (f_2(d^*), \ldots, f_T(d^*))'$.

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

$$\bar{\mathbf{M}}_{T_1}(d^*) = \mathbf{I}_{T_1} - \bar{\mathbf{z}}(d^*, d^*)(\bar{\mathbf{z}}'(d^*, d^*)\bar{\mathbf{z}}(d^*, d^*))^- \bar{\mathbf{z}}'(d^*, d^*), \qquad (9)$$

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 (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.

Introduce the infeasible projection matrix based on unobserved factors,

$$\mathbf{M}_F(d^*) = \mathbf{I}_{T_1} - \mathbf{F}(d^*)(\mathbf{F}(d^*)'\mathbf{F}(d^*))^- \mathbf{F}(d^*)'.$$

Then adopting Pesaran (2006)'s argument under the rank conditions in Assumption 1.2 and 1.4, we have as $(n, T)_j \to \infty$ that

$$\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) \approx \mathbf{M}_F(d^*)\mathbf{F}(d^*) = 0, \qquad (10)$$

implying that both projection matrices can be employed interchangeably for factor removal in the asymptotics provided that the rank condition holds.

Based on (8) and using (10), the defactored observed series for each $i = 1, \ldots, n$,

$$\tilde{\mathbf{z}}_i(d^*, d^*) \approx \tilde{\mathbf{x}}_i(d^*)\beta_{i0}\zeta' + \tilde{\mathbf{E}}_i\left(d^* - d_{i0}, d^* - \vartheta_{i0}\right), \tag{11}$$

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 given in (10) is of order $O_p\left(n^{-1} + (nT)^{-1/2}\right)$ as shown in Appendix A.1.

3.2   Estimation of Linear Model Parameters

Writing (11) for $i = 1, \ldots, n$, and $t = 2, \ldots, T$ we now integrate the defactored series back by d^* to their original integration orders, disregarding the projection errors that are negligible under Assumption 1.1 as $n \to \infty$, see Ergemen and Velasco (2015), to perform estimations as

$$\tilde{z}_{it}^*(d_i, \vartheta_i) = \zeta \beta_{i0}' \tilde{x}_{it}^*(d_i) + \tilde{\epsilon}_{it}^* \left(d_i - d_{i0}, \vartheta_i - \vartheta_{i0}\right), \tag{12}$$

where the first and second equations of (12) 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 (3), (12) can be written as

$$\tilde{z}_{it}^*(d_i, \vartheta_i) - \sum_{j=1}^{p} B_j(\theta_i)\tilde{z}_{it-j}^*(d_i, \vartheta_i) \tag{13}$$

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

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

$$\tilde{x}_{it}^*(\vartheta_i) - \sum_{j=1}^{p} B_{2j}(\theta_i)\tilde{z}_{it-j}^*(d_i, \vartheta_i) = -\sum_{j=1}^{p} B_{2j}(\theta_i)\zeta \beta_{i0}' \tilde{x}_{it-j}^*(d_i) + \tilde{v}_{2it}^* \left(\vartheta_i - \vartheta_{i0}\right) \tag{14}$$

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

$$\tilde{y}_{it}^*(d_i) = \beta_{i0}'\tilde{x}_{it}^*(d_i) + \rho_i'\tilde{x}_{it}^*(\vartheta_i) + \sum_{j=1}^{p}\left(B_{1j}(\theta_i) - \rho_i'B_{2j}(\theta_i)\right)\tilde{z}_{it-j}^*(d_i,\vartheta_i) \tag{15}$$

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

with B_{1j} denoting the first row of B_j constituting the $1 \times (k+1)$ vector and B_{2j} denoting the remaining $k \times (k+1)$ matrix that correspond to the first and second equations of $\tilde{z}_{it}^*(d_i,\vartheta_i)$, respectively, and $\rho_i = E[\tilde{v}_{2it}^*\tilde{v}_{2it}^{*\prime}]^{-1}E[\tilde{v}_{2it}^*\tilde{v}_{1it}^*]$.

Under (15), cointegration (i.e. when $\vartheta_{i0} > d_{i0}$) 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 setting in which $d_{i0} > \vartheta_{i0}$ since the error term in (15) 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 to impose is $\vartheta_{i0} \neq d_{i0}$ because when $\vartheta_{i0} = d_{i0}$, this leads to collinearity in (15) thus rendering the identification of β_{i0} and ρ_i impossible. The 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 $\rho_i \equiv 0$, corresponding to exogenous regressors, has been developed by Ergemen and Velasco (2015), 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 $q \times [2k(p+1)+p]$ restriction matrix Q that is $I_{2k(p+1)+p}$ when there are no prior zero restrictions on B_j, and a $q < 2k(p+1)+p$ 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 (15) as

$$\tilde{y}_{it}^*(d_i) = \omega_i'Q\tilde{Z}_{it}^*(d_i,\vartheta_i) + \tilde{v}_{1it}^*(d_i - d_{i0}) - \rho_i'\tilde{v}_{2it}^*(\vartheta_i - \vartheta_{i0}) \tag{16}$$

with

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

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

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

$$\hat{\omega}_i(\tau_1, \tau_2) := M_i(\tau_1, \tau_2)^{-1} m_i(\tau_1, \tau_2) \tag{17}$$

with

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

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 (17) as

$$\hat{\beta}_i(\tau_1, \tau_2) = \psi_\beta' \hat{\omega}_i(\tau_1, \tau_2), \quad \psi_\beta = \left(\overbrace{1, \ldots, 1}^{k-many}, 0, \ldots, 0\right)'. \tag{18}$$

While $\hat{\beta}_i$ in (18) 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 (2015) 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

$$\hat{\rho}_i(\tau_1, \tau_2) = \psi_\rho' \hat{\omega}_i(\tau_1, \tau_2), \quad \psi_\rho = \left(\overbrace{0, \ldots, 0}^{k-many}, \overbrace{1, \ldots, 1}^{k-many}, 0, \ldots, 0\right)'.$$

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. $\psi_\theta = \left(\underbrace{0, \ldots, 0}_{2k-many}, 1, \ldots, 1\right)'$.

13

Finally, in case of interest, fixed effects α_i and μ_i in (1) can also be consistently estimated after estimating other model parameters and applying proper whitening to the time-varying model components. We do not explore this further in this paper since first differencing as part of our methodology removes fixed effects with the advantage of not requiring any restrictions on them.

