Fixed-b estimation and inference in heterogenous dynamic cointegrated panels

Samuel Sender*

Working Paper, 2015

*Researcher, Tilburg University, CentER, and Netspar. Email: s.j.sender@uvt.nl

Fixed-b estimation and inference in heterogenous dynamic cointegrated panels.

Abstract

We study semi-parametric estimation and inference in cointegrated panels with endogenous feedback, allowing for general time-series and cross-section dependence and heterogeneity.

In the class of semi-parametric estimators that use spectral [non-parametric] estimation of the asymptotic covariance matrix of the errors to eliminate the bias, we propose a simple pooled estimator which is T-consistent (with bias $o(1/T)$) in the *fixed-b* asymptotics (Kiefer and Vogelsang, 2005) where the ratio of the bandwidth to the panel time length tends to a positive constant (strictly greater than zero, lower than one) in the *T*-limit.

The reference, the Fully Modified OLS of Phillips and Hansen (1990), do not achieve T-consistency under *fixed-b*. Despite its aim, FM-OLS test statistics are non-pivotal, and always reject in the *N*-limit (and size distortions increase with N).

We propose tests that include $\mathscr{O}(b)$ bounds to account for the variance of the semi-parametric correction. Empirically, our tests exhibit minor distortions compared to that with FM-OLS. Our *fixed-b* correction is also shown to considerably improve the size statistics, and reduce the bias and RMSE compared to other estimators in finite samples.

Keywords: Panel cointegration, Fully modified, Fixed-b asymptotics, Dynamic Least Squares, FM-OLS, DOLS, DGLS, IM-OLS.

JEL: B23, C01, C13, C18, C32, C33, C52.

1 Introduction

Motivation

Cointegration analysis has an important place in econometrics, notably because the estimators are super-convergent even when the residuals are not weakly exogenous (a situation which is called endogenous feedback). Integrated time series also keep the memory of the history of innovations, so working without differentiating time-series should permit long-term analyses that are crucial in economics. Unfortunately, these promises are often nullified by distorted size and power statistics in finite samples, thus improving existing estimators without imposing restrictive conditions remains an important challenge.

We study a panel where in each cross-section a single cointegrated relationship exists between time-series, with identical cointegrated slopes for all cross-sections, but where the stationary residual is not weakly exogenous to the integrated regressors, which is similar to Phillips and Moon (1999). By contrast, however, we permit cross-section heterogeneity of the short-run dynamics, weak-cross-sectional dependence. Mostly, unlike in most of the literature on cointegration, we do not assume that non-parametric estimators of the asymptotic covariance matrix based on residuals from a first-step OLS are necessarily T-consistent.

OLS are asymptotically unbiased with cointegrated variables, even in the presence of endogenous feedback: the OLS bias vanishes at an $\mathcal{O}(1/T)$ rate. OLS however do not permit nuisance-parameter-free test statistics: volatility and bias being equally $\mathcal{O}(1/T)$, the *t*-statistic for a scalar β is not centered and for a single time series has non-normal distribution (unless the error term is truly exogenous). This is well known in the literature on the asymptotic theory of (co)integrated time-series, and with notably the FCLT and resulting functional Brownian motions used in Park and Phillips (1988). This is also the starting point of the semi-parametric correction proposed by Phillips and Hansen (1990) in the fully modified OLS (FM-OLS).

The core idea of FM-OLS is to exogenise the error term u to X, using a non-parametric estimator of the asymptotic covariance matrix Ω_{vu}. But such estimators, applied to the residual of a first-step OLS regression, are only unbiased in the limit where the relative bandwidth size

$M/T \xrightarrow{M,T\to\infty} 0$.

In any given estimation problem, a bandwidth $M > 0$ is chosen, and the *fixed-b* asymptotic of Kiefer and Vogelsang (2005), where $M/T \xrightarrow{T\to\infty} b > 0$ (or $M = [b \cdot T]$), captures in a realistic manner the departure from the traditional kernel limits where $M/T \xrightarrow{T\to\infty} 0$.

Under *fixed-b*, the nuisance terms of FM-OLS are described in Vogelsang and Wagner (2014), which show that FM-OLS is not T-consistent and that the complex functional Brownian motions involved have unknown volatility. For a single time-series, or a panel with a small width N, *fixed-b* asymptotics can be seen as a simple approximation which relevance would vanish as the T sample size gets very large. For panels, however, size distortions increase with N and test statistics always reject in the limit where $N \to \infty$.

In addition, simulations (Phillips and Loretan, 1991; Pedroni, 2001; Vogelsang and Wagner, 2014) have shown dramatic size and power distortions of the FMOLS estimator when T is small or $b \neq 0$, so an open question is whether *fixed-b* corrections are available for the cointegrated model (leading to at least the same improvement in empirical size as is known for stationary series with exogenous regressors).

Entirely different approaches exist that do not make use of non-parametric estimators of the asymptotic covariance. This is the case of course with parametric approaches (Stock and Watson, 1993), its semi-parametric extension via augmentation of the regressor space (the D-OLS of Saikkonen, 1991), or partial summing of the variables in equation (plus addition of the original regressors) with the IM-OLS approach of Vogelsang and Wagner (2014). Each has strengths and weaknesses we will not analyse extensively (but reported simulations show poor performance in finite samples).

Our contribution is to show that the parsimonious semi-parametric estimators can be made T-consistent under *fixed-b* asymptotics (and of course to show the gains in size and RMSE associated to such refined asymptotics and estimation techniques). To our knowledge, so far, no such non-parametric correction have been proposed.

The difficulty of this exercise has been underlined first by Pedroni (1996, 2001) and Hansen and Phillips (1990), stating that feasible FM-OLS based on estimated residuals 'unfortunately works much less well' than if based on the true residuals, and that, because the scaled OLS bias is not eliminated asymptotically in presence of endogeneity, basing the estimation of the long-run covariance matrix on residuals from a first-step OLS regression is not sufficient to eliminate size distortions.

The long-run covariance matrix cannot be T-consistently estimated from the first-step OLS regression, however the bias in non-parametric estimators is linked to the first-step OLS estimation error, an observation which will permit us to T-consistently estimate the OLS bias, and therefore to center *t*- of *chi-square*- test statistics, which is important for time-series and crucial for panels.

To do so, we naturally rely on spectral estimators with notably Andrews (1991), Andrews and Monahan (1992), and Newey and West (1987), and on the *fixed-b* asymptotic theory of Kiefer and Vogelsang (2005). More specifically, we use *fixed-b* asymptotics in the context of a cointegration model, which thus closely relates Vogelsang and Wagner (2014). We however use new notations for all these techniques both for notational economy and to be able to disentangle the impact of the first-step estimation error on the non-parametric estimators.

We study the $(N,T) \to \infty$ with $N = o(T)$ asymptotics of panels where in each cross-section, time-series are cointegrated with identical slopes β. Our assumptions (see Section A.1 for rigour) are minimal: we require **(A1)** a functional central limit theorem for the error term and innovations in the regressor, **(A2)** weak convergence of sample covariance matrices to functional Brownian motion; we allow cross-section heterogeneity and weak cross-section dependence as long as **(A3)** all relevant quantities in the estimation process have finite first and second moments.

Then, it is remarkable that the nuisance terms that affect the scaled bias can be eliminated and a T-consistent estimator of the slope constructed under *fixed-b* asymptotics. The *t*-statistics are centered, and the variance of the error known up to a proportional $\mathcal{O}(b)$ perturbation, which permits testing with bounds on the critical values. With the median of the bounds, the empirical size in our study is in the 2%-10% range (corresponding to the no-feedback and strong feedback situations), whereas the FM-OLS has size 4%-80% in the same scenarios (where 80% or above happens for small T, large N; with strong feeback, T=90,N=50, our sizes is 6% vs for FM 50%).

Organisation of the paper

This paper is organised as follows: Section 2 details the model and notations, Section 3 analyses limits of the POLS estimator and non-parametric estimators under *fixed-b*, Section 4 develops a T-consistent pooled panel estimator, and Section 5 shows numerical results.

Appendix A.1 states more formally regularity conditions, A.3 gives more detailed proofs.

