An exact finite-sample correction to fully-modified estimators of cointegrated panels.

Samuel Sender*

Working Paper, 2014

*Researcher, Tilburg University, CentER, and Netspar. Email: s.j.sender@uvt.nl
I thank Stéphane Grégoir for extremely useful feedback and advice.

An exact finite-sample correction to fully-modified estimators of cointegrated panels.

Abstract

Though ordinary least square (OLS) estimates are super-consistent with cointegrated variables, their finite-T bias can be large in the presence of endogenous feedback. Fully modified OLS (FMOLS) are parsimonious tools to measure the cointegrating [long-run] slope between integrated variables in the presence of endogenous feedback, and correct the first-order OLS bias to the extent necessary to provide a nuisance parameter-free asymptotic distribution.

Yet because FMOLS rely on a first-step OLS estimator that is biased, and has weak power and size, FMOLS also has poor finite-T properties. I show that FMOLS asymptotically leave an $\mathbb{O}(h/T)$ fraction of the OLS bias, where h is the selected bandwidth.

I also propose an improved estimator, which corrects the first-order conditional bias and removes the residual $\mathbb{O}(h/T)$ FMOLS bias.

I establish the maximal speed at which N can grow simultaneously to T for the tests statistics of panel group-mean of time-series estimators to be (asymptotically) nuisance parameter-free. My improved estimator permits analysing wider panels.

In the scenarios reviewed by previous research, my finite-T FMOLS has lower bias and RMSE than the 'asymptotic' FMOLS for all values of T and N.

Keywords: Panel cointegration, Fully modified, Dynamic Least Squares, FMOLS, DOLS, DGLS, IM-OLS, stationarity tests.

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

1 Introduction

1.1 Motivation

Though time-series econometric theory nested in a T-asymptotic framework gives elegant closed-form equations and teachable intuitions, it often gives little insight on its application in realistic settings and in particular when T is finite and in the presence of endogeneity of an unknown form.

Though the finite-T properties of mainly time-series estimators have been documented by means of simulations, estimators that are designed within a finite-T framework are mainly restricted to parametric models and panel data. Bao and Ullah (2007) showed the finite-sample properties of estimators can be improved by identifying and removing the 'second-order [asymptotic] bias', because this bias may not be negligible in finite samples.

Although these results do not apply to two step-estimators or general models with endogenous bias of an unknown form.

Vogelsang and Wagner (2014), in a related approach, describe the behaviour of non-parametric estimators when $T \to \infty$ and $b = h/T$ is kept fixed. The *fixed-b* theory enables one to assess capture some aspect of the relevant empirical situation when T is given (and finite) and h is non zero, and to understand that the second-order bias is not entirely removed by FMOLS. This approach however does not permit further improvement to the FMOLS estimator.

This paper lays a stone to the finite-sample improvement of consistent estimators in the presence of endogeneity of an unknown form.

As in Bao and Ullah (2007), we correct for an asymptotically vanishing bias that matters in finite samples. Our proposed approach is straightforward since it relies on the correction of the 'first-step' bias of two-step estimators.

1.2 Related literature

This research relates to and use results from the following strands of literature: the asymptotic time-series (and panel) cointegration literature in the presence of endogenous feedback, with Park and Phillips (1988), Phillips and Hansen (1990), Saikkonen (1991), Stock and Watson (1993), Pesaran and Smith (1995), Kao and Chiang (2001), Pedroni (2001), Mark and Sul (2003), Moon and Perron (2004); spectral estimators and automatic bandwidth selection, with Andrews (1991), Andrews and Monahan (1992), as well as Newey and West (1994); *fixed-b* asymptotics with Vogelsang and Wagner (2014); small-sample inefficiency with Banerjee et al. (1986), and finite-T bias with Kiviet and Phillips (1993), Bao and Ullah (2007).

We study a univariate integrated model $\mathbf{Y_{i,t}} = \alpha_\mathbf{i} + \mathbf{X_{i,t}}\beta + \mathbf{u_{i,t}}$ (M1) where $\beta \neq 0$ is a scalar, Y_i and X_i are unidimensional cointegrated variables, $X_{i,t} = \sum_{j=1}^{t} \varepsilon_{i,j}$, and ξ_i the vector that stacks u_i and ε_i satisfies the multivariate invariance principle [functional central limit theorem], and has continuous spectral density. $(t = 1, \ldots, T)$ indexes time observations and $(i = 1, \ldots, N)$ indexes individuals.

Denote $b^{OLS}_{(T)}$the OLS bias for a given DGP is a function of T. The 'super-consistency' of OLS with integrated variables relies on the fact that in the bias $b^{OLS}_{(T)} = \mathbb{E}[\frac{X_i' u_i}{X_i' X_i}]$, the denumerator $X_i' X_i = \mathbb{O}(T^2)$ while $X_i' u_i = \mathbb{O}(T)$, so $b^{OLS}_{(T)} = \mathbb{O}(1/T)$. Test statistics are not nuisance-parameter free (the t-statistics for $H_0 : \widehat{\beta_i^{OLS}} = \beta$ is not a standard Gaussian) firstly because the volatility of $\widehat{\beta_i^{OLS}}$ also is $\mathbb{O}(1/T)$.

This property of differing levels of integration between regressor and nuisance term has firstly been used to show that OLS is 'super-consistent', then with Pedroni and Philipps to build nuisance parameter-free test statistics: fully modified ordinary least squares (FMOLS), formalised by Phillips and Hansen (1990) for time series, were designed as a parsimonious[1] answer to this issue of testing. I use this property one step further to correct for the first-order conditional bias of FMOLS.

FMOLS achieves nuisance parameter-free test statistics by applying non-parametric correction to a parametric first-step OLS. The correction involves exogeneising the regressand Y on the regressor X, and (sufficiently) correcting for the first-order OLS bias by means of a spectral estimate of the long-term covariance between the error term u and the increments $\Delta X = \varepsilon$.

FMOLS estimators are asymptotically efficient.[2] However FMOLS relies on a first-step estimator that has weak power and size, and Pedroni (2001) shows that this leaves the econometrician

[1]FMOLS only estimates one slope coefficient and the long-run covariance matrix of ξ_i. Parametric models, by contrast, require modelling the full dependence structure. Dynamic ordinary least squares (DOLS) estimate each of the slopes of Y on a number $h \to \infty$ of lead/lag values of X_i. In finite sample, DOLS truncates the sample, a potential source of significant volatility when T is small.

[2]Though the notion of efficiency for the slope of integrated processes is not standard, this vocabulary is used by both Phillips and Saikkonen, see for instance the title of Saikkonen (1991).

with similarly poor finite-T properties.

Indeed, the asymptotic FMOLS correcttion only ensures that the bias of the FMOLS slope vanishes at an $\mathbb{o}(1/T)$ rate, against $\mathbb{O}(1/T)$ for OLS. The question of how to increase the speed of convergence [the speed at which the bias vanishes] appears crucial because of poor finite-sample properties of OLS and FMOLS.

We show that the conditional first-step OLS bias indeed impacts the FMOLS estimator, and that this conditional bias present in the FMOLS calculations can be entirely removed, thus improving not only the asymptotic speed of convergence and the finite-sample bias (*in all likelihood*, *i.e.*, expect in the implausible situation where the conditional and unconditional FMOLS cancel out for a given process, bandwidth and value of T). The volatility of the correction is also controlled to ensure good finite-sample properties. Empirically, our conditionally unbiased estimator bring great improvement over FMOLS when T is small.

More generally, strictly speaking, with any two-step estimator, as long as the first-step bias is not evaluated (and corrected), the ability to apply second-order corrections to the estimators is limited.

1.3 Contributions

The first contribution is to clarify that FMOLS asymptotically leaves an $\mathbb{O}(h/T)$ fraction of the $\mathbb{O}(1/T)$ OLS bias uncorrected, where h is the selected bandwidth and the OLS bias for a given DGP is a function of T denoted $b^{OLS}_{(T)}$.

Under the standard FMOLS assumptions, I explicitly introduce the bias in the OLS residuals conditional on X and demonstrate how this conditional bias impacts the FMOLS estimators; I build finite-T FMOLS estimators that are free of conditional bias, by means of a non-parametric bandwidth-related correction similar to a correction for the degrees of freedom in parametric estimates.

I fully correct for the impact of the first-step conditional OLS bias on the FMOLS calculations, but do not correct for the unconditional bias. I improve the asymptotic speed of convergence, this without nuisance term associated to new projections, thus diminishing the finite-T bias *in all* *likelihood*[4]

In addition, we control the volatility brought by the estimation of the needed correction.

Since panel FMOLS makes use of time series estimators with panel data, I analyse the joint asymptotic properties of panel group-mean estimators when $N = d(T)$ and $T \to \infty$, and establish the maximal speed at which N can grow simultaneously to T for the for the tests statistics to be asymptotically nuisance parameter-free. The improvement in the speed of convergence rate of my FT [FT the conditionally unbiased finite-T version of FMOLS] estimator enables one to use it for wider panels than existing time-series estimators.

1.4 Organisation of the paper

This paper is organised as follows: Section 2 gives stylised results and proofs. Section 3 details the model and notations, Section 4 develops finite-T versions of FMOLS 'time-series' estimators. Section 5 rationalises the results of earlier literature, and shows the benefits of the methodologies developed in this paper by means of additional simulations.

Appendix A.1 lists the proofs. Appendix A.2 summarises modern approaches to cointegration. It details the steps of FMOLS, an estimator more complex than that presented in Section 2 below. It also reviews alternative estimators such as DOLS and IM-OLS. Alternative estimators have greater RMSE that FMOLS and their results are not reported in the tables of Section 5. A.3 sets out simple cases to refine intuitions about the bias correction.

2 Illustration of the theory with a toy model

Toy Model assumptions and notations.

Consider $\mathbf{Y_{i,t}} = \mathbf{X_{i,t}}\beta_\mathbf{i} + \mathbf{u_{i,t}}$ (M2) a Toy version of the univariate cointegration model **(M1)**. Denote $X_{i,t} = \sum_{j=1}^{t} \varepsilon_{i,j}$ denoted $X_i = S\varepsilon_i$ in vector form. The main Toy model assumptions is that ε_i is *i.i.d.* standard normal and that $X_{i,0} = 0$ (or that it is an observed constant). We then write in vector form $u_i = \sum_{k=0}^{T-1} \psi_k L^k \varepsilon + v_i$ with v_i *i.i.d.* standard normal and $v_i \perp \varepsilon_i$.

[4] Again, *in all likelihood* because the unconditional bias in FMOLS itself could by chance compensate the conditional bias in some situations.

This Toy model does no aim to restrict the theory but simply to provide a sketch of the proof, notably because, denoting L the matrix lag operator, $\mathbb{E}[X_i' L^k \varepsilon_i] = T - k$.

Noting that for the i$^{\text{ieth}}$ panel, OLS estimates $\widehat{\beta_i^{OLS}} = (X_i'X_i)^{-1}X_i'Y_i = \beta_i + (X_i'X_i)^{-1}X_i'u_i$, the i$^{\text{ieth}}$ OLS bias is the scalar $b_i^{OLS} = \mathbb{E}\left[\frac{X_i'u_i}{X_i'X_i}\right]$.