3.3   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_{i0} and ϑ_{i0}. Estimation of long-range dependence parameters in the panel data context is a relatively new topic. Robinson and Velasco (2015) 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 (2015) 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 (13). Assuming an upper-triangular structure for $B_j(\theta_i)$ in (3), we write (14) as

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

with

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

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

$$\hat{\phi}_i(\vartheta) := G_i(\vartheta)^{-1} g_i(\vartheta) \tag{19}$$

where

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

Having obtained (19), ϑ_{i0} can be estimated, from any of the k equations of x_{it} since the model assumes a common ϑ_{i0} for all components of x_{it}, based on the criterion whose multivariate version is given by

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

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

Then d_{i0} can be estimated from (16) 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 $O_p(T^{-\underline{d}})$ and $O_p(T^{-\underline{\vartheta}})$. 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 $\sqrt{T}$-consistent initial estimates, $\hat{\tau}_i = (\hat{d}_i, \hat{\vartheta}_i)$, as

$$\tau_i = \hat{\tau}_i - \mathbf{H}_T^{-1}(\hat{\tau}_i)\mathbf{h}_T(\hat{\tau}_i), \tag{20}$$

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

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

3.4 Asymptotic Results for Slope and Long-Range Dependence Parameters

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 (2007) 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 (2007) 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 other model parameters. This contrasts with the results derived by Robinson and Velasco (2015), 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 $(n, T)_j \to \infty$,*

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

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 A.4.

This joint estimation result differs from the one by Robinson and Hualde (2003) but is similar to that by Hualde and Robinson (2007) 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 (2007). Under lack of autocorrelation and endogeneity induced by the idiosyncratic innovations, Ergemen and Velasco (2015) 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.

3.5 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,

$$\hat{\beta}_{CCMG}\left(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\vartheta}}\right) := \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i\left(\hat{d}_i, \hat{\vartheta}_i\right) \tag{21}$$

where bold denotes the parameter vectors.

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.

For the analysis of the mean-group estimate, we introduce the following extra condition.

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

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

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

Assumption 3.1 is a regularity condition ensuring the existence of the least-square estimate in (17) and thus of the slope estimate in (18) while Assumption 3.2 is used in the derivation of asymptotic results of the common-correlation mean group estimate defined in (21). These conditions are also used by Pesaran (2006), but based on stationary $I(0)$ variables.

We finally establish the asymptotic behaviour of the common correlated mean-group slope estimate in the next theorem.

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

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

This theorem extends the results by Pesaran (2006) and Kapetanios et al. (2011) on $I(0)$ and $I(1)$ variables, where this GLS-type estimate now converges at the $\sqrt{n}$ rate without requiring any restrictions 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.

4   Deterministic Trends

While our model in (1) can accommodate both deterministic and stochastic unobserved trends via the common factor vector 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 the elements of λ_i and γ_i. To relax such a restriction and allow for separate time trends, we extend the model in (1) as

$$y_{it} = \alpha_i + \alpha_i^1 q(t) + \beta_{i0}' x_{it} + \lambda_i' f_t + \Delta_t^{-d_{i0}} \epsilon_{1it}, \tag{22}$$
$$x_{it} = \mu_i + \mu_i^1 r(t) + \gamma_i' f_t + \Delta_t^{-\vartheta_{i0}} \epsilon_{2it},$$

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

The case in which $q(t)$ and $r(t)$ in (22) are linear, possibly with drifts, can be studied 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 (9) to include ones at its first column, which then mirrors fixed-effects estimation in first differences. In both of these approaches, possibly with additional rate restrictions on n and T, asymptotics remain the same: 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 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 (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.

Furthermore, structural breaks can be featured in the common factors under additional assumptions that control the size of $\mathbf{F}(d^*)'\mathbf{F}(d^*)/T$. Such common breaks could be dealt with by projections and the asymptotic results would not change. Additionally, local or global breaks in trends with stationary or nonstationary errors can be studied cf. Perron and Zhu (2005), but under more restrictive assumptions, which we do not explore in this paper.

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, considering a scalar y_{it} and x_{it} for simplicity. 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. Considering different cross-section sizes and time-series lengths, we fix 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.

For $n = 10$ and $T = 50$, Tables 1 and 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 since increasing δ introduces further noise for the estimation. 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 3 presents the case of endogeneity without short-memory dynamics. Compared to the results in Table 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 4, slope estimates perform similarly to the results in Table 1. The performance of fractional parameter estimates $\hat{\vartheta}$ and $\hat{d}$, however, are slightly worsened compared to the results in Table 1. A further comparison between Tables 2 and 3 reveals that under endogeneity, short-memory dynamics help both the feasible and infeasible slope estimates in terms of performance in some cases which we conjecture to be due to the availability of lagged-variable instrumentation. 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 and 4.

We also explore the behaviour of the estimates under autocorrelation and endogeneity taking $n = 20$ and $T = 54$, which matches the sample size of the panel data set used in the empirical application in Section 6, as well as $n = 50$ and $T = 100$, which is generically used in simulation designs in panel data literature. These results are reported in Tables 5 and 6, respectively. In both cases, bias profiles of the estimates generally improve with the long-range dependence parameters being estimated with much less bias. The performance of the estimates is ameliorated compared to the results in Table 2 as a reflection of the increase in the sample size.

Finally, in Figure 1, we show that heterogeneous slope estimates individually behave quite well taking $\vartheta_{i0} = 1, d_{i0} = 1, \delta = 1, \theta_1 = \theta_2 = 0.5, \rho_i = 0.5$ under the sample size $n = 20$ and

$T = 54$, which is used in the empirical application presented in the next section. Also for different parameter values, the behaviour of heterogeneous slope parameters are similar so we do not report extensive results due to space considerations.

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

In structural estimation, using comparable level data, such as GDP and debt, leads to easy-to-interpret results. An additional advantage of using such level data is that the estimation results have clear interpretations. With this in mind, in this empirical application we study the long-run relationship between real GDP and debt in logs, whose persistence characteristics we expect to be similar.

There is a vast literature on the nature of the relationship between GDP and debt studying whether it is just correlation between the series or there is causality from one to the other. As one recent reference, Panizza and Presbitero (2014) show that there is a negative correlation between GDP growth and public debt but there is no causal relationship between the two in OECD countries. On the other hand, Puente-Avojin and Sanso-Navarro (2015) present evidence for causality from debt to GDP, GDP to debt or in both directions for some of the OECD countries. While it would certainly be interesting to delve more into the question of whether or not there is causality between these indicators, in this empirical application we will only focus on cointegration between GDP and debt, which suits our methodology. Also following the literature, we treat log-GDP as the dependent variable and log-debt as the endogenous explanatory variable.

