Fractionally Integrated Panel Data Systems∗

Yunus Emre Ergemen†

Department of Economics, Universidad Carlos III de Madrid

September 25, 2014

Abstract

We consider large $n, T$ panel data models with fixed effects, persistent common factors allowing for cross-section dependence, and possibly correlated persistent innovations that are assumed fractionally integrated. In the model, a) persistence in the innovations and the common factor allows for cointegrating relationships in the unobserved idiosyncratic components; b) the vector of innovations, whose elements can be contemporaneously correlated, can also exhibit short-memory dynamics in the form of a finite-order vector autoregressive process; c) deterministic linear and nonlinear trends can be nested, and observable series can co-break through the common factor. We perform estimations on the defactored observed series where the projections are based on the sample averages of fractionally differenced data. In the estimation, we use a computationally convenient equation-by-equation conditional-sum-of-squares (CSS) criterion, leading to GLS-type estimates for slope parameters. The CSS estimates of individual slope and long-range dependence parameters are $\sqrt{T}$ consistent while the mean-group slope estimate is $\sqrt{n}$ consistent, all with centered asymptotic normal distributions, irrespective of cointegration. A study of small-sample performance is carried out showing desirable properties for our estimation method. Finally, an empirical application to the long-run relationship between real GDP growth and debt-to-GDP growth as well as real GDP and debt at levels is included.

JEL Classification: C22, C23

Keywords: Fractional integration, factor models, short memory, system estimation, fixed effects, debt and economic growth.

∗I am greatly indebted to Prof. Dr. Carlos Velasco for his guidance and constant support. I also would like to thank Manuel Arellano, Miguel Delgado, Juan José Dolado, Jesús Gonzalo, Javier Hualde, Bent Nielsen, Peter M. Robinson, Enrique Sentana, Abderrahim Taamouti and the participants in the IV Workshop in Time Series Econometrics, CEMFI PhD Metrics Workshop and UC3M Student Seminars for their helpful comments and discussions.

†Corresponding Address: Department of Economics, Universidad Carlos III de Madrid, Calle Madrid, 126, 28903 Getafe (Madrid), SPAIN. yergemen@eco.uc3m.es. Tel:(+34)691737583
1 Introduction

In economics, fractional long-range dependence can arise due to aggregation. Several aggregate economic variables have been shown to exhibit long-memory properties although their individual components are usually well described by dynamic autoregressive models. Granger (1980) argues that aggregating such models can lead to models that have dramatically different correlation structures for both dependent and independent individual series as is the case when aggregating micro variables such as total personal income, unemployment, consumption of non-durable goods, inventories, and profits. Chambers (1998) shows that U.K. macroeconomic series exhibit fractional long-range dependence when the dynamic models describing the series are temporally aggregated. 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.

In order to get a solid empirical perspective, several indicators are frequently organized in a panel data structure to incorporate the characteristics of different units, such as countries or assets, while describing their time-series dynamics. The examples of macroeconomic panel data indicators include GDP, interest rates, inflation rates, unemployment rates, and in finance, it is standard to use a panel data structure in portfolio performance evaluations and risk management. Analysis of such panel indicators has been carried out using both static and dynamic models. To be more realistic, recent research in panel data theory focuses on developing inference based on stationary \( I(0) \) variables when unobserved heterogeneity and interactions between cross-section units are present; see e.g. Pesaran (2006). The research on nonstationary panel data models, on the other hand, has typically developed in an autoregressive framework with \( I(1) \) variables. As one well-known reference, Phillips and Moon (1999) develop limit theory for heterogeneous panel data models with \( I(1) \) series. Different nonstationary settings have also been considered to account for individual cross-section characteristics and interactions between cross-section units. For example, Bai and Ng (2004) and Bai (2010) propose unit-root testing procedures when idiosyncratic innovations and the common factor are both \( I(1) \), and Moon and Perron (2004) propose the use of dynamic factors to test for unit roots in cross-sectionally dependent panels.

In a pure time-series context, 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). Since empirical studies have repeatedly shown that many economic and financial time series exhibit fractional long-range dependence (possibly due to aggregation) and several macroeconomic and financial indicators are presented in the form of panels, panel data models should also be able to accommodate such behaviour. 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 (2014) 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 (2014) 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 develop inference for panel data models where we allow for fractionally integrated long-range dependence in both idiosyncratic innovations and the common factor while letting the innovation terms be contemporaneously correlated to introduce the possibility of cointegrated system estimation in the classical sense. In these models, a fractional integration or memory parameter describes the persistence, which constitutes an alternative to dynamic autoregressive (AR) panel data models. The model setup requires that both the cross-section size, \( n \), and the time-series length, \( T \), grow in the asymptotics, departing from short panel data models (fixed \( T \)) and multivariate time series models (fixed \( n \)) due to the need to control for cross-section dependence modelled by a common-factor structure. The setup differs from the ones by Hassler et al. (2011) and Robinson and Velasco (2014) as we explicitly model cross-section dependence and allow for cointegrating relationships in the unobserved idiosyncratic 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). The setup also differs from the one by Ergemen and Velasco (2014) due to the allowance for contemporaneous correlation in the idiosyncratic innovations, which enables system estimation on the defactored observed series. Allowing for endogeneity via the idiosyncratic innovations enables the model to achieve wider empirical applicability, especially in cases where endogeneity induced by the unobserved common factor is not the only source of contemporaneous correlation. For example, empirical analyses of endogenous growth theories and the purchasing power parity hypothesis generally require that the idiosyncratic errors be correlated even after the factor structure is removed due to prevailing two-way endogeneity in data.

Our model can successfully address the cases in which a time series approach would lead to invalid results. The observable series can display the same memory level when the integration order of the common factor is greater than those of the idiosyncratic innovations. Thus a pure time-series approach may falsely conclude the relationship between the observables spurious. In this case, possible cointegrating relationships can only be disclosed after the common factor structure is projected out, implying that accounting for individual unit characteristics and cross-section interactions is essential in obtaining valid inference, as is the case under our setup.

In the estimation of the slope and long-range dependence parameters, we use a computationally convenient 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. Although the resulting estimates are not readily as efficient as Gaussian maximum likelihood estimates due to the system information not being exploited at once, they can easily be updated to efficiency by taking a single Newton step from the previously obtained initial estimates. We explore the small-sample performance of our estimates by means of Monte Carlo experiments both when autocorrelations and/or endogeneity are lacking and present, which show desirable results in terms of the performance of our estimates.

In the empirical application, we investigate the long-run relationship between real GDP and debt-to-GDP growth rates as well as debt and real GDP at levels for 20 high-income OECD countries for the time period 1955-2008. We find that GDP growth responds negatively to a change in the debt-to-GDP ratio significantly for 5 out of 20 countries while the relationship is insignificant for other countries at the 5% level. None of these significant relationships are cointegrating, which suggests that there is no long-lasting equilibrium relationship between debt-to-GDP growth and real GDP growth. On the other hand, real GDP and debt at levels have a significant relationship for all countries but Greece, New Zealand and Sweden, and this relationship is cointegrating for several countries, which we can find using our panel approach and not using a pure time series methodology as we show comparing our results to those that would be obtained by Hualde and Robinson (2007).

The remainder of the paper proceeds as follows. Next section contains estimation details of slope and fractional integration parameters. Section 3 lists all the conditions needed 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)” denotes joint asymptotics in which both the cross-section size and time-series length are growing; “→p” denotes convergence in probability; and “→d” denotes convergence in distribution. All mathematical proofs and technical lemmas are collected in an appendix at the end of the paper.

2 Model and Parameter Estimation

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

\[
\begin{align*}
    y_{it} &= \alpha_i + x_{it}\beta_i + f_i \lambda_i + \Delta^{-d_0} \epsilon_{1it}, \\
    x_{it} &= \mu_i + f_i \gamma_i + \Delta^{-\theta_0} \epsilon_{2it},
\end{align*}
\]

where \(y_{it}\) and \(x_{it}\) are scalars whose idiosyncratic innovations have unknown true integration orders \(d_0\) and \(\theta_0\) for \(i = 1, \ldots, n\) and \(t = 1, \ldots, T\), and \(f_i \sim I(\delta)\) where \(\delta\) is also unknown. While vector \(x_{it}\) may also be analyzed allowing for a multiple regression setting, we consider the simplest case to focus on the main ideas. Throughout the paper, the subscript at the fractional differencing operator attached to a vector or scalar \(\epsilon_{it}\)
(i.e. a type-II process) has the meaning

\[ \Delta_t^{-d} \epsilon_{it} = \Delta^{-d} \epsilon_{it} 1(t > 0) = \sum_{j=0}^{t-1} \pi_j(-d) \epsilon_{it-j}, \tag{2} \]

\[ \pi_j(-d) = \frac{\Gamma(j + d)}{\Gamma(j + 1) \Gamma(d)}, \]

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

\[ B(L; \theta_i) \epsilon_{it} \equiv \left( I_2 - \sum_{j=1}^{p} B_j(\theta_i) L^j \right) \epsilon_{it} = v_{it}, \tag{3} \]

where \(L\) is the lag operator, \(\theta_i\) the short-memory parameters, \(I_2\) the \(2 \times 2\) identity matrix, \(B_j\) the \(2 \times 2\) upper-triangular short-memory matrices, and \(v_{it}\) is a bivariate sequence that is identically and independently distributed across \(i\) and \(t\) with zero mean and covariance matrix \(\Omega_i > 0\). The upper-triangularity assumption on the short-memory matrices, \(B_j\), provides a great deal of parsimony in the asymptotics as it further develops the triangular structure of the system, and it is in line with the long-run VAR restriction of Blanchard and Quah (1989) and the short-run VAR restriction of Sims (1987). The arrays \(\{\alpha_i, i \geq 1\}\) and \(\{\mu_i, i \geq 1\}\) are unobserved individual fixed effects; \(\{f_t, t > 0\}\) is the \(I(\delta)\) unobserved common factor that induces cross-section dependence and possibly further endogeneity in the system; \(\{\lambda_i, i \geq 1\}\) and \(\{\gamma_i, i \geq 1\}\) are unobserved factor loadings indicating how much each cross-section unit is affected by \(f_t\). In addition to these general dynamics, autoregressive conditional heteroskedasticity can also be featured in the common factor so that the model can be suitable also for applications in empirical finance such as volatility analysis.

We assume initial conditions in the idiosyncratic innovations away since they are asymptotically negligible under a heterogeneous setup; see Ergemen and Velasco (2014). In a standard way, we first-difference (1) to remove the fixed effects,

\[ \Delta y_{it} = \Delta x_{it} \beta_{i0} + \Delta f_t \lambda_i + \Delta \Gamma^{1-d_{i0}} \epsilon_{1it}, \]
\[ \Delta x_{it} = \Delta f_t \gamma_i + \Delta \Gamma^{1-d_{i0}} \epsilon_{2it}, \tag{4} \]

for \(i = 1, \ldots, n\) and \(t = 2, \ldots, T\). Setting

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

(4) can be prewhitened from idiosyncratic long-range dependence for some fixed exogenous differencing choice, \(d^*\), satisfying \(d^* > 1/4\), using which
all variables become asymptotically stationary with their sample means converging to
population limits.

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

\begin{align*}
y_{it}(d^*) &= x_{it}(d^*) \beta_{i0} + f_1(d^*) \lambda_i + \varepsilon_{1it}(d^* - d_{i0}), \\
x_{it}(d^*) &= f_1(d^*) \gamma_i + \varepsilon_{2it}(d^* - \vartheta_{i0}).
\end{align*}

(5)

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

\begin{align*}
z_{it}(d^*, d^*) &= \zeta x_{it}(d^*) \beta_{i0} + F_t(d^*) L_i + \varepsilon_{it}(d^* - d_{i0}, d^* - \vartheta_{i0}),
\end{align*}

(6)

where $\zeta = (1, 0)'$, $F_t(d^*) = f_1(d^*) \otimes I_2$, and $L_i = (\lambda_i', \gamma_i')'$.

The structure $F_t(d^*) L_i$ in (6) induces cross-section correlation between units $i$ through
$F_t(d^*)$. The common factor may also feature breaks both at levels and in persistence
thus making the model more flexible for empirical applications, e.g. studies containing
global-crisis periods. 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 com-
mon factor structure. A different version of this procedure has been recently adopted in
case of persistent common factors by Ergemen and Velasco (2014). 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 prin-
cipal components approach (PCA) 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 lead-
ing to lower finite-sample performance in estimation as pointed out by Pesaran (2006).
Moreover, PCA estimation of factors with fractional long-range dependence has not
been explored in the literature yet. Bearing in mind this fact, we project out the com-
mon factor structure using the cross-section averages of fractionally differenced data, by
which the projection errors vanish asymptotically.

2.1 Projection of the Common Factor Structure

To give the details about projection, let us write (6) in cross-section averages as

\begin{align*}
\bar{z}_t(d^*, d^*) &= \zeta x_t(d^*) \beta_{i0} + F_t(d^*) \bar{L} + \bar{\varepsilon}_t(d^* - d_{i0}, d^* - \vartheta_{i0}),
\end{align*}

(7)

where $\bar{\varepsilon}_t(d^* - d_{i0}, d^* - \vartheta_{i0})$ is $O_p(n^{-1/2})$ for large enough $d^*$. Thus, $\bar{z}_t(d^*, d^*)$ asymptoti-
cally captures all the information provided by the common factor – since it also contains
$\bar{x}_t(d^*)$. 

Let us write the time-stacked observed series as \( x_{it}(d^*) = (x_{i1}(d^*), \ldots, x_{iT}(d^*))' \) and \( z_{it}(d^*, d^*) = (z_{i1}(d^*, d^*), \ldots, z_{iT}(d^*, d^*))' \) for \( i = 1, \ldots, n \). Then, for each \( i = 1, \ldots, n \),

\[
    z_{it}(d^*, d^*) = x_{it}(d^*)b_{i0}\zeta^t + F(d^*)L_i + E_i(d^* - d_{i0}, d^* - \vartheta_{i0}),
\]

where \( E_i = (d^* - d_{i0}, d^* - \vartheta_{i0}, \ldots, \epsilon_{iT}(d^* - d_{i0}, d^* - \vartheta_{i0}))' \) and \( F(d^*) = (\text{vec}[F_2(d^*)], \ldots, \text{vec}[F_T(d^*)])' \).