By construction: $\hat{u}_i = u_i - (\widehat{\beta_i^{OLS}} - \beta_i) \cdot X_i$ (1), and in two-step estimators, it is important to note that relying on the residual of a biased first-step regression leads to biased estimates. We denote by $\widehat{\widehat{(\cdot)}}$ estimators based on the residuals from the first-step OLS regression, and by $\widehat{(\cdot)}$ estimators directly based on observed variables.

Simple FMOLS estimator.

The simple FMOLS estimator used in Phillips and Loretan (1991) involves estimating the i$^{\text{ieth}}$ OLS bias as $\widehat{\widehat{b_i^{FM}}} = \frac{\hat{\hat{\Omega}}^{(h)}_{21,i}}{\frac{1}{T}X_i'X_i}$ where $\hat{\hat{\Omega}}^{(h)}_{21,i}$ denotes a h-bandwidth kernel estimator of $\frac{1}{T}X_i'\hat{u}_i$, then computing the group-mean FM slope $\hat{\hat{\beta}}^{GM,FM} = \mathbb{E}_i[\widehat{\beta_i^{OLS}} - \widehat{\widehat{b_i^{FM}}}]$.

FMOLS unconditional bias.

I call the *unconditional bias* of FMOLS the bias that would arise if the FMOLS technique was applied on the unobserved u. Denoting $\widehat{\Omega}^{(h)}_{21,i}$ an unfeasible estimator of $\frac{1}{T}X_i'u_i$ (since u_i is not observed), in the Toy model, the *unconditional measurement error* is $\widehat{\Omega}^{(h)}_{21,i} - \Omega_{21,i} = \frac{1}{T}\sum_{k=0}^{T-1}(1 - w(k))X_i'L^k\varepsilon_i$ and would only be nil in expectation if $w = 1$ and $h = T - 1$ (or h sufficient to capture all non-nil covariances).

Generally, the unconditional bias in FMOLS arises in finite-T samples from three factors:

- The structure of kernel weights
- The choice of the bandwidth
- The fact that FMOLS estimates separately the numerator and the denumerator of a stochastic ratio rather than directly the ratio.[6]

From the property of spectral estimators, the *unconditional* bias of the *unfeasible* estimator is $\mathbb{E}[\widehat{b}^{FM}|X] - b^{OLS}(X) = \wp(T^{-2})$

[6] For instance, FMOLS estimates $\hat{\Omega}_{22,i}$ then its inverse rather than directly its inverse, so it relied on the unfeasible $\frac{\hat{\Omega}_{21,i}}{\hat{\Omega}_{22,i}}$ rather than on $\widehat{\Omega_{21,i}/\Omega_{22,i}}$, yielding an $\wp(1/T^2)$ approximation. Of course, for a single time-series, not only is this approximation is of the same order of magnitude than the unconditional bias that results from the two first factors, *i.e.*, from the kernel estimation of $\hat{\Omega}_i$.

Draft 2014 v1

FMOLS conditional bias.

Our technique does not correct the *unconditional bias* but only the *conditional bias* of FMOLS linked to the reliance on biased residuals from a first-step OLS. After all, with endogeneity of an unknown form in a single time-series, spectral estimators would permit *optimal inference* (in the vocabulary of Phillips) if based on the unobserved $\xi_i = (u_i', \varepsilon_i')'$.

The *conditional measurement error* of $\widehat{\widehat{\Omega}}_{21}$ is $\widehat{\widehat{\Omega}}_{21}^{(h)} - \widehat{\Omega}_{21}^{(h)} = -(\hat{\beta}^{OLS} - \beta)\frac{1}{T}\sum_{k=0}^{h} w(k)X'L^k\varepsilon$,

Since, $\widehat{\widehat{b_i^{FM}}} = \frac{\widehat{\widehat{\Omega}}_{21,i}^{(h)}}{\frac{1}{T}X_i'X_i}$ (3), we define $\hat{e}(h,X) = \sum_{k=0}^{h} w(k)X_i'L^k\varepsilon_i/X_i'X_i$ which can be interpreted as the fraction of the observations used in kernel estimates.

The main equation of interest to analyse the properties of the FM estimator becomes

or $\widehat{\widehat{b_i^{FM}}} = \widehat{b}_i^{FM} + (\hat{\beta}_i^{OLS} - \beta)\cdot\hat{e}(h,X))$ (5)

Nuisance terms in FMOLS.

The FMOLS estimator involves very complex nuisance terms when T is finite. Of course, it depends on the bandwidth and of the form of the kernel via the function $\hat{e}(h,X)$, in a way that is asymptotically described by the *fixed-b* theory exposed in Hashimzade and Vogelsang (2008) and Vogelsang and Wagner (2014).

When T is finite, however, another source of complexity and nuisance associated with fourth order moments in ε_i arises because the conditional measurement error in $\widehat{\widehat{\Omega}}_{21,i}^{(h)}$ reads, when the variance of ε is unknown, $\underbrace{\left[\frac{X_i'\Psi\varepsilon_i}{X_i'X_i}\cdot\sum w(k)\cdot\frac{X_i'L^k\varepsilon_i}{\varepsilon_i'(L^k)'L^k\varepsilon_i}\right]}_{4^{th}-order\ \ terms} + \underbrace{\left[\frac{X_i'v_i}{X_i'X_i}\cdot\sum w(k)\cdot\frac{X_i'L^k\varepsilon_i}{\varepsilon_i'(L^k)'L^k\varepsilon_i}\right]}_{mean\text{-}zero\ as\ v\perp X}$.

Impact on the speed of convergence.

One sees easily that $\mathbb{E}[\hat{e}(h,X)] = \mathbb{O}(h/T)$. For instance, if a flat/truncated kernel is used, in the Toy model, $\mathbb{E}[\hat{e}(h,X)] = \frac{\sum_{k=0}^{h}(T-k)}{T(T+1)/2} + o(T^{-2}) = \frac{\sum_{k=0}^{h}(T-k)}{T(T+1)/2} + o(T^{-2}) = \frac{(h+1)(2T-h)}{T(T+1)} + o(T^{-2})$ so $\mathbb{E}[\hat{e}(h,X)] = \mathbb{O}(h/T)$ and $\frac{h+1}{T} < \mathbb{E}[\hat{e}(h,X)] < 2\frac{h+1}{T}$ (for $h < T-1$).

The residual bias from $\widehat{\widehat{b_i^{FM}}}$ is $\mathbb{E}[\widehat{\widehat{b_i^{FM}}} - b_i^{OLS}] = \mathbb{E}[\widehat{\widehat{b_i^{FM}}} - \widehat{b_i^{FM}}] + \mathbb{E}[\widehat{b_i^{FM}} - b_i^{OLS}] = \mathbb{E}[(\widehat{\beta_i^{OLS}} - \beta_i)\cdot\hat{e}(h,X)] + o(T^{-2}) = \mathbb{O}(\frac{h}{T})b^{OLS} + o(T^{-2})$. Since $b^{OLS} = \mathbb{O}(1/T)$, the residual bias is $\mathbb{E}[\widehat{\widehat{b_i^{FM}}} - b_i^{OLS}] = \mathbb{O}(\frac{h}{T^2})$. When a Bartlett kernel is used and $h = o(T^{2/9})$ then $b^{FM} = o(T^{-16/9})$.

Correction of the conditional residual bias in FMOLS.

Define $\widehat{df_{corr}}(h,X) = (1-\hat{e}(h,X))^{-1}$ to remind its interpretation as a correction for the number of degrees of freedom and note that since $\hat{e}(h,X) = \mathbb{O}(h/T)$, with spectral estimators (which imply that $h/T \to 0$), $\widehat{df_{corr}} \to^{\mathbb{P}} 1$. Then $\widehat{\widehat{b_i^{FT}}} = \widehat{df_{corr}}(h,X) \cdot \widehat{\widehat{b_i^{FM}}}$ denotes a conditionally unbiased estimator of b^{OLS}. Since the unfeasible estimator $\mathbb{E}[\widehat{b}^{FM}|X] = b^{OLS}(X) + \mathscr{o}(T^{-2})$, we have that $\mathbb{E}[\widehat{\widehat{b_i^{FT}}}|X_i] = b_i^{OLS}(X_i) + \mathscr{o}(T^{-2})$ is *(asymptotically)* unbiased conditionally to X_i to a T^{-2} order.

RMSE and regularity conditions of conditionally unbiased estimators.

Though the set under which $1-\hat{e}(h,X)$ is arbitrarily close to 0 is asymptotically of zero measure, because the finite-T behaviour of $\hat{e}$ is unknown, we work with a modified estimator $\widehat{df_{corr}}(h,X)$ that is bounded in order to ensure that the conditional volatility $\sigma(\widehat{\beta}^{FT}|\widehat{df_{corr}}) = \widehat{df_{corr}} \cdot \sigma(\widehat{\beta}^{FM})$ is bounded for any finite-T.

We use $\widehat{df_{corr}}(h) = 10$ when $|\hat{e}(h_i,X_i) - 1| < 10^{-2}$, and $\frac{1}{1-\hat{e}(h,X)}$ otherwise. Such control appears necessary for $T = 10$ as the probability that $\hat{e}$ is in the neighbourhood of 1 is not necessarily negligible. Results however are not very sensitive to the truncation point, chosen in our simulations because it seems already lax to allow multiplying the volatility by 10.

Conclusion for a single time-series.

Strictly speaking, the FMOLS calculations are unequivocally improved by the correction proposed here, but one can still imagine a situation where this deteriorates the performance of FMOLS, for instance if for a specific DGP and length of the time-series, the conditional and unconditional bias cancel out.

Divergence rate of N relative to T for panel FM group-means.

The test statistic of interest, also the focus of Pedroni (2001), is the distribution of the t-statistic under the null that the estimated cointegrating slope is the true slope: $\widehat{\beta}^{(GM,\cdot)} = \beta_0$ where $(GM,\cdot)$ denotes the group-mean of any estimator studied, under the assumption of *i.i.d.* cross-sections.

Having characterised the residual bias and variance of FM and FT estimates makes it possible to study at which maximal speed N can grow simultaneously to T for the tests statistics to remain asymptotically nuisance parameter-free.

Since $t^{GM,(\cdot)} \equiv \sqrt{N} \cdot \mathbb{E}_i[t_i^{(\cdot)}]$, where $\mathbb{E}_i$ denotes a (panel) group-mean estimate, a nuisance parameter-free statistic requires $\mathbb{E}[t^{GM,(\cdot)}] \xrightarrow[(N,T)\to\infty]{} {}^{\mathbb{P}}\, 0$. Since the volatility of all time-series estimator considered are $\mathbb{O}(T^{-1})$ and consistently estimated, and since the group-mean t-statistic of *i.i.d.* variables is asymptotically normal, the necessary and sufficient conditions for *consistent* test statistics (that have asymptotically the adequate size under the null) are:

- Panel OLS test statistics are degenerate ($\forall T, \forall \alpha$, the nominal size tends to 1 if $N \to \infty$).
- FMOLS group-mean test statistics are consistent *iff* applied to panels when $N = \mathbb{O}(T^{14/9})$.
- FT group-mean test statistics are consistent *iff* applied to panels where $N = \mathbb{O}(T^2)$.