In the analysis, 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. To construct the debt data, we use the PPP-based GDP data for the sample period.

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 log-GDP and log-debt in Figures 2 and 3, 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 memory estimations based on a parametric conditional-sum-of-squares criterion used in Ergemen and Velasco (2015)'s basic model, and report the results in Table 7. The estimation results show that log-GDP and log-debt are integrated of an order around unity, which is in line with the literature in that they are treated as $I(1)$ variables.

Furthermore, we note that in the analysis of log-GDP and log-debt, our method can nest two unobserved common factors that on average capture other relevant indicators for the study. Allowance for these common factors can be considered a more flexible way of modelling the relationship between log-GDP and log-debt as opposed to adding separate observable series, e.g. inflation as in Chudik et al. (2013). Having allowed for persistence in the common factors, we can estimate the maximum integration order of the two. The maximum memory of the common factors of log-GDP and log-debt, which is estimated based on the cross-section averages of the time-stacked series, is found to be of order 1.0079 indicating that removing the common factors is essential for disclosing possible cointegrating relationships. To justify this, we provide benchmark estimation results based on the pure time-series estimation approach, which can contain the factor structure only in the error term, by Hualde and Robinson (2007) assuming a VAR(1) structure. Along this line, we are interested in identifying nontrivial cointegrating relationships: such relationships between log-GDP and log-debt exist if a) estimated slope coefficients are significantly different from zero; b) 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 8.

According to the results in Table 8, parameter estimates are significant for all countries except Australia and Canada, with mixed signs. From these results, it is further indicated that log-GDP and log-debt 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, estimating our model in (22), we check the long-run relationship between log-GDP and log-debt, again assuming a VAR(1) structure. These estimation results are reported in

Table 9.

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 9, we find that log-GDP and log-debt 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, Japan, 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, see Elmendorf and Mankiw (1999), Reinhart and Rogoff (2010) and Chudik et al. (2013), a positive effect can be, for example, due to the debt increasing because of government spending while also fuelling real GDP, see DeLong and Summers (2012).

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.

A further comparison between the estimation results in Tables 8 and 9 shows the reversal of the slope estimates for several countries. This can be explained by the fact that the results in Table 8 are obtained under a time-series setup that neglects country-specific heterogeneity (institutions, geographical location, etc.) as well as cross-country dependence (OECD membership, high income, etc.) whereas the results in Table 9 are obtained incorporating those.

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 8 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 organizations or unions, as in a pure time-series estimation, identifying the true nature of the relationships between these variables will not be possible.

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. (2011). While this is a theoretical possibility in (1) with additional conditions on the common factor vector, f_t, we do not analyze spatial dependence explicitly, only allow for contemporaneous correlations in the innovations to partially account for it. Parametric modelling of spatial dependence, see e.g. Pesaran and Tosetti (2011), may provide further insights. Moreover, the multiple regression framework can be extended to allow for x_{it} whose elements display different degrees of persistence. 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. PC estimation of fractionally integrated factor models is yet to be explored in the literature.

A Technical Appendix

A.1 Proof of Theorem 1

Projections are carried out based on (9). Denoting $\bar{\mathbf{z}}(d^*, d^*) \equiv \bar{\mathbf{z}}(d^*)$, let us write

$$\mathbf{x}_i'(d^*)\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) = \mathbf{x}_i'(d^*)\mathbf{I}_{T_1}\mathbf{F}(d^*) - \mathbf{x}_i'(d^*)\bar{\mathbf{z}}(d^*)(\bar{\mathbf{z}}'(d^*)\bar{\mathbf{z}}(d^*))^-\bar{\mathbf{z}}'(d^*)\mathbf{F}(d^*), \qquad (23)$$

with

$$\bar{\mathbf{z}}(d^*) = \mathbf{F}(d^*)\bar{\mathbf{C}} + \bar{\mathcal{E}}\left(\mathbf{d}^* - \mathbf{d}, \mathbf{d}^* - \boldsymbol{\vartheta}\right) \qquad (24)$$

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

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

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 (23) can be expressed as

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

$$
T_1^{-1}\bar{\mathbf{z}}'(d^*)\bar{\mathbf{z}}(d^*) = T_1^{-1}\bar{\mathbf{C}}'\mathbf{F}'(d^*)\mathbf{F}(d^*)\bar{\mathbf{C}} + T_1^{-1}\bar{\mathbf{C}}'\mathbf{F}'(d^*)\bar{\mathcal{E}} + T_1^{-1}\bar{\mathcal{E}}'\mathbf{F}(d^*)\bar{\mathbf{C}} + T_1^{-1}\bar{\mathcal{E}}'\bar{\mathcal{E}}
$$

$$
T_1^{-1}\bar{\mathbf{z}}'(d^*)\mathbf{F}(d^*) = T_1^{-1}\bar{\mathbf{C}}'\mathbf{F}'(d^*)\mathbf{F}(d^*) + T_1^{-1}\bar{\mathcal{E}}'\mathbf{F}(d^*).
$$

By Assumption 2,

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

with $\sum_{j=1}^{\infty} j\,\|B_j\| < K$, 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 the simpler case to motivate the main ideas.

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

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

Then, by the idempotence of the projection matrix, this result implies that

$$
\mathbf{x}_i'(d^*)\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) = \mathbf{x}_i'(d^*)\mathbf{M}_F\mathbf{F}(d^*) + O_p\left(\frac{1}{n} + \frac{1}{\sqrt{nT}}\right), \tag{25}
$$

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

$$
T_1^{1/2}\mathbf{x}_i'(d^*)\bar{\mathbf{M}}_{T_1}(d^*)\mathbf{F}(d^*) = T_1^{1/2}\mathbf{x}_i'(d^*)\mathbf{M}_F\mathbf{F}(d^*) + O_p\left(\frac{\sqrt{T}}{n}\right). \tag{26}
$$

After prewhitening by d^*, the assumptions imposed by Robinson and Hidalgo (1997) are satisfied so the case in which finite $p \neq 0$ can be established using their method with the only difference being the details of our method's projection algebra that we detail in the $p = 0$ case to emphasize the contributions by using our method. Along this line, Ergemen and Velasco (2015) show the asymptotic behaviour of their estimates when $p \neq 0$ so we do not repeat those details in the proofs as they loom largely on this already lengthy paper.