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

\[
    M_{T_1}(d^*) = I_{T_1} - \tilde{z}(d^*, d^*)(\tilde{z}'(d^*, d^*)\tilde{z}(d^*, d^*))^{-1}\tilde{z}'(d^*, d^*),
\]

where \( \tilde{z}(d^*, d^*) = n^{-1} \sum_{i=1}^n z_{it}(d^*, d^*) \), and \( P^- \) denotes the generalized inverse of a matrix \( P \). When the projection matrix is built with the original (possibly nonstationary) series, it is not possible to make sure that the cross-section averages can asymptotically replace the factor structure. On the other hand, using some \( d^* > \max\{\vartheta_{max}, d_{max}, \delta\} - 1/4 \) for prewhitening guarantees that the projection errors vanish asymptotically.

Based on (8), the defactored observed bivariate series for each \( i = 1, \ldots, n \),

\[
    \tilde{z}_i(d^*, d^*) = \tilde{x}_i(d^*)b_{i0}\zeta^t + \tilde{E}_i(d^* - d_{i0}, d^* - \vartheta_{i0}),
\]

where \( \tilde{z}_i(d^*, d^*) = M_{T_1}(d^*)z_i(d^*, d^*) \), \( \tilde{x}_i(d^*) = M_{T_1}(d^*)x_i(d^*) \) and \( \tilde{E}_i(d^*) = M_{T_1}(d^*)E_i(d^*) \). The projection error, \( M_{T_1}(d^*)F(d^*) \), is of order \( O_p(n^{-1} + (nT)^{-1/2}) \) and is readily contained in each of the defactored series.

### 2.2 Estimation of Linear Model Parameters

Writing (10) for \( i = 1, \ldots, n \), and \( t = 2, \ldots, T \) as

\[
    \tilde{z}_{it}(d^*, d^*) = \zeta \tilde{x}_{it}(d^*)b_{i0} + \epsilon_{it}(d^* - d_{i0}, d^* - \vartheta_{i0}),
\]

we now integrate the defactored series back by \( d^* \) to their original integration orders, to perform estimations, as

\[
    \tilde{z}_{it}^*(d_i, \vartheta_i) = \zeta \tilde{x}_{it}^*(d_i)b_{i0} + \epsilon_{it}^*(d_i - d_{i0}, \vartheta_i - \vartheta_{i0}),
\]

where the first and second equations of (11) are obtained, respectively, by

\[
    \tilde{y}_{it}^*(d_i) = \Delta^{d_i-d_t} \tilde{y}_{it}(d^*) \quad \text{and} \quad \tilde{x}_{it}^*(\vartheta_i) = \Delta^{\vartheta_i-\vartheta_t} \tilde{x}_{it}(d^*).
\]

To explicitly show the short-memory dynamics in the model based on (3), (11) can be written as

\[
    \tilde{z}_{it}^*(d_i, \vartheta_i) - \sum_{j=1}^p B_j(\vartheta_i)\tilde{z}_{it-j}^*(d_i, \vartheta_i) = \left\{ \zeta \tilde{x}_{it}^*(d_i) - \sum_{j=1}^p \sum_{j=1}^p B_j(\vartheta_i)\zeta \tilde{x}_{it-j}^*(d_i) \right\} \beta_{i0} + \tilde{v}_{it}^*(d_i - d_{i0}, \vartheta_i - \vartheta_{i0}),
\]
whose second equation, noting that
\[ \hat{\beta}_{it}(d_{i}, \vartheta_{i}) = (\hat{y}_{it}(d_{i}), \hat{x}_{it}(\vartheta_{i}))' \]
is
\[ \hat{x}_{it}(\vartheta_{i}) - \sum_{j=1}^{p} B_{2j}(\theta_{i}) \hat{x}_{it-j}(d_{i}, \vartheta_{i}) = \left( - \sum_{j=1}^{p} B_{2j}(\theta_{i}) \hat{x}_{it-j}(d_{i}) \right) \beta_{i0} + \hat{v}_{2it}(\vartheta_{i} - \vartheta_{0}) \]  
(13)
and the first equation can be organized to account for the contemporaneous correlation if we write
\[ \hat{y}_{it}(d_{i}) - \hat{x}_{it}(\vartheta_{i}) \rho_{i} + \sum_{j=1}^{p} (B_{1j}(\theta_{i}) - \rho_{i} B_{2j}(\theta_{i})) \hat{x}_{it-j}(d_{i}, \vartheta_{i}) \]
\[ - \left( \sum_{j=1}^{p} (B_{1j}(\theta_{i}) - \rho_{i} B_{2j}(\theta_{i})) \hat{x}_{it-j}(d_{i}) \right) \beta_{i0} + \hat{v}_{1it}(d_{i} - d_{0}) - \rho_{i} \hat{v}_{2it}(\vartheta_{i} - \vartheta_{0}) \]  
(14)
with \( B_{kj} \) denoting the \( k \)-th row of \( B_{j} \), and \( \rho_{i} = E[\hat{v}_{1it} \hat{v}_{2it}]/E[\hat{v}_{2it}^2] \).

Under (14), cointegration (i.e. \( \vartheta_{0} > d_{0} \)) is useful in the estimation of \( \beta_{i0} \) since the signal that can be extracted from \( \hat{x}_{it}(d_{i}) \) is stronger than that from \( \hat{x}_{it}(\vartheta_{i}) \). However, identification of \( \beta_{i0} \) is still possible in a spurious regression where \( d_{0} > \vartheta_{0} \) since the error term in (14) is orthogonal to \( \hat{v}_{2it}(\cdot) \) given that \( v_{it} \) are identically and independently distributed. The only exclusion we have under a spurious setting is the case in which \( \vartheta_{0} = d_{0} \), which leads to collinearity in (14) thus rendering the identification of \( \beta_{i0} \) impossible. The spurious estimation case in which \( d_{0} > \vartheta_{0} \) is evidently more useful 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 readily leads to the identification of \( \beta_{i0} \) and \( \rho_{i} \) so long as \( \vartheta_{0} \neq d_{0} \), 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 \( B_{kj} \).

In this paper, short-memory dynamics are not our main concern so we treat \( B_{j}(\cdot) \) as nuisance parameters.

The case in which \( p = 0 \) and \( \rho_{i} = 0 \), corresponding to white noise \( \epsilon_{it} \) and no endogeneity respectively, has been developed by Ergemen and Velasco (2014), where estimation is carried out for the parameters only in the first equation and \( \vartheta_{i} \) are treated as nuisance parameters. In the present paper, while the main parameter of interest is still \( \beta_{i0} \), we can also obtain the estimates of \( d_{0}, \vartheta_{0}, \rho_{i} \) and \( B_{j}(\theta_{i}) \).

First, we use a \( q \times (3p+2) \) restriction matrix \( Q \) that is \( I_{3p+2} \) when there are no prior zero restrictions on \( B_{j} \), and a \( q < 3p+2 \) matrix \( Q \) with prior zero restrictions that is obtained by dropping rows corresponding to restrictions, which may improve efficiency by eliminating some lagged values of the series. Then, write (14) as
\[ \hat{y}_{it}(d_{i}) = \omega'(Q \hat{Z}_{it}(d_{i}, \vartheta_{i}) + \hat{v}_{1it}(d_{i} - d_{0}) - \rho_{i} \hat{v}_{2it}(\vartheta_{i} - \vartheta_{0}) \]
(15)
with
\[ \hat{Z}_{it}(d_{i}, \vartheta_{i}) = (\hat{x}_{it}(d_{i}), \hat{x}_{it}(\vartheta_{i}), \hat{u}_{it-1}(d_{i}, \vartheta_{i}), ..., \hat{u}_{it-p}(d_{i}, \vartheta_{i}))' \]
\[ \hat{u}_{it-k}(d_{i}, \vartheta_{i}) = (\hat{x}_{it-k}(d_{i}), \hat{x}_{it-k}(\vartheta_{i}), \hat{y}_{it-k}(d_{i}))' \]
\( \forall k = 1, \ldots, p \).
and \( \omega_i \) being the vector of coefficients that are functions of \( \beta_i, \rho_i \) and \( B_{kj}(\theta_i) \) whose least-squares estimate is given by

\[
\hat{\omega}_i(\tau_1, \tau_2) := M_i(\tau_1, \tau_2)^{-1} m_i(\tau_1, \tau_2) \tag{16}
\]

with

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

where \((\tau_1, \tau_2)\) denotes the infeasible cases of \((d_{i0}, \vartheta_{i0})\), \((\hat{d}_i, \vartheta_{i0})\), \((d_{i0}, \hat{\vartheta}_i)\) and the feasible case of \((\hat{d}_i, \hat{\vartheta}_i)\).

In most empirical work, the main parameter of interest is \( \beta_{i0} \), for which the estimate can simply be obtained from (16) as

\[
\hat{\beta}_i(\tau_1, \tau_2) = \psi' \hat{\omega}_i(\tau_1, \tau_2), \quad \psi = (1, 0, \ldots, 0)'. \tag{17}
\]

While (17) is less efficient than the estimate obtained by joint CSS estimation in the VAR \( \epsilon_{it} \) case, it is computationally much simpler in practice. The case of \( p = 0 \) has been discussed by Ergemen and Velasco (2014), where (17) is efficient when \( \text{Cov}(\epsilon_{1it}, \epsilon_{2it}) = 0 \).

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

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

Short-memory parameters can be estimated similarly taking e.g. \( \psi_\theta = (0, 0, 1, \ldots, 1)' \) when \( B_j(\cdot) \) is known, otherwise \( B_j(\theta_i) \) are estimated.

### 2.3 Estimation of Long-Range Dependence Parameters

For the estimation of long memory or fractional integration parameters, we only consider the empirically relevant case of unknown \( d_i \) and \( \vartheta_i \). Estimation of long-range dependence parameters in the panel data context is a relatively new topic. Robinson and Velasco (2014) 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 (2014) propose two fractional panel data models with cross-section dependence in both of 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 a equation-by-equation CSS approach that is computationally advantageous in comparison to a joint CSS approach, whose parsimony would be more pronounced under a multiple regression setting. To distinguish clearly between these two approaches, we first write the relevant part of the joint CSS criterion function (concentrated out of \( \Omega \)
such a problem, a single Newton step may be taken from an initial estimate of $\tau$ where locating an approximate minimum is computationally demanding. To overcome this problem, a single Newton step may be taken from an initial $\sqrt{T}$-consistent estimate of $\tau_i = (d_i, \vartheta_i, \theta_i)$ as in Robinson and Hualde (2003). Denoting a $\sqrt{T}$-consistent estimate of $\tau_i$ by $\tau^*_i$, consider

$$
\tau_i = \tau^*_i - H_T^{-1}(\tau^*_i)h_T(\tau^*_i),
$$

where

$$
H_T(\tau) = \left( \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial \hat{\nu}^*_u(\tau)}{\partial \tau'} \right)' \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\nu}^*_u(\tau) \hat{\nu}^*_u(\tau)' \right)^{-1} \frac{\partial \hat{\nu}^*_u(\tau)}{\partial \tau'} \right),
$$

and

$$
h_T(\tau) = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial \hat{\nu}^*_u(\tau)}{\partial \tau'} \right)' \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\nu}^*_u(\tau) \hat{\nu}^*_u(\tau)' \right)^{-1} \hat{\nu}^*_u(\tau).
$$

This approach may prove computationally more convenient for obtaining an estimate of $\tau_{i0}$ given its $\sqrt{T}$-consistent initial estimate, though finding such an initial estimate is challenging itself. We therefore follow a step-by-step approach at whose first step we write (13) as

$$
B \hat{x}_{it}(\vartheta_i) = \frac{\partial \hat{x}_{it}(\vartheta_i)}{\partial \tau} R \hat{X}_{ii}(\vartheta_i) = \hat{v}^*_t(\vartheta_i - \vartheta_i)
$$

with

$$
\hat{X}_{ii}(\vartheta_i) = (\hat{x}_{i-1}(\vartheta_i), \ldots, \hat{x}_{i-p}(\vartheta_i))',
$$

the $r \times p$ matrix $R = I_p$ for $r = p$, but for $r < p$, $R$ is obtained by dropping rows from $I_p$, and $\phi_i$ collecting the $B_{23j}$ that are nonzero a priori. Then an estimate of $\phi_i$,

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

(20)
where
\[ G_i(\cdot) = R \frac{1}{T} \sum_{t=p+1}^{T} \tilde{X}_{it}(\cdot)\tilde{X}_{it}'(\cdot)R' \quad \text{and} \quad g_i(\cdot) = R \frac{1}{T} \sum_{t=p+1}^{T} \tilde{X}_{it}(\cdot)\tilde{x}_{it}(\cdot). \]

Having obtained (20), \( \vartheta_{i0} \) can be estimated by
\[ \hat{\vartheta}_i = \arg \min_{\vartheta \in \mathcal{V}} \sum_{t=p+1}^{T} \left\{ \tilde{x}_{it}(\vartheta) - \hat{\phi}_i(\vartheta)'R\tilde{X}_{it}(\vartheta) \right\}^2, \]
with \( \mathcal{V} = [\vartheta, \bar{\vartheta}] \subset (0, \frac{3}{2}). \)

Then \( d_{i0} \) can be estimated from (15) by
\[ \hat{d}_i = \arg \min_{d \in \mathcal{D}} \sum_{t=p+1}^{T} \left\{ \tilde{y}_{it}(d) - \hat{\omega}_i(d, \hat{\vartheta}_i)'Q\tilde{Z}_{it}(d, \hat{\vartheta}_i) \right\}^2, \]
with \( \mathcal{D} = [d, \bar{d}] \subset (0, \frac{3}{2}). \)

2.4 Common Correlated Mean-Group Slope Estimates

In many empirical applications, there is also an interest in obtaining the mean effect of the slope estimates. Given the linearity of the model in \( \beta_i \), we consider the common-correlation mean-group estimate,
\[ \hat{\beta}_{CCMG} (d, \vartheta) := \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i (d, \vartheta). \]  \hspace{1cm} (21)

This estimate is essentially a GLS mean-group estimate based on the average of individual feasible slope estimates. An OLS mean-group estimate can also be considered based on the prewhitening parameter \( d^* \) as
\[ \hat{\beta}_{CCMG} (d^*) := \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i (d^*). \]

While the common correlated pooled estimate for \( \beta_i \) may also be considered, its analysis requires further conditions and is left for future research. For the asymptotic analysis of the mean-group estimate, it is standard to use a random coefficients model as in
\[ \beta_i = \beta_0 + w_i, \quad w_i \sim iid (0, \Omega_w), \]
with \( w_i \) independent of all other model variables.