2 Model and notations

2.1 Generic Notations

$\mathbb{E}$ denotes the expectation of a random variable. By contrast, $\widehat{\mathbb{E}}_i[\cdot] = \sum_{i=1}^{N}[\cdot_i]$ denotes the sample cross-section (denoted CS) average of a quantity $(\cdot)$. Subscripts i are often dropped in the notations, as the FCLT applies for each time-series independently.

The T-scaled estimation error is denoted Θ with the estimator as subscript, or simply β for OLS. Thus $\Theta_\beta = T\cdot(\widehat{\beta}^{POLS} - \beta)$.

Variables: $\widetilde{(\cdot)}$ denotes a demeaned variable, *e.g.* $\widetilde{X} = X - \overline{X}$, while $\widehat{(\cdot)}$ denotes an estimated variable from the first-step pooled regression, *e.g.* $\widehat{u} = \tilde{u} - \widetilde{X}\cdot\widehat{\Theta}_\beta$.

Estimators: $\widehat{(\cdot)}$ denotes an estimator, $\widehat{(\cdot)}^b$ indicates a kernel estimator with a $b = M/T$ relative bandwidth, $\widehat{(\cdot)}^b_{\tilde{u}}$ indicates that the estimator is computed based on the demeaned $\tilde{u}$ (even if $\tilde{u}$ is unobserved), while $\widehat{(\cdot)}^b_{\hat{u}}$ uses the residuals $\hat{u}$ estimated from the first-step (pooled) OLS regression.

Convergence: as in most of the literature, the symbol $\Rightarrow$ denotes weak convergence of random variables on their respective probability measures. Specifically, a construct $X_T(r) = \frac{1}{\sqrt{T}}\sum_{\tau=1}^{[rT]} v_\tau$ is defined over $D[0,1]^T$ the product metric space of all real valued *càdlàg* functions at each point in $[0,1]$, with the the component space $D[0,1]$ endowed with the Skorohod metric.

We use the symbol $\rightarrow$ to indicate degenerate limits (convergence[1] to a constant or divergence to ∞). $X \xrightarrow[N,T\rightarrow\infty]{} c$ is a shortcut for $\lim_{N,T\rightarrow\infty} X = c$.

[1]In the case of a degenerate constant or dirac mass limit, weak convergence, convergence in probability, in distribution, and almost sure convergence are interchangeable concepts.

The notations for the spectral covariance matrix and its decomposition are identical to those of Vogelsang and Wagner (2014), see Appendix A.2 for more details.

We use shortcut notations for univariate integrals of Brownian motions in $[0,1]$ we write $\int B$ for $\int_0^1 B(r)dr$, and $\int B \cdot dB' := \int_0^1 B(r)dB'(r)$, and $\int B \cdot B'(1) := \int_0^1 B(r)B'(1)dr$. We clarify the integration space when not over $[0,1]$ (typically $\int_0^b (\dots)$) and specify the arguments in double integrals (typically $\int\int(\dots)dr \cdot ds$) (or in integrals locally defined).

2.2 Model and summary of assumptions

We study the integrated model

$$Y_{i,t} = \alpha + X_{i,t}\beta + u_{i,t} \tag{1}$$

and $\beta \neq 0$ is the identical cointegrating slope, Y_i and X_i are $I(1)$ cointegrated variables, no cointegration relationship exists between the X_i, but where u is not necessarily weakly exogenous to $X_{i,t} = \sum_{j=1}^{t} v_{i,j}$. No condition is imposed on the scalar α_i. Main assumptions are:

(A1): for each time-series (subscript i dropped), $\eta_t = [u_t,\ v_t']'$ is a vector of $I(0)$ processes which satisfies a multivariate invariance principle:

$T^{-1/2}\sum_{t=1}^{[rT]}\eta_t \Rightarrow B(r) = \begin{bmatrix} B_u(\mathrm{r}) \\ B_v(\mathrm{r}) \end{bmatrix} = \Omega^{1/2}W(r)$, $r \in [0,1]$, with $\mathbb{E}[B_v B_u'] = \Omega_{vu}$

(A2): as in Phillips and Durlauf (1986), we also require the weak convergence of the following sample covariance matrices to matrix stochastic integrals: $T^{-1}\sum_1^T X_t \cdot u_t' \Rightarrow \int_0^1 B_v dB_u' + \Delta_{vu}$

(A3) regularity conditions ensure that a central limit theorem can be applied to the bias and to the mean of all spectral covariance matrices and their estimators, which rules out common factors[2] and requires that $\widehat{\Omega}^b_{vv}$ be ‘qualitatively’ bounded away from zero (see Appendix A.1).

We will denote $\widehat{\mathbb{E}}_i\Omega_i \xrightarrow[N\to\infty]{} \overline{\Omega}$, and $\widehat{\overline{\Omega}}$ is a shortcut notations for the estimate of the mean, while $\widehat{\overline{\Omega}_{vv}^{-1}}$ stands for the estimate of the inverse of the mean covariance matrix of the regressors.

We then build a pooled estimator under the assumption of a cross-section of $N \to \infty$ individuals

[2]Note that the absence of common factor between the cross-section is needed for convergence, but this is a natural assumption to the modeler who aims to include common exogenous factors in the regression, see Gregoir (2005). Without such effort or assumption, the cross-section dimension of the panel is of virtually no use for convergence. In practise we here rule out unobserved factors.

indexed by $(i = 1, \ldots, N)$ possibly weakly dependent, where the form of endogenous 'feedback' and related endogeneity bias varies in the cross-section but where **(A3)** holds.

I also work with symmetric kernels. While all $\mathcal{O}(b)$ terms are valid for any symmetric kernels (the proof is however only given for truncating kernels), we only develop the $\mathcal{O}(b^2)$ correction for the Bartlett kernel. Automatic bandwidth selection has not been explicitly studied in the literature the presence of endogeneity, so for the theory we assume in this paper the bandwidth to be exogenously given (and consistent for all panel members); we use $M = 4 \cdot \left(\frac{T}{100}\right)^{2/9}$ in simulations.

2.3 New notations for non-parametric covariance matrix estimation

To disentangle the components of $\hat{u}$, we propose a new representation of the *fixed-b* theory and of the local functional Brownian motions involved in the limits of the non-parametric estimators of the asymptotic covariance matrices.

Definition 2.1. *fixed-b* asymptotics

The *fixed-b* theory owes its name to Kiefer et al. (2000), Kiefer and Vogelsang (2005), and *fixed-b asymptotics* or *fixed-b* limit denote asymptotics involving a non-parametric estimator of the asymptotic covariance matrix at a bandwidth M when $M/T \xrightarrow[M,T\to\infty]{} b$.

Remark 2.2. Kiefer et al. (2000), Kiefer and Vogelsang (2005) study inference in linear models with stationary regressors and exogenous error, so in their work the nuisance term is the HAC estimation of the variance of the residual, which does not converge to a constant, so, in their model, the t-statistic asymptotically follows a distribution similar to a student distribution rather than a standard normal.

We study estimation and inference under *fixed-b* when the weak exogeneity assumption is violated – this implies bias in non-parametric estimators, on top of the volatility already documented.

Lemma 2.3. *fixed-b limits of estimators of the asymptotic covariance matrix.*

*Under assumptions **(A1-A2)**, with a truncating kernel, with $k(\cdot)$ the kernel weighting function and M the bandwidth, $\widehat{\Delta}^b_{\eta,i}$ the estimator of $\Delta_{\eta,i}$ (based on the true η_i) has the following fixed-b* *limit (subscript i dropped) :*

$$\begin{aligned}
\widehat{\Delta}^b_\eta &= \frac{1}{T}\sum_{i,j\leq i}\eta_j\cdot k(\frac{i-j}{M})\eta_i' \\
&= \frac{1}{\sqrt{T}}\sum_i\left(\frac{1}{\sqrt{T}}\sum_{l=0}^{min(M,i-1)}k(\frac{l}{M})\eta_{i-l}\right)\eta_i' \\
&\Rightarrow \Delta_\eta + \int B^{b/2}_\eta\cdot dB'_\eta
\end{aligned} \tag{2}$$

with the constructed variable $X^{b/2}$ and the Brownian it weakly converges to defined as:

$$\frac{1}{\sqrt{T}}X^{b/2}([rT]) = \frac{1}{\sqrt{T}}\sum_{l=0}^{min(M,[rT]-1)}k(\frac{l}{M})\eta_{[rT]-l}\Rightarrow B^{b/2}_\eta(r) = \int_0^{b\wedge r}k^*(s)dB_\eta(r-s) \tag{3}$$

where $k^*(\cdot)$ normalised kernel weights (such that $\int k^* = b$, e.g. for Bartlett $k^*(s) = 1-s/b$). See proof on page 21.