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 (15), denoting $\sum_t = \sum_{t=2}^T$,

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

from which we can write

$$\hat{\beta}_i(d, \vartheta) - \beta_{i0} = \frac{\sum_t \tilde{x}_{it}^*(d)\tilde{v}_{1.2it}^* \sum_t \tilde{x}_{it}^*(\vartheta)'\tilde{x}_{it}^*(\vartheta) - \sum_t \tilde{x}_{it}^*(\vartheta)\tilde{v}_{1.2it}^* \sum_t \tilde{x}_{it}^*(d)'\tilde{x}_{it}^*(\vartheta)}{\sum_t \tilde{x}_{it}^*(d)'\tilde{x}_{it}^*(d) \sum_t \tilde{x}_{it}^*(\vartheta)'\tilde{x}_{it}^*(\vartheta) - (\sum_t \tilde{x}_{it}^*(d)'\tilde{x}_{it}^*(\vartheta))}, \tag{27}$$

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

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

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

$$L_{i,T}(\vartheta) = \frac{1}{T}\tilde{\mathbf{x}}_i^*(\vartheta)'\tilde{\mathbf{x}}_i^*(\vartheta). \tag{28}$$

Now,

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

where $\Delta^{d^* - 1}\Delta\tilde{\mathbf{x}}_i = \Delta^{d^* - 1}\Delta\mathbf{x}_i - \bar{\mathbf{z}}(d^*)\hat{\varsigma}_x$ with $\hat{\varsigma}_x = (\bar{\mathbf{z}}(d^*)'\bar{\mathbf{z}}(d^*))^{-1}\bar{\mathbf{z}}(d^*)'\Delta^{d^* - 1}\Delta\mathbf{x}_i$ so that

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

Next, to be able to make use of (25), let us write

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

with

$$I_x = \Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i} - \mathbf{F}(\vartheta)\left(\mathbf{F}(d^*)'\mathbf{F}(d^*)\right)^{-1}\mathbf{F}'(d^*)\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i},$$

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

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

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

$$M_z := M_z(\vartheta) = \bar{\mathbf{z}}(\vartheta)(\bar{\mathbf{z}}(d^*)'\bar{\mathbf{z}}(d^*))^{-1}\bar{\mathbf{z}}'(d^*),$$

we can write (28) as

$$\frac{1}{T}\left\{\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i} - M_f\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i} + (M_f - M_z)\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\right\}$$

$$\times \left\{\Delta^{\vartheta-\vartheta_{i0}}\mathbf{v}_{2i} - M_f\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i} + (M_f - M_z)\Delta^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\right\}',$$

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 (2015) and Theorem 1 of Ergemen and Velasco (2015). To show that the second squared term is negligible, write

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

28

where

$$M_f'M_f = \mathbf{F}(d^*)\left(\mathbf{F}(d^*)'\mathbf{F}(d^*)\right)^{-1}\mathbf{F}(\vartheta)'\mathbf{F}(\vartheta)\left(\mathbf{F}(d^*)'\mathbf{F}(d^*)\right)^{-1}\mathbf{F}(d^*)' \tag{29}$$

satisfying under Assumption 1 that

$$\frac{\mathbf{F}(d^*)'\mathbf{F}(d^*)}{T} \to_p \Sigma_f > 0$$

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

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

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

and applying (29), 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

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

for which it is shown in Lemma 4 that

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

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

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

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

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 (2007), 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

$$\hat{\beta}_i(\hat{\tau}) - \beta_{i0} = \hat{\beta}_i(\hat{\tau}) - \hat{\beta}_i(\tau) + \hat{\beta}_i(\tau) - \beta_{i0} \quad \text{with} \quad \tau = (d_{i0}, \vartheta_{i0}), \tag{30}$$

where

$$\hat{\beta}_i(\hat{\tau}) - \hat{\beta}_i(\tau) = o_p(1)$$

using that $\hat{\tau} - \tau = O_p\left(T^{-1/2}\right)$, which is stronger than the one used in Theorem 1 of Robinson and Hualde (2003) cf. their Lemmas 4 and 5. $\square$

A.2 Proof of Theorem 2

Asymptotic normality of the slope estimates can readily be established based on (30), (27) and (26)

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

where Σ_β is the variance-covariance matrix obtained from (27) 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

$$\sqrt{T}\,\frac{\partial L_{i,T}(\vartheta)}{\partial \vartheta}\bigg|_{\vartheta=\vartheta_{i0}} = \frac{2}{\sqrt{T}}\left\{\mathbf{v}_{2i} - M_{f,0}\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i} + (M_{f,0} - M_{z,0})\,\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\right\}$$

$$\times \left\{(\log\Delta_t)\,\mathbf{v}_{2i} - \dot{M}_{f,0}\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i} + \left(\dot{M}_{f,0} - \dot{M}_{z,0}\right)\Delta_t^{d^*-\vartheta_{i0}}\mathbf{v}_{2i}\right\}'$$

where

$$M_{f,0} := M_f(\vartheta_{i0}) = \mathbf{F}(\vartheta_{i0})\left(\mathbf{F}(d^*)'\mathbf{F}(d^*)\right)^{-1}\mathbf{F}(d^*)',$$

$$M_{z,0} := M_z(\vartheta_{i0}) = \bar{\mathbf{z}}(\vartheta_{i0})(\bar{\mathbf{z}}(d^*)'\bar{\mathbf{z}}(d^*))^{-1}\bar{\mathbf{z}}(d^*)',$$

$$\dot{M}_{f,0} := \dot{M}_f(\vartheta_{i0}) = \dot{\mathbf{F}}(\vartheta_{i0})\left(\mathbf{F}(d^*)'\mathbf{F}(d^*)\right)^{-1}\mathbf{F}(d^*)',$$

$$\dot{M}_{z,0} := \dot{M}_z(\vartheta_{i0}) = \dot{\bar{\mathbf{z}}}(\vartheta_{i0})(\bar{\mathbf{z}}(d^*)'\bar{\mathbf{z}}(d^*))^{-1}\bar{\mathbf{z}}(d^*)',$$

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

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

applying a central limit theorem for martingale difference sequences as shown by Robinson and Velasco (2015).

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

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

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 (2015) 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 The-

orem 2 of Hualde and Robinson (2011), and the proof is then complete. $\square$

A.3   Proof of Theorem 3

The asymptotic behaviour of the mean-group slope estimate is readily shown in Pesaran (2006) 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. $\square$

A.4   Covariance Matrix Estimate $\hat{\mathbf{A}}_i \hat{\mathbf{B}}_i \hat{\mathbf{A}}_i'$

Definitions of the variance-covariance matrix components are comparable to those obtained by Hualde and Robinson (2007). 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 (2007).