All asymptotic results along with the regularity conditions used to derive them are presented in the next section.
3 Assumptions and Main Results

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

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

1. The fractional integration parameters, with true values \( \vartheta_{i0} \neq d_{i0} \), satisfy \( \max\{\vartheta_{\max}, d_{\max}, \delta\} - \min\{\vartheta, d\} < 1/2 \), 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, of size at most two, satisfies \( f_t = \alpha_f + \Delta_t^{-\delta} z_f^t \), where \( z_f^t = \sum_{k=0}^{\infty} \Psi_f^k \varepsilon_{f,t-k} \) with \( \sum_{k=0}^{\infty} k \| \Psi_f^k \| < \infty \), \( \det \left( \sum_{k=0}^{\infty} \Psi_f^k s^k \right) \neq 0 \) for \( |s| = 1 \) and \( \varepsilon_f^t \sim iid(0, \Omega_f) \), \( E \left\| \varepsilon_f^t \right\|^4 < \infty \).

3. \( f_t \) and \( \varepsilon_{it} \) are independent, and independent of factor loadings \( \lambda_i \) and \( \gamma_i \) for all \( i \) and \( t \).

4. Factor loadings \( \lambda_i \) and \( \gamma_i \) are independent across \( i \), and the matrix containing sample averages of factor loadings is full column rank.

Assumption 1.1 is a fairly general version of the assumptions used by e.g. [Hualde and Robinson (2011)] and [Nielsen (2013)], additionally ensuring that the projection errors asymptotically vanish with the prescribed choice of \( d^* \). This condition also requires that the lower bounds of the sets \( V \) and \( D \) not be too apart from other memory parameters when \( d_{i0} \in D \) and \( \vartheta_{i0} \in V \), in which case it is further implied that \( \vartheta_{i0} - d_{i0} < 1/2 \), i.e. at most weak fractional cointegration.

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

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. This is straightforwardly satisfied in case of one common factor.

Assumption 2 (System errors). The process \( \varepsilon_{it} \) has the representation

\[
\varepsilon_{it} = \Psi(L; \theta_i) v_{it}
\]
where

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

and the $2 \times 2$ matrices $\Psi_j$ satisfy that

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

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

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

Assumptions 2.1-2.3 are quite standard in the analysis of stationary VAR processes, as were also used by [Robinson and Hualde (2003)](Robinson2003), constituting the counterpart conditions for $B_j$. The first condition rules out possible collinearity in the innovations imposing a standard summability requirement and ensures well-defined functional behaviour at zero frequency, allowing for invertibility. The second condition is needed for the uniform convergence of the Hessian in the asymptotic distribution, and finally the moment requirement in the third condition is in general easily satisfied under Gaussianity. The iid requirement in the last condition may be relaxed to martingale difference innovations whose conditional and unconditional third and fourth order moments are equal, which would then indicate iid behaviour up to fourth moments.

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

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

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

Assumption 3.1 is a regularity condition ensuring the existence of (16) and thus of (17) while Assumption 3.2 is used in the derivation of asymptotic results of (21).

Under our setup, the common-factor structure that accounts for cross-sectional dependence is projected out, and this restricts the analysis to weak cointegration where $d_{i0} - d_{i0} < 1/2$. In a pure time-series context, [Hualde and Robinson (2007)](Hualde2007) derive joint asymptotics for memory and slope parameters without accounting for individual characteristics of the series. Although the results by [Hualde and Robinson (2007)](Hualde2007) are similar to ours, showing our results relies heavily on the projection algebra due to the allowance of cross-section dependence by means of a common-factor structure.

The next theorem presents the consistency of slope and long-range dependence parameter estimates since these are the main parameters of interest in structural estimation.
**Theorem 1.** Under Assumptions 1-3, as \((n,T) \to \infty\),
\[
\begin{aligned}
\left\{ \hat{\beta}_i(\hat{d}_i, \hat{\vartheta}_i) - \beta_{i0} \right. \\
\left. d_i - d_{i0} \right. \\
\left. \hat{\vartheta}_i - \vartheta_{i0} \right. \\
\to \rho 0.
\end{aligned}
\]

This result does not require a rate condition on \(n\) and \(T\) so long as they jointly grow in the asymptotics, and it can be readily extended to include also the other model parameters. This contrasts with the results derived by Robinson and Velasco (2014), 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 in the heterogeneous setup, as is needed here, but projection on cross-section averages for factor structure removal further requires that \(n\) grow.

Next, we show the joint asymptotic distribution of the parameters, where a rate condition is imposed on \(n\) and \(T\) to remove the projection error with an a priori differencing choice of \(d^*\). To simplify the presentation in this paper, we consider a large enough \(d^*\) prescribed in Assumption 1.1 without pointing out a fixed value although for most applications \(d^* = 1\) would be enough anticipating \(\vartheta_{i0}, \delta, d_{i0} < 5/4\).

**Theorem 2.** Under Assumptions 1-3, and if \(\sqrt{T}/n \to 0\) as \((n,T) \to \infty\),
\[
\sqrt{T} \left\{ \hat{\beta}_i(\hat{d}_i, \hat{\vartheta}_i) - \beta_{i0} \right. \\
\left. d_i - d_{i0} \right. \\
\left. \hat{\vartheta}_i - \vartheta_{i0} \right. \\
\to_d N (0, A_i B_i A_i') .
\]

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 \(T/n^{-2} \to 0\) 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 (2014) 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.

The variance-covariance matrix in Theorem 2 has a highly involved analytic expression, but definitions of the estimates \(\hat{A}_i, \hat{B}_i\), thus forming the positive semi-definite covariance matrix estimate \(\hat{A}_i \hat{B}_i \hat{A}_i'\), are provided in Appendix A.5.

The estimates that are outcomes of univariate optimizations in Theorem 2 lack efficiency since the information provided by the panel system is not completely exploited. However, parameter estimates can be updated to efficiency by taking a single Newton-step as described in (19).

We finally consider the asymptotic behaviour of the common correlated mean-group slope estimates.
Theorem 3. Under Assumptions 1-3, as \((n, T) \to \infty\),

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

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 conditions on the relative growth of \(n\) to \(T\). The asymptotic variance-covariance matrix, \(\Omega_w\), can be estimated nonparametrically along the lines of Pesaran (2006) by

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

since variability only depends on the heterogeneity of the \(\beta_i\).

Corollary of Theorem 3. Under Assumptions 1-3, as \((n, T) \to \infty\),

\[
\sqrt{n} \left( \hat{\beta}_{CCMG} \left( d^* \right) - \beta_0 \right) \to_d N \left( 0, \Omega_w \left( d^* \right) \right).
\]

Since this estimate does not include the estimates of the long-range dependence parameters but is based only on the prewhitening parameter, \(d^*\), it can be viewed as an OLS estimator. The asymptotic variance-covariance matrix, \(\Sigma_w\), can be estimated in a similar way to \(\hat{\Omega}_w\) by

\[
\hat{\Omega}_w \left( d^* \right) = \frac{1}{n-1} \sum_{i=1}^{n} \left( \hat{\beta}_i \left( d^* \right) - \hat{\beta}_{CCMG} \left( d^* \right) \right) \left( \hat{\beta}_i \left( d^* \right) - \hat{\beta}_{CCMG} \left( d^* \right) \right)'.
\]

4 Deterministic Trends

While our model in (1) can accommodate both deterministic and stochastic unobserved trends via the common factor \(f_t\), this imposes that the trending behaviour be shared by some cross-section units, in particular by those with nonzero factor loadings. This then indicates that among those cross-section units sharing the same trend, the difference is only up to a constant, based on \(\lambda_i\) and \(\gamma_i\). To relax such a restriction and allow for separate time trends, we extend the model in (1) as

\[
y_{it} = \alpha_i + \alpha_1 q(t) + x_{it} \beta_{i0} + f_t \lambda_i + \Delta_t^{-d_{i0}} \epsilon_{1it}, \\
x_{it} = \mu_i + \mu_1 r(t) + f_t \gamma_i + \Delta_t^{-\theta_{i0}} \epsilon_{2it},
\]

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

The case in which \(q(t)\) and \(r(t)\) in (22) are linear, possibly with drifts, can be straightforwardly analyzed in second differences, at whose first and second differences the time trends are reduced to constants and removed, respectively. Otherwise, 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, asymptotics remain the same under the conditions prescribed in Section 3, although the series may be overdifferenced in the beginning, they are integrated back by the order of their initial differencing orders after projections to their original integration orders, e.g. for double differencing, as in

\[ \Delta_i^{d-2} \Delta^2 y_{it} = \Delta_i^d y_{it} \quad \text{and} \quad \Delta_i^{d-2} \Delta^2 x_{it} = \Delta_i^d x_{it}. \]

In cases of (possibly fractional) nonlinearity in \( q(t) \) and \( r(t) \), such as \( t^2, t^3 \) and \( \Delta^{-\varphi}1 \) with \( \varphi > 1/2 \), removal or estimation of trends may become more complicated as opposed to the linear case. When the orders of trend polynomials are known, the first column of the projection matrix in (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 possible, though some non-parametric GLS detrending approach might be used. This case is beyond the scope of the present paper and is not further explored in what follows.

5 Simulations

In this section, we investigate the finite-sample behaviour of our estimates, \( \hat{\beta}_i(d_0, \vartheta_0) \), \( \hat{d}_i \), \( \hat{\vartheta}_i \) and \( \hat{\beta}_i(d_0, \vartheta_0) \), by means of Monte Carlo experiments. While we estimate the parameters for each \( i \) separately, we report the mean-group estimates to save on space. We draw the mean zero Gaussian idiosyncratic innovations vector \( v_{it} \) with covariance

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

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 \( B_j(\theta_i) = \text{diag} \{ \theta_{1i}, \theta_{2i} \} \) to generate \( \epsilon_{it} \).

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

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 lack of autocorrelation and endogeneity, with the biases further decreasing in \( \vartheta \) although their magnitudes increase in \( \delta \). 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 \( \vartheta \), and the estimate \( \hat{d} \) has a bias generally decreasing in \( d \).

In terms of performance, slope estimates behave well both under lack 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 \( \hat{d} \) show a great deal of robustness.

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. 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 finite-sample behaviour of our estimates under (22) taking \( r(t), q(t) = t \). As before, estimations are performed in first differences, but the projection matrix in (9) is now augmented to include ones in its first column. This way, the estimation method mimics fixed-effects estimation in first differences, and the corresponding bias and RMSE profiles are shown in Tables 5 and 6. The results in Tables 5 and 1 are comparable as are the results in Tables 6 and 2. With the inclusion of linear trends, while both the infeasible and feasible slope estimates have positive and small biases, the fractional integration parameter estimates appear to have been underestimated in general. In fact, considering the fact that first-differencing and projection remove the trend completely, Tables 5 and 6 can be read as the bias and RMSE profiles pertaining to fixed-effects estimation.

Finally, we replicate the results in Table 2 taking \( n = 5 \) and \( T = 25 \) to explore the small-sample behaviour of the estimates, which is reported in Table 7. In terms of performance, the standard errors roughly double while the bias profiles of slope estimates remain more or less the same. However, fractional integration parameter estimates generally suffer from larger biases compared to the results in Table 2.
Table 1: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0$ and $\rho = 0$)

<table>
<thead>
<tr>
<th></th>
<th>$\vartheta = 0.75$</th>
<th></th>
<th>$\vartheta = 1$</th>
<th></th>
<th>$\vartheta = 1.25$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d = 0.5$</td>
<td>$d = 0.75$</td>
<td>$d = 1$</td>
<td>$d = 0.5$</td>
<td>$d = 0.75$</td>
<td>$d = 1$</td>
</tr>
<tr>
<td>$\delta = 0.4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>-0.0015</td>
<td>-0.0016</td>
<td>-0.0015</td>
<td>-0.0007</td>
<td>-0.0011</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>-0.0017</td>
<td>-0.0018</td>
<td>-0.0016</td>
<td>-0.0007</td>
<td>-0.0012</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0194</td>
<td>0.0187</td>
<td>0.0160</td>
<td>-0.0072</td>
<td>-0.0070</td>
</tr>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>0.0052</td>
<td>-0.0092</td>
<td>-0.0201</td>
<td>0.0107</td>
<td>-0.0131</td>
</tr>
<tr>
<td>RMSE</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>0.0497</td>
<td>0.0526</td>
<td>0.0510</td>
<td>0.0421</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>0.0496</td>
<td>0.0526</td>
<td>0.0511</td>
<td>0.0419</td>
<td>0.0493</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0320</td>
<td>0.0316</td>
<td>0.0303</td>
<td>0.0256</td>
<td>0.0255</td>
</tr>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>0.0435</td>
<td>0.0435</td>
<td>0.0466</td>
<td>0.0489</td>
<td>0.0445</td>
</tr>
<tr>
<td>$\delta = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>-0.0018</td>
<td>-0.0018</td>
<td>-0.0016</td>
<td>-0.0015</td>
<td>-0.0016</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>-0.0020</td>
<td>-0.0019</td>
<td>-0.0017</td>
<td>-0.0018</td>
<td>-0.0018</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0526</td>
<td>0.0519</td>
<td>0.0495</td>
<td>-0.0025</td>
<td>-0.0027</td>
</tr>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>0.0704</td>
<td>0.0184</td>
<td>-0.0118</td>
<td>0.0708</td>
<td>0.0133</td>
</tr>
<tr>
<td>RMSE</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>0.0629</td>
<td>0.0547</td>
<td>0.0514</td>
<td>0.0536</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>0.0570</td>
<td>0.0542</td>
<td>0.0515</td>
<td>0.0489</td>
<td>0.0510</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0644</td>
<td>0.0638</td>
<td>0.0620</td>
<td>0.0249</td>
<td>0.0250</td>
</tr>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>0.0906</td>
<td>0.0487</td>
<td>0.0431</td>
<td>0.0921</td>
<td>0.0479</td>
</tr>
</tbody>
</table>
Table 2: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$)