Since $\hat{e}_i(h, X_i)$ is bounded, the same results are obtained with heterogenous but independently distributed cross-sections, provided that a central limit theorem can be applied on the bias and its estimate.

3 Model and notations

3.1 Model

We study a univariate model $\mathbf{Y_{i,t} = \alpha_i + X_{i,t}\beta_i + u_{i,t}(M1)}$ where $u_{i,t}$ is second-order stationary, and $\beta_i \neq 0$ is a scalar, Y_i and X_i are unidimensional cointegrated variables. In $(t = 1, \dots, T)$, $(i = 1, \dots, N)$, t indexes time observations and i indexes individuals.[7] In theory, individual β_i can differ in the cross-section as long as a central limit theorem on the β_i permit a consistent definition of the group-mean, since the OLS and FMOLS bias of each time-series is estimated independently. In simulations, however, $\beta_i = \beta_0 \forall i$.

We then study the convergence of the group-mean of time-series estimators under the assumption that in a cross-section of N *independent* individuals indexed by i, X_i and Y_i are $I(1)$ and cointegrated, with an identical cointegrating slope,[8] while the constant as well as the form of endogenous 'feedback' and of the endogeneity bias may vary in independent cross-section as long as regularity conditions ensure that a central limit theorem can be applied to the bias and to its time-series

[7]Stacking Y, X, u and ε (defined according to their DGP) in $N \cdot T$ vectors, α and β as a N-vector, the equation of interest is $Y_{1:T\cdot N} = (I_N \otimes \iota_T) \cdot \alpha_{1:N} + X_{[1:T\cdot N, 1:T\cdot N]} \cdot ((I_N \otimes \iota_T)\beta_{1:N}) + u_{1:T\cdot N}$. For our group-mean estimators, in the absence of a common factor to panels, we do not use Kronecker notations but rather use an index i for each cross-section, and omit this index when not necessary. For pooled estimators and a common exogenous Z, Kronecker notations are the most straightforward: $Y_{1:T\cdot N} = (I_N \otimes \iota_T) \cdot \alpha_{1:N} + X_{1:T\cdot N} \cdot \beta + (I_N \otimes Z_{1:T}) \cdot \delta + u_{1:T\cdot N}$.

[8]Heterogeneous β_i are possible provided that a central limit thorem can be applied to the group-mean of β_i and that some hypothesis can be formed on the value of that group-mean. following Phillips and Moon (1999), we could argue that panels with 'near-homogenous' slopes have the same property.

estimates.[9]

Note that for any cointegrated pair that has either a VAR or VECM representations, **(M1)** can be deducted; the causality from the error term $u_{i,t-1}$ and past increments in $\Delta Y_{i,t-k}$ to $\Delta X_{i,t} = \varepsilon_{i,t}$ involves in **(M1)** a covariance between $u_{i,t}$ and $\varepsilon_{i,t+k}$ with $k > 0$, and vice-versa. The representation **(M1)** leaves unspecified the dependence structure, and is thus more natural for parsimonious semi-parametric methods such as FMOLS.[10]

The $\mathbf{Y_{i,t}} = \alpha_\mathbf{i} + \mathbf{X_{i,t}}\beta_\mathbf{i} + \mathbf{u_{i,t}}$ **(M1)** cointegration models cover the case of additional exogenous regressors, identified deterministic trends and observable common factors, a multivariate X, but restricts the theory to identified cointegration relationships where Y belongs to the cointegration space (with a non-nil coefficient) and where no cointegration relationship exists between I(1) regressors. This restriction is needed to ensure that the first-step OLS regression is superconsistent. Multiple cointegration (TS or CS) is in general not possible in such a framework, and I also work without unobserved factors.

The model however extends straightforwardly to a uniquely multivariate cointegration relationship with additional deterministic trends and exogenous variables (see Section A.3.1 on page 28).

3.2 Notations

L denotes the matrix lag operator (one-band matrix, 1st off-diag =1), and $L_k = L^k(\text{ if } k \geq 0)$, $L_k = L^{|k|'}(\text{ if } k < 0)$ is a generic lead/lag operator.

$\mathbb{E}_i = \underset{i \in 1:N}{\mathbb{E}}$ denotes the cross-section (denoted CS) average for $i \in [1:N]$.

Matrix notations: M denotes the demeaning matrix, L_k the generic k-order lead ($k > 0$)/lag ($k < 0$) operator, S the partial sum matrix, and tr the trace operator.

Estimators: $\widehat{(\cdot)}$ denotes an estimator based on observed data (or an unfeasible estimator if based

[9] Note that the absence of common factor between the cross-section is needed for convergence, but that this is a natural assumption to the modeler who aims to include common exogenous factors in the regression, see Gregoir (2005). We also note that the assumption of CS homogeneity is often used for expository purposes when N is fixed and $T \to \infty$, see for instance Pedroni (2001).

[10] FMOLS is a two-stage estimator involving an OLS regression in the first step and a non-parametric estimate of the bias in the second step. In this paper, we thus use the term non-parametric to denote the use of kernel/spectral estimators to identify the bias, without formally identifying the data-generating process nor estimating the slopes associated to each lead/lag terms of the VECM representation.

on the unobserved u), $\widehat{\widehat{(\cdot)}}$ an estimator based on the residuals from the first-step OLS regression, $\widetilde{(\cdot)}$ a demeaned variable, *e.g.* $\widetilde{X} = MX$.

Thus $\Omega_{21,i} = \mathbb{E}X_i'Mu_i$ is accompanied by $\widehat{\widehat{\Omega}}_{21,i}^{(h)}$ a spectral estimator at a h bandwidth, but based on first-step OLS residuals, while $\widehat{\Omega}_{21,i}^{(h)}$ denotes an unfeasible h-bandwidth kernel estimator of $\Omega_{21,i}$ that would be obtained if based directly on the unobserved u_i (or $M \cdot u_i$).

b represents a bias, thus b^{OLS} is the OLS bias, and b^{FM} the bias from FMOLS estimators. $\widehat{\widehat{b^{FM}}}$ is by contrast is the FMOLS estimator of the OLS bias.

Subscripts i are often dropped in the notations, and, for ease of interpretation, the results are displayed with the Bartlett kernel rather than for a general kernel.

3.3 Assumptions and convergence properties

(A1) We assume that, for all i, each vector ξ_i is non-deterministic and ergodic with $\mathbb{E}[\xi_{i,t}] = 0$ and $\mathbb{E}[|\xi_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$, and satisfies the multivariate functional central limit theorem or multivariate invariance principle, *i.e.*, $T^{-1/2}\sum_{t=1}^{[Tr]} \xi_{i,t} \Rightarrow B_i(r,\Omega_i)$, 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 ξ_i, can be decomposed as $\Omega_i = \Sigma_i + \Gamma_i + \Gamma_i'$ with $\Sigma_i = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \mathbb{E}\xi_{i,t}\xi_{i,t}'$ and with $\Gamma_i = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T}\sum_{\tau=2}^{t-1} \mathbb{E}\xi_{i,t}\xi_{i,\tau}'$.

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

(A2) The estimation theory of FMOLS also relies on the convergence of sample covariance matrices used for estimation must towards matrix stochastic integrals: $T^{-1}\sum_1^T B_{i,t-1}\xi_{i,t}' \to \int_0^1 B_i dB_i' + \Lambda$. We may assume continuous spectral density at frequency zero to ensure the convergence of non-parametric estimates of the long-term covariance matrix.[11]

Necessary conditions for ***(A1)*** and ***(A2)*** are discussed with details in Phillips and Durlauf

[11]Continuous spectral density at frequency zero is also a natural requirement for automatic bandwidth selection since the procedure relies on the derivation of the spectral density; in most of this literature, however, the bandwidth is simply selected with a consistent rule of the thumb, *i.e.*, $h = 4 \cdot (\frac{T}{100})^{2/9}$. I argue that this choice is made because automatic selection procedures are not optimal when based on the (biased) residual of an OLS regression.

(1986),[12] Park and Phillips (1988), and Phillips and Moon (1999).[13] We note in particular that:

- Martingale difference sequences satisfying $\mathbb{E}[\xi|\mathscr{F}_{t-1}] = 0$ finite covariance matrix $\Sigma > 0$, and finite fourth-order moments, where $\mathscr{F}_{t-1}$ is the $\sigma - field$ generated by $\{\xi_{t-\tau}, \tau = 1, 2, ...\}$, satisfy the two properties above, see Phillips and Durlauf (1986, p 475) for details.

- Strictly stationary processes also satisfy these conditions, simply requiring that $\mathbb{E}[\xi\xi'] < \infty$.[14]

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

(A3) The initialization of the system described in **(M1)** happens at $t = 0$ and $X_{i,o}$ and α_i may be any random variables, without any distributional requirement such as finite fourth-order moments of α_i. This happens because we rely on an explicit demeaning of the system (and because we do not seek to make inference on α_i or the distribution/mean thereof). We assume that $\varepsilon_{i,1}$ (and thus $X_{i,0}$) are observed, so that in our simulations no observation is lost when computing when exogeneising *w.r.t.* ε_i in FMOLS.

(A4) Finally, assumptions need to be made to ensure the convergence of group-mean estimators under heterogeneous feedback, and, in particular, as underlined by Phillips and Moon (1999), 'some moments conditions on $(\Omega_{22,i})^{-1}\Omega_{21,i}$ [are needed] to avoid heavy tails in the density of $(\Omega_{22,i})^{-1}$ ' (p1077). Requiring that $\exists(c_{min}, c_{max})$ such as $det(\Omega_{22,i}) > c_{min} \forall i$ and $tr(\Omega_{22,i}) < c_{max} \forall i$ simplifies these requirements.[15] These conditions ensure that a central limit theorem can be applied to the bias and to its time-series estimates.

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

[13] Phillips and Moon (1999) assume that ξ is generated by a random coefficient linear process $\xi_{i,t} = \sum_{s=0}^{\infty} C_{i,s} V_{i,t-s}$ where C are random matrices and define regularity conditions on C. While we leave the definition of needed restrictions over the moments of C and V further research, we note that their pooled estimator requires more homogeneity on $C_{i,\cdot}$ than the group-mean estimator we study in this paper.

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

[15] The exact conditions are that when $f(\Omega) = \mathbb{O}(e^{tr(-c\Omega)})$ for some $c > 0$ when $tr(\Omega) \to \infty$ and that $f(\Omega) = \mathbb{O}(det(\Omega)^{\gamma})$ for some $\gamma > 7$ when $det(\Omega) \to 0$.

With such conditions, convergence of the group-mean of time series estimators of all estimators is almost sure when $(T, N = d(T)) \to \infty$ with a suitable divergence rate for N.

4 Removing the conditional bias in FMOLS

4.1 Conditional finite-T FMOLS bias

Note that a Toy version of the FMOLS technique is presented in Section 2 on page 5, and the full estimator is reminded in Section A.2.4 on page 26.