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

$$\hat{\omega}_{i\tau_1} = \hat{M}_i^{-1}\left(\hat{m}_{i\tau_1} - \hat{M}_{i\tau_1}^{-1}\hat{\omega}_i\right), \quad \hat{\omega}_{i\tau_2} = \hat{M}_i^{-1}\left(\hat{m}_{i\tau_2} - \hat{M}_{i\tau_2}^{-1}\hat{\omega}_i\right),$$

$$\hat{m}_{i\tau_1} = Q\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{Z}_{it\tau_1}^*(\hat{d}_i)\tilde{y}_{it}^*(\hat{d}_i) + \tilde{Z}_{it}^*(\hat{d}_i,\hat{\vartheta}_i)\tilde{y}_{it\tau_1}^*(\hat{d}_i)\right\},$$

$$\hat{M}_{i\tau_1} = Q\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{Z}_{it\tau_1}^*(\hat{d}_i)\tilde{Z}_{it}^{*\prime}(\hat{d}_i,\hat{\vartheta}_i) + \tilde{Z}_{it}^*(\hat{d}_i,\hat{\vartheta}_i)\tilde{Z}_{it\tau_1}^{*\prime}(\hat{d}_i)\right\}Q',$$

$$\hat{m}_{i\tau_2} = Q\frac{1}{T}\sum_{t=p+1}^{T}\tilde{Z}_{it\tau_2}^*(\hat{\vartheta}_i)\tilde{y}_{it}^*(\hat{d}_i),$$

$$\hat{M}_{i\tau_2} = Q\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{Z}_{it\tau_2}^*(\hat{\vartheta}_i)\tilde{Z}_{it}^{*\prime}(\hat{d}_i,\hat{\vartheta}_i) + \tilde{Z}_{it}^*(\hat{d}_i,\hat{\vartheta}_i)\tilde{Z}_{it\tau_2}^{*\prime}(\hat{\vartheta}_i)\right\}Q'$$

with the parameter subscripts denoting the first partial derivative as in

$$\tilde{y}_{it\tau_1}^*(\hat{d}_i) = (\log\Delta)\tilde{y}_{it}^*(\hat{d}_i),$$

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

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

and also

$$\hat{s}_{i\tau_1\tau_1} = \frac{1}{T}\sum_{t=p+1}^{T}\hat{v}_{it\tau_1}^{*}{}^2, \quad \hat{s}_{i\tau_1\tau_2} = \frac{1}{T}\sum_{t=p+1}^{T}\hat{v}_{it\tau_1}^*\hat{v}_{it\tau_2}^*, \quad \hat{s}_{i\tau_2\tau_2} = \frac{1}{T}\sum_{t=p+1}^{T}\hat{w}_{it\tau_2}^{*\prime}\hat{w}_{it\tau_2}^*,$$

$$\hat{v}_{it\tau_1}^* = \tilde{y}_{it\tau_1}^*(\hat{d}_i) - \hat{\omega}_{i\tau_1}'Q\tilde{Z}_{it}^*(\hat{d}_i,\hat{\vartheta}_i) - \hat{\omega}_i'Q\tilde{Z}_{it\tau_1}^*(\hat{d}_i),$$

$$\hat{v}_{it\tau_2}^* = -\hat{\omega}_{i\tau_2}'Q\tilde{Z}_{it}^*(\hat{d}_i,\hat{\vartheta}_i) - \hat{\omega}_i'Q\tilde{Z}_{it\tau_2}^*(\hat{\vartheta}_i),$$

$$\hat{w}_{it\tau_2}^* = \tilde{x}_{it\tau_2}^*(\hat{\vartheta}_i) - \hat{\phi}_{i\tau_2}'R\tilde{X}_{it}^*(\hat{\vartheta}_i) - \hat{\phi}_i'R\tilde{X}_{it\tau_2}^*(\hat{\vartheta}_i),$$

$$\tilde{x}_{it\tau_2}^*(\hat{\vartheta}_i) = (\log\Delta)\tilde{x}_{it}^*(\hat{\vartheta}_i), \quad \tilde{X}_{it\tau_2}^*(\hat{\vartheta}_i) = (\log\Delta)\tilde{X}_{it}^*(\hat{\vartheta}_i),$$

$$\hat{\phi}_{i\tau_2} = \hat{G}_i^{-1}\left(\hat{g}_{i\tau_2} - \hat{G}_{i\tau_2}\hat{\phi}_i\right),$$

$$\hat{g}_{i\tau_2} = R\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{X}_{it\tau_2}^*(\hat{\vartheta}_i)\tilde{x}_{it}^*(\hat{\vartheta}_i)' + \tilde{X}_{it}^*(\hat{\vartheta}_i)\tilde{x}_{it\tau_2}^*(\hat{\vartheta}_i)'\right\},$$

$$\hat{G}_{i\tau_2} = R\frac{1}{T}\sum_{t=p+1}^{T}\left\{\tilde{X}_{it\tau_2}^*(\hat{\vartheta}_i)\tilde{X}_{it}^*(\hat{\vartheta}_i)' + \tilde{X}_{it}^*(\hat{\vartheta}_i)\tilde{X}_{it\tau_2}^*(\hat{\vartheta}_i)'\right\}R'.$$

Finally,

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

where

$$\hat{v}_{1.2,it}^*(\hat{d}_i, \hat{\vartheta}_i) = \hat{v}_{1it}^*(\hat{d}_i) - \rho_i'\hat{v}_{2it}^*(\hat{\vartheta}_i),$$

$$\hat{v}_{2it}^*(\hat{\vartheta}_i) = \tilde{x}_{it}^*(\hat{\vartheta}_i) - \hat{\phi}_i'R\tilde{X}_{it}^*(\hat{\vartheta}_i).$$

B 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 detailed proof for all three subparts is given in the proof of Theorem 1 by Ergemen and Velasco (2015).

34

Lemma 2. *Under Assumption 1,*

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

Proof of Lemma 2. The result follows from the arguments in the proofs of Theorems 4-6 by Ergemen and Velasco (2015).

Lemma 3. *Under Assumption 1,*

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

Proof of Lemma 3. The result follows from the arguments in the proofs of Theorems 4-6 by Ergemen and Velasco (2015).

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 by Ergemen and Velasco (2015).

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 the proof of Theorem 7 by Ergemen and Velasco (2015).

References

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

BAI, J. AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

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

BLANCHARD, O. J. AND D. QUAH (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79(4), 655–73.

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.

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

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*.

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.

DELONG, J. B. AND L. H. SUMMERS (2012): "Fiscal Policy in a Depressed Economy," *Brookings Papers on Economic Activity*, 233–297.