<table>
<thead>
<tr>
<th></th>
<th>$\vartheta = 0.75$</th>
<th></th>
<th>$\vartheta = 1$</th>
<th></th>
<th>$\vartheta = 1.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d = 0.5$</td>
<td>$d = 0.75$</td>
<td>$d = 1$</td>
<td>$d = 0.5$</td>
<td>$d = 0.75$</td>
</tr>
<tr>
<td>$\delta = 0.4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>-0.0150</td>
<td>-0.0171</td>
<td>-0.0132</td>
<td>-0.0122</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>-0.0088</td>
<td>-0.0168</td>
<td>-0.0239</td>
<td>-0.0071</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0368</td>
<td>0.0364</td>
<td>0.0336</td>
<td>0.0234</td>
</tr>
<tr>
<td></td>
<td>$\hat{\hat{d}}$</td>
<td>-0.0016</td>
<td>-0.0189</td>
<td>-0.0407</td>
<td>-0.0009</td>
</tr>
<tr>
<td>RMSE</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>0.0450</td>
<td>0.0486</td>
<td>0.0468</td>
<td>0.0379</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>0.0440</td>
<td>0.0485</td>
<td>0.0513</td>
<td>0.0374</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0423</td>
<td>0.0420</td>
<td>0.0397</td>
<td>0.0290</td>
</tr>
<tr>
<td></td>
<td>$\hat{\hat{d}}$</td>
<td>0.0357</td>
<td>0.0408</td>
<td>0.0551</td>
<td>0.0349</td>
</tr>
<tr>
<td>$\delta = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>-0.0162</td>
<td>-0.0168</td>
<td>-0.0106</td>
<td>-0.0107</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>-0.0138</td>
<td>-0.0166</td>
<td>-0.0215</td>
<td>-0.0122</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0437</td>
<td>0.0432</td>
<td>0.0403</td>
<td>0.0246</td>
</tr>
<tr>
<td></td>
<td>$\hat{\hat{d}}$</td>
<td>0.0277</td>
<td>-0.0072</td>
<td>-0.0336</td>
<td>0.0244</td>
</tr>
<tr>
<td>RMSE</td>
<td>$\hat{\beta}_{MG}(d, \vartheta)$</td>
<td>0.0486</td>
<td>0.0482</td>
<td>0.0449</td>
<td>0.0414</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>0.0473</td>
<td>0.0482</td>
<td>0.0492</td>
<td>0.0417</td>
</tr>
<tr>
<td></td>
<td>$\hat{\vartheta}$</td>
<td>0.0497</td>
<td>0.0493</td>
<td>0.0468</td>
<td>0.0300</td>
</tr>
<tr>
<td></td>
<td>$\hat{\hat{d}}$</td>
<td>0.0493</td>
<td>0.0373</td>
<td>0.0498</td>
<td>0.0465</td>
</tr>
</tbody>
</table>
Table 3: Bias and RMSE Profiles with \( n = 10 \) and \( T = 50 \) (\( \theta_1 = \theta_2 = 0 \) and \( \rho = 0.5 \))

<table>
<thead>
<tr>
<th>( \delta = 0.4 )</th>
<th>( \vartheta = 0.75 )</th>
<th>( \vartheta = 1 )</th>
<th>( \vartheta = 1.25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d = 0.5 )</td>
<td>( d = 0.75 )</td>
<td>( d = 1 )</td>
<td>( d = 0.5 )</td>
</tr>
<tr>
<td>Bias ( \hat{\beta}_{MG}(d, \vartheta) )</td>
<td>-0.0109</td>
<td>-0.0158</td>
<td>-0.0155</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>-0.0115</td>
<td>-0.0155</td>
<td>-0.0200</td>
</tr>
<tr>
<td>( \hat{\vartheta} )</td>
<td>0.0202</td>
<td>0.0197</td>
<td>0.0165</td>
</tr>
<tr>
<td>( \hat{d} )</td>
<td>0.0211</td>
<td>0.0007</td>
<td>-0.0153</td>
</tr>
<tr>
<td>RMSE ( \hat{\beta}_{MG}(d, \vartheta) )</td>
<td>0.0443</td>
<td>0.0477</td>
<td>0.0463</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>0.0449</td>
<td>0.0477</td>
<td>0.0485</td>
</tr>
<tr>
<td>( \hat{\vartheta} )</td>
<td>0.0334</td>
<td>0.0332</td>
<td>0.0317</td>
</tr>
<tr>
<td>( \hat{d} )</td>
<td>0.0432</td>
<td>0.0369</td>
<td>0.0400</td>
</tr>
<tr>
<td>( \delta = 1 )</td>
<td>( \vartheta = 0.75 )</td>
<td>( \vartheta = 1 )</td>
<td>( \vartheta = 1.25 )</td>
</tr>
<tr>
<td>( d = 0.5 )</td>
<td>( d = 0.75 )</td>
<td>( d = 1 )</td>
<td>( d = 0.5 )</td>
</tr>
<tr>
<td>Bias ( \hat{\beta}_{MG}(d, \vartheta) )</td>
<td>-0.0230</td>
<td>-0.0276</td>
<td>-0.0215</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>-0.0261</td>
<td>-0.0247</td>
<td>-0.0274</td>
</tr>
<tr>
<td>( \hat{\vartheta} )</td>
<td>0.0540</td>
<td>0.0534</td>
<td>0.0505</td>
</tr>
<tr>
<td>( \hat{d} )</td>
<td>0.0917</td>
<td>0.0352</td>
<td>0.0014</td>
</tr>
<tr>
<td>RMSE ( \hat{\beta}_{MG}(d, \vartheta) )</td>
<td>0.0664</td>
<td>0.0567</td>
<td>0.0494</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>0.0593</td>
<td>0.0539</td>
<td>0.0526</td>
</tr>
<tr>
<td>( \hat{\vartheta} )</td>
<td>0.0654</td>
<td>0.0649</td>
<td>0.0627</td>
</tr>
<tr>
<td>( \hat{d} )</td>
<td>0.1048</td>
<td>0.0538</td>
<td>0.0369</td>
</tr>
</tbody>
</table>
Table 4: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0$)

<table>
<thead>
<tr>
<th>$\vartheta$</th>
<th>$d$</th>
<th>Bias $\hat{\beta}_{MG}(d, \vartheta)$</th>
<th>$\hat{\beta}_{MG}(d, \hat{\vartheta})$</th>
<th>RMSE $\hat{\beta}_{MG}(d, \vartheta)$</th>
<th>$\hat{\beta}_{MG}(d, \hat{\vartheta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.75$</td>
<td>$0.5$</td>
<td>-0.0008</td>
<td>-0.0017</td>
<td>0.0586</td>
<td>0.0660</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>-0.0006</td>
<td>-0.0018</td>
<td>0.0612</td>
<td>0.0702</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>0.0347</td>
<td>0.0345</td>
<td>0.0403</td>
<td>0.0402</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.0487</td>
<td>-0.0585</td>
<td>0.0659</td>
<td>0.0730</td>
</tr>
<tr>
<td>$1$</td>
<td>$0.5$</td>
<td>-0.0010</td>
<td>-0.0018</td>
<td>0.0657</td>
<td>0.0677</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>-0.0009</td>
<td>-0.0018</td>
<td>0.0667</td>
<td>0.0714</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>0.0420</td>
<td>0.0416</td>
<td>0.0479</td>
<td>0.0476</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.0208</td>
<td>-0.0496</td>
<td>0.0523</td>
<td>0.0656</td>
</tr>
<tr>
<td>$1.25$</td>
<td>$0.5$</td>
<td>-0.0010</td>
<td>-0.0018</td>
<td>0.0657</td>
<td>0.0677</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>-0.0009</td>
<td>-0.0018</td>
<td>0.0667</td>
<td>0.0714</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>0.0420</td>
<td>0.0416</td>
<td>0.0479</td>
<td>0.0476</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.0208</td>
<td>-0.0496</td>
<td>0.0523</td>
<td>0.0656</td>
</tr>
</tbody>
</table>
Table 5: Bias and RMSE Profiles with \( n = 10 \) and \( T = 50 \) (\( \theta_1 = \theta_2 = 0 \) and \( \rho = 0 \) with linear trends)

<table>
<thead>
<tr>
<th>( \delta = 0.4 )</th>
<th>( \dot{\vartheta} = 0.75 )</th>
<th>( \dot{\vartheta} = 1 )</th>
<th>( \dot{\vartheta} = 1.25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( d = 0.5 )</td>
<td>( d = 0.75 )</td>
<td>( d = 1 )</td>
</tr>
<tr>
<td>Bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(d, \dot{\vartheta}) )</td>
<td>0.0011</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>0.0011</td>
<td>0.0013</td>
<td>0.0014</td>
</tr>
<tr>
<td>( \hat{\dot{\vartheta}} )</td>
<td>0.0078</td>
<td>0.0068</td>
<td>0.0041</td>
</tr>
<tr>
<td>( \hat{\hat{d}} )</td>
<td>-0.0136</td>
<td>-0.0573</td>
<td>-0.0804</td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(d, \dot{\vartheta}) )</td>
<td>0.0507</td>
<td>0.0511</td>
<td>0.0490</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>0.0510</td>
<td>0.0516</td>
<td>0.0495</td>
</tr>
<tr>
<td>( \hat{\dot{\vartheta}} )</td>
<td>0.0311</td>
<td>0.0310</td>
<td>0.0309</td>
</tr>
<tr>
<td>( \hat{\hat{d}} )</td>
<td>0.0447</td>
<td>0.0728</td>
<td>0.0931</td>
</tr>
<tr>
<td>( \delta = 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(d, \dot{\vartheta}) )</td>
<td>0.0002</td>
<td>0.0009</td>
<td>0.0012</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>0.0003</td>
<td>0.0009</td>
<td>0.0012</td>
</tr>
<tr>
<td>( \hat{\dot{\vartheta}} )</td>
<td>0.0217</td>
<td>0.0208</td>
<td>0.0184</td>
</tr>
<tr>
<td>( \hat{\hat{d}} )</td>
<td>0.0281</td>
<td>-0.0350</td>
<td>-0.0708</td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(d, \dot{\vartheta}) )</td>
<td>0.0563</td>
<td>0.0522</td>
<td>0.0489</td>
</tr>
<tr>
<td>( \hat{\beta}_{MG}(\hat{d}, \hat{\vartheta}) )</td>
<td>0.0553</td>
<td>0.0528</td>
<td>0.0495</td>
</tr>
<tr>
<td>( \hat{\dot{\vartheta}} )</td>
<td>0.0389</td>
<td>0.0387</td>
<td>0.0381</td>
</tr>
<tr>
<td>( \hat{\hat{d}} )</td>
<td>0.0582</td>
<td>0.0591</td>
<td>0.0853</td>
</tr>
</tbody>
</table>
Table 6: Bias and RMSE Profiles with $n = 10$ and $T = 50$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$ with linear trends)

<table>
<thead>
<tr>
<th>$\hat{\vartheta}$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d = 0.5$</td>
<td>$d = 0.75$</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(d, \hat{\vartheta})$</td>
<td>-0.0146</td>
<td>-0.0173</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>-0.0067</td>
<td>-0.0171</td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0121</td>
<td>0.0116</td>
</tr>
<tr>
<td>$\hat{\hat{d}}$</td>
<td>-0.0343</td>
<td>-0.0709</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(d, \hat{\vartheta})$</td>
<td>0.0474</td>
<td>0.0493</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>0.0462</td>
<td>0.0495</td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0257</td>
<td>0.0256</td>
</tr>
<tr>
<td>$\hat{\hat{d}}$</td>
<td>0.0504</td>
<td>0.0814</td>
</tr>
<tr>
<td>$\delta = 1$</td>
<td>$\hat{\beta}_{MG}(d, \hat{\vartheta})$</td>
<td>-0.0147</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>-0.0098</td>
<td>-0.0166</td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0145</td>
<td>0.0138</td>
</tr>
<tr>
<td>$\hat{\hat{d}}$</td>
<td>-0.0175</td>
<td>-0.0618</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(d, \hat{\vartheta})$</td>
<td>0.0480</td>
<td>0.0481</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \hat{\vartheta})$</td>
<td>0.0471</td>
<td>0.0487</td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0274</td>
<td>0.0273</td>
</tr>
<tr>
<td>$\hat{\hat{d}}$</td>
<td>0.0433</td>
<td>0.0739</td>
</tr>
</tbody>
</table>
Table 7: Bias and RMSE Profiles with $n = 5$ and $T = 25$ ($\theta_1 = \theta_2 = 0.5$ and $\rho = 0.5$)