Theorem 4.1. *Rate of convergence of FMOLS*
$\mathbb{E}\left[\widehat{b_i^{FM}}(T) - (\hat{\beta}_i^{OLS} - \beta)\right] = \mathbb{E}\left[\widehat{b_i^{FM}}(T)\right] - b_i^{OLS}(T) = \mathbb{O}(h_i/T) \cdot b_i^{OLS}(T)$ *with h_i the (consistent) kernel bandwidth.*

If as usual for the Bartlett kernel $h_i = o(T^{2/9})$, $\quad \mathbb{E}\left(\widehat{b_i^{FM}} - b_i^{OLS}\right) = o(T^{-7/9}) \cdot b_i^{OLS} = o(T^{-16/9})$.
The proof, similar in spirit to that of page 7, can be found in Appendix A.1.1 on page 22.

4.2 Conditionally unbiased time-series estimators

Remark 4.2. Theorem 4.3 extends the Toy model notably by taking into account the exact nature of the FMOLS corrections to compute $\tilde{e}$, and uses a modified estimator $\widehat{df_{corr}}(h, X)$ that is (artificially) $\mathbb{O}(1)$ for any given estimate and any T and retains the interpretation as a correction for the number of degrees of freedom in a non-parametric context. Proof can be found on page 22.

Theorem 4.3. *Construction of an unbiased time-series estimator from the FMOLS estimator and the relation between its conditional bias and the 'first step' OLS bias.*[16]
$\widehat{b^{FT}} = \widehat{df_{corr}}(h, X) \cdot \widehat{b^{FM}}$ *denotes a finite-T, estimator of the bias corrected for the conditional bias in $\hat{u}$, with*

$$\widehat{df_{corr}}(h, X) = \begin{cases} 10 & \textit{if} \quad |\tilde{e}(h_i, X_i) - 1| < 10^{-2} \\ \frac{1}{1 - \tilde{e}(h, X)} & \textit{Otherwise} \end{cases}$$

[16]The use of the demeaned $\tilde{x}$ allows for univariate notations.

and $\tilde{e}(h,X) = (\tilde{x}'\tilde{x})^{-1}\cdot\left\{\left(\sum_{k=0}^{h} w(k)\tilde{x}' L_k \varepsilon\right) - \left(\sum_{k=-h}^{h} w(|k|)\varepsilon' L_k \tilde{x}\right)\frac{\widehat{\Sigma}_{22}+\widehat{\Gamma}_{22}^{(h)}-\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}_{22}^{(h)}}\right\}$*, or equivalently*

$$\tilde{e}(h,X) = (\tilde{x}'\tilde{x})^{-1}\left\{\left(\sum_{k=-h}^{h} w(|k|)\varepsilon L_k \tilde{x}\right)\frac{\widehat{\Gamma}_{22}^{(h)}+\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}_{22}^{(h)}} - \sum_{k=1}^{h} w(|k|)\varepsilon' L_k \tilde{x}\right\}.$$

4.3 Divergence rate of N and T for nuisance parameter-free hypothesis tests

Although FMOLS is nested in a T-asymptotic framework, which supposes that $T \to \infty$ before $N \to \infty$, a typical recommendation is to use the FMOLS technique with panel data, sometimes without considering explicitly the relative width and length of the panels. Pedroni (2001), for instance, suggests (p8) that although the theory is based on sequential limits arguments, N and T may in practice grow large concurrently.[17]

Yet the use of panel data call for a formal study of the divergence rate of N and T when $T \to \infty$. We thus consider $N = d(T)$ and $T \to \infty$ and Property 4.4 formally studies necessary and sufficient conditions for nuisance-free hypothesis tests based on the panel group-mean of TS estimators.

Property 4.4. *Asymptotic size of tests*

- *OLS. Whatever the relative size of T and N, the nominal size of the group-mean of OLS estimators of* β *tends to 1.*

- *FMOLS. (With a Bartlett kernel,)* $N = \mathbb{O}(T^{14/9})$ *or* $d(T)/T^{14/9} = \mathbb{O}(1)$ *is necessary for the* $t-stat$ *of the group-mean to converge in probability to a* $\mathbb{N}(0,1)$*, i.e., for nuisance parameters to be removed from the asymptotic distribution.*

- *FT.* $N = \mathbb{O}(T^2)$ *or* $d(T)/T^2 = \mathbb{O}(1)$ *is required for the removal of the nuisance parameters. Here, the requirement that* $N < T$ *can be expected to be sufficiently restrictive in practice.*

[17] The typical observation is that the sequential limit argument may be interpreted as the future consistence of the estimator after additional collection of observation in the time dimension for a fixed panel size, which also implicitly overlooks the consistency of the estimates made at any given point in time.

5 Numerical simulations

To assess the performance of estimators, researchers usually first perform experiments in which they simulate their bias, sample volatility, (R)MSE, and size.[18]

For consistency with FMOLS literature, namely the experiments performed by Phillips and Loretan 1991 (hereafter PL91) and Pedroni 2001 (hereafter P01), we rely on a $MA(1)$ specification for ξ, more precisely scenario(s) IV identified by PL91 as those under which FMOLS has poor performance – high residual bias and distorted size statistics. The data is simulated with constant betas, $\beta_i = \beta_0 = 2$; the individual constant is a uniform random $\alpha_i \rightsquigarrow \mathbb{U}(2,4)$; the feedback term is $\xi \rightsquigarrow MA(1)$, written as $\xi_t = \eta_t + \Theta\eta_{t-1}$ and $\eta \rightsquigarrow \mathbb{N}(0,\Omega_\eta)$, with $\Theta = \begin{pmatrix} 0.3 & 0.4 \\ \Theta_{21} & 0.6 \end{pmatrix}$, and $\Omega_\eta = \begin{pmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix}$.

Scenario IV is characterised by $\Theta_{21} = -0.8$; in the worse-case scenario is characterised by $\sigma_{21} = 0.5$, a simple implementation of FMOLS, instead of correcting the bias, increases it. In the second-worse scenario with $\sigma_{21} = 0$, FMOLS only reduces marginally the bias.

In scenario IV the feedback is homogenous in the cross-section – randomly heterogenous CS do not qualify as worse case scenario. The second-worse case is the baseline scenario of P01, who add cross-section heterogeneity by having random σ_{21} and Θ_{21}

The size statistics are shown relative to (the true) $H_0 : \hat{\beta}^{GM} = \beta_0 = 2$ for a target size of 5%. To make the tables comparable with previous experiments, we use a t-statistic rather than a Wald test, arguably more relevant since it permits testing the true null that $\beta_i = \beta_0 = 2 \forall i$; this test is computed in a straightforward manner as $\sqrt{N} \cdot \mathbb{E}_i \left[\hat{\beta}_1^{(\cdot)}\right] \,/\, \mathbb{E}_i \left[\widehat{\sigma}(\hat{\beta}_1^{(\cdot)})\right]$. Variations using other weighing schemes for $\hat{\beta}^{GM}$, and/or the group-mean of individual t-statistics, $\sqrt{N} \cdot \mathbb{E}_i \left[\hat{\beta}_1^{(\cdot)} / \widehat{\sigma}(\hat{\beta}_1^{(\cdot)})\right]$, for testing, are available upon request.

The rest of this Section is organised as follows: the first sub-section explains and rationalises the results of the experiments performed by PL91; the second assesses the performance of the

[18]The size and power can be displayed for for various nominal sizes. The power requires the specification of an alternative. In a second step the exact value of the statistics is documented.

asymptotic bias correction in the PL91 case, and compares the small-sample performance of all estimators studied. An heterogenous CS specification as in P01 is also presented.

The alternative estimators we know, DOLS and IM-OLS have significantly larger RMSE than FMOLS and FT for all values of T and N and are not reported in the summary tables. These alternative estimators are summarised in Appendix A.2.4.

5.1 Rationalisation of existing FMOLS experiments

When T is small, non-parametric kernel estimators are highly dependent on the weights given to the first lead/lag, thus on the choice of the panel and of the bandwidth. Table 1 shows the variations in weights for realistic choices of bandwidth, when T=10.

Table 1: Kernel weights given to the first lead/lags.

Bandwidth	Kernel	w(1)	w(2)	w(3)
1.3	Bartlett	0.23	0.0	0.0
	Quadratic Spectral	0.38	-0.09	0.03
	Parzen	**0.02**	0.0	0.0
2	Bartlett	0.5	0.0	0.0
	Quadratic Spectral	**0.69**	0.14	-0.09
	Parzen	0.25	0.0	0.0

PL91 and P01 used simulations to assess the bias from OLS and a simple version of FMOLS. For a given DGP, since PL91 and P01 use a deterministic bandwidth[19] with known kernel weights, theses biases can be approximated as in our Toy model (or more precisely derived as expectation of ratios of quadratic forms).[20]

Qualitatively, noting that $\psi_0 = \mathbb{E}[\xi_t \xi_t'] = \Omega_\eta + \Theta\Omega_\eta\Theta'$, that $\psi_1 = \mathbb{E}[\xi_t \xi_{t-1}'] = \Omega_\eta\Theta'$ and that $\psi_{-1} = \mathbb{E}[\xi_{t-1}\xi_t'] = \Theta\Omega_\eta$; in both cases of scenario IV in PL91, $\Theta_{21} = -0.8$; so $(\Theta\Omega_\eta\Theta')_{12} \approx 0$, the mean FMOLS bias correction depends on the weight that is given to first lead/lag, thus on the choice of the kernel. The bias correction is partial and volatile, and the worse-case scenario is qualitatively similar to example *(b)* on page 27: since $\psi_{-1} \approx -\psi_1$ and $\psi_0 \approx 0$, demeaning has little

[19]We note that PL91 use a Bartlett kernel with a fixed bandwidth of 5, while P01 uses a T-dependent rule $h = 4\left(\frac{T}{100}\right)^{2/9}$, and the kernel he uses is not specified.

[20]Note that the actual results are sensitive to the very implementation chosen. In particular, in the modern implementation of FMOLS, the orthogonalisation – which was not performed by PL91 – changes the outcome.

effect, and the application of kernel estimators on biased residuals means that FMOLS amplify the bias, when T is small and the kernel estimator weights insufficiently the first lag covariance.

By contrast, in table 1 of P01 as in table 2 of PL91, $\Theta_{21} = 0.4$. Pedroni considers random σ_{21} with mean value 0. With these mean values, $(\Theta\Theta')_{21} = 0.36$, so $\psi_0 \prec 1.36$; $\psi_1 = \psi_{-1} \prec \Theta_{21} = \Theta_{12} = 0.4$. The impact of lead/lag feedback almost cancel out, and the bias arises almost entirely from the instantaneous feedback $\psi_0 >> \frac{T-2}{T}\psi_1 - \frac{T-1}{T}\psi_{-1}$. Then the FMOLS implementation is not very sensitive to the choice of the kernel, and its corrections fare well even when T is small. As kernel estimates have small bias in this scenario, our additional correction also fares well.

5.2 Experiments with the F-T FMOLS correction

This section details experiments with the same processes as in PL91, and where only time-series corrections are made to the estimators. The size statistics are computed for a nominal size of 5%, and two hundred thousand simulations are made. Notations for this section are provided in table 2.