ELMENDORF, D. W. AND G. N. MANKIW (1999): *Government Debt*, Elsevier, chap. Volume 1, Part C, 1615–1669.

ERGEMEN, Y. E. AND C. VELASCO (2015): "Estimation of Fractionally Integrated Panels with Fixed-Effects and Cross-Section Dependence," *CREATES Working Paper 35*.

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.

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

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.

HUALDE, J. AND P. M. ROBINSON (2007): "Root-N-Consistent Estimation of Weak Fractional Cointegration," *Journal of Econometrics*, 140, 450–484.

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

KAPETANIOS, G., M. H. PESARAN, AND T. YAMAGATA (2011): "Panels with Non-Stationary Multifactor Error Structures," *Journal of Econometrics*, 160(2), 326–348.

MICHELACCI, C. AND P. ZAFFARONI (2000): "(Fractional) Beta Convergence," *Journal of Monetary Economics*, 45, 129–153.

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.

PANIZZA, U. AND A. F. PRESBITERO (2014): "Public Debt and Economic Growth: Is there a Causal Relationship?" *Journal of Macroeconomics*, 41, 21–41.

PERRON, P. AND X. ZHU (2005): "Structural Breaks with Deterministic and Stochastic Trends," *Journal of Econometrics*, 129, 65–119.

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

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

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

PUENTE-AVOJIN, M. AND M. SANSO-NAVARRO (2015): "Granger Causality between Debt and Growth: Evidence from OECD Countries," *International Review of Economics & Finance*, 35, 66–77.

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

ROBINSON, P. M. (1978): "Statistical Inference for a Random Coefficient Autoregressive Model," *Scandinavian Journal of Statistics*, 5, 163–168.

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

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

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

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

Table 1: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0$ and $\rho = 0$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\delta = 0.4:$										
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
$\delta = 1:$										
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: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\delta = 0.4:$										
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
$\delta = 1:$										
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 3: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0$ and $\rho = 0.5$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\delta = 0.4$:										
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
$\delta = 1$:										
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 4: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\delta = 0.4:$										
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
$\delta = 1:$										
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 5: Bias and RMSE Profiles with $n = 20$ and $T = 54$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$)

		$\vartheta = 0.75$			$\vartheta = 1$			$\vartheta = 1.25$		
		$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$	$d = 0.5$	$d = 0.75$	$d = 1$
$\delta = 0.4:$										
Bias	$\hat{\beta}_{MG}(d,\vartheta)$	-0.0150	-0.0168	-0.0133	-0.0122	-0.0210	-0.0196	-0.0101	-0.0275	-0.0399
	$\hat{\beta}_{MG}(\hat{d},\hat{\vartheta})$	-0.0085	-0.0167	-0.0241	-0.0071	-0.0133	-0.0192	-0.0096	-0.0211	-0.0313
	$\hat{\vartheta}$	0.0394	0.0392	0.0366	0.0242	0.0255	0.0260	-0.0001	-0.0000	-0.0000
	$\hat{d}$	0.0003	-0.0164	-0.0366	0.0010	-0.0175	-0.0390	-0.0049	-0.0213	-0.0421
RMSE	$\hat{\beta}_{MG}(d,\vartheta)$	0.0324	0.0347	0.0328	0.0277	0.0357	0.0369	0.0221	0.0373	0.0504
	$\hat{\beta}_{MG}(\hat{d},\hat{\vartheta})$	0.0309	0.0347	0.0385	0.0270	0.0326	0.0367	0.0230	0.0338	0.0440
	$\hat{\vartheta}$	0.0423	0.0421	0.0399	0.0268	0.0280	0.0285	0.0205	0.0207	0.0211
	$\hat{d}$	0.0247	0.0293	0.0441	0.0254	0.0297	0.0462	0.0276	0.0316	0.0487
$\delta = 1:$										
Bias	$\hat{\beta}_{MG}(d,\vartheta)$	-0.0163	-0.0167	-0.0112	-0.0118	-0.0190	-0.0156	-0.0099	-0.0254	-0.0342
	$\hat{\beta}_{MG}(\hat{d},\hat{\vartheta})$	-0.0129	-0.0167	-0.0225	-0.0122	-0.0132	-0.0157	-0.0138	-0.0216	-0.0270
	$\hat{\vartheta}$	0.0480	0.0477	0.0451	0.0252	0.0255	0.0249	-0.0000	-0.0000	-0.0000
	$\hat{d}$	0.0265	-0.0065	-0.0307	0.0226	-0.0092	-0.0343	-0.0140	-0.0135	-0.0373
RMSE	$\hat{\beta}_{MG}(d,\vartheta)$	0.0350	0.0347	0.0313	0.0295	0.0347	0.0341	0.0239	0.0358	0.0451
	$\hat{\beta}_{MG}(\hat{d},\hat{\vartheta})$	0.0339	0.0348	0.0369	0.0308	0.0326	0.0342	0.0269	0.0342	0.0403
	$\hat{\vartheta}$	0.0518	0.0515	0.0493	0.0277	0.0279	0.0275	0.0213	0.0214	0.0217
	$\hat{d}$	0.0405	0.0263	0.0396	0.0368	0.0268	0.0427	0.0328	0.0278	0.0449

Table 6: Bias and RMSE Profiles with $n = 50$ and $T = 100$ ($\theta_1 = \theta_2 = 0.5$ 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	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0094	-0.0112	-0.0094	-0.0060	-0.0127	-0.0128	-0.0036	-0.0150	-0.0259
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0051	-0.0112	-0.0162	-0.0042	-0.0082	-0.0127	-0.0053	-0.0121	-0.0211
	$\hat{\vartheta}$	0.0267	0.0267	0.0254	0.0143	0.0153	0.0165	-0.0000	-0.0000	-0.0000
	$\hat{d}$	0.0030	-0.0072	-0.0187	0.0062	-0.0077	-0.0201	0.0072	-0.0095	-0.0221
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0160	0.0180	0.0167	0.0121	0.0182	0.0193	0.0087	0.0183	0.0295
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0143	0.0180	0.0213	0.0118	0.0158	0.0192	0.0101	0.0165	0.0254
	$\hat{\vartheta}$	0.0277	0.0277	0.0265	0.0151	0.0160	0.0172	0.0110	0.0105	0.0112
	$\hat{d}$	0.0113	0.0127	0.0215	0.0148	0.0128	0.0226	0.0189	0.0140	0.0243
$\underline{\delta = 1:}$										
Bias	$\hat{\beta}_{MG}(d, \vartheta)$	-0.0093	-0.0111	-0.0082	-0.0051	-0.0115	-0.0105	-0.0034	-0.0143	-0.0225
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	-0.0070	-0.0112	-0.0164	-0.0079	-0.0082	-0.0104	-0.0082	-0.0132	-0.0185
	$\hat{\vartheta}$	0.0408	0.0406	0.0391	0.0150	0.0149	0.0148	-0.0000	-0.0000	-0.0000
	$\hat{d}$	0.0247	0.0003	-0.0143	0.0231	-0.0029	-0.0182	-0.0225	-0.0040	-0.0192
RMSE	$\hat{\beta}_{MG}(d, \vartheta)$	0.0170	0.0180	0.0157	0.0129	0.0175	0.0174	0.0195	0.0178	0.0262
	$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$	0.0162	0.0180	0.0213	0.0151	0.0159	0.0174	0.0129	0.0175	0.0230
	$\hat{\vartheta}$	0.0439	0.0437	0.0424	0.0158	0.0157	0.0156	0.0113	0.0111	0.0119
	$\hat{d}$	0.0310	0.0117	0.0180	0.0293	0.0115	0.0210	0.0302	0.0122	0.0219