<table>
<thead>
<tr>
<th>$\vartheta$</th>
<th>$d$</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{MG}(d, \dot{\vartheta})$</td>
<td>Bias</td>
<td>-0.0149</td>
<td>-0.0175</td>
<td>-0.0112</td>
<td>-0.0155</td>
<td>-0.0262</td>
<td>-0.0200</td>
<td>-0.0192</td>
<td>-0.0429</td>
<td>-0.0514</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \dot{\vartheta})$</td>
<td>-0.0082</td>
<td>-0.0169</td>
<td>-0.0271</td>
<td>-0.0079</td>
<td>-0.0152</td>
<td>-0.0191</td>
<td>-0.0141</td>
<td>-0.0303</td>
<td>-0.0366</td>
<td></td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0405</td>
<td>0.0400</td>
<td>0.0361</td>
<td>0.0280</td>
<td>0.0296</td>
<td>0.0286</td>
<td>0.0031</td>
<td>-0.0028</td>
<td>-0.0029</td>
<td></td>
</tr>
<tr>
<td>$\hat{\dot{d}}$</td>
<td>-0.0133</td>
<td>-0.0442</td>
<td>-0.0841</td>
<td>-0.0149</td>
<td>-0.0445</td>
<td>-0.0871</td>
<td>-0.0290</td>
<td>-0.0496</td>
<td>-0.0899</td>
<td></td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0405</td>
<td>0.0400</td>
<td>0.0361</td>
<td>0.0280</td>
<td>0.0296</td>
<td>0.0286</td>
<td>0.0031</td>
<td>-0.0028</td>
<td>-0.0029</td>
<td></td>
</tr>
<tr>
<td>$\hat{\dot{d}}$</td>
<td>-0.0133</td>
<td>-0.0442</td>
<td>-0.0841</td>
<td>-0.0149</td>
<td>-0.0445</td>
<td>-0.0871</td>
<td>-0.0290</td>
<td>-0.0496</td>
<td>-0.0899</td>
<td></td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0405</td>
<td>0.0400</td>
<td>0.0361</td>
<td>0.0280</td>
<td>0.0296</td>
<td>0.0286</td>
<td>0.0031</td>
<td>-0.0028</td>
<td>-0.0029</td>
<td></td>
</tr>
<tr>
<td>$\hat{\dot{d}}$</td>
<td>-0.0133</td>
<td>-0.0442</td>
<td>-0.0841</td>
<td>-0.0149</td>
<td>-0.0445</td>
<td>-0.0871</td>
<td>-0.0290</td>
<td>-0.0496</td>
<td>-0.0899</td>
<td></td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0405</td>
<td>0.0400</td>
<td>0.0361</td>
<td>0.0280</td>
<td>0.0296</td>
<td>0.0286</td>
<td>0.0031</td>
<td>-0.0028</td>
<td>-0.0029</td>
<td></td>
</tr>
<tr>
<td>$\hat{\dot{d}}$</td>
<td>-0.0133</td>
<td>-0.0442</td>
<td>-0.0841</td>
<td>-0.0149</td>
<td>-0.0445</td>
<td>-0.0871</td>
<td>-0.0290</td>
<td>-0.0496</td>
<td>-0.0899</td>
<td></td>
</tr>
</tbody>
</table>

$\delta = 1$:

<table>
<thead>
<tr>
<th>$\vartheta$</th>
<th>$d$</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{MG}(d, \dot{\vartheta})$</td>
<td>Bias</td>
<td>-0.0149</td>
<td>-0.0163</td>
<td>-0.0081</td>
<td>-0.0127</td>
<td>-0.0222</td>
<td>-0.0148</td>
<td>-0.0167</td>
<td>-0.0381</td>
<td>-0.0444</td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \dot{\vartheta})$</td>
<td>-0.0119</td>
<td>-0.0164</td>
<td>-0.0246</td>
<td>-0.0105</td>
<td>-0.0132</td>
<td>-0.0148</td>
<td>-0.0167</td>
<td>-0.0285</td>
<td>-0.0314</td>
<td></td>
</tr>
<tr>
<td>$\dot{\vartheta}$</td>
<td>0.0452</td>
<td>0.0448</td>
<td>0.0414</td>
<td>0.0298</td>
<td>0.0306</td>
<td>0.0290</td>
<td>-0.0029</td>
<td>-0.0028</td>
<td>-0.0031</td>
<td></td>
</tr>
<tr>
<td>$\dot{\hat{d}}$</td>
<td>0.0798</td>
<td>0.0919</td>
<td>0.1173</td>
<td>0.0776</td>
<td>0.0903</td>
<td>0.1193</td>
<td>0.0803</td>
<td>0.0901</td>
<td>0.1203</td>
<td></td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0641</td>
<td>0.0643</td>
<td>0.0628</td>
<td>0.0503</td>
<td>0.0511</td>
<td>0.0513</td>
<td>0.0213</td>
<td>0.0211</td>
<td>0.0219</td>
<td></td>
</tr>
<tr>
<td>$\dot{d}$</td>
<td>0.0831</td>
<td>0.0887</td>
<td>0.1145</td>
<td>0.0805</td>
<td>0.0882</td>
<td>0.1171</td>
<td>0.0791</td>
<td>0.0873</td>
<td>0.1171</td>
<td></td>
</tr>
</tbody>
</table>

RMSE $\hat{\beta}_{MG}(d, \dot{\vartheta})$

<table>
<thead>
<tr>
<th>$\vartheta$</th>
<th>$d$</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{MG}(d, \dot{\vartheta})$</td>
<td>0.0951</td>
<td>0.107</td>
<td>0.0994</td>
<td>0.0851</td>
<td>0.0979</td>
<td>0.1033</td>
<td>0.0745</td>
<td>0.0948</td>
<td>0.1097</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}_{MG}(\hat{d}, \dot{\vartheta})$</td>
<td>0.0973</td>
<td>0.1023</td>
<td>0.1047</td>
<td>0.0879</td>
<td>0.0985</td>
<td>0.1047</td>
<td>0.0771</td>
<td>0.0931</td>
<td>0.1051</td>
<td></td>
</tr>
<tr>
<td>$\dot{\vartheta}$</td>
<td>0.0604</td>
<td>0.0604</td>
<td>0.0584</td>
<td>0.0492</td>
<td>0.0504</td>
<td>0.0506</td>
<td>0.0210</td>
<td>0.0206</td>
<td>0.0212</td>
<td></td>
</tr>
<tr>
<td>$\dot{\hat{d}}$</td>
<td>0.0798</td>
<td>0.0919</td>
<td>0.1173</td>
<td>0.0776</td>
<td>0.0903</td>
<td>0.1193</td>
<td>0.0803</td>
<td>0.0901</td>
<td>0.1203</td>
<td></td>
</tr>
<tr>
<td>$\hat{\vartheta}$</td>
<td>0.0641</td>
<td>0.0643</td>
<td>0.0628</td>
<td>0.0503</td>
<td>0.0511</td>
<td>0.0513</td>
<td>0.0213</td>
<td>0.0211</td>
<td>0.0219</td>
<td></td>
</tr>
<tr>
<td>$\dot{d}$</td>
<td>0.0831</td>
<td>0.0887</td>
<td>0.1145</td>
<td>0.0805</td>
<td>0.0882</td>
<td>0.1171</td>
<td>0.0791</td>
<td>0.0873</td>
<td>0.1171</td>
<td></td>
</tr>
</tbody>
</table>

RMSE $\hat{\beta}_{MG}(\hat{d}, \dot{\vartheta})$


6 Fractional Panel Analysis of the Economic Growth and Debt Relationship

6.1 Related Literature

The relationship between debt and economic growth has been extensively analyzed based on several different approaches leading to mixed results. Among many others, Elmendorf and Mankiw (1999) argue for the negative effect of public debt on growth. Reinhart and Rogoff (2010) use a debt-bracketing approach coupled with threshold estimation to conclude that high debt hinders economic growth in developed countries. Baglan and Yoldas (2013) show that nonlinearities caused by a common debt-level threshold is insignificant and suggest grouping the countries according to their debt-to-GDP ratios to conclude a common negative relationship between GDP growth and debt for countries with chronically high debt. In line with these findings, Chudik et al. (2013) show that debt has a negative and significant effect on growth in the long-run and that debt-level thresholds have no significant effects thus refuting the nonlinearity arguments based on thresholding in debt dynamics. Contrary to these views, DeLong and Summers (2012) argue that hysteresis arising from recessions can lead to a situation in which expansionary fiscal policies may have positive effect on long-run GDP growth.

Overall the existing literature has provided ambiguous conclusions as to whether the relationship between debt and GDP growth is negative or positive due to large differences in their estimation methodologies. Except for the econometric specification by Chudik et al. (2013), which constitutes the AR alternative of ours, all others rely on homogeneous slope estimation methods, completely disregarding country characteristics and possible interactions between countries. Such homogeneity assumption on the slope parameter implies that different countries converge to their equilibrium at the same rate and that there is no debt overhang from one country to another, which is implausible given the increasing interdependencies between economies. Although Chudik et al. (2013) can address these issues in their cross-sectionally augmented autoregressive distributed lag estimation strategy, they restrict their analyses to $I(0)$ and $I(1)$ cases. Just like in the other references, their decision on the stationarity of the dynamics of debt-to-GDP ratio and GDP growth is merely based on unit-root testing. However, as is well known by now, rejecting the null of a unit root does not imply $I(0)$ stationarity in the series; see Dolado et al. (2002), which is why allowing for non-integer orders of integration in the analysis is required.

We analyze the relationship between real GDP and debt-to-GDP growth rates and the relationship between real GDP and debt at levels separately in the following subsections.
6.2 Empirical Analysis of the GDP Growth and Debt-to-GDP Ratio Relationship

We examine the long-term effects of debt on economic growth using our fractionally integrated panel data estimation methodology. Using our approach, we incorporate country-specific characteristics and the interactions between countries while also allowing for endogeneity without having to restrict our analysis to $I(0)$ and $I(1)$ cases, by which we are able to detect stationarity and nonstationarity of fractional orders that will lead to more accurate inference.

In our 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. Real GDP growth rates and debt-to-GDP ratios for each country are plotted in Figures 1 and 2, respectively.

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

---

1 We use data on central government debt since this is the only available data for all the high-income OECD countries that we consider.
First, we estimate the fractional integration orders of real GDP growth rates and debt-to-GDP ratios using local Whittle estimation based on Robinson (1995) with bandwidth choices of $m = 10, 14$. Given that the sample contains 54 time-series data points, choosing larger bandwidths will lead to higher-frequency and short-memory contamination in the estimates. The estimation results are collected in Table 8.

The results in Table 8 suggest that real GDP growth rates may in fact be integrated of fractional orders and even be mildly nonstationary\footnote{Chudik et al. (2013) also point out that growth rates may be mildly nonstationary and use this information to select sufficiently many lags in their ARDL specification.} although they are always considered to be $I(0)$ variables in the literature. While the null of $I(0)$ stationarity in GDP growth rates cannot be rejected for several countries given the standard errors of their memory estimates, there are also other countries in our sample whose growth rates are significantly fractionally integrated of different orders, thus justifying a heterogeneous approach.

The integration order estimates of debt-to-GDP ratios presented in Table 8 are all significant and around unity, indicating high persistence of varying orders. For several countries, debt-to-GDP ratios are integrated of an order larger than one and for others less than one although in the literature they are immediately considered to be $I(1)$ variables due to the unrealistic homogeneity assumption in inter-country dynamics. Chudik et al. (2013) use debt-to-GDP growth rates in their analysis, for which we present the integration orders also in Table 8. These fractional integration or memory estimates suggest that debt-to-GDP growth can still be persistent for some countries with varying magnitudes.

We also estimate the fractional integration order of the common factor based on the cross-section averages of the series, which proxy the factor structure well as is evident from \footnote{Chudik et al. (2013) also point out that growth rates may be mildly nonstationary and use this information to select sufficiently many lags in their ARDL specification.}. The common factor is integrated of orders 0.7577 and 0.7067 for $m = 10, 14$,
respectively, providing evidence that the cross-section dependence is persistent itself, which has not been considered in this literature so far.


<table>
<thead>
<tr>
<th>Real GDP Growth</th>
<th>Debt-to-GDP Ratio</th>
<th>Debt-to-GDP Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 10$</td>
<td>$m = 14$</td>
</tr>
<tr>
<td>Australia</td>
<td>0.4020</td>
<td>0.1109</td>
</tr>
<tr>
<td>Austria</td>
<td>0.5601</td>
<td>0.3823</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.5381</td>
<td>0.3680</td>
</tr>
<tr>
<td>Canada</td>
<td>0.1561</td>
<td>0.1935</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.2710</td>
<td>0.2308</td>
</tr>
<tr>
<td>Finland</td>
<td>0.1762</td>
<td>0.1521</td>
</tr>
<tr>
<td>France</td>
<td>0.5129</td>
<td>0.4893</td>
</tr>
<tr>
<td>Germany</td>
<td>0.7708</td>
<td>0.3244</td>
</tr>
<tr>
<td>Greece</td>
<td>0.4891</td>
<td>0.4299</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.4383</td>
<td>0.4777</td>
</tr>
<tr>
<td>Italy</td>
<td>0.3190</td>
<td>0.4618</td>
</tr>
<tr>
<td>Japan</td>
<td>0.8071</td>
<td>0.6454</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.5373</td>
<td>0.2805</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.1095</td>
<td>0.1641</td>
</tr>
<tr>
<td>Norway</td>
<td>0.2428</td>
<td>0.1299</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.3924</td>
<td>0.3498</td>
</tr>
<tr>
<td>Spain</td>
<td>0.3323</td>
<td>0.4371</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.5035</td>
<td>0.3662</td>
</tr>
<tr>
<td>UK</td>
<td>-0.2749</td>
<td>-0.1820</td>
</tr>
<tr>
<td>US</td>
<td>-0.2500</td>
<td>-0.1440</td>
</tr>
<tr>
<td>s.e.</td>
<td>(0.1581)</td>
<td>(0.1336)</td>
</tr>
</tbody>
</table>

Having obtained the integration order estimates for GDP growth and debt-to-GDP ratio as well as debt-to-GDP ratio growth, we now only focus on the long-run relationship between real GDP growth rates and debt-to-GDP growth rates, as is the case in Chudik et al. (2013), for two reasons: 1. Regressing GDP growth, which is stationary for most countries, on debt-to-GDP ratio, which is highly nonstationary, is completely uninformative whereas a regression based on the change in the debt-to-GDP ratio, which is almost as stationary as GDP growth, can prove insightful; 2. Interpretation of the results is more useful since our primary focus is on determining how economic growth responds to a change in the debt-to-GDP ratio.