Table 2: Notations for table(s) in Section 5

Notation	Description
b^{OLS}	$\mathbb{E}_{sim}\left(\mathbb{E}_{i=1:N}[\hat{\beta}^{OLS}\right)]$ the simulated group-mean OLS bias
$\sigma(\widehat{\beta^{OLS}})$	Simulated sample standard deviation of the OLS slope
$\widehat{\sigma}^{OLS}(\widehat{\beta^{OLS}})$	White's HC3 estimator of $\sigma(\widehat{\beta^{OLS}})$
$^{Rmse}_{OLS}$	Root Mean Square Error (RMSE) of the OLS slope
$\alpha^{H_0}_{(OLS)}$	Nominal size for standard group-mean OLS: $\mathbb{P}\left\lvert(\widehat{\beta^{OLS}} - \beta_0)/\widehat{\sigma}^{OLS}(\widehat{\beta^{OLS}})\right\rvert > t(\alpha)$
FM	Denotes a measure (σ, $\widehat{\sigma}$, RMSE, α^{H0}) linked to a FMOLS estimator
FT or $FT-FM$	Measure linked to a bandwidth-corrected, *finite-T* FM estimator

The time-series estimators use White (1980) HC3 estimator as model estimates of the volatility of the slope, because when T is small FMOLS residuals are orthogonalised by but exogeneised to the regressors, so the asymptotic calculation of the FMOLS variance will underestimate the variance of the FMOLS slope – detailed experiments are available on request.

Following Theorem 4.3 on page 13 we truncate the value of the multiplier to 10. Results are little sensitive to the truncation point (and virtually insensitive when $T \geq 20$).

Tables in this section shows that even in the most complex situation described by PL91, FT always has smaller bias and RMSE than FM. Not shown in the tables, it also has lower RMSE than

Table 3: Improved TS estimators: worse-case scenario of PL91 ($\sigma_{21} = 0.5$)

		OLS			FMOLS			Finite-T		
N	T	b^{OLS}	$RMSE_{ols}$	$\alpha^{H_0}_{OLS}$	b^{FM}	$RMSE_{FM}$	$\alpha^{H_0}_{FM}$	b^{FT}	$RMSE_{FT}$	$\alpha^{H_0}_{FT}$
10	10	0.303	0.322	0.895	0.218	0.249	0.620	0.104	0.195	0.287
	20	0.176	0.187	0.955	0.101	0.120	0.621	0.048	0.086	0.318
	30	0.124	0.132	0.966	0.061	0.075	0.572	0.030	0.054	0.313
	40	0.095	0.101	0.969	0.042	0.053	0.533	0.022	0.039	0.301
	50	0.077	0.082	0.972	0.031	0.040	0.496	0.017	0.030	0.292
20	10	0.303	0.313	0.992	0.218	0.234	0.850	0.104	0.157	0.359
	20	0.176	0.181	0.999	0.101	0.111	0.838	0.048	0.070	0.409
	30	0.124	0.128	0.999	0.061	0.068	0.787	0.030	0.044	0.404
	40	0.095	0.098	0.999	0.042	0.048	0.742	0.022	0.031	0.390
	50	0.078	0.080	0.999	0.032	0.036	0.700	0.017	0.024	0.381
30	10	0.303	0.309	1	0.218	0.229	0.947	0.104	0.141	0.420
	20	0.176	0.180	1	0.101	0.108	0.935	0.048	0.063	0.487
	30	0.124	0.126	1	0.061	0.066	0.897	0.030	0.040	0.481
	40	0.095	0.097	1	0.042	0.046	0.863	0.022	0.029	0.470
	50	0.078	0.079	1	0.032	0.035	0.824	0.017	0.022	0.457

Table 3, computed with 200 000 simulations, shows that in the worse-case scenario of PL91, the OLS bias is so large (at 0.3 for $T = 10$) that the 5% tests have a nominal size of virtually 100% for all values of T and N. FMOLS estimators provide only a very partial correction (a bias of greater than 0.2 for $T = 10$ and than 0.1 for $T = 20$ remains), so that the 5% tests have a nominal size of more than 75% for all T when $N = 30$.

Even in this complex scenario, multiplying the FMOLS correction by $\widehat{df_{corr}}$ consistently improves the estimation of the slope for all (N,T). The size of finite-T estimators improves but remains heavily distorted (*e.g.* above 40% for $N = 30$ and $T = 50$.)

Table 4: Improved TS estimators: 2nd worse scenario IV of PL91 ($\sigma_{21} = 0$)

		OLS			FMOLS			Finite-T		
N	T	b^{OLS}	$RMSE_{ols}$	$\alpha^{H_0}_{OLS}$	b^{FM}	$RMSE_{FM}$	$\alpha^{H_0}_{FM}$	b^{FT}	$RMSE_{FT}$	$\alpha^{H_0}_{FT}$
10	10	0.169	0.189	0.633	0.122	0.151	0.363	0.064	0.138	0.218
	20	0.092	0.102	0.760	0.055	0.071	0.395	0.031	0.060	0.245
	30	0.064	0.070	0.791	0.033	0.045	0.370	0.020	0.037	0.239
	40	0.048	0.053	0.804	0.023	0.032	0.347	0.015	0.027	0.232
	50	0.039	0.043	0.811	0.017	0.025	0.324	0.011	0.021	0.224
20	10	0.170	0.180	0.885	0.122	0.138	0.571	0.064	0.108	0.269
	20	0.092	0.097	0.952	0.055	0.064	0.597	0.031	0.048	0.323
	30	0.063	0.067	0.963	0.033	0.039	0.554	0.020	0.030	0.318
	40	0.048	0.051	0.968	0.023	0.028	0.519	0.015	0.022	0.311
	50	0.039	0.041	0.970	0.017	0.021	0.487	0.011	0.017	0.301
30	10	0.169	0.176	0.970	0.122	0.132	0.726	0.064	0.095	0.314
	20	0.092	0.096	0.992	0.055	0.061	0.739	0.032	0.043	0.390
	30	0.063	0.066	0.994	0.033	0.037	0.691	0.020	0.027	0.389
	40	0.048	0.050	0.995	0.023	0.026	0.653	0.015	0.019	0.383
	50	0.039	0.040	0.996	0.017	0.020	0.614	0.011	0.015	0.371

Table 4 is second worse case scenario of PL91, CS homogenous, with 200 000 simulations.
As in table 3, multiplying the FMOLS correction by $\frac{1}{1-\tilde{e}(h)}$ consistently improves the estimation of the slope for all measures (and when $T = 10$, it is useful to impose bound on the multiplier).

The bias is lesser than in the worse-case, and the nominal size of the 5% tests less distorted: it is below 25% for $N = 10$.

other estimators we have reviewed in the literature (see Section A.2.4).

Conform to Property 4.4, the size of the FM converges slowly to the nominal size, and that of FT more quickly, but a greater N leads to higher nominal size.

In table 3 which represents the most complex scenario of PL 91, estimators remain severely distorted from all standpoint; FMOLS estimators, in this implementation reduce the bias by 25%, while finite-T correction allow a bias reduction of two-thirds.

Table in 5 shows that introducing heterogeneity does not substantially change the message since TS estimators are by design robust to CS heterogeneity. In fact, since cross-section heterogeneity diminishes the probability of being in a worse-case scenario, the performance of all estimators studied here improves in such specification.

Table 5: TS estimators, Scenario IV of PL91 with σ_{21} and Θ_{21} random.

		OLS				FMOLS				Finite-T			
N	T	b^{OLS}	$\underset{ols}{RMSE}$	$(within)$	$\alpha^{H_0}_{OLS}$	b^{FM}	$\underset{FM}{RMSE}$	$(within)$	$\alpha^{H_0}_{FM}$	b^{FT}	$\underset{FT}{RMSE}$	$(within)$	$\alpha^{H_0}_{FT}$
10	10	0.165	0.192	0.189	0.521	0.116	0.153	0.152	0.332	0.056	0.139	0.139	0.199
	20	0.092	0.106	0.104	0.667	0.052	0.071	0.070	0.368	0.026	0.058	0.058	0.223
	30	0.064	0.073	0.072	0.706	0.031	0.044	0.044	0.341	0.016	0.036	0.036	0.214
	40	0.049	0.056	0.055	0.725	0.022	0.032	0.032	0.323	0.012	0.026	0.026	0.210
	50	0.040	0.045	0.044	0.733	0.016	0.024	0.024	0.301	0.009	0.020	0.020	0.202
20	10	0.165	0.179	0.178	0.780	0.116	0.136	0.135	0.512	0.056	0.106	0.106	0.243
	20	0.092	0.099	0.098	0.896	0.052	0.063	0.063	0.549	0.026	0.045	0.045	0.281
	30	0.064	0.069	0.068	0.920	0.031	0.038	0.038	0.507	0.016	0.028	0.028	0.272
	40	0.049	0.053	0.053	0.930	0.022	0.027	0.027	0.474	0.012	0.020	0.020	0.266
	50	0.040	0.042	0.042	0.935	0.016	0.021	0.021	0.441	0.009	0.016	0.016	0.255
30	10	0.165	0.174	0.173	0.907	0.116	0.129	0.128	0.654	0.056	0.092	0.092	0.281
	20	0.092	0.097	0.096	0.971	0.052	0.059	0.059	0.683	0.026	0.040	0.040	0.332
	30	0.064	0.067	0.067	0.981	0.031	0.036	0.036	0.635	0.016	0.025	0.025	0.323
	40	0.049	0.051	0.051	0.985	0.022	0.025	0.025	0.597	0.012	0.018	0.018	0.317
	50	0.040	0.042	0.042	0.986	0.016	0.019	0.019	0.558	0.009	0.014	0.014	0.303

In the spirit of Pedroni (2001), table 5 reproduces the experiments of scenario 4 of Phillips and Loretan (1991), but adding cross-sectional heterogeneity by adding randomness in σ_{21} and Θ_{21} in the range of value studied by PL91: $\sigma_{21} \rightsquigarrow \mathbb{U}([-0.85, 0.85])$ and $\Theta_{21} \rightsquigarrow \mathbb{U}([-0.8, 0])$. In this table (200 k simulations), the random elements of the MA process are those that PL91 varied (by means of scenarios).

We note that for each simulation, $\theta_i = (\sigma_{21,i}, \Theta_{21,i})$ varies randomly, so that the total variance of the slope across simulations $V(\hat{\beta}) = RMSE^2(\hat{\beta})$ is the sum of the *within* and *between* variance: $V(\hat{\beta}) = \mathbb{E}_s\left[\mathbb{E}_i(\hat{\beta}_i - b(\theta_s)|\theta_s)^2\right] + \mathbb{E}_s\left[b(\theta_s)^2\right]$ where s represent possible parameter values and b the bias associated to θ_s. The column $(within)$ estimates the mean of the *within* variance across simulations (and takes its square root): $(within)^2 = \mathbb{E}_s\left[\mathbb{E}(\hat{\beta}_s - b(\theta_s)|\theta_s)^2\right]$

6 Conclusion

This paper proposes a finite-T framework for panel cointegration, and uses the breakthrough T-asymptotic time series cointegration works of Hansen, Parker, Pedroni, Phillips, Saikkonen, Stock and Watson as a stepping stone.