Table 7: Parametric CSS Estimates of the Integration Orders

	Log(GDP)	Log(Debt)
Australia	1.0101	0.9979
Austria	1.0128	1.0142
Belgium	1.0080	1.0059
Canada	1.0112	1.0034
Denmark	1.0093	1.0076
Finland	1.0111	1.0138
France	1.0098	1.0107
Germany	1.0101	1.0063
Greece	1.0169	1.0206
Ireland	1.0087	1.0106
Italy	1.0105	1.0106
Japan	1.0161	1.0156
Netherlands	1.0090	1.0039
New Zealand	1.0105	1.0026
Norway	1.0114	1.0097
Portugal	1.0130	1.0132
Spain	1.0139	1.0044
Sweden	1.0090	1.0090
UK	1.0049	0.9959
US	1.0058	1.0007

Note: *This table reports the parametric conditional-sum-of-squares (CSS) estimation results of the indicators across countries. Standard error of these estimates is 0.1061.*

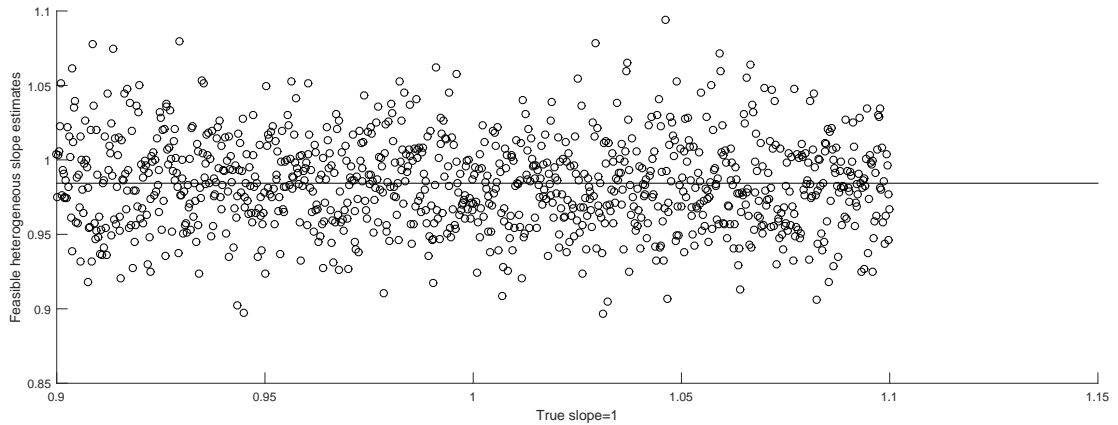Figure 1: Feasible Heterogeneous Slope Estimates, $n = 20$ and $T = 54$.

Table 8: Benchmark Estimation Results for Slope and Memory Parameters based on Hualde and Robinson (2007)

	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 (2007) that disregards individual country characteristics and cross-country dependence. Robust standard errors are reported in parentheses. Bold indicates significance up to the 5% level.*

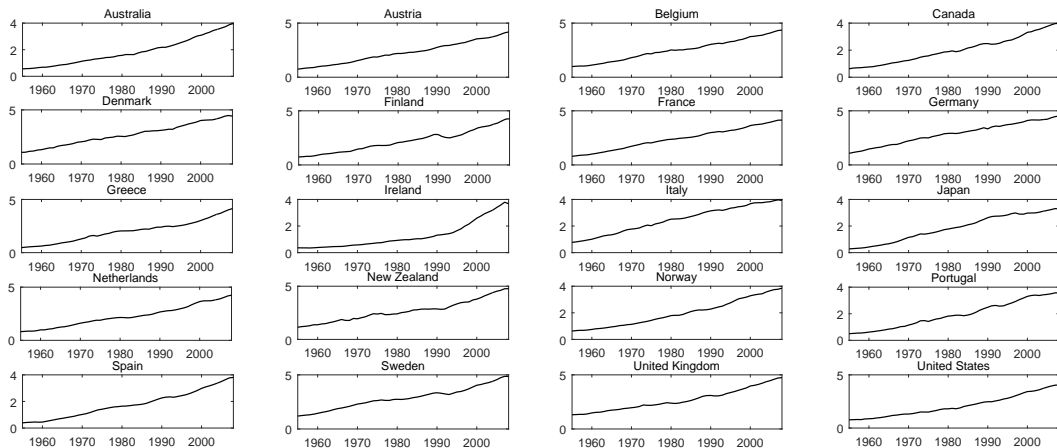Figure 2: Real GDP in Logs, 1955-2008.