We therefore estimate [1] taking $y_{it}$ as the real GDP growth and $x_{it}$ as the debt-to-GDP ratio growth of country $i$, based on our methodology in which we account for country-specific characteristics, such as institutions and geographical location, as well as characteristics that are common for all countries – OECD membership, high income, etc. Our estimation methodology also allows for the two-way endogeneity between the debt-to-GDP ratio and real GDP growth since the idiosyncratic innovations are allowed to be correlated in the model, which is a necessary trait in this analysis as discussed...
by Baglan and Yoldas (2013) and Chudik et al. (2013). The estimation results, taking $d^* = 1.25$ and assuming a VAR(1) structure in the idiosyncratic innovations, are reported in Table 9. For all countries, slope coefficient estimates are negative but the negative relationship between GDP growth and debt-to-GDP ratio growth is significant only for Belgium, Finland, New Zealand, Germany, Ireland and the United States. Considering the fact that our estimation methodology takes into account the fractional integration orders of the series, which in general leads to larger but more accurate standard errors for slope estimates, we do not find a significant relationship between real GDP growth and debt-to-GDP growth for the rest of the countries. The VAR(1) structure we pre-imposed is rejected for most of the countries, which is why the reported standard errors of long-range dependence parameter estimates are all the same.

Our methodology also allows for the identification of nontrivial cointegrating relationships, in particular if the slope estimate is significant and $\hat{d}_i > d_i$. However, we find that there is no statistically significant evidence for a cointegrating relationship between economic growth and debt growth for any of the countries. This leads to the conclusion that there is no long-lasting equilibrium relationship between GDP growth and debt growth while there is a parallel behaviour between them for several time periods, e.g. as reported by Chudik et al. (2013) where they perform their analysis taking 3 lags. Along with most of the claims in the literature, this could be due to the net direction of the causality between these variables being undetermined in the longer run: while high debt burden may have an adverse impact on economic growth, low GDP growth (by reducing tax revenues and increasing public expenditures) could also lead to high debt-to-GDP ratios.

We finally consider the common correlated mean-group slope estimate to investigate whether on average there is a negative and significant relationship between real GDP growth and debt-to-GDP ratio growth for the 20 high-income OECD countries that we consider. The mean-group slope estimate is -0.1912 with a standard deviation of 0.1448, which we calculate nonparametrically based on the variance-covariance matrix defined in Theorem 3, resulting in a p-value of 0.1924. Therefore, the average impact of an increase in the debt-to-GDP ratio growth on real GDP growth is not statistically significant at the 5% level for these 20 high-income OECD countries.

To sum up, using our methodology that takes into account country fixed effects, interdependencies between countries and noninteger integration orders, we find that for most countries in our sample, debt-to-GDP ratio growth rates do not have a statistically significant impact on real GDP growth rates although for Belgium, Finland, New Zealand, Germany, Ireland and the United States, this impact is significant at the 1% level. The mean-group estimate is insignificant indicating that on average there is no significant relationship between the real GDP and debt-to-GDP growth rates.
Table 9: Estimation Results for the Slope and Long-Range Parameters

<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>Austria</th>
<th>Belgium</th>
<th>Canada</th>
<th>Denmark</th>
<th>Finland</th>
<th>France</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_i$</td>
<td>-0.2132</td>
<td>-0.1122</td>
<td>-0.6226</td>
<td>-0.0999</td>
<td>-0.0209</td>
<td>-0.5008</td>
<td>-0.0189</td>
</tr>
<tr>
<td>s.e.($\hat{\beta}_i$)</td>
<td>(0.1217)</td>
<td>(0.1028)</td>
<td>(0.1159)</td>
<td>(0.1087)</td>
<td>(0.1060)</td>
<td>(0.0851)</td>
<td>(0.0608)</td>
</tr>
<tr>
<td>$\hat{\vartheta}_i$</td>
<td>0.6590</td>
<td>0.6310</td>
<td>0.6807</td>
<td>0.4485</td>
<td>0.6333</td>
<td>0.4936</td>
<td>0.3166</td>
</tr>
<tr>
<td>$\hat{d}_i$</td>
<td>0.0680</td>
<td>0.8910</td>
<td>0.7840</td>
<td>0.7420</td>
<td>0.9140</td>
<td>0.5220</td>
<td>0.7780</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Italy</th>
<th>Japan</th>
<th>Netherlands</th>
<th>New Zealand</th>
<th>Norway</th>
<th>Portugal</th>
<th>Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_i$</td>
<td>-0.1033</td>
<td>-0.0903</td>
<td>-0.0966</td>
<td>-0.3645</td>
<td>-0.0275</td>
<td>-0.0094</td>
<td>-0.1538</td>
</tr>
<tr>
<td>s.e.($\hat{\beta}_i$)</td>
<td>(0.1008)</td>
<td>(0.0711)</td>
<td>(0.0960)</td>
<td>(0.1305)</td>
<td>(0.1192)</td>
<td>(0.0977)</td>
<td>(0.1036)</td>
</tr>
<tr>
<td>$\hat{\vartheta}_i$</td>
<td>0.6628</td>
<td>0.8257</td>
<td>0.8856</td>
<td>0.7009</td>
<td>0.6508</td>
<td>0.5088</td>
<td>0.4492</td>
</tr>
<tr>
<td>$\hat{d}_i$</td>
<td>0.2420</td>
<td>0.6170</td>
<td>0.4790</td>
<td>0.4250</td>
<td>0.7240</td>
<td>0.4310</td>
<td>0.8170</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Germany</th>
<th>Sweden</th>
<th>Greece</th>
<th>Ireland</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_i$</td>
<td>-0.3185</td>
<td>-0.0699</td>
<td>-0.0194</td>
<td>-0.5229</td>
<td>-0.0446</td>
<td>-0.7529</td>
</tr>
<tr>
<td>s.e.($\hat{\beta}_i$)</td>
<td>(0.0909)</td>
<td>(0.0527)</td>
<td>(0.1193)</td>
<td>(0.0962)</td>
<td>(0.1003)</td>
<td>(0.1090)</td>
</tr>
<tr>
<td>$\hat{\vartheta}_i$</td>
<td>0.5828</td>
<td>0.7782</td>
<td>0.5790</td>
<td>1.0122</td>
<td>0.7174</td>
<td>0.7290</td>
</tr>
<tr>
<td>$\hat{d}_i$</td>
<td>0.7700</td>
<td>0.0001</td>
<td>0.7690</td>
<td>0.8910</td>
<td>0.8080</td>
<td>0.8010</td>
</tr>
</tbody>
</table>

Note: This table reports the estimation results of the individual slope and memory parameters across countries. Estimations are performed based on (7) where the projections are carried out with $d^* = 1.25$. Robust standard errors are reported in parentheses. Standard error of the memory estimates is 0.1071.
6.3 Empirical Analysis of the Relationship between GDP and Debt at Levels

In structural estimation, using comparable level data, such as GDP and consumption, leads to easy-to-interpret results. With this in mind, we repeat the analysis in the previous subsection using real GDP and debt at levels, whose characteristics we expect to be similar, so that we can identify possible long-run relationships. Using the same datasets as in the previous subsection, we construct the debt data based on the PPP-based GDP data, which we used earlier, for the time period 1955-2008. This way, we can guarantee that the results have clear interpretations.

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

For both series, there is a clear trending behaviour, leading us to think that they are both nonstationary series. To verify this, we carry out local Whittle estimations with \( m = 10, 14 \) as we did in the previous section. The results are collected in Table 10.

The estimation results show that real GDP and debt at levels are integrated of an order around unity, which is in line with the literature where they are treated as \( I(1) \) variables. The common factor of real GDP and debt is estimated based on the cross-section averages of the series and is integrated of orders 0.9193 and 0.9496 for \( m = 10, 14 \), respectively, indicating that removing the common factor is essential for disclosing possible cointegrating relationships. To verify this statement, we provide benchmark estimation results based on the pure time-series estimation approach by Hualde and Robinson (2007) assuming a VAR(1) structure. Nontrivial cointegrating relationships will exist if a) the slope coefficients are significantly different from zero;
Figure 4: Debt at Levels, 1955-2008.


<table>
<thead>
<tr>
<th>Real GDP (Level)</th>
<th>Debt (Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 10$</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>$m = 14$</td>
</tr>
<tr>
<td>Australia</td>
<td>0.9728</td>
</tr>
<tr>
<td>Austria</td>
<td>0.9482</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.9678</td>
</tr>
<tr>
<td>Canada</td>
<td>0.9763</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.9464</td>
</tr>
<tr>
<td>Finland</td>
<td>0.9172</td>
</tr>
<tr>
<td>France</td>
<td>0.9765</td>
</tr>
<tr>
<td>Germany</td>
<td>0.9363</td>
</tr>
<tr>
<td>Greece</td>
<td>0.9361</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.9639</td>
</tr>
<tr>
<td>Italy</td>
<td>0.9883</td>
</tr>
<tr>
<td>Japan</td>
<td>1.0156</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.9533</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.9250</td>
</tr>
<tr>
<td>Norway</td>
<td>1.0039</td>
</tr>
<tr>
<td>Portugal</td>
<td>1.0306</td>
</tr>
<tr>
<td>Spain</td>
<td>0.9676</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.9039</td>
</tr>
<tr>
<td>UK</td>
<td>0.9653</td>
</tr>
<tr>
<td>US</td>
<td>0.9925</td>
</tr>
</tbody>
</table>

s.e. (0.1581) (0.1336) (0.1581) (0.1336)
b) the estimated integration orders of covariates are larger than those of the estimation residuals, i.e. $\hat{\nu}_i > \hat{d}_i$. These benchmark estimation results are collected in Table 11.

Table 11: Benchmark Estimation Results for the Slope and Long-Range Parameters based on Hualde and Robinson (2007)

<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>Austria</th>
<th>Belgium</th>
<th>Canada</th>
<th>Denmark</th>
<th>Finland</th>
<th>France</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_i$</td>
<td>0.0024</td>
<td>-0.0112</td>
<td>-0.1831</td>
<td>-0.0027</td>
<td>-0.1458</td>
<td>-0.2002</td>
<td>-0.1010</td>
</tr>
<tr>
<td>s.e.$(\hat{\beta}_i)$</td>
<td>(0.0054)</td>
<td>(0.0052)</td>
<td>(0.0065)</td>
<td>(0.0055)</td>
<td>(0.0090)</td>
<td>(0.0090)</td>
<td>(0.0054)</td>
</tr>
<tr>
<td>$\hat{\nu}_i$</td>
<td>1.2499</td>
<td>1.2500</td>
<td>1.2499</td>
<td>1.1611</td>
<td>1.2500</td>
<td>1.2500</td>
<td>1.2500</td>
</tr>
<tr>
<td>$\hat{d}_i$</td>
<td>1.4900</td>
<td>1.4890</td>
<td>1.4600</td>
<td>1.5000</td>
<td>1.4040</td>
<td>1.4800</td>
<td>1.5000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Italy</th>
<th>Japan</th>
<th>Netherlands</th>
<th>New Zealand</th>
<th>Norway</th>
<th>Portugal</th>
<th>Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_i$</td>
<td>0.0094</td>
<td>-0.0630</td>
<td>-0.0223</td>
<td>0.0067</td>
<td>-0.0032</td>
<td>0.0370</td>
<td>-0.0584</td>
</tr>
<tr>
<td>s.e.$(\hat{\beta}_i)$</td>
<td>(0.0056)</td>
<td>(0.0052)</td>
<td>(0.0070)</td>
<td>(0.0102)</td>
<td>(0.0057)</td>
<td>(0.0074)</td>
<td>(0.0068)</td>
</tr>
<tr>
<td>$\hat{\nu}_i$</td>
<td>1.2499</td>
<td>1.2499</td>
<td>1.2499</td>
<td>1.2500</td>
<td>1.0698</td>
<td>1.1626</td>
<td>1.2499</td>
</tr>
<tr>
<td>$\hat{d}_i$</td>
<td>1.4520</td>
<td>1.4990</td>
<td>1.5000</td>
<td>1.3760</td>
<td>1.0698</td>
<td>1.1626</td>
<td>1.2499</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Germany</th>
<th>Sweden</th>
<th>Greece</th>
<th>Ireland</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_i$</td>
<td>-0.0382</td>
<td>-0.0422</td>
<td>0.0143</td>
<td>-0.1917</td>
<td>0.0502</td>
<td>-0.0718</td>
</tr>
<tr>
<td>s.e.$(\hat{\beta}_i)$</td>
<td>(0.0085)</td>
<td>(0.0069)</td>
<td>(0.0063)</td>
<td>(0.0138)</td>
<td>(0.0194)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>$\hat{\nu}_i$</td>
<td>1.2499</td>
<td>1.2500</td>
<td>1.2369</td>
<td>1.2500</td>
<td>1.2499</td>
<td>1.2507</td>
</tr>
<tr>
<td>$\hat{d}_i$</td>
<td>1.3370</td>
<td>1.4833</td>
<td>1.4982</td>
<td>1.4998</td>
<td>1.3670</td>
<td>1.5101</td>
</tr>
</tbody>
</table>

**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. Standard error of the memory estimates is 0.1071.

According to the results in Table 11, real GDP and debt at levels do not have a cointegrating relationship for any of the countries although there are several significant slope estimates. Now, using our model allowing for linear time trends, we check the long-run relationship between real GDP and debt assuming a VAR(1) structure. The cointegration estimation results are reported in Table 12.

According to the estimation results in Table 12, we find that debt and real GDP have a significant relationship at the 5% level for all countries except Greece, New Zealand, and Sweden. The significant effect of debt on GDP is positive for Belgium, Italy, Netherlands and United Kingdom, and it is negative and significant for the remaining countries. While a negative and significant effect of debt on real GDP is generally reported in the literature, a positive effect can be, for example, due to the debt increasing because of government spending since it also fuels real GDP; also see DeLong and Summers (2012).

The relationship between real GDP and debt does not have a cointegration nature for Belgium, Ireland, Japan and Portugal, which suggests that the significant interplay between the variables has a short-term focus. On the other hand, we find a cointegrating
relationship between real GDP and debt for Australia, Austria, Canada, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Spain, United Kingdom and United States. The pre-imposed VAR(1) structure is rejected for most of the countries, which is why the standard errors of long-range dependence parameter estimates are the same.

Finally, we check the average effect of debt on GDP in these 20 OECD countries. The common correlated mean-group estimate is -0.0344 with a standard deviation of 0.1740 suggesting that on average there is no significant effect of debt on GDP. This suggests that the widely used assumption in the literature of all countries converging to their equilibrium at the same rate does not hold since there are significant results for individual countries.