Table 6: Summary table: 5% nominal size of TS

		Worse-Case			2nd Worse-Case		
N	T	α^{H0}_{ols}	α^{H0}_{FM}	α^{H0}_{FT}	α^{H0}_{ols}	α^{H0}_{FM}	α^{H0}_{FT}
10	10	0.895	0.620	0.287	0.633	0.363	0.218
	20	0.955	0.621	0.318	0.760	0.395	0.245
	30	0.966	0.572	0.313	0.791	0.370	0.239
20	10	0.972	0.496	0.292	0.811	0.324	0.224
	20	0.992	0.850	0.359	0.885	0.571	0.269
	30	0.999	0.838	0.409	0.952	0.597	0.323
30	10	0.999	0.700	0.381	0.970	0.487	0.301
	20	1	0.947	0.420	0.970	0.726	0.314
	30	1	0.935	0.487	0.992	0.739	0.390

In table 6, the nominal size for a 5% target size is presented. One sees that the size statistics is degenerate for OLS; that the size of finite-T FMOLS estimators converges faster that that of the asymptotic FMOLS, but that the size of all TS estimators is strongly distorted.

FMOLS estimates may be highly biased when T is small, and the technique becomes progressively useless when T becomes very big. Thus the situations in which FMOLS correctly correct the bias may be unknown to the econometrician, and require further analysis of the properties of the panel studied.[21]

The finite-T correction needed to the FMOLS estimators can be understood both as a correction for the (first-order) conditional bias and as a correction for the number of degrees of freedom in a non-parametric context. This correction significantly improves upon the FMOLS in all situations studied, and allows the use of the group-mean estimator in wider panels.

The techniques proposed extend directly to the addition of common exogenous regressors, and can be used for a variety of empirical issues in finance such as so-called replicating products; in financial economics such as purchasing power parity estimates;[22] in household finance where the

[21] When T is low, the time-series minimisation of the RMSE of the spectral covariance matrix may not minimise the RMSE of the group-average bias-corrected mean slope.

[22] The techniques we use provide more precise slope estimates that those that currently exist. However, the tests

cross-section is large but large individual data sets short.

Although we markedly improve the speed of convergence of the estimator, the inherent limitation is that spectral estimators are only defined in a (T-)asymptotic framework. When T is small, spectral density estimators may leave an unknown 'unconditional' bias, resulting in large size distortions. As FMOLS is often used with panel data, an open question is whether one can rely on the cross-sectional dimension of the panel to optimally estimate and correct for the endogeneity bias in panel data.

7 Extension - draft

We propose to use this FT technique to revisit the purchasing power parity (PPP) hypothesis. While the above analysis shows that with independent cross sections, FT significantly improves the size of the test statistics, one cannot exclude some cross-section dependence between economic series (and in particular between direct exchange rates measured against a common currency).

We will thus rely on modified wild bootstrap (heafter MWB) of Smeekes and Urbain (2014), who show that, as in (Shao, 2011), bootstrapping the entire cross-section of a panel preserves the unspecified cross-sectional dependence of the panel, thus yielding bootstrap statistics robust to (unspecified) cross-sectional dependence. We show that these techniques, studied in Smeekes and Urbain (2014) for unit-root tests, are also valid to test a hypothesis on the slope of cointegrated time-series.

Finally, we note that, while for estimation purposes, FT is the natural choice for the bias reduction it permits, there are several possible candidates for bootstrap-based inference, namely, the first version of FMOLS of Phillips and Loretan (1991), the modern FMOLS (Phillips and Hansen 1990 and Pedroni 1996), and FT (Sender (2014)). We thus study the properties of the bootstrapped test statistics of these different estimators.

performed in this paper assume independent cross-sections. Testing in the presence of cross-sectional dependency requires the use of additional techniques, such as the use of factors analysis or of specific forms of bootstrap.

A Appendices

A.1 Proofs

A.1.1 Proofs for FMOLS residual bias

Proof. Proof of theorem 4.1 on page 13.

Noting that the orthogonalisation of Y_i *wrt* ε_i leaves the expected FMOLS bias unchanged, FMOLS uses a spectral estimator of $\frac{1}{T}X_i'\hat{u}_i/\frac{1}{T}X_i'MX_i$.

The kernel estimator having at a consistent bandwidth an $\mathbb{O}^{\mathbb{P}}(\frac{1}{T})$ bias of an unknown form, we have $\mathbb{E}[\widehat{\Omega}_{21,i}^{(h_i)}-\Omega_{21,i}]=\mathbb{O}^{\mathbb{P}}(\frac{1}{T})$, then as $\mathbb{E}[\hat{u}_i-u_i|X_i]=-MX_ib_i^{OLS}(X_i)$, we have $\mathbb{E}\left[\widehat{\widehat{\Omega}}_{21,i}^{(h_i)}-\Omega_{21,i}|X_i\right]=\mathbb{E}[-\sum_{k=0}^{h_i}X_i'ML^k\varepsilon_iw(k)\cdot b_i^{OLS}]+\mathbb{O}^{\mathbb{P}}(\frac{1}{T})$.

So, $\mathbb{E}\left[\widehat{\widehat{\Omega}}_{21,i}^{(h_i)}-\Omega_{21,i}\right]=b_i^{OLS}\cdot\mathbb{O}^{\mathbb{P}}(h_i)$ as the weights of the kernel are absolutely summable. And $\mathbb{E}[\frac{1}{T}X_i'MX_i]=\mathbb{O}^{\mathbb{P}}(T)$ since X is $I(1)$. Then, $\mathbb{E}[\frac{\widehat{\widehat{\Omega}}_{21,i}^{(h_i)}-\Omega_{21,i}}{\frac{1}{T}X_i'MX_i}]=\mathbb{O}^{\mathbb{P}}(\frac{h_i}{T})\left(\mathbb{O}^{\mathbb{P}}(\frac{1}{T})+o(1/T)\right)$. So $\mathbb{E}[\frac{\widehat{\widehat{\Omega}}_{21,i}^{(h_i)}-\Omega_{21,i}}{\frac{1}{T}X_i'MX_i}]=\mathbb{O}^{\mathbb{P}}(\frac{h_i}{T^2})$. N.B.: for the group-mean, $b^{FM}=\mathbb{E}\left[\mathbb{E}_i[\frac{\widehat{\widehat{\Omega}}_{21,i}^{(h_i)}-\Omega_{21,i}}{\frac{1}{T}X_i'MX_i}]\right]=\mathbb{O}^{\mathbb{P}}(\frac{\mathbb{E}[h_i]}{T^2})$. □

A.1.2 Proofs for time-series correction of FM estimators

The FMOLS approach – see A.2.4 – can be summarised as (6), where, dropping $_i$ subscripts from notations, the spectral density of $\hat{\xi}=(\hat{u}^{OLS},\varepsilon)$ is $\widehat{\widehat{\Omega}}^{(h)}=\widehat{\widehat{\Sigma}}+\widehat{\widehat{\Gamma}}^{(h)}+\widehat{\widehat{\Gamma}}'^{(h)}$ with Σ the instantaneous covariance, $\Gamma^{(h)}$ the lag covariance, and $\Gamma_{21}^{(h)}$ the covariance between u and lag values of ε.[24]

$$\widehat{\widehat{\beta^{FM}}}=\widehat{\beta^{OLS}}+(\tfrac{1}{T}\tilde{x}'\tilde{x})^{-1}\widehat{\widehat{\Omega}}_{21}^{(h)}\frac{\widehat{\Sigma}_{22}+\widehat{\Gamma}_{22}^{(h)}-\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}_{22}^{(h)}}-(\tfrac{1}{T}\tilde{x}'\tilde{x})^{-1}(\widehat{\widehat{\Sigma}}_{21}+\widehat{\widehat{\Gamma}}_{21}^{(h)}) \qquad (6)$$

FMOLS estimate the OLS bias as: $(\frac{1}{T}\tilde{x}'\tilde{x})^{-1}(\widehat{\widehat{\Sigma}}_{21}+\widehat{\widehat{\Gamma}}_{21}^{(h)})-(\frac{1}{T}\tilde{x}'\tilde{x})^{-1}\widehat{\widehat{\Omega}}_{21}^{(h)}\frac{\widehat{\Sigma}_{22}+\widehat{\Gamma}_{22}^{(h)}-\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}_{22}^{(h)}}$ where conform the $\widehat{\widehat{(\cdot)}}$ notation, only $\widehat{\widehat{\Omega}}_{21}^{(h)}$ and its components $\widehat{\widehat{\Sigma}}_{21}$ and $\widehat{\widehat{\Gamma}}_{21}^{(h)}$ are subject to a bias that relates directly to the first-step measurement error $(\widehat{\beta^{OLS}}-\beta_0)$ of the OLS slope.

[24] Remind that $\widehat{\widehat{m}}=\widehat{\widehat{\Sigma}}_{21}+\widehat{\widehat{\Gamma}}_{21}^{(h)}-\frac{\widehat{\widehat{\Omega}}_{21}^{(h)}}{\cdot}\widehat{\Omega}_{22}^{(h)}(\widehat{\Sigma}_{22}+\widehat{\Gamma}_{22}^{(h)})$ and that $\widehat{\widehat{\beta^{FM}}}=(\tilde{x}'\tilde{x})^{-1}\left[\tilde{x}'y^*-T\widehat{\widehat{m}}\right]=(\tilde{x}'\tilde{x})^{-1}\left[\tilde{x}'(y-\widehat{\widehat{\Omega}}_{21}^{(h)}/\widehat{\Omega}_{22}\varepsilon)-T\widehat{\widehat{m}}\right]=\widehat{\beta^{OLS}}-(\tilde{x}'\tilde{x})^{-1}\left[\widehat{\widehat{\Omega}}_{21}^{(h)}/\widehat{\Omega}_{22}(\tilde{x}'\varepsilon)+T\widehat{\widehat{m}}\right]=\widehat{\beta^{OLS}}-(\frac{1}{T}\tilde{x}'\tilde{x})^{-1}\left[\widehat{\widehat{\Omega}}_{21}^{(h)}/\widehat{\Omega}_{22}(\frac{1}{T}\tilde{x}'\varepsilon)+\widehat{\widehat{m}}\right]$;

The conditional measurement error is:
$(\frac{1}{T}\tilde{x}'\tilde{x})^{-1}\cdot(\widehat{\beta^{OLS}}-\beta_0)\cdot\left(\sum_{k=-h}^{h}\frac{w(|k|)}{T}\varepsilon' L_k\tilde{x}\right)\frac{\widehat{\Sigma}_{22}+\widehat{\Gamma}^{(h)}_{22}-\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}^{(h)}_{22}}-(\frac{1}{T}\tilde{x}'\tilde{x})^{-1}\cdot(\widehat{\beta^{OLS}}-\beta_0)\cdot\left(\sum_{k=0}^{h}\frac{w(|k|)}{T}\tilde{x}' L_k\varepsilon\right)$

Which, denoting $\tilde{e}(h,X)=(\tilde{x}'\tilde{x})^{-1}\cdot\left\{\left(\sum_{k=0}^{h}w(k)\tilde{x}'L_k\varepsilon\right)-\left(\sum_{k=-h}^{h}w(|k|)\varepsilon' L_k\tilde{x}\right)\frac{\widehat{\Sigma}_{22}+\widehat{\Gamma}^{(h)}_{22}-\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}^{(h)}_{22}}\right\}$, simplifies as $(\beta_0-\widehat{\beta^{OLS}})\cdot\tilde{e}(h,X)$. Schematically: $\left(\beta_0-\widehat{\beta^{OLS}}\right)\longleftarrow\left(\widehat{\widehat{\beta^{FM}}}-\widehat{\beta^{OLS}}\right)+\left(\beta_0-\widehat{\beta^{OLS}}\right)\tilde{e}(h,X)$, and, denoting $\widehat{\widehat{b^{FT}}}=\frac{\widehat{\widehat{b^{FM}}}}{1-\tilde{e}(h,X)}=\frac{\widehat{\widehat{\beta^{FM}}}-\widehat{\beta^{OLS}}}{1-\tilde{e}(h,X)}$, and $\widehat{\widehat{\beta^{FT}}}=\widehat{\beta^{OLS}}-\widehat{\widehat{b^{FT}}}$, we have in expectation that $\mathbb{E}\left[\widehat{\widehat{b^{FT}}}|(X,h,w)\right]=b^{OLS}(X)$ and $\mathbb{E}\left[\widehat{\widehat{\beta^{FM}}}\right]=\beta$.