The key to our analysis is to recognise that, for a given r time index, $1/\sqrt{T}$ times the 'locally weighted' partial sum in (2.3) converges to a Brownian motion under the same hypothesis than for the convergence of $1/\sqrt{T}$ times the partial sum of η: the arguments of equation (2.5) on page 1062 and following Lemma 3 of Phillips and Moon, 1999 apply to our modified partial sum.

Since $B^{b/2}_\eta(r)$ is the weighted sum of $dB_\eta(r-s)$ in an interval local to r, of length $min(b,r) = b\wedge r$, we name $B^{b/2}_\eta(r)$ a *local* Brownian motion.

Defining $B^{-b/2}_\eta(r) = \int_0^{b\wedge(1-r)}k^*(s)dB_\eta(r+s)$, we have $\widehat{\Omega}^b_\eta \Rightarrow \Omega_\eta + \int\left(B^{b/2}_\eta + B^{-b/2}_\eta\right)\cdot dB'_\eta$; by abuse of notation,[3] we write:

$$\widehat{\Omega}^b_\eta \Rightarrow \Omega_\eta + \int B^b_\eta\cdot dB'_\eta$$

Thus, while $\widehat{\Omega}^b_\eta$ (and $\widehat{\Omega}^b_{\tilde{\eta}}$) has the same $\mathscr{O}(b)\cdot\Omega_\eta$ variance as in Kiefer and Vogelsang (2005), functional Brownian motions involved have a classical functional Brownian motion representation.

[3] $B^{-b/2}_\eta(r)$ is a time reversed Brownian motion, which could be seen as an abstraction; however we can write $\int B^{-b/2}_\eta(r)\cdot dB'_\eta(r) = \left(\int B^{b/2}_\eta(r+b)\cdot dB'_\eta(r+b)\right)'$. But we seek notational economy.

Main result (see theorem 4.1 on page 13)

Assume **(A1-A3)** and denote

$$\widehat{\Theta}_{\beta} = \frac{6}{1-3/2b+b^2/2} \cdot \left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\Omega}^{b}_{vv,i}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\Delta}^{b}_{v\hat{u},i} - \frac{1}{2}\widehat{\Omega}^{b}_{v\hat{u},i}\right)\right) \tag{4}$$

Then

$$\lim_{T,N\to\infty}\widehat{\Theta}_{\beta} = \lim_{T,N\to\infty}\Theta_{\beta} = 6\cdot\frac{\overline{\Delta}_{vu} - \frac{1}{2}\overline{\Omega}_{vu}}{\overline{\Omega}_{vv}} \tag{5}$$

where $\widehat{\Omega}^{b}_{vv,i}$ is the kernel estimator at the bandwidth $M = b\cdot T$ of the spectral autocovariance[4] of v_i , and $\widehat{\Delta}^{b}_{v\hat{u},i}$ and $\widehat{\Omega}^{b}_{v\hat{u},i}$ are kernel estimators of the 'half' and full spectral covariances of v_i and u_i, estimated using $\hat{u}_i$ is the residual from a first step simple pooled OLS regression.

A different kernel would yield another multiplicative constant $m(b) = \frac{6}{1-3/2b+b^2/m2}$ with $m2 \neq 2$ in (4); so would identified deterministic trends and observable common factors.

Because of the semi-parametric correction, the volatility is always comprised between that of an ideal one-step estimator that would yield the same residuals (which can be thought of that of POLS, if it was unbiased which we denote $\sigma(\beta^{POLS})$) by abuse of notation, and that $\sigma(\beta^{POLS})$ plus the $\mathscr{O}(b)$ variance of the semi-parametric correction, which can be easily simulated ($2b\cdot\Omega_{vu}/\Omega_{vv}$ with *i.i.d.* panels). This permits testing, see Section 5 to see the gain in size from our refined *fixed-b* asymptotics.

3 Bias of pooled panel OLS

3.1 Limit of the pooled panel OLS estimator

Usually, time-series or panel estimators in the presence of homogenous feedback are based on a first-step OLS regression, then one tries to recover the bias. Focusing on the slope, we have

$\hat{\beta}^{POLS} = (\widehat{\mathbb{E}}_i \frac{1}{T}\tilde{X}'\tilde{X})^{-1}\left(\widehat{\mathbb{E}}_i \frac{1}{T}\tilde{X}'y\right)$, and $(\hat{\beta}^{POLS} - \beta) = (\widehat{\mathbb{E}}_i \frac{1}{T}\tilde{X}'\tilde{X})^{-1}\left(\widehat{\mathbb{E}}_i \frac{1}{T}\tilde{X}'u\right)$

[4] We use the terms spectral covariance and asymptotic covariance interchangeably. This means in practice a AC estimator that can be found in any statistical package, for instance the Newey-West estimator.

From Phillips and Durlauf (1986), we have (the two notations are equivalent by Frish-Waugh): $\frac{1}{T}\tilde{X}'u \Rightarrow \Delta_{vu} + \int_0^1 \widetilde{B}_v(r)dB'_u(r) = \Delta_{vu} + \int_0^1 B_v(r)dB'_{\tilde{u}}(r)$,

so, provided that $(N \to \infty, N = o(T))$, we have (see Phillips and Moon, 1999),

$\underset{N\to\infty}{lim}\ \widehat{\mathbb{E}}_i \frac{1}{T}\tilde{X}'u \Rightarrow \underset{N\to\infty}{lim}\ \frac{1}{N}\sum_{i=1}^N[\mathbb{E}\frac{1}{T}\tilde{X}_i'u_i] = \overline{\Delta}_{vu} - \frac{1}{2}\overline{\Omega}_{vu}$.

To see this, note that $\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]}\tilde{u}_j \Rightarrow B_{\tilde{u}}(r) = B_u(r) - rB_u(1)$, so that $dB_{\tilde{u}}(r) = dB_u(r) - B_u(1)dr$.

So, $\int B_v dB'_{\tilde{u}} = \int B_v dB'_u - \int B_v B'_u(1)$, and $\int B_v dB'_u = \frac{1}{2}[B_v(1)B'_u(1) - \Omega_{vu}]$ has mean 0, while $\mathbb{E}\int B_v B'_u(1) = \Omega_{vu}\int r dr = \frac{1}{2}\Omega_{vu}$

It is well known that $\frac{1}{T^2}\tilde{X}'\tilde{X} \Rightarrow \int \widetilde{B}_v\widetilde{B}'_v$, where $B_v = B_v - \int B_v$, has expectation $\frac{1}{6}\Omega_{vv}$.

As a consequence, the scaled limit of the pooled panel bias is

$\Theta_\beta = T(\hat{\beta} - \beta) \xrightarrow[N,T\infty]{} 6 \cdot \left(\overline{\Omega}_{vv}\right)^{-1}\left(\overline{\Delta}_{vu} - \frac{1}{2}\overline{\Omega}_{vu}\right)$

Although exogenisation of the error term would permit the construction of consistent estimators under *fixed-b* asymptotics, this is a practical and theoretical challenge because the $\Omega_{vu,i}$ are not observed, and because a prior T-consistent estimate of Θ_β would be needed.