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 could be both positive and negative across countries. This contrasts with the relationship between real GDP growth and debt-to-GDP ratio growth where we do not find any significant cointegrating relationship. We also show by our mean-group estimate that on average there is no significant effect of debt on real GDP, which is also the case for real GDP growth and debt-to-GDP ratio growth. That is to say, unless heterogeneity and interdependencies across countries are taken into account, as in a pure time-series
estimation, identifying the true nature of the relationship 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 cointegration 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:

1. The parametric factor structure inducing cross-section dependence in our model may be assumed to have been approximated by infinitely many 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, $f_t$, we do not analyze spatial dependence explicitly. Parametric modelling of spatial dependence, see e.g. Pesaran and Tosetti (2011), may provide further insights.

2. The VAR dynamics in (1) may be extended to more general VARMA dynamics possibly involving nonfundamentalness arguments, with which there could be some insight gains.

3. A multiple regression framework can be considered through the allowance of vector $x_{it}$ whose elements display different degrees of persistence. While the extension is trivial when the entire vector displays the same persistence characteristics, the treatment of unit-varying persistence is likely to complicate the uniformity arguments shown in this paper. This extension, however, may allow for the identification of multiple cointegrating relationships.

4. The fractionally integrated unobserved factor structure may be estimated up to a rotation and those estimates may be used as plug-in estimates in drawing inference on other model parameters, thus allowing the model to be used in forecasting studies. PCA estimation of fractionally integrated factor models are yet to be explored in the literature.

5. Testing procedures either on the form of the panel, e.g. linearity, or on model parameters, e.g. breaks and stability, can be proposed.
A Technical Appendix

A.1 Projection of the Factor Structure

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

\[
x'_i(d^*) \bar{M}_{T_1}(d^*) F(d^*) = x'_i(d^*) I_{T_1} F(d^*) - x'(d^*) \bar{z}(d^*) (\bar{z}'(d^*) \bar{z}(d^*))^{-\frac{1}{2}} \bar{z}'(d^*) F(d^*), \tag{23}
\]

where

\[
\bar{z}(d^*) = F(d^*) \bar{C} + \bar{E} (d^* - d_0, d^* - \vartheta_0) \tag{24}
\]

with

\[
\bar{C} = \begin{pmatrix}
\tilde{\gamma} \beta + \tilde{\lambda} & 0 \\
0 & 0 \\
0 & 0 \\
0 & \tilde{\gamma}
\end{pmatrix}
\]

and

\[
\bar{E} (d^* - d_0, d^* - \vartheta_0) = \bar{e} (d^* - d_0, d^* - \vartheta_0) + \bar{e}_2 (d^* - \vartheta_0) \tilde{\beta} \tilde{\varsigma}'.
\]

Suppressing the terms \( \bar{E} (d^* - d_0, d^* - \vartheta_0) \equiv \bar{E} \), (23) can be opened up as

\[
T_{1}^{-1} x'_i(d^*) \bar{z}(d^*) = T_{1}^{-1} x'_i(d^*) F(d^*) \bar{C} + T_{1}^{-1} x'_i(d^*) \bar{E}
\]

\[
T_{1}^{-1} \bar{z}'(d^*) \bar{z}(d^*) = T_{1}^{-1} \bar{C}' F'(d^*) F(d^*) \bar{C} + T_{1}^{-1} \bar{C}' F'(d^*) \bar{E} + T_{1}^{-1} \bar{E}' F(d^*) \bar{C} + T_{1}^{-1} \bar{E}' \bar{E}
\]

\[
T_{1}^{-1} \bar{z}'(d^*) F(d^*) = T_{1}^{-1} \bar{C}' F'(d^*) F(d^*) + T_{1}^{-1} \bar{E}' F(d^*).
\]

By Assumption 2,

\[
\bar{e}_t = \Psi(L; \theta) \bar{v}_t
\]

with \( \sum_{j=1}^{\infty} j \| \Psi_j(\theta) \| < K \), where \( K \) is a positive constant. Thus, projections based on \( \bar{v}_t \) and \( \bar{e}_t \) incur errors of the same size.

Then, by Lemma [1], the projection error is of size

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

Denote the projection matrix containing the factors \( M_F \). This result implies that

\[
x'_i(\cdot) \bar{M}_{T_1}(\cdot) F(\cdot) = x'_i(\cdot) M_F F(\cdot) + O_p \left( \frac{1}{n} + \frac{1}{\sqrt{nT}} \right), \tag{25}
\]

indicating that \( M_F \) can replace \( \bar{M}_{T_1} \) as \( n \to \infty \), which may be useful for the asymptotic covariance matrix derivations. Furthermore,

\[
T_{1}^{1/2} x'_i(\cdot) \bar{M}_{T_1}(\cdot) F(\cdot) = T_{1}^{1/2} x'_i(\cdot) M_F F(\cdot) + O_p \left( \frac{\sqrt{T}}{n} \right).
\]
A.2 Proof of Theorem 1

We begin by showing the consistency of \( \hat{\beta}_i(d, \vartheta) \), taking for simplicity \( p = 0 \) together with \( d = d_0 \) and \( \vartheta = \vartheta_0 \). Then in [14], denoting \( \sum_t = \sum_{t=1}^T \),

\[
\hat{\beta}_i(d, \vartheta) = \frac{\sum_t \bar{x}_{it}^2(\vartheta) \sum_t \bar{x}_{it}^*(d) \tilde{y}_u(d) - \sum_t \bar{x}_{it}^*(d) \bar{x}_{it}^*(\vartheta) \sum_t \bar{x}_{it}^*(\vartheta) \tilde{y}_u(d)}{\sum_t \bar{x}_{it}^2(d) \sum_t \bar{x}_{it}^2(\vartheta) - (\sum_t \bar{x}_{it}^*(d) \bar{x}_{it}^*(\vartheta))^2},
\]

from which we can write

\[
\hat{\beta}_i(d, \vartheta) - \beta_{i0} = \frac{\sum_t \bar{x}_{it}^2(\vartheta) \sum_t \bar{x}_{it}^*(d) \tilde{v}_{1,2it} - \sum_t \bar{x}_{it}^*(d) \bar{x}_{it}^*(\vartheta) \sum_t \bar{x}_{it}^*(\vartheta) \tilde{v}_{1,2it}}{\sum_t \bar{x}_{it}^2(d) \sum_t \bar{x}_{it}^2(\vartheta) - (\sum_t \bar{x}_{it}^*(d) \bar{x}_{it}^*(\vartheta))^2},
\]

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

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

We next show the consistency of \( \hat{\beta}_i \) taking \( p = 0 \), which is because the proof follows the exact same steps for other \( p \) values, and the projection approach is better motivated in this less complicated setting. Write the time-stacked CSS as

\[
L_{i,T}(\vartheta) = \frac{1}{T} \tilde{x}_i^*(\vartheta) \tilde{x}_i^{**}(\vartheta),
\]

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

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

where

\[
\Delta^{d^*-1} \Delta \tilde{x}_i = \Delta^{d^*-1} \Delta x_i - \zeta_s \tilde{z}(d^*)
\]

\[
= \Delta^{d^*-1} \Delta x_i - \Delta^{d^*-1} \Delta x_i \tilde{z}(d^*) (\tilde{z}(d^*) \tilde{z}'(d^*))^{-1} \tilde{z}(d^*)
\]

so that

\[
\Delta^{\vartheta-d^*} \Delta^{d^*-1} \Delta \tilde{x}_i = \Delta^{\vartheta-1} \Delta x_i - \zeta_s \tilde{z}(\vartheta).
\]

Next, to be able to make use of [25], let us write

\[
\Delta^{\vartheta-1} \Delta \tilde{x}_i = I_x + J_x
\]

with

\[
I_x = \Delta^{\vartheta-d^*} \mathbf{v}_{2i} - \Delta^{d^*-\vartheta} \mathbf{v}_{2i} \mathbf{F}'(d^*) \mathbf{(F}(d^*) \mathbf{F}'(d^*))^{-1} \mathbf{F(\vartheta)},
\]

\[
J_x = \Delta^{d^*-\vartheta} \mathbf{v}_{2i} \left\{ \mathbf{F}'(d^*) \mathbf{(F}(d^*) \mathbf{F}'(d^*))^{-1} \mathbf{F(\vartheta)} - \tilde{z}'(d^*) (\tilde{z}(d^*) \tilde{z}'(d^*))^{-1} \tilde{z}(\vartheta) \right\}
\]

37
where \( \mathbf{F}(d^*) = (f_2(d^*), \ldots, f_T(d^*))' \). Then using the notation
\[
M_f := M_f(\vartheta) = \mathbf{F}'(d^*) (\mathbf{F}(d^*)\mathbf{F}'(d^*))^{-1} \mathbf{F}(\vartheta),
\]
\[
M_z := M_z(\vartheta) = \mathbf{z}'(d^*) (\mathbf{z}(d^*)\mathbf{z}'(d^*))^{-1} \mathbf{z}(\vartheta),
\]
we can write (26) as
\[
\frac{1}{T} \left\{ \Delta^{d-\vartheta_{i0}} v_{2i} - \Delta^{d-\vartheta_{i0}} v_{2i} M_f + \Delta^{d-\vartheta_{i0}} v_{2i} (M_f - M_z) \right\}
\times \left\{ \Delta^{d-\vartheta_{i0}} v_{2i} - \Delta^{d-\vartheta_{i0}} v_{2i} M_f + \Delta^{d-\vartheta_{i0}} v_{2i} (M_f - M_z) \right\}',
\]
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,
\[
\frac{1}{T} \Delta^{d-\vartheta_{i0}} v_{2i} \Delta^{d-\vartheta_{i0}} v_{2i}',
\]
converges uniformly in \( \vartheta \) and is minimized for \( \vartheta = \vartheta_{i0} \) as in the proof of Theorem 3.3 of Robinson and Velasco (2014) and Theorem 1 of Ergemen and Velasco (2014). To show that the second squared term is negligible, write
\[
\frac{1}{T} \Delta^{d-\vartheta_{i0}} v_{2i} M_f M_f' \Delta^{d-\vartheta_{i0}} v_{2i}
\]
where
\[
M_f M_f' = \mathbf{F}'(d^*) (\mathbf{F}(d^*)\mathbf{F}'(d^*))^{-1} \mathbf{F}(\vartheta) \mathbf{F}(\vartheta)' (\mathbf{F}(d^*)\mathbf{F}'(d^*))^{-1} \mathbf{F}(d^*)
\]
satisfying
\[
\frac{\mathbf{F}(d^*)\mathbf{F}'(d^*)}{T} \rightarrow_p \Sigma_f > 0
\]
\[
\sup_{\vartheta \in \mathcal{V}} \left| \frac{\mathbf{F}(\vartheta)\mathbf{F}(\vartheta)'}{T} \right| = O_p \left( 1 + T^{2(\delta - \bar{\vartheta})-1} \right) = O_p(1)
\]
because \( \delta - \bar{\vartheta} < 1/2 \). Now since
\[
\frac{\Delta^{d-\vartheta_{i0}} v_{2i} \mathbf{F}'(d^*)}{T} = O_p \left( T^{-1/2} + T^{\vartheta_{i0}+\delta - 2d^*-1} \right) = o_p(1),
\]
we have that
\[
\sup_{\vartheta \in \mathcal{V}} \left| \frac{1}{T} \Delta^{d-\vartheta_{i0}} v_{2i} M_f M_f' \Delta^{d-\vartheta_{i0}} v_{2i}' \right| = o_p(1).
\]
The third squared term
\[
\sup_{\vartheta \in \mathcal{V}} \left| \frac{1}{T} \Delta^{d-\vartheta_{i0}} v_{2i} (M_f - M_z) (M_f - M_z)' \Delta^{d-\vartheta_{i0}} v_{2i}' \right| = o_p(1)
\]
because
\[ F(d^*)M_z M_z' F'(d^*) = F(d^*)Z'(d^*) (Z(d^*)Z'(d^*))^{-1} Z(v) Z'(v) (Z(d^*)Z'(d^*))^{-1} Z(v) F'(d^*) \]
for which it can easily be shown using (24) and the projection details above that
\[
\sup_{\vartheta \in V} \left| F(d^*)M_z M_z' F'(d^*) \right| = O_p \left( \frac{1}{n} + \frac{1}{\sqrt{nT}} + \frac{T^{2(\vartheta_{\text{max}} - \vartheta) - 1}}{\sqrt{n}} + \frac{T^{\vartheta_{\text{max}} + \delta - 2\vartheta - 1}}{\sqrt{n}} \right) = o_p(1)
\]
since \( \vartheta_{\text{max}} - \vartheta < 1/2 \) and \( \delta - \vartheta < 1/2 \). The proof of consistency for \( \hat{\vartheta}_i \) is then complete.

The consistency of \( \hat{d}_i \) in the time-stacked CSS
\[
\hat{d}_i = \arg \min_{d \in D} \frac{1}{T} \left( \hat{y}_i^*(d) - \hat{\omega}_i(d, \hat{\vartheta}_i) Q \tilde{Z}_i^*(d, \hat{\vartheta}_i) \right) \left( \hat{y}_i^*(d) - \hat{\omega}_i(d, \hat{\vartheta}_i) Q \tilde{Z}_i^*(d, \hat{\vartheta}_i) \right)'
\]
with \( D = [d, \bar{d}] \subset (0, \frac{3}{2}) \) can be shown using exactly the same line of reasoning as above additionally incorporating the estimation effects of \( \hat{\omega}_i \) that are \( O_p(T^{-1/2}) \), and thus the proof is omitted.