Table 9: Estimation Results for Slope and Memory Parameters based on (22)

	Australia	Austria	Belgium	Canada	Denmark	Finland	France
$\hat{\beta}_i$	**-0.0439**	**-0.1072**[†]	**0.0228**	**0.0423**	**-0.0258**[†]	**0.3100**	**0.0155**[†]
s.e.$(\hat{\beta}_i)$	(0.0030)	(0.0040)	(0.0033)	(0.0036)	(0.0048)	(0.0045)	(0.0024)
$\hat{\vartheta}_i$	**1.4160**	**1.0281**	**1.4900**	**1.0992**	**1.4900**	**1.2860**	**1.1394**
s.e.$(\hat{\vartheta}_i)$	(0.2766)	(0.0713)	(0.0388)	(0.2037)	(0.0946)	(0.1026)	(0.1054)
$\hat{d}_i$	**1.1680**	**0.4040**	**1.4999**	**0.7260**	**1.0820**	**1.1830**	**0.9740**
s.e.$(\hat{d}_i)$	(0.0298)	(0.0276)	(0.0212)	(0.0255)	(0.0380)	(0.0360)	(0.0180)

	Italy	Japan	Netherlands	New Zealand	Norway	Portugal	Spain
$\hat{\beta}_i$	**-0.0846**[†]	**0.0388**[†]	**-0.2530**	-0.0158	**-0.1023**	**-0.0800**[†]	**0.0920**[†]
s.e.$(\hat{\beta}_i)$	(0.0039)	(0.0037)	(0.0038)	(0.0099)	(0.0028)	(0.0046)	(0.0041)
$\hat{\vartheta}_i$	**1.1774**	**1.4789**	**1.1380**	**1.2312**	**1.1339**	**1.0690**	**1.2614**
s.e.$(\hat{\vartheta}_i)$	(0.0451)	(0.0519)	(0.1086)	(0.2044)	(0.2047)	(0.1302)	(0.1002)
$\hat{d}_i$	**0.9710**	**0.6110**	**1.2220**	**0.9100**	**1.0300**	**0.8590**	**0.7520**
s.e.$(\hat{d}_i)$	(0.0295)	(0.0258)	(0.0264)	(0.0677)	(0.0241)	(0.0333)	(0.0296)

	Germany	Sweden	Greece	Ireland	UK	US
$\hat{\beta}_i$	**0.2521**[†]	**0.1203**[†]	**-0.0247**[†]	**0.6950**[†]	**-0.2115**	0.0044
s.e.$(\hat{\beta}_i)$	(0.0050)	(0.0032)	(0.0048)	(0.0047)	(0.0042)	(0.0036)
$\hat{\vartheta}_i$	**1.0540**	**1.4900**	**0.9648**	**1.1796**	**1.2656**	**1.4900**
s.e.$(\hat{\vartheta}_i)$	(0.0898)	(0.0748)	(0.0777)	(0.1194)	(0.3260)	(0.0432)
$\hat{d}_i$	**0.2470**	**0.9620**	**0.1800**	**0.3430**	**1.2630**	**1.3210**
s.e.$(\hat{d}_i)$	(0.0345)	(0.0242)	(0.0364)	(0.0337)	(0.0335)	(0.0306)

Note: *This table reports the estimation results of the individual slope and memory parameters across countries. Estimations are performed based on (22) where the projections are carried out with $d^* = 1$. 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 3: Debt in Logs, 1955-2008.

Research Papers
2016

CREATES
Center for Research in Econometric
Analysis of Time Series
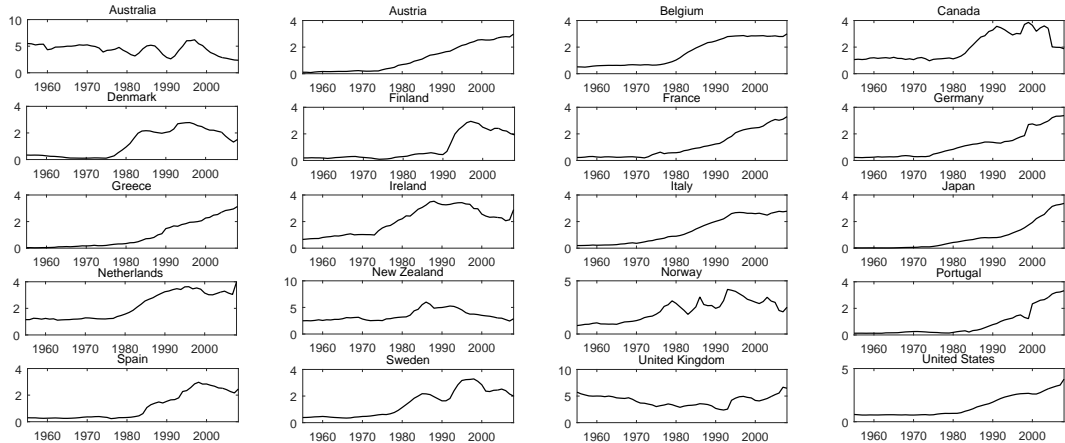
2015-45:	Kim Christensen, Mark Podolskij, Nopporn Thamrongrat and Bezirgen Veliyev: Inference from high-frequency data: A subsampling approach
2015-46:	Asger Lunde, Anne Floor Brix and Wei Wei: A Generalized Schwartz Model for Energy Spot Prices - Estimation using a Particle MCMC Method
2015-47:	Annastiina Silvennoinen and Timo Teräsvirta: Testing constancy of unconditional variance in volatility models by misspecification and specification tests
2015-48:	Harri Pönkä: The Role of Credit in Predicting US Recessions
2015-49:	Palle Sørensen: Credit policies before and during the financial crisis
2015-50:	Shin Kanaya: Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach
2015-51:	Tommaso Proietti: Exponential Smoothing, Long Memory and Volatility Prediction
2015-52:	Mark Podolskij, Christian Schmidt and Mathias Vetter: On U- and V-statistics for discontinuous Itô semimartingale
2015-53:	Mark Podolskij and Nopporn Thamrongrat: A weak limit theorem for numerical approximation of Brownian semi-stationary processes
2015-54:	Peter Christoffersen, Mathieu Fournier, Kris Jacobs and Mehdi Karoui: Option-Based Estimation of the Price of Co-Skewness and Co-Kurtosis Risk
2015-55	Kadir G. Babaglou, Peter Christoffersen, Steven L. Heston and Kris Jacobs: Option Valuation with Volatility Components, Fat Tails, and Nonlinear Pricing Kernels
2015-56:	Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij: Limit theorems for stationary increments Lévy driven moving averages
2015-57:	Andreas Basse-O'Connor and Mark Podolskij: On critical cases in limit theory for stationary increments Lévy driven moving averages
2015-58:	Yunus Emre Ergemen, Niels Haldrup and Carlos Vladimir Rodríguez-Caballero: Common long-range dependence in a panel of hourly Nord Pool electricity prices and loads
2015-59:	Niels Haldrup and J. Eduardo Vera-Valdés: Long Memory, Fractional Integration, and Cross-Sectional Aggregation
2015-60:	Mark Podolskij, Bezirgen Veliyev and Nakahiro Yoshida: Edgeworth expansion for the pre-averaging estimator
2016-01:	Matei Demetrescum, Christoph Hanck and Robinson Kruse: Fixed-b Inference in the Presence of Time-Varying Volatility
2016-02:	Yunus Emre Ergemen: System Estimation of Panel Data Models under Long-Range Dependence