3.2 Nuisance term in non-parametric estimates

Lemma 3.1. *nuisance in $\widehat{\Omega}^b_{v\hat{u},i}$ (resp. $\widehat{\Delta}^b_{v\hat{u},i}$), the feasible kernel estimator of $\Omega_{vu,i}$ (resp. $\Delta_{vu,i}$)*

*Under hypothesis **(A1)-(A2)** (weak convergence of partial sums and of sample covariance matrices to functional Brownian motions), the feasible estimators based on $\hat{u}_i$ the residuals from the first-step OLS regression weakly converge as follows:*

$$\widehat{\Omega}^b_{v\hat{u},i} \Rightarrow \Omega_{vu,i} + \int B^b_{v,i} \cdot dB'_{\tilde{u},i} \tag{6}$$
$$- \int B^b_{v,i} \cdot \widetilde{B}'_{v,i} \cdot \Theta_\beta \tag{7}$$

$$\widehat{\Delta}^b_{v\hat{u},i} \Rightarrow \Delta_{vu,i} + \int B^b_{v,i} \cdot dB'_{\tilde{u},i} \tag{6}$$
$$- \int B^b_{v,i} \cdot \widetilde{B}'_{v,i} \cdot \Theta_\beta \tag{7}$$

Proof. The proof involves three arguments.

- $\hat{u}_i = \tilde{u}_i - \frac{1}{T}\tilde{X}_i \cdot \Theta_\beta$. Thus, from the multivariate principle, $\frac{1}{\sqrt{T}}\hat{u}_i \Rightarrow dB_{\tilde{u}_i} - \frac{1}{T}\widetilde{B}_{v,i \cdot \Theta_\beta}$
- The $dB_{\tilde{u}_i}$ part yields the limit (6) following Lemma (**2.3**) on page 7.
- The $\frac{1}{T}\widetilde{B}_{v,i} \cdot \Theta_\beta$ part yields (7), following the continuous mapping theorem (the sum product variables that converge towards Brownian motion is the limit of a Riemann Stieltjes sum).

□

Remark 3.2. Link to the usual *fixed-b* results.

The nuisance from demeaning in (6) is studied in the *fixed-b* literature. Focussing on the bias,[5] and dropping the index i from notations, we have

$$\mathbb{E}[\int B_v^{b/2} \cdot dB_{\tilde{u}}'] = \mathbb{E}[\int B_v^{b/2} \cdot dB_u'] - \mathbb{E}[\int B_v^{b/2}(r) \cdot B_u'(1)dr] = 0 - \frac{b}{2} \cdot \Omega_{vu} + O(b^2)$$

for any truncating symmetric kernel. This comes because

$\mathbb{E}[B_v^{b/2}(r) \cdot B_u'(1)] = \int_0^b k^*(s)ds = b\int_0^1 k(s)ds = b/2$ for $r \geq b$

(k^* integrates to $\frac{1}{2}$ over the positive or negative half line). The $\mathscr{O}(b^2)$ term is kernel specific. For the Bartlett kernel with normalised wrights $k^*(s) = 1 - s/b$, defining the standard Browian Motion $W = \Omega^{-1/2}B$ we have:

$$\mathbb{E}[\int W^{b/2} \cdot W'(1)] = \int_0^1 b/2 \cdot db - \int_0^b \int_0^r s/b \cdot ds \cdot dr = \frac{1}{2}(b - b^2/3)$$

By symmetry, $\mathbb{E}[\int B_v^b \cdot dB_{\tilde{u}}'] = (b - b^2/3)\Omega_{vu}$

Remark 3.3. An impossible (direct) exogenisation:

The form $\int B_{v,i}^b \cdot \widetilde{B}_{v,i}' \cdot \Theta_\beta$ in equation (7) means that the first-step scaled bias Θ_β is multiplied by a random functional Brownian motion, which results in a bias that is not simply proportional to Ω_{vu}. A simple scaling of $\widehat{\Omega}_{v\hat{u}}^b$ does not suffice for bias-free estimates of Ω_{vu}. So *(direct) asymptotic exogenisation of the error term is not possible* (and endogeneity complicates the *fixed-b* analysis).

[5] (6) is known to have variance $(\frac{4b}{3} + O(b^2)) \cdot \Omega$, and test statistics can usually be built to take this nuisance parameter into account, see for instance Kiefer et al. (2000) for a correction of the bias and tabulation of the critical value of the t-statistic for a single time series. In Kiefer and Vogelsang, 2005, p1142, 1146 it is found that for a univariate $\tilde{u}$, the variance is $V(\widehat{\Omega}_{\tilde{u}}^b)/\Omega_{\tilde{u}} = 4/3b - 7/3b^2 + 14/15b^3 + 2/9b^4$ for $b < 1/1$; the critical value for (symmetric) t-statistic at the 5% size is $cv(b, 97.5\%) = 1.96 + 2.9694b + 0.41b^2 - 0.5324b^3$

3.3 Estimation and inference with FM-OLS

Estimation

For a single time-series, the Fully Modified OLS estimator reads:

$$\widehat{\beta^{FM}} = (\tfrac{1}{T}\tilde{X}'\tilde{X})^{-1} \cdot \left[\tfrac{1}{T}(\tilde{X}'Y^*) - \hat{m}\right] \text{ with } \hat{m} = \widehat{\Delta}^b_{v\hat{u}} - \frac{\widehat{\Omega}^b_{v\hat{u}}}{\widehat{\Omega}^b_{vv}} \cdot \widehat{\Delta}^b_{v\tilde{v}} \text{ and } Y^* = Y - \frac{\widehat{\Omega}^b_{v\hat{u}}}{\widehat{\Omega}^b_{vv}} \cdot v$$

This can easily be adapted to panel, with in the case of heterogenous dynamics a specific correction for each time-series, and for homogenous panels $\widehat{\Omega}^b_{v\hat{u}}$and $\widehat{\Omega}^b_{vv}$can be identical for all time-series.

With the same decomposition than in Lemma 3.1, following the steps in A.3, one can see that $\Theta_{\beta^{FM}} \xrightarrow[N,T\to\infty]{} \mathscr{O}(b) + \mathscr{O}(b) \cdot \Theta_\beta$.

Inference under *fixed-b*

Then the t-statistic (for a univariate X) diverges (the numerator is $\mathscr{O}(b)$ and the denominator or standard deviation tends to 0 as $N \to \infty$).

Practical implications for panels

The typical rule for bandwidth selection for the Bartlett kernel is $M = \mathscr{O}(T^{2/9})$. This means that in mainly applications, $M/T = o(T^{-7/9})$ and the bias is $\mathscr{O}(M/T^2) = \mathscr{O}(T^{-16/9})$, while the volatility is $\left(\mathscr{O}(T^{-1}) + \mathscr{O}(M/T^2)\right) \cdot \mathscr{O}(N^{-1/2}) = \mathscr{O}(T^{-1} \cdot N^{-1/2})$.

The OLS bias is known to be possibly large, and, when the bandwidth is clearly distinct from zero (eg 10% of more), FM-OLS will also leave a substantial biased. Then, for a given (M,T), increasing the width of the panel always increases size distortions. This is very clearly illustrated in empirical tests with FM-OLS in section 5.

4 A simple pooled panel estimator

4.1 Pooled-panel estimator

Theorem 4.1. *Simple pooled panel*
Define $\widehat{\Theta}_\beta = \frac{6}{1-3/2b+b^2/2} \cdot \left[\left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\Omega}_{vv,i}^{b} \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\widehat{\Delta}_{v\hat{u},i}^{b} - \frac{1}{2} \widehat{\Omega}_{v\hat{u},i}^{b} \right) \right) \right]$

*Under assumptions **(A1-A3)**, $\widehat{\Theta}_\beta$ is a $\sqrt{N}T$-consistent scaled estimator of the scaled bias under fixed-b asymptotics.*

*Assumptions **(A1-A2)** imply the weak convergence to the relevant functional Brownian motions for each cross-section, and **(A3)** permits the point-convergence of the empirical cross-sectional mean of this quantities. See proof in Section A.3 on page 21.*

Remark 4.2. We thus note that the (relative) bandwidth size b magnifies the volatility of the estimator, and has thus a role similar to that of degrees of freedom with parametric estimators.

Remark 4.3. This estimator can be implemented without any programming skill, using standard statistical packages for kernel estimators.

