Working Papers / Documents de travail

Estimation and Testing for Fractional Cointegration

Marcel Aloy, Aix-Marseille Université, Greqam
Gilles de Truchis, Aix-Marseille Université, Greqam

WP 2012 - Nr 15

Estimation and Testing for Fractional Cointegration

Marcel Aloy [*] and Gilles de Truchis [†]

June 5, 2012

Abstract

Estimation of bivariate fractionally cointegrated models usually operates in two steps: the first step is to estimate the long run coefficient (β) whereas the second step estimates the long memory parameter (d) of the cointegrating residuals. We suggest an adaptation of the maximum likelihood estimator of Hualde and Robinson (2007) to estimate jointly β and d, and possibly other nuisance parameters, for a wide range of integration orders when regressors are $I(1)$. The finite sample properties of this estimator are compared with various popular estimation methods of parameters β (LSE, ADL, DOLS, FMLS, GLS, MLE, NBLS, FMNBLS), and d (LPE,LWE,LPM,FML) through a Monte Carlo experiment. We also investigate the crucial question of testing for fractional cointegration (that is, $d < 1$). The simulation results suggest that the one-step methodology generally outperforms others methods, both in terms of estimation precision and reliability of statistical inferences. Finally we apply this methodology by studying the long-run relationship between stock prices and dividends in the US case.

JEL classification: C32, C15, C53, C58.
Keywords: Fractional cointegration; Long memory; Monte Carlo experiment; Cointegration test.

1 INTRODUCTION

This paper deals with estimation of a generalized form of the standard triangular cointegration model:

$$(1-L)^d(y_t - \beta x_t) = \varepsilon_{1t}, \qquad (1-L)^\delta x_t = \varepsilon_{2t}, \qquad t = 1, ..., n \tag{1}$$

In equation (1), we assume that $\epsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}]^\top$ is a vector of $I(0)$ variables, $\omega_t = [y_t, x_t]$ a vector of $I(\delta)$ variables and we define $(1-L)^d$ by its binomial expansion

[*]Corresponding author. Aix-Marseille School of Economics, Aix-Marseille University, Château La Farge, Route des Milles, 13290 Aix-en-Provence, France (marcel.aloy@univ-amu.fr)

[†]Aix-Marseille School of Economics, Aix-Marseille University, 13290 Aix-en-Provence, France (gilles.de-truchis@univ-amu.fr)

$$
\begin{aligned}
(1-L)^d &= 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - ... \\
&= \sum_{k=0}^{+\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} L^k \qquad (2) \\
\Gamma(z) &= \int_0^{+\infty} t^{z-1} e^{-t} dt,
\end{aligned}
$$

where L is the lag operator and $(1-L)^d$ the fractional difference operator (denoted Δ^d). Equation (1) extends the traditional cointegration framework since we allow $d \in (0,1)$ and $\delta \in (0,1)$ to be real numbers, while $d = 0$ and $\delta = 1$ corresponds to the standard cointegration model. According to Granger (1986) and Engle and Granger (1987), cointegration arises when $d < \delta$. The case where $d < \delta$ implies a dimensionality reduction, revealing a near-zero-frequency correlation between the two series. Denoting $\lambda = \delta - d$ the reduction parameter and $C(\delta, \lambda)$ the cointegration relationship we can identify five cases:

i) strong fractional cointegration $C(\delta, \lambda)$: $1/2 < \delta \leq 1$, $0 \leq d < 1/2$ and $\lambda > 1/2$;

ii) weak fractional cointegration $C(\delta, \lambda)$: $0 < \delta \leq 1$, $0 < d < \delta$ and $\lambda < 1/2$;

iii) stationary fractional cointegration $C(\delta, \lambda)$: $0 < \delta < 1/2$, $0 \leq d < \delta$ and $\lambda < 1/2$;

iv) standard cointegration $C(1, 1)$: $\delta = 1$, $d = 0$ and $\lambda = 1$;

v) spurious regression: $1/2 \leq \delta \leq 1$, $1/2 \leq d \leq \delta$ and $\lambda = 0$.

The concepts of stationary and weak fractional cointegration result from properties of fractionally integrated process demonstrated by Granger and Joyeux (1980) and Hosking (1981): an $I(d)$ process is covariance stationary when $d < 0.5$ and covariance non-stationary when $d \geq 0.5$. As long a $d < 1$ the process is mean reverting. Consequently , in equation (1), $d < \delta \leq 1$ implies the existence of a reversion mechanism toward the long run equilibrium.

In the standard cointegration framework the long run coefficient β is the only parameter of interest and the literature provides many studies dealing with estimation of β in this specific case. Panopoulou and Pittis (2004) lead a finite sample comparison of the most commonly used estimators. In the more general case of fractional cointegration, parameters of interest in equation (1) are β, d and δ. Consequently, the joint estimation of parameters in equation (1) is tedious and existing studies focuse on only specific cases. The seminal paper of Cheung and Lai (1993) suggests a two-step methodology carrying out the Least Squares Estimator (LSE) and the Log-Periodogram regression Estimator (LPE) of d, developed by Geweke and Porter-Hudak (1983). It consists in estimating β in a first step and d, upon collected residuals, in a second step. Their approach cover the fractional cointegration case but they restrict $\delta = 1$. Marinucci and Robinson (2001) provide a good overview of fractional cointegration and investigate asymptotic and finite sample properties of the Narrow-Band Least Squares estimator of β (introduced