As $\Omega=\Sigma+\Gamma+\Gamma'$ and $\Gamma_{22}=\Gamma'_{22}$ (here Γ_{22} is a scalar), we can rewrite $\tilde{e}(h,X)$ as $(\tilde{x}'\tilde{x})^{-1}\left\{\left(\sum_{k=-h}^{h}w(|k|)\varepsilon' L_k\tilde{x}\right)\frac{\widehat{\Gamma}^{(h)}_{22}+\frac{1}{T}\tilde{x}'\varepsilon}{\widehat{\Omega}^{(h)}_{22}}-\sum_{k=1}^{h}w(|k|)\varepsilon' L_k\tilde{x}\right\}$. Because the finite-T behaviour of $\tilde{e}(h,X)$ is unknown, we work with a modified estimator $\widehat{df_{corr}}(h)=10$ when $|\tilde{e}(h_i,X_i)-1|<10^{-2}$ and $\frac{1}{1-\tilde{e}(h,X)}$ otherwise. □

A.1.3 Proofs for size properties

Size properties: *Proof of Property 4.4 on page 14*:

From standard arguments, the volatility of all estimators considered is $\mathbb{O}(T^{-1})$ (and thus asymptotically correctly estimated).

With *i.i.d.* cross-sections, that $N\to\infty$ guarantees that the $t-statistic$ of the group-mean converges in probability towards a normal with variance 1, so the only asymptotic nuisance parameter is the mean of this normal distribution, itself driven by the speed of reduction of the bias in the time-series to the speed at which the width of the panel increases simultaneously to T.

Theorem 4.1 on page 13 states that the bias for a single time-series is $\mathbb{O}(T^{-16/9})$ for FMOLS[25] and $o(T^{-2})$ for FT (and it is well-known that it is $\mathbb{O}(T^{-1})$ for OLS).

Then as a direct consequence that the OLS bias shrinks at the same speed that the volatility, the expectation of the t-statistic for $\widehat{\beta}^{OLS}$ (against the true null $H_0:\widehat{\beta}^{OLS}=\beta_0$ where β_0 is the true cointegrating slope) is $\mathbb{E}[t^{H_0}_{\beta^{OLS}}]=\mathbb{O}(1)$ thus the size statistic is degenerate (for any target size α).

For FMOLS, we must have $\frac{N o(T^{-16/9})}{\sqrt{N}\mathbb{O}(T^{-1})}\to 0$, which yields $\sqrt{N}=\mathbb{O}(T^{7/9})$ and $N=o(T^{14/9})$.

For FT, we must have $\frac{N o(T^{-2})}{\sqrt{N}\mathbb{O}(T^{-1})}\to 0$, which yields $\sqrt{N}=o(T)$ and in turn $N=o(T^2)$.

[25] $b^{FM}=\mathbb{O}^p(h/T^2)$, thus $b^{FM}=o^p(T^{-16/9})$ if $h=o(T^{2/9})$ with a Bartlett kernel.

The regularity conditions set-out in Section 3.3 ensure that the same rate of convergence applies with (*inid*) heterogenous and (*i.i.d.*) homogenous cross-sections. Following the logic of Phillips and Moon (1999), the assumption that a CLT applies on both the bias and on its estimate permits the same rate of convergence for the group-mean of OLS and FM estimators. The same results are obtained for FT by noting that the $\mathbb{O}^{\mathbb{P}}(\frac{h}{T})$ in theorem 4.1 is lesser than 1 (see also page 7), and that $\widehat{df_{corr}}$ is equally bounded (in expectation and artificially for any value of T). □

A.2 Cointegration – summary of main approaches

This section summarises the main time-series estimators to the cointegrating slope with I(1) variables. We chose OLS and FMOLS as references in the tables, DOLS and IM-OLS having larger RMSE in the scenarios reviewed. CU-FM has undefined properties and is not shown either.

A.2.1 The OLS approach

In the presence of endogenous feedback, OLS regressions with stationary variables are in general biased, and there may be no way of recovering the bias.[26]

With dynamic feedback in I(1) variables, however, OLS is asymptotically unbiased because the regressors X have more persistent autocovariances than the feedback; then the asymptotically unbiased residuals allow the recovery of the true asymptotic covariance structure and further bias correction.

OLS – whether a first-step analysis of **(M1)** is complemented with a semi-parametric correction, **(M1)** is complemented with lead/lag of ΔX, its variables integrated, or a VAR/VECM system is analysed – is the backbone of cointegration analysis with a known functional form, in line with a literature that starts encompasses Stock (1987), Engle and Granger (1987), Phillips and Hansen (1990) for time series, and, Pedroni (2001) for panel approaches.[27]

[26]Trivially, if X is *i.i.d.* and $\Psi = \psi_0 I$ (instantaneous feedback). Then, $\hat{\beta} \underset{T\to\infty}{\to} \beta + \psi_0$ and instantaneous feedback can be distinguished by no means from increased exposure of Y to X (except if one can recourse to instrumental variable of course).

[27]An alternative line of thought is that of Johansen (1988), which is mainly useful when the number of cointegrated vectors is *priori* unknown.

A.2.2 The DOLS-DGLS approach

The dynamic least square approach (Saikkonen 1991 and independently by Stock and Watson 1993) suppresses dynamic feedback by adding the lead and lagged increments in X to the right-hand side of the equation, so that:

$$\hat{\beta}_i^{DLS} = \left(\sum_{t=1}^{T} (z_{it} z_{it}')^{-1} (\sum_{t=1}^{T} z_{it} \tilde{y}_{it}) \right)_1$$

where z_{it} are the $2q+1$ regressors $(\tilde{x}_i = x_i - \bar{x}_i, \Delta x_{i,t-q}, \ldots, \Delta x_{i,t+q})$, where $\tilde{y}_i = y_i - \bar{y}_i$, and the subscript 1 indicates that only the first element of the $(2q+1)$ vector of regressors is taken.[28]

This approach, asymptotically equivalent to the FMOLS approach,[29] is naturally not fit to small-T samples, because it involves both over-parametrisation and an important truncation.

In small samples, any estimation technique that is not parsimonious leads the volatility of the estimates to explode. In addition, the distribution of the estimates may be highly skewed due to the greater probability of reaching a (nearly) singular covariance matrix of Z , thus the expected volatility may not be a very useful nor reliable indication.[30]

A.2.3 The IM-OLS approach

Vogelsang and Wagner (2014) propose an additional integration of the variables in **(M1)** for OLS to achieve nuisance-parameter free test statistics, without the need to choose or tune a bandwidth parameter. This additional integration however has a cost in terms of volatility (and RMSE) of the integrated-modified OLS (IM-OLS) estimator.

With our notations, if $Y = \alpha + X\beta + u$, then $SY = S\mathbb{1} \cdot \alpha + SX \cdot \beta + Su$ where $(S\mathbb{1})_t = t$ is a trend. From the functional central limit theorem, $T^{-1/2} \sum_{t=1}^{[Tr]} \xi_{i,t} \Rightarrow B_i(r, \Omega_i)$, so that the dependence between $T^{-1/2} Su$ and $T^{-1/2} X$ is (asymptotically) only characterised by an instantaneous correlation.

[28] Of course, if x is p-dimensional, then not 1 but p slope coefficients are of interest.

[29] The group-mean $\hat{\bar{\beta}} = \frac{1}{N} \sum_i \hat{\beta}_i$

[30] See the Frish-Waugh decomposition in table (1). Monte-carlo simulations show that, with $T = 8$, $\psi_0 = 0.2$ and $v \rightsquigarrow \mathbb{N}(0, 1/4)$, and a supposedly known zero intercept, adding a useless lag in the regression leads to the volatility of the beta to increase from 10% to 70%. With N=12, goes from 10% to 20%. But more than 50% of the rise of the volatility as the T-size shrinks is due to those simulations where the expanded set of regressors is nearly colinear.
Time trends in small sample generate quite substantial volatility because they are not properly disentangled from stochastic trends.

Then, the equation $SY = S\mathbb{1}\cdot\alpha + SX\beta + X\beta_\Omega + Su$ can be estimated 'efficiently' by an OLS regression (and the test-statistics are nuisance-parameter free). In this regression, $\beta_\Omega = \Omega_{21}/\Omega_{22}$ does not require a non-parametric estimation. However, since the error term Su is $I(1)$, the variance of the estimators of this equation is greater than that of the FMOLS (and FT) estimators.

A.2.4 The FMOLS approach

The FMOLS does not have the drawbacks of cutting the sample.[31] It relies on a semi-parametric approach in which the nuisance parameters are estimated from the residuals of a first-step OLS regression. See Phillips and Hansen (1990) for time series and Pedroni (2001) for a panel approach. As in Pedroni (2001) we use element-wise notations for model **(M1)**.

Pedroni (1996) proposes the group-mean: $\widehat{\widehat{\beta}}^{G,FM} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\widehat{\beta}}_i^{FM}$, which steps are as follows:

1. For each i (subscripts are dropped in the notations), regress Y on X (and other variables)

2. Estimate the spectral density of $\hat{\xi} = (\hat{u}^{OLS}, \varepsilon)$, denoted $\widehat{\widehat{\Omega}} = \widehat{\widehat{\Sigma}} + \widehat{\widehat{\Gamma}} + \widehat{\widehat{\Gamma}}'$, relying on a 'consistent'[32] bandwidth selection and a kernel/spectral estimator (but on biased $\hat{u}$).