Proof. sketch of the proof (see A.3 on page 21)

In pooled estimators, the nuisance term for the bias is in the nominator, and has expectation $\frac{1}{N} \sum_{i=1}^{N} \left(\Delta_{vu,i} - \frac{1}{2} \Omega_{vu,i} \right)$. Even though both $\widehat{\Delta}_{v\hat{u},i}^{b}$ and $\widehat{\Omega}_{v\hat{u},i}^{b}$ have complex asymptotics, their combination $\frac{1}{N} \sum_{i=1}^{N} \left(\widehat{\Delta}_{v\hat{u},i}^{b} - \frac{1}{2} \widehat{\Omega}_{v\hat{u},i}^{b} \right)$ permits to build a T-consistent estimator. More precisely, for each cross-section, we have:

$$\widehat{\Delta}_{v\hat{u},i}^{b} - \frac{1}{2} \widehat{\Omega}_{v\hat{u},i}^{b} \Rightarrow \Delta_{vu,i} - \frac{1}{2} \Omega_{vu,i} \tag{8}$$

$$+ \frac{1}{2} \int \left(B_{v,i}^{b/2} - B_{v,i}^{-b/2} \right) \cdot dB_{\tilde{u},i}' \tag{9}$$

$$- \frac{1}{2} \int \left(B_{v,i}^{b/2} - B_{v,i}^{-b/2} \right) \cdot \widetilde{B}_{v,i}' \cdot \Theta_\beta \tag{10}$$

where (9) has a zero-mean process and where (10) has a known mean[6] conditional on the unknown Θ_β. As (8) is also directly linked to $\underset{N,T\to\infty}{lim} \Theta_\beta$, taking the cross-section averages permits

[6] $\mathbb{E}[(10)] \Rightarrow \frac{1}{4}(b - b^2/3)\Omega_{vv,i}\Theta_\beta$

a one-to-one relationship between $\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\Delta}^{b}_{v\hat{u},i}-\frac{1}{2}\widehat{\Omega}^{b}_{v\hat{u},i}\right)$ and Θ_β.

Note that the proof from the decomposition (8) – (10) follows immediately from Lemma (**3.1**) on page 10. See section A.3 on page 21 for a more detailed proof.

Remark 4.4. Boundaries for testing

$\widehat{\Theta}_\beta$ is an estimator of the scaled bias $\mathbb{E}[\Theta_\beta]$ with volatility $\sigma(\widehat{\Theta}_\beta)=\mathscr{O}(b)/\sqrt{N}$, which contrasts with $\sigma(\Theta_\beta)=\mathscr{O}(1)/\sqrt{N}$.

In semi-parametric models (Newey, 1990), the nonparametric correction is a nuisance, so that the volatility of the estimator is greater than that of a hypothetical parametric model that would have generated the same estimate.

The lower bound on the variance is thus the HAC estimator, based on the residual from $\hat{\beta}^{simple}$, but using $\widetilde{X}$ as a simple instrument. We denote this lower bound (resp. its estimate) $\sigma(\Theta_\beta)$ (resp. $\widehat{\sigma}(\Theta_\beta)$) by abuse of notation.

From Jensen's inequality, $\sigma\left(T\cdot(\hat{\beta}^{simple}-\beta)\right)=\sigma(\Theta_\beta-\widehat{\Theta}_\beta)\leq\sigma(\Theta_\beta)+\sigma(\widehat{\Theta}_\beta)$, which provides an upper bound for testing.

We note that as b shrinks, the nuisance terms vanish, so do the boundaries (and the sizes become asymptotically normal, which is not the case with FM-OLS).[7]

Remark 4.5. Tests with exact size

It is possible to obtain exact confidence intervals by simulation, provided some restrictions on the cross-section heterogeneity and dependence are given.

We leave here the form of a general or specific exact test statistic, be it asymptotic and finite-sample for further research, since specific restrictions would also permit more specific estimation techniques: trivially, with *i.i.d.* CS, exact size statistics can be simulated, but one may note that with such conditions, the nuisance term from *fixed-b* non-parametric can arguably be eliminated by averaging.[8]

[7]We can write: $\sigma(\Theta_\beta-\widehat{\Theta}_\beta)=\sigma\left(T\cdot(\hat{\beta}^{simple}-\beta)\right)\xrightarrow[N,T\to\infty,b\to 0]{}\mathscr{O}(1)/\sqrt{N}\xleftarrow[N,T\to\infty]{}\sigma(\Theta_\beta)$

[8]If the dispersion in the cross-section can be summarised by a parametric quantity (*e.g.* random effects for Ω_i, knowing that the nuisance term in the time series i is adequately summarised by Ω_i, the distribution of the slope estimator can in theory be tabulated. This is yet another restriction that needs testing.

5 Numerical Illustrations

For consistency with FM-OLS literature, we repeat the experiments performed by Phillips and Loretan 1991 (hereafter PL91) and later Pedroni 2001, relying on a $MA(1)$ specification for η.

We here report specifically a the results for scenario where size distortions are known to occur, the 'worse-case' scenario n^0 IV of PL91, the data is simulated with constant betas, $\beta_i = \beta_0 = 2$, a individual constant is a uniform random $\alpha_i \rightsquigarrow \mathbb{U}(2,4)$; the feedback term reads:

$\eta_t = \eta_t + \Theta\eta_{t-1}$ and $\eta \rightsquigarrow \mathbb{N}(0,\Omega_\eta)$, with $\Theta = \begin{pmatrix} 0.3 & 0.4 \\ -0.8 & 0.6 \end{pmatrix}$, and $\Omega_\eta = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

The results for 6 000 simulations over various panel sizes are reported on Table 1 on the following page.

The size statistics are shown relative to (the true) $H_0 : \widehat{\beta}^{(\cdot)} = 2$ for a target size of 5%.

In this implementation, we take $\left(\widehat{\overline{\Omega_{vv}}^{-1}}\right) = \left(\widehat{\overline{\Omega_{vv}}}\right)^{-1} \cdot (I + (\overline{\Omega_{vv}})^{-1} \cdot var(\widehat{\Omega}^b_{vv,i})/N)$ as estimator of the inverse of the variance.

This very basic first-order correction for the variability in $\widehat{\Omega^b_{vv,i}}$ improves the estimator when N is small ($N \leq 40$), but one can notice that this correction is imperfect when $N = 10$ or 20.

For P-OLS and FM-OLS, we use the HC3 variance estimator, for it is the best able to capture the covariance between the biased residual and the regressors (as well as the resulting conditional heteroskedasticity). Nevertheless, P-OLS has empirical size close to 1 for all (N,T), and, even if the distortions of FM-OLS size statistics decrease as $(T \to \infty, b \to 0)$, they always increase with N, and are always above 30%.

For $\widehat{\sigma}(\widehat{\beta}^{simple})$, in the current table we are using the simplest possible approximation, with the Newey and West (1987) traditional HAC estimator, multiplied by the following coefficient $\frac{1+2/\sqrt{N}}{1-3/4\cdot b+b^2/4}$, (half the leverage is $3/4 \cdot b + b^2/4$, and $2/\sqrt{N}$ reflects of our small-N finite-sample adjustment to reflect our empirical estimator $\left(\widehat{\overline{\Omega_{vv}}^{-1}}\right)$).

Size distortions occur for small N and T; they only increase in N when T is small (remind that, in theory, $N = o(T)$ is required which may be reflected in the distortions for small T and very large N).[9]

[9] Tests here are oversized, they would be undersized in the absence of endogenous feedback, the only situation in

In table 2, we show the opposite situation where the error term is exogenous and OLS is the linear unbiased estimator. Since no correction for nuisance parameters is needed, our correction is the nuisance. This scenario serves to show that our approximate sized tests can also be undersized (and to what extent is what is likely to represent the worse case scenario for our estimator). They can be around 2%.

Other tables are available on demand.