Finally, establishing
\[
\hat{\beta}_i(\hat{d}_i, \hat{\vartheta}_i) - \beta_{i0} = O_p \left( \frac{1}{\sqrt{T}} + \frac{1}{n} + \frac{1}{\sqrt{nT}} \right) = o_p(1)
\]
follows from the Mean Value Theorem writing
\[
\hat{\beta}_i(\hat{\tau}) - \beta_{i0} = \hat{\beta}_i(\hat{\tau}) - \hat{\beta}_i(\tau) + \hat{\beta}_i(\tau) - \beta_{i0} \quad \text{with} \quad \tau = (d_{i0}, \vartheta_{i0}),
\]
where
\[
\hat{\beta}_i(\hat{\tau}) - \hat{\beta}_i(\tau) = \hat{\beta}_i(\tau^t) (\hat{\tau} - \tau)
\]
with \( \left| \hat{\beta}_i(\tau^t) \right| < K \) for some intermediate-value vector \( \tau^t \), and using that \( \hat{\tau} - \tau = O_p \left( T^{-1/2} \right) \). □

**A.3 Proof of Theorem 2**

Asymptotic normality of the slope estimates can readily be established based on the arguments above since
\[
\sqrt{T} \left( \hat{\beta}_i(\hat{d}_i, \hat{\vartheta}_i) - \beta_{i0} \right) = O_p(1) + O_p \left( \frac{\sqrt{T}}{n} \right)
\]
where the second \( O_p \) term appears due to the projection, and is removed if \( \sqrt{T}/n \to 0 \) as \( n \to \infty \).
Showing the asymptotic normality of $\hat{\theta}_i$ and $\hat{d}_i$ follows the same steps, which is why we only prove the result for $\hat{\theta}_i$ to focus on the main ideas. The $\sqrt{T}$-normalized score evaluated the true value, $\vartheta_0$, is given by

$$\sqrt{T} \frac{\partial L_{i,T}(\vartheta)}{\partial \vartheta} \bigg|_{\vartheta = \vartheta_0} = \frac{2}{\sqrt{T}} \left\{ \mathbf{v}_{2i} - \Delta_t^{d^* - \vartheta_0} \mathbf{v}_{2i} M_{f,0} + \Delta_t^{d^* - \vartheta_0} \mathbf{v}_{2i} (M_{f,0} - M_{z,0}) \right\} \times \left\{ (\log \Delta_t) \mathbf{v}_{2i} - \Delta_t^{d^* - \vartheta_0} \mathbf{v}_{2i} \hat{M}_{f,0} + \Delta_t^{d^* - \vartheta_0} \mathbf{v}_{2i} (\hat{M}_{f,0} - \hat{M}_{z,0}) \right\}$$

where

$$M_{f,0} := M_f(\vartheta_0) = F'(d^*) (F(d^*) F'(d^*))^{-1} F(\vartheta_0),$$

$$M_{z,0} := M_z(\vartheta_0) = \tilde{z}'(d^*) (\tilde{z}(d^*) \tilde{z}'(d^*))^{-1} \tilde{z}(\vartheta_0),$$

$$\hat{M}_{f,0} := \hat{M}_f(\vartheta_0) = F'(d^*) (F(d^*) F'(d^*))^{-1} \hat{F}(\vartheta_0),$$

$$\hat{M}_{z,0} := \hat{M}_z(\vartheta_0) = \tilde{z}'(d^*) (\tilde{z}(d^*) \tilde{z}'(d^*))^{-1} \tilde{z}(\vartheta_0),$$

and $\hat{F}(\vartheta) = (\partial/\partial \vartheta) F(\vartheta)$. Taking $n = 1$, the term

$$\frac{2}{\sqrt{T}} \mathbf{v}_{2i} \left[ (\log \Delta_t) \mathbf{v}_{2i} \right]'$$

yields the asymptotic normal distribution applying a central limit theorem for martingale difference sequences.

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. First,

$$\frac{2}{\sqrt{T}} \mathbf{v}_{2i} \hat{M}_{f,0} \Delta_t^{d^* - \vartheta_0} \mathbf{v}'_{2i} = \frac{2}{\sqrt{T}} \mathbf{v}_{2i} F'(d^*) (F(d^*) F'(d^*))^{-1} \hat{F}(\vartheta_0) \Delta_t^{d^* - \vartheta_0} \mathbf{v}'_{2i} = o_p(1)$$

since

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

$$\frac{F(d^*) F'(d^*)}{T} \rightarrow_p \sum_f > 0$$

$$\frac{\hat{F}(\vartheta_0) \Delta_t^{d^* - \vartheta_0} \mathbf{v}'_{2i}}{T} = O_p \left( T^{-1/2} + T^{d^* - 1} \log T \right).$$

The term dealing with the projection approximation,

$$\frac{2}{\sqrt{T}} \mathbf{v}_{2i} \left( \hat{M}_{f,0} - M_{z,0} \right) \Delta_t^{d^* - \vartheta_0} \mathbf{v}'_{2i}$$

can easily be shown to be $o_p(1)$ following the same steps described earlier. All other cross terms are negligible using similar arguments to what has been discussed until now so the result follows.

Finally, convergence of the Hessian can be shown following [Hualde and Robinson (2011)], and the proof is then complete. □
A.4 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. We therefore refrain from giving the same proof here. □

A.5 Covariance Matrix Estimate $\hat{A}_i\hat{B}_i\hat{A}_i'$

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

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

with

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

$$\hat{a}_{i3} = (1, 0, \ldots, 0)'\hat{\omega}_{i\tau_1} \hat{s}_{i\tau_1}^{-1} \hat{s}_{i\tau_1} \hat{s}_{i\tau_2}^{-1} - (1, 0, \ldots, 0)'\hat{\omega}_{i\tau_2} \hat{s}_{i\tau_2}^{-1},$$

$$\hat{a}_{i4} = -\hat{s}_{i\tau_1}^{-1}, \quad \hat{a}_{i5} = \hat{s}_{i\tau_1}^{-1} \hat{s}_{i\tau_1} \hat{s}_{i\tau_2}^{-1}, \quad \hat{a}_{i6} = -\hat{s}_{i\tau_2}^{-1},$$

where

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

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

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

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

$$\hat{M}_{i\tau_2} = Q \frac{1}{T} \sum_{t=p+1}^{T} \left\{ \hat{Z}_{i\tau_2}(\hat{\theta}_i) \hat{Z}_{\tau_2}(\hat{\theta}_i, \hat{\theta}_i) + \hat{Z}_{i\tau_2}(\hat{\theta}_i, \hat{\theta}_i) \hat{Z}_{i\tau_2}(\hat{\theta}_i) \right\} Q'$$

with the parameter subscripts denoting the first partial derivative as in

$$\hat{y}_{i\tau_1}(\hat{d}_i) = (\log \Delta) \hat{y}_{i\tau}(\hat{d}_i),$$

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

$$\hat{Z}_{i\tau_2}(\hat{\theta}_i) = (\log \Delta) \left\{ 0, \hat{x}_{i\tau}^*(\hat{\theta}_i), 0, \hat{x}_{i\tau-1}^*(\hat{\theta}_i), 0, \ldots, 0, \hat{x}_{i\tau-p}^*(\hat{\theta}_i), 0 \right\}.$$
and also
\[
\hat{s}_{i\tau r_1} = \frac{1}{T} \sum_{t=p+1}^{T} \hat{v}_{i\tau r_1}^* \hat{\varphi}_i, \quad \hat{s}_{i\tau r_2} = \frac{1}{T} \sum_{t=p+1}^{T} \hat{v}_{i\tau r_2}^* \hat{\varphi}_i, \quad \hat{s}_{i\tau r_2} = \frac{1}{T} \sum_{t=p+1}^{T} \hat{w}_{i\tau r_2}^* \hat{\varphi}_i,
\]
\[
\hat{v}_{i\tau r_1} = \hat{y}_{i\tau r_1}^*(d_i) - \hat{\omega}_{i\tau r_1}^* Q \hat{Z}_d^*(\hat{d}_i, \hat{\varphi}_i) - \hat{\omega}_{i\tau r_1}^* Q \hat{Z}_{i\tau r_1}^*(\hat{d}_i),
\]
\[
\hat{v}_{i\tau r_2} = -\hat{\omega}_{i\tau r_2}^* Q \hat{Z}_d^*(\hat{d}_i, \hat{\varphi}_i) - \hat{\omega}_{i\tau r_2}^* Q \hat{Z}_{i\tau r_2}^*(\hat{\varphi}_i),
\]
\[
\hat{w}_{i\tau r_2} = \hat{\bar{x}}_{i\tau r_2}^*(\hat{\varphi}_i) - \hat{\delta}_{i\tau r_2}^* R \hat{X}_d^*(\hat{\varphi}_i) - \hat{\delta}_{i\tau r_2}^* R \hat{X}_{i\tau r_2}^*(\hat{\varphi}_i),
\]
\[
\hat{x}_{i\tau r_2}^*(\hat{\varphi}_i) = (\log \Delta) \hat{\bar{x}}_{i\tau r_2}^*(\hat{\varphi}_i), \quad \hat{X}_{i\tau r_2}^*(\hat{\varphi}_i) = (\log \Delta) \hat{X}_d^*(\hat{\varphi}_i),
\]
\[
\hat{\phi}_{i\tau r_2} = \hat{G}_{i}^{-1} \left( \hat{g}_{i\tau r_2} - \hat{G}_{i\tau r_2} \hat{\varphi}_i \right),
\]
\[
\hat{g}_{i\tau r_2} = R \frac{1}{T} \sum_{t=p+1}^{T} \left\{ \hat{X}_{i\tau r_2}^*(\hat{\varphi}_i) \hat{\bar{x}}_d^*(\hat{\varphi}_i) + \hat{X}_d^*(\hat{\varphi}_i) \hat{X}_{i\tau r_2}^*(\hat{\varphi}_i) \right\},
\]
\[
\hat{G}_{i\tau r_2} = R \frac{1}{T} \sum_{t=p+1}^{T} \left\{ \hat{X}_{i\tau r_2}^*(\hat{\varphi}_i) \hat{\bar{x}}_d^*(\hat{\varphi}_i)' + \hat{X}_d^*(\hat{\varphi}_i) \hat{X}_{i\tau r_2}^*(\hat{\varphi}_i)' \right\} R'.
\]

Finally,
\[
\hat{B}_i = \frac{1}{T} \sum_{t=p+1}^{T} \left[ \begin{array}{c}
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) Q \hat{Z}_d^*(\hat{d}_i, \hat{\varphi}_i) \\
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) \hat{v}_{i\tau r_1}^* \\
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) \hat{w}_{i\tau r_2}^*
\end{array} \right] \left[ \begin{array}{c}
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) Q \hat{Z}_d^*(\hat{d}_i, \hat{\varphi}_i) \\
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) \hat{v}_{i\tau r_1}^* \\
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) \hat{w}_{i\tau r_2}^*
\end{array} \right]',
\]

where
\[
\hat{v}_{1,2,i}^*(\hat{d}_i, \hat{\varphi}_i) = \hat{v}_{1,i}^*(\hat{d}_i) - \rho_i \hat{v}_{2,i}^*(\hat{\varphi}_i),
\]
\[
\hat{v}_{2,i}^*(\hat{\varphi}_i) = \hat{x}_d^*(\hat{\varphi}_i) - \hat{\delta}_{1,i}^* R \hat{X}_d^*(\hat{\varphi}_i).
\]

**B Lemmata**

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

a. \[
T_1^{-1} \hat{\mathcal{E}}' \hat{\mathcal{E}} = O_p \left( \frac{1}{n} + \frac{1}{\sqrt{nT}} \right),
\]

b. \[
T_1^{-1} \hat{\mathcal{E}}' F(d^*) = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

c. \[
T_1^{-1} \hat{\mathcal{E}}_2'(d^* - \vartheta_0) \hat{\mathcal{E}} = O_p \left( \frac{1}{n} + \frac{1}{\sqrt{nT}} \right),
\]
where \( \bar{\epsilon} = (\bar{\epsilon}_2, \ldots, \bar{\epsilon}_T)' \).

**Proof of Lemma 1.a.** Let us write

\[
\bar{\epsilon}_t = \left( \frac{\Delta t - \delta_0}{\Delta t - \vartheta_0} \epsilon_{1t} + \frac{\Delta t - \vartheta_0}{\Delta t - \vartheta_0} \epsilon_{2t} \right).
\]

Then,

\[
T_1^{-1} \left( \sum_{t=2}^{T} \bar{\epsilon}_t' \bar{\epsilon}_t \right) = T_1^{-1} \sum_{t=2}^{T} \left( \frac{\Delta t - \delta_0}{\Delta t - \vartheta_0} \epsilon_{1t} \right)^2 + T_1^{-1} \sum_{t=2}^{T} \left( \frac{\Delta t - \vartheta_0}{\Delta t - \vartheta_0} \epsilon_{2t} \right)^2 + T_1^{-1} \sum_{t=2}^{T} \left( \frac{\Delta t - \vartheta_0}{\Delta t - \vartheta_0} \epsilon_{2t} \right)^2
\]

\[
+ 2T_1^{-1} \sum_{t=2}^{T} \frac{\Delta t - \delta_0}{\Delta t - \vartheta_0} \epsilon_{1t} \frac{\Delta t - \vartheta_0}{\Delta t - \vartheta_0} \epsilon_{2t},
\]

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

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

**b.** The expression has zero expectation. Using the independence of \( f_t \) and \( \bar{\epsilon}_t \),

\[
\text{Var} \left( \frac{\sum_{t=2}^{T} \bar{\epsilon}_t' f_t}{T_1} \right) = \frac{\sum_{t=2}^{T} \sum_{t'=2}^{T} E(f_t f_{t'}) E(\bar{\epsilon}_t \bar{\epsilon}_{t'})}{T_1^2},
\]

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

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

(27)

Take with no loss of generality, \( \vartheta_{\text{max}} > d_{\text{max}} \). Then (27) becomes

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

Thus, \( \frac{\sum_{t=2}^{T} \bar{\epsilon}_t' f_t}{T_1} = O_p ((nT)^{-1/2}) \).

**c.** The expectation of \( T_1^{-1} \left( \sum_{t=2}^{T} \bar{\epsilon}_t \bar{\epsilon}_{2t} \right) \) is \( O(n^{-1}) \) and its variance is \( O((nT)^{-1/2}) \),

which can be shown as in Lemma 1.a. Thus, \( T_1^{-1} \left( \sum_{t=2}^{T} \bar{\epsilon}_t \bar{\epsilon}_{2t} \right) = O_p \left( n^{-1} + (nT)^{-1/2} \right) \).

\(\square\)
References