3. Triangularise $\widehat{\widehat{\Omega}}$ as follows: $\widehat{\widehat{L}}_{11} = (\widehat{\Omega}_{11} - \widehat{\widehat{\Omega}}_{21}^{2}/\widehat{\Omega}_{22})^{1/2}$, $\widehat{L}_{12} = 0$,
and $\widehat{\widehat{L}}_{21} = \widehat{\widehat{\Omega}}_{21}/\widehat{\Omega}_{22}^{1/2}$ and $\hat{L}_{22} = \widehat{\Omega}_{22}^{1/2}$ (7)

4. Build y^* asymptotically orthogonal[33] to $\varepsilon_{(1:T)}$ as follows: $y^* = y - \frac{\widehat{\widehat{L}}_{21}}{\widehat{L}_{22}}\varepsilon = y - \frac{\widehat{\widehat{\Omega}}_{21}}{\widehat{\Omega}_{22}}\varepsilon$ (8)

5. Compute the additional correction term $\widehat{\widehat{m}} = \widehat{\widehat{\Sigma}}_{21} + \widehat{\widehat{\Gamma}}_{21} - \frac{\widehat{\widehat{L}}_{21}}{\widehat{L}_{22}}(\widehat{\Sigma}_{22} + \widehat{\Gamma}_{22})$ (9)

6. Estimate the modified slope: $\widehat{\widehat{\beta^{FM}}} = (\tilde{x}'\tilde{x})^{-1}\left[(\tilde{x}'y^*) - T\widehat{\widehat{m}}\right]$ (10)

7. Compute the group mean:
$$\widehat{\widehat{\beta}}^{G,FM} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2\right)^{-1}\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)y_{it}^* - T\widehat{\widehat{m}}_i\right)\right] = \mathbb{E}_i\left[\frac{\frac{1}{T}X_i'MY_i^* - \widehat{\widehat{m}}_i}{\frac{1}{T}X_i'MX_i}\right]$$

[31] The first observation is lost however when one needs to differentiate X to find its increments ε.

[32] Automatic bandwidth selection would be optimal for a single time series if u was not biased. Automatic selection as well as T-dependent selection criteria remain consistent however even when u is biased. Note that only components related to $\hat{u}$ are based on biased time series and thus noted with $\widehat{\widehat{\ }}$ rather than $\widehat{\ }$.

[33] The idea is to suppress all nuisance terms so as to retrieve a normal estimator.

8. Pedroni (2001) shows that asymptotically: $\bar{t}_{\widehat{\widehat{\beta}}^{G,FM}} \to N(0,1)$ with

$$\begin{aligned}\bar{t}_{\widehat{\widehat{\beta}}^{G,FM}} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} t_{\widehat{\widehat{\beta}}_i^{FM}} \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\left(\widehat{\widehat{\beta}}_i^{FM} - \beta_0\right)\cdot \widehat{\widehat{L}}_{11i}^{-1}\left(\sum_{t-1}^{T}(x_{it}-\bar{x}_i)^2\right)^{1/2}\right] \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\widehat{\widehat{\beta}}_i^{FM} - \beta_0}{\sigma_{LT}(\hat{u}^{FM})(\tilde{x}_i\tilde{x}_i')^{-1/2}} = \sqrt{N}\mathbb{E}_i\frac{\widehat{\widehat{\beta}}_i^{FM} - \beta_0}{\sigma_{LT}(\hat{u}^{FM})(X_i'MX_i)^{-1/2}}\end{aligned}$$

A.2.5 Continuously updated FMOLS

An alternative methodology to the correction I propose involves relying on a continuously updated FMOLS estimator (CU-FM),[34] however, unfortunately, little can be said about the convergence and finite-T properties of CU-FM because of the complexity involved in the FMOLS corrrection. Its empirical performance is also difficult to analyse and seems somehow random.[35]

A.3 Two simple illustrations

To clarify the partial ability and sometimes failure of FMOLS to correct for the bias, we look at the simplest possible setting where the kernel estimator is based on the dirac kernel (and only the instantaneous covariance is used to correct for the bias). For this informal example, we simply approximate the expcetation of the ratio of random variables by ratio of expectations, thus all stated equalities are in fact true to an $\mathcal{o}(\frac{1}{T^2})$. We also denote the partial sum operator (lower-triangular unitary) by $S = \sum_{k=0}^{T-1} L_k$, so that $X_t = \sum_{k=0}^{t-1}\varepsilon_{t-k}$ reads $X = S\varepsilon$ in matrix form.

a) Instantaneous feedback

Keeping the same approximations, with only instantaneous feedback of strength ψ_0, the bias in the OLS regression is $\frac{\mathbb{E}[\psi_0\varepsilon' S\varepsilon]}{\mathbb{E}[\varepsilon' S' S\varepsilon]} = \frac{2\psi_0}{T+1}$, so the residual
$\hat{u}^{OLS} \rightsquigarrow u - 2\frac{\psi_0}{T+1}S\varepsilon = v + \psi_0(I - \frac{2}{T+1}S)\varepsilon$.

[34] The OLS residual serves to compute a first FMOLS, and analysing the properties of $\hat{u}^{FM}$ the residual computed with the FM slope serves to compute an updated FMOLS, etc...

[35] For some parameter values, it brings little improvement over FMOLS, in others it outperforms my finite-T estimator.

Using the dirac kernel with weight 1 at 0 lag and zero otherwise, $\widehat{\widehat{\Omega}} = \widehat{\widehat{\Sigma}}$ and $\frac{\widehat{\widehat{L}}_{2,1}}{\widehat{L}_{2,2}} = \frac{\widehat{\widehat{\Sigma}}_{2,1}}{\widehat{\Sigma}_{2,2}} = \psi_0 \frac{T-1}{T+1}$ and $y^* = y - \psi_0 \frac{T-1}{T+1} \varepsilon = (S + 2\frac{\psi_0}{T+1} I)\varepsilon$. The second OLS regression will generate $b^{FM} = 4\frac{\psi}{(T+1)^2}$. This bias grows with $\gimel$ and reaches the full OLS bias for $\gimel = T-1$.

b) Long-lasting feedback

Suppose that the feeback is characterised by a single $\psi_k \neq 0$. This leads to $b^{OLS} = 2\psi_k \frac{T-k}{T(T+1)}$, it induces a bias in the measure of the covariance of $\sum_{q=0}^{Q} [-b^{OLS} \cdot (T-q)] = -b^{OLS} \cdot \frac{(Q+1)(2T-Q)}{2}$.

We have $\mathbb{E}\widehat{\widehat{\Gamma}}_{2,1}^{(h)} = -2\psi_k \frac{(T-h)(T-k)}{T(T+1)}$ for $k \neq h \geq 0$. With a Dirac kernel, $\mathbb{E}\frac{\widehat{\widehat{\Omega}}_{2,1}^{OLS}}{\widehat{\Omega}_{2,2}} = -b = -2\psi_k \frac{T-k}{T+1}$, then $\widehat{\widehat{y}}^* = y + b\varepsilon = x + \psi_k \left(2\frac{T-k}{T+1} I + L_k\right) \varepsilon$.

Thus $\widehat{\widehat{\beta^{FM}}}$ has (an absolute) bias greater than OLS: $b^{FM} = 2\psi_k \frac{(T-k)(T+3)}{(T+1)^2} = b^{OLS} \cdot [1 + \frac{2}{T+1}]$ (keeping the same $\wp(1/T^2)$ approximation above). This explains the findings of PL91 that FMOLS sometimes increase rather than correct for the bias.

A.3.1 Multivariate applications

The methodology applies straightforwardly to a generic multivariate which in vector form reads:

$$\begin{cases} Y = \alpha + X\beta + Z\delta + u \\ \Delta X = \varepsilon \end{cases} \qquad \text{(M3)}$$

In this model, X can be multivariate (as long as there is no cointegration between regressors), Z may be a set of deterministic trends and constant, and possibly include additional exogenous variables. It is straightforward to eliminate deterministic trends, the constant and stationary exogenous variables by Frish-Waugh projections. In this case, the **(MA.3.1)** model projects into $\tilde{Y} = \tilde{X}\beta + \tilde{u}$ where $\tilde{Y} = R_{MZ}Y$ with $R_{MZ} = I - (MZ)\left(Z'MZ\right)^{-1} Z'M$

Because it is needed that there are no cointegration relationships between right-hand scale variables X and Z, simultaneous/multiple multivariate cointegration relationships cannot in general be analysed with within this framework. With three variables, if there are two cointegration relationships, all can be rewritten as pair-wise cointegration, but the analysis of a multivariate cointegration relationship would not be possible because projecting X and Y on the complement of Z would yield two stationary variables. It is thus needed that the different cointegration relationships are defined on different subsets of (Y,X).

We leave for an extension the possibility to extend the time-series techniques developed in this paper to the case of integrated instrumental variables, as well as with non-trivial forms of cross-sectional dependency.

Focusing now on the single-equation with a multivariate X that contains p $I(1)$ regressors, the bias correction previously developed reads in matrix and vector form for the Toy FMOLS: $\hat{e}(h,X) = \frac{1}{T}\sum_{k=0}^{h} w(k)X'ML^k\varepsilon$ where one sees easily that $\hat{e}(h,X)$ is a (p,p) matrix. The correction $\tilde{e}(h,X)$ is also straighforward to modify for FMOLS. Then $\widehat{\widehat{b^{FT}}} = (I-\tilde{e}(h,X))^{-1}\cdot\widehat{\widehat{b^{FM}}} = \widehat{df_{corr}}\cdot\widehat{\widehat{b^{FM}}}$, yields estimators with no conditional bias in the multivariate X (and can be applied to any super-consistent first-step estimation).

References

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

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

Banerjee, A., J. Dolado, D. Hendry, and G. Smith (1986). Exploring equilibrium relationships in econometrics through static models: Some monte carlo evidence. *Oxford Bulletin of Economics and Statistics 48*(3), 253–277.

Bao, Y. and A. Ullah (2007). The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics 140*(2), 650–669.

Breitung, J. and M. Pesaran (2008). Unit roots and cointegration in panels. *The Econometrics of Panel Data*, 279–322.

Engle, R. and C. Granger (1987). Cointegration and error-correction: Representation, estimation, and testing. *Econometrica 55*(2), 251–276.

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

Hashimzade, N. and T. J. Vogelsang (2008). Fixed-b asymptotic approximation of the sampling behaviour of nonparametric spectral density estimators. *Journal of Time Series Analysis 29*(1), 142–162.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control 12*(2-3), 231–254.

Kao, C. and M. Chiang (2001). On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics 15*, 179–222.

Kiviet, J. and G. Phillips (1993). Alternative bias approximations in regressions with a lagged-dependent variable. *Econometric Theory 9*, 62–62.

Mark, N. and D. Sul (2003). Cointegration vector estimation by panel dols and long-run money demand. *Oxford Bulletin of Economics and Statistics 65*(5), 655–680.

Moon, H. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of econometrics 122*(1), 81–126.

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. Phillips (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory 4*, 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.

Pesaran, M. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics 68*(1), 79–113.

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

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

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

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

Sender, S. (2014). *An exact finite-sample correction to fully-modified estimators of cointegrated panels.* Working paper.

Shao, X. (2011). A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics 162*(2), 213–224.

Smeekes, S. and J. Urbain (2014). A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing. Technical report.

Stock, J. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica: Journal of the Econometric Society*, 1035–1056.

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

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

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.