Table 1: Worse-case scenario of PL91

			P-OLS			FM-OLS			Simple		
T	N	*b=M/T*	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
10	20	24%	0,218	0,227	0,97	0,118	0,134	0,63	0,018	0,084	0,043
	30			0,224	1	0,116	0,128	0,77	0,029	0,072	0,068
	40			0,222	1	0,115	0,123	0,85	0,035	0,065	0,088
	50			0,222	1	0,115	0,122	0,92	0,040	0,063	0,128
	60			0,221	1	0,115	0,121	0,96	0,043	0,061	0,158
	70			0,221	1	0,114	0,119	0,97	0,044	0,060	0,200
30	20	10%	0,080	0,083	0,98	0,028	0,036	0,44	-0,007	0,029	0,037
	30			0,082	1	0,027	0,032	0,54	0,000	0,022	0,039
	40			0,081	1	0,027	0,031	0,63	0,003	0,019	0,049
	50			0,081	1	0,027	0,030	0,70	0,005	0,018	0,073
	60			0,081	1	0,027	0,029	0,76	0,007	0,017	0,088
	70			0,081	1	0,027	0,029	0,82	0,008	0,016	0,109
50	20	7%	0,048	0,051	0,98	0,014	0,019	0,35	-0,008	0,019	0,053
	30			0,050	1	0,014	0,017	0,44	-0,003	0,014	0,044
	40			0,049	1	0,014	0,016	0,53	0,000	0,011	0,049
	50			0,049	1	0,014	0,016	0,60	0,001	0,010	0,060
	60			0,049	1	0,013	0,015	0,65	0,002	0,009	0,068
	70			0,049	1	0,013	0,015	0,71	0,003	0,009	0,084
70	20	5%	0,035	0,037	0,99	0,009	0,013	0,30	-0,008	0,015	0,074
	30			0,036	1	0,009	0,011	0,39	-0,003	0,010	0,059
	40			0,036	1	0,009	0,011	0,45	-0,001	0,008	0,052
	50			0,036	1	0,009	0,01	0,52	-	0,007	0,053
	60			0,036	1	0,009	0,01	0,58	-	0,006	0,056
	70			0,036	1	0,009	0,01	0,62	-	0,006	0,071
90	20	4%	0,027	0,029	0,99	0,006	0,01	0,29	-0,007	0,012	0,095
	30			0,028	1	0,006	0,008	0,36	-0,003	0,008	0,062
	40			0,028	1	0,006	0,008	0,42	-0,002	0,007	0,063
	50			0,028	1	0,006	0,008	0,47	-	0,006	0,057
	60			0,028	1	0,006	0,007	0,52	-	0,005	0,066
	70			0,028	1	0,006	0,007	0,58	-	0,005	0,063

Table 1 has 6 000 simulations. In the worse-case scenario of PL91, the 5% tests have empirical size of virtually 100% for OLS; FM-OLS permits a partial correction of bias and size, and β^{simple} improves size and RMSE.

which POLS and FMOLS hae exact size.

Table 2: Case with exogenous error term

			P-OLS			FM-OLS			Simple		
T	N	b=*M/T*	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
10	20	24%	0	0,057	0,05	0	0,060	0,06	0	0,067	0,008
	30			0,045	0,04		0,048	0,05		0,053	0,010
	40			0,039	0,03		0,041	0,05		0,045	0,010
	50			0,035	0,04		0,037	0,05		0,040	0,013
	60			0,032	0,04		0,034	0,05		0,037	0,014
	70			0,030	0,04		0,032	0,05		0,034	0,019
50	20	7%	0	0,011	0,05	0	0,011	0,06	0	0,012	0,009
	30			0,009	0,05		0,009	0,05		0,010	0,010
	40			0,008	0,05		0,008	0,06		0,008	0,011
	50			0,007	0,05		0,007	0,06		0,007	0,014
	60			0,006	0,05		0,007	0,06		0,007	0,015
	70			0,006	0,05		0,006	0,06		0,006	0,017
90	20	4%	0	0,006	0,05	0	0,006	0,05	0	0,006	0,007
	30			0,005	0,05		0,005	0,05		0,005	0,008
	40			0,004	0,05		0,004	0,05		0,004	0,012
	50			0,004	0,05		0,004	0,05		0,004	0,014
	60			0,004	0,05		0,004	0,06		0,004	0,015
	70			0,003	0,05		0,003	0,05		0,003	0,018

Table 2 has 6 000 simulations, and u is exogenous to X. Then OLS are BLUE, and the use of semi-parametric estimator only results in loss of efficiency and size distortions. In this case, (contrary to when dynamic feedback generates bias), our volatility estimates are too large and the empirical sizes too small. This ‘best-case’ scenario for OLS (and FM-OLS) is arguably a worse-case scenario for our estimator.

6 Conclusion

Semi-parametric estimators are parsimonious thus usually efficient, but when the first-step estimator is biased, they also are biased when the bandwidth is a non-zero fractions of the panel time length T. In *fixed-b* asymptotics, this implies degenerate test statistics for panels, and size distortions that increase with N, possibly nullifying the advantages of gathering panel data.

This paper proposes a pooled panel estimator for heterogenous panels. In the class of semi-parametric estimators we have studied, this is by far the simplest, and within the cointegrated panel model, the only required restriction for the used of the pooled panel estimator is an identical long-term slopes for each of the cross-sections.[10]

Finally, heterogenous cointegration slopes in general require time-series estimators, which necessarily entail significantly more complex asymptotics, and is also left for further research.

[10]Pooled DOLS would no work under these assumptions.

A Appendices

A.1 Assumptions and convergence properties

A.1.1 Assumptions

(A1) We assume that, for all i, each vector η_i is non-deterministic and ergodic with $\mathbb{E}[\eta_{i,t}] = 0$ and $\mathbb{E}[|\eta_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$, satisfies the multivariate functional central limit theorem or multivariate invariance principle, *i.e.*, $T^{-1/2}\sum_{t=1}^{[Tr]} \eta_{i,t} \Rightarrow B_i(r, \Omega_i)$, where $\Rightarrow$ denotes weak convergence and $B_i(r, \Omega_i)$ is a Brownian motion defined over $r \in [0,1]$ and with covariance matrix Ω_i, where Ω_i, of full rank, is the long-term covariance matrix of $\eta_{i,t}$, can be decomposed as $\Omega_i = \Sigma_i + \Lambda_i + \Lambda_i'$ with $\Sigma_i = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \mathbb{E}\eta_{i,t}\eta_{i,t}'$ and with $\Lambda_i = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T}\sum_{\tau=1}^{t-1} \mathbb{E}\eta_{i,t}\eta_{i,\tau}'$.

(A2) We assume convergence of sample covariance matrices used for estimation towards matrix stochastic integrals: $T^{-1}\sum_1^T(\sum_{\tau=1}^t \eta_{i,\tau})\eta_{i,t}' \to \int_0^1 B_i dB_i' + \Delta_i$

(A3) We assume that a central limit theorem can be applied to the asymptotic covariance matrices of the cross-sections and to their estimates. For the unobserved true matrices, we assume that $\widehat{\mathbb{E}}_i\Omega_{(\cdot,i)} \xrightarrow[N\to\infty]{} \overline{\Omega}_{(\cdot)}$, that $\widehat{\mathbb{E}}_i\left[(\Omega_{vv,i})^{-1}\Omega_{vu,i}\right] \xrightarrow[N\to\infty]{} M_{vvu}(<\infty)$, and $\widehat{\mathbb{E}}_i\left[(\Omega_{vv,i})^{-1}\right] \xrightarrow[N\to\infty]{} M_{vv}(<\infty)$.

(A4) The initialization of the system described in **(1)** happens at $t = 0$, without any distributional assumption on $X_{i,o}$ and α_i since these are eliminated my demeaning of the system (by contrast, finite fourth-order moments of α is often assumed when one seeks to make inference on α or the distribution/mean thereof).[11]

A.1.2 Necessary or sufficient conditions

Necessary conditions for ***(A1)*** and ***(A2)*** are discussed with details in Phillips and Durlauf (1986),[12] Park and Phillips (1988), and Phillips and Moon (1999).[13] We note in particular that:

[11]The initial value of the error term $u_{i,1}$ is in general a nuisance term. The assumption that a multivariate functional central limit theorem applies guarantees that we adequately characterise the asymptotic distribution.

[12]Phillips and Durlauf (1986, pp 475-476) clarify that for general processes, $\mathbb{E}[(\sum_{\tau=1}^T \eta_{i,\tau})(\sum_{\tau=1}^T \eta_{i,\tau})'] \to \Omega$, η uniformly integrable, $sup\mathbb{E}[|\eta_{it}|^{2+\delta}] < \infty$, and $\mathbb{E}[T^{-1}(\sum_{\tau=k+1}^{k+T} \eta_{i,\tau})(\sum_{\tau=k+1}^{k+T} \eta_{i,\tau})'] \to \Omega$ are sufficient conditions.

[13]Phillips and Moon (1999) assume that η is generated by a random coefficient linear process $\eta_{i,t} = \sum_{s=0}^{\infty} C_{i,s}V_{i,t-s}$ where C are random matrices and define regularity conditions on C. We leave the definition of needed restrictions over the moments of C and V further research.

- Theorem 2.1 of Phillips and Durlauf (1986, p 475) details sufficient conditions for ***(A1)*** and ***(A2)***; we here summarise those in Phillips (1988) for brevity: (a) $\mathbb{E}[\eta_t] = 0 \quad \forall t$; (b) η has finite moments of order $\beta + \delta$ with $(\beta \geq 2, \delta > 0)$; (c) $\mathbb{E}[\frac{1}{T}(\sum_{\tau=1}^{T} \eta_\tau) \cdot (\sum_{\tau=1}^{T} \eta_\tau)'] \to \Omega > 0$, (d) η is strong mixing with mixing numbers α_m that satisfy: $\sum_1^\infty \alpha_m^{1-2/\beta} < \infty$

- Stationarity simplifies the assumptions: strictly stationary ergodic processes also satisfy these conditions, then essentially requiring that $\mathbb{E}[\eta \eta'] < \infty$;[14] the limit of $\mathbb{E}[\frac{1}{T}(\sum_{\tau=1}^{T} \eta_\tau) \cdot (\sum_{\tau=1}^{T} \eta_\tau)']$ then exists, it is denoted Ω and must be positive definite (see Corrolary 2.2 of Phillips and Durlauf, 1986)

- Martingale difference sequences satisfying $\mathbb{E}[\eta_t|\mathscr{F}_{t-1}] = 0$ finite covariance matrix $\Sigma > 0$, and finite fourth-order moments, where $\mathscr{F}_{t-1}$ is the $\sigma - field$ generated by $\{\eta_{t-\tau}, \tau = 1, 2, \ldots\}$, also satisfy the two properties above (see for example Gregoir, 2005).

- Note that the assumption of finite fourth order moments is often made in the literature (Andrews, 1991, p824), relative to the first formulation of Donsker's theorem; that of summable fourth cumulants is also often made in the automatic bandwidth selection literature (Newey and West, 1994, p 636).

For ***(A3)***, as underlined by Phillips and Moon (1999), 'some moments conditions on $(\Omega_{vv,i})^{-1}\Omega_{vu,i}$ [are needed] to avoid heavy tails in the density of $(\Omega_{vv,i})^{-1}$ ' (p1077). They show that it sufficies that $f(\Omega) = \mathscr{O}(e^{tr(-c\Omega)})$ for some $c > 0$ when $tr(\Omega) \to \infty$ and that $f(\Omega) = \mathscr{O}(det(\Omega)^\gamma)$ for some $\gamma > 7$ when $det(\Omega) \to 0$.

A.1.3 Discussion/interpretation

The set of assumptions ***(A1)*** is made to ensure that OLS is super-consistent,[15] and that the $\mathscr{O}(T^{-1})$ OLS bias has the form defined in Phillips and Hansen (1990).

[14]Under the convenient assumption of second-order stationarity, the invariance principle and notations above for the long-term covariance matrix reduce to $\Sigma_i = \mathbb{E}\eta_i\eta_i'$ and $\Lambda_i = \sum_{k=2}^{\infty} \mathbb{E}[\eta_{i,k}\eta_{i,1}^{'}]$, and η_i has continuous spectral density matrix $f_{\eta_i\eta_i}(\beta)$, with $\Omega_i = 2\pi f_{\eta_i\eta_i}(0)$.

[15]Then we find the well known result that for a single time series, $\hat{\beta} - \beta \Rightarrow \frac{1}{T} \cdot \frac{\Delta_{vu} + \int \widetilde{B}_v dB_u'}{\int \widetilde{B}_v \widetilde{B}_v'}$

The estimation theory of FMOLS also relies on ***(A2)***. These conditions are sufficient to ensure that they apply, replacing B_i by $B_i^{b/2}$

(A3) is made to ensure the convergence of group-mean and/or pooled estimators under heterogeneous feedback. One can simply require $\exists(c_{min}, c_{max})$ such as $det(\Omega_{vv,i}) > c_{min} \forall i$ and $tr(\Omega_{vv,i}) < c_{max} \forall i$.

A.1.4 Practical set-up

In simulations, we will draw $u_{i,1}$ from its stationary distribution. We assume $v_{i,1}$ (and thus $X_{i,0}$) observed, so that in our simulations no observation is lost when exogenising u *w.r.t.* X in FMOLS.

A.2 Functional central limit theorem and related notations

For this section, we have tried to borrow the notations of Vogelsang and Wagner (2014). For each cross-section (subscript i dropped), $\eta_t = [u_t,\ v_t']'$ is assumed to be a vector of $I(0)$ processes that satisfies a functional central limit theorem (FCLT):

$T^{-1/2}\sum_{t=1}^{[rT]}\eta_t \Rightarrow B(r) = \begin{bmatrix} B_u(\text{r}) \\ B_v(\text{r}) \end{bmatrix} = \Omega^{1/2}W(r)\,, r \in [0,1],$ where $[rT]$ denotes the integer part of rT and $W(r)$ is $(k+1)-$ dimensional vector of independent standard Brownian motions.

$$\Omega = \sum_{j=-\infty}^{\infty} \text{E}(\eta_t\eta_{t-j}') = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} > 0$$

Here, $\Omega_{\text{vu}} = \Omega_{uv}'$ and $\Omega_{vv} > 0$ rules out cointegration in X. The long-run covariance matrix reads $\Omega = \Sigma + \Lambda + \Lambda'$, with $\Sigma = \text{E}(\eta_t\eta_t')$

and $\Lambda = \sum_{j=1}^{\infty}\mathbb{E}(\eta_{t-j}\eta_t')$. We also have $\Delta = \Sigma + \Lambda$, and of particular interest is $\Delta_{vu} = \sum_{j=0}^{\infty}\mathbb{E}(v_{(t-j)}u_t')$. We partition $W(r)$ as $W(r) = [W_{u\perp v}(\text{r}),\ W_v'(\text{r})]'$, where $W_{u\perp v}(r)$ and $W_v(r)$ are a scalar and a k-dimensional standard Brownian motion respectively. Denote $\Omega^{1/2}$ the Cholesky decomposition where $\sigma_{u\perp v}^2 = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} = \Omega_{uv}(\Omega_{vv}^{-1/2})'$, *i.e.*, $\Omega^{1/2} = \begin{bmatrix} \sigma_{u\perp v} & \lambda_{uv} \\ 0 & \Omega_{vv}^{1/2} \end{bmatrix}$

Then $B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \begin{bmatrix} \sigma_{u\perp v}W_{u\perp v} + \lambda_{uv}W_v(r) \\ \Omega_{vv}^{1/2}W_v(r) \end{bmatrix}$ is used.

A.3 Proofs

Proof. of Lemma 2.3 on page 7

In addition to the multivariate asymptotic principle, we have required as in Phillips and Durlauf (1986) the weak convergence of the following sample covariance matrices to matrix stochastic integrals (from A2):

$$T^{-1}\sum_{1}^{T}(\sum_{\tau=1}^{t} v_\tau)u'_t \Rightarrow \int_0^1 B_v dB'_u + \Delta_{vu}$$

The very arguments of Phillips and Durlauf (1986, pp253-254) can be repeated by replacing in their notations $X_T(r) = \frac{1}{\sqrt{T}}\sum_{\tau=1}^{[rT]} v_\tau$ by $X_T^{b/2}(r) = \frac{1}{\sqrt{T}}\sum_{l=0}^{min(M,[rT]-1)} k(\frac{l}{M})v_{[rT]-l}$. As they put it, this limit functional Brownian motion is not the simple result of the continuous mapping theorem[16] applied on $X_T(r) \Rightarrow B_v(r)$ (resp., for us, on $X_T^{b/2}(r) \Rightarrow B_v(r)$) and $u_{[rT]} \Rightarrow dB_u(r)$.

The non-trivial element Δ_{vu} in both limiting distributions arises because, when u_t are not Martingale difference sequences, $\mathbb{E}[X_t^{b/2} \cdot u_{t'}] = \mathbb{E}[X_t \cdot u_{t'}] = \mathbb{E}[\frac{1}{T}X' \cdot u] = \Delta_{vu} \neq 0$.

Qualitatively, one understands that this expectation depends on the serial covariance properties of u_t, v_t, where since the covariances are absolutely summable, their value tend to zero beyond a certain truncation M, so that this expectation is valid for $X^{b/2}$ computed with any kernel – in fact, Δ_{vu} is the limit of non-parametric estimator $\widehat{\Delta}_{vu}^b$ when $M/T \xrightarrow{M,T\to\infty} 0$. □

Proof. of theorem 4.1 on page 13

The theorem required first the limit of the numerator. We had on page 13:

$$\begin{aligned}\widehat{\Delta}^b_{v\hat{u},i} - \frac{1}{2}\widehat{\Omega}^b_{v\hat{u},i} \Rightarrow \Delta_{vu,i} - \frac{1}{2}\Omega_{vu,i} & \\ + \frac{1}{2}\int\left(B_{v,i}^{b/2} - B_{v,i}^{-b/2}\right)\cdot dB'_{\tilde{u},i} & \quad (9)\\ - \frac{1}{2}\int\left(B_{v,i}^{b/2} - B_{v,i}^{-b/2}\right)\cdot \widetilde{B}'_{v,i}\cdot\Theta_\beta & \quad (10)\end{aligned}$$

(9) is a mean-zero process by construction.[17]for independent panels).

[16] After all, the sample covariance matrix is not a continuous functional of $X_t^{b/2}$; in addition $\int B_v^{b/2}\cdot dB'_u$ is a matrix stochastic integral; $B(r)$ is a Brownian motion, almost surely of unbounded variations, so $\int BdB'$ is not the mean square limit of a Riemann Stieltjes sum.

[17] It is easy to see that if v is one-dimensional, $\left(B_{v,i}^{b/2}(r) - B_{v,i}^{-b/2}(r)\right)\cdot dB'_{v,i}(r) \equiv 0$ by symmetry: $v_t \cdot v'_u = v_u \cdot v'_t =$

To show in (10) that $\mathbb{E}\left[\frac{1}{2}\cdot\int\left(B_{v,i}^{b/2}-B_{v,i}^{-b/2}\right)\cdot\widetilde{B}_{v,i}'\right]=\frac{1}{4}(b-b^2/3)\Omega_{vv,i}$, consider:

$$
\begin{aligned}
\mathbb{E}[\int B_v^{b/2}\cdot\widetilde{B}_v'] &= \mathbb{E}[\int B_v^{b/2}\cdot B_v'] - \mathbb{E}[\int B_v^{b/2}\cdot\int B_v'] && \\
\mathbb{E}[\int B_v^{b/2}\cdot\int B_v'] &= \mathbb{E}[\int B_v^{-b/2}\cdot\int B_v'] && \text{by symmetry, so} \\
\mathbb{E}[\int\left(B_{v,i}^{b/2}-B_{v,i}^{-b/2}\right)\cdot\widetilde{B}_{v,i}'] &= \mathbb{E}[\int\left(B_{v,i}^{b/2}-B_{v,i}^{-b/2}\right)\cdot B_{v,i}'] && \text{follows} \\
\mathbb{E}[\int B_v^{-b/2}\cdot B_v'] &= 0 && \text{by construction, and} \\
\mathbb{E}[\int B_v^{b/2}\cdot B_v'] &= \frac{1}{2}(b-b^2/3)\Omega_{vv} && \textit{(then divide by 2)}
\end{aligned}
$$

Then $\mathbb{E}\left[\widehat{\Delta}_{v\hat{u},i}^b-\frac{1}{2}\widehat{\Omega}_{v\hat{u},i}^b|X_i\right]\Rightarrow\Delta_{vu,i}-\frac{1}{2}\Omega_{vu,i}-\frac{1}{4}(b-b^2/3)\cdot\Omega_{vv,i}\cdot\Theta_\beta$

So $\frac{1}{N}\sum_{i=1}^N\left(\widehat{\Delta}_{v\hat{u},i}^b-\frac{1}{2}\widehat{\Omega}_{v\hat{u},i}^b\right)\Rightarrow\overline{\Delta}_{vu}-\frac{1}{2}\overline{\Omega}_{vu}-\frac{1}{4}(b-b^2/3)\cdot\overline{\Omega}_{vv}\cdot\Theta_\beta$

As a consequence, given that $\widehat{\Omega}_{vv,i}^b$ is unbiased,[18] and given the convergence of the mean of sample matrices $\frac{1}{N}\sum_{i=1}^N\widehat{\Omega}_{vv,i}^b\Rightarrow\overline{\Omega}_{vv}$, and $\Theta_\beta\Rightarrow\mathbb{E}[\Theta_\beta]=\frac{\overline{\Delta}_{vu}-\frac{1}{2}\overline{\Omega}_{vu}}{\frac{1}{6}\overline{\Omega}_{vv}}$ by assumption **(A3)**, we have

$$
\mathbb{E}\left[\left(\frac{1}{6\cdot N}\sum_{i=1}^N\widehat{\Omega}_{vv,i}^b\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N\left(\widehat{\Delta}_{v\hat{u},i}^b-\frac{1}{2}\widehat{\Omega}_{v\hat{u},i}^b\right)\right)\right]\Rightarrow\Theta_\beta\cdot\left(1-\frac{3}{2}(b-b^2/3)\right)
$$

□

Remark A.1. Heterogenous CS and CS dependence can easily be tested independently. For instance, with CS *in*dependence but heterogeneous CS, the empirical variance of $\widehat{\Omega}_{vv,i}^b$ (or $\widehat{\Omega}_{v\hat{u},i}^b$ at the bias-corrected slope) will be greater than $4/3b\cdot\Omega$; with homogeneous but dependent cross-sections, the reverse happens. The joint hypothesis is thus not trivial to test (this is left for further research).

$v_u\cdot v_t$ for scalars. So, (9)= $\left(B_{v,i}^{b/2}-B_{v,i}^{-b/2}\right)\cdot dB_{\tilde{u}\perp v,i}'$ is a mean-zero normal.

[18] $\widehat{\Omega}_{vv,i}^b$ is unbiased when $X=\sum v$ has no trend a correction would be needed if X is I(1) with a deterministic trend

References

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica : Journal of the Econometric Society 59*(3), 817–858.

Andrews, D. and C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* : *Journal of the Econometric Society 60*(4), 953–966.

Gregoir, S. (2005). Representation and statistical analysis of weakly linearly exchangeable dynamic panels. Technical report, Working Paper, CREST.

Hansen, B. E. and P. C. Phillips (1990). Estimation and inference in models of cointegration: A simulation study. *Advances in Econometrics 8*, 225–248.

Kiefer, N. and T. Vogelsang (2005, 12). A new first-order asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(06), 1130–1164.

Kiefer, N., T. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Newey, W. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics 5*(2), 99–135.

Newey, W. and K. West (1987, May). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–08.

Newey, W. and K. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies 61*(4), 631–653.

Park, J. and P. C. Phillips (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory 4*(03), 468–497.

Pedroni, P. (1996). Fully modified ols for heterogeneous cointegrated panels and the case of purchasing power parity (working paper in economics no. 96-020). *Bloomongton: Indiana University*.

Pedroni, P. (2001). Fully modified OLS for heterogeneous cointegrated panels. *Advances in econometrics 15*, 93–130.

Phillips, P. and M. Loretan (1991). Estimating long-run economic equilibria. *The Review of Economic Studies 58*(3), 407–436.

Phillips, P. C. (1988). Weak convergence to the matrix stochastic integral? $\int_0^1 bdb'$. *Journal of Multivariate Analysis 24*(2), 252–264.

Phillips, P. C. and S. Durlauf (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies 53*(4), 473–495.

Phillips, P. C. and B. Hansen (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies 57*(1), 99–125.

Phillips, P. C. and H. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*(5), 1057–1111.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory 7*(1), 1–21.

Stock, J. and M. Watson (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica: Journal of the Econometric Society 61*(4), 783–820.

Vogelsang, T. and M. Wagner (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics 178*(2), 741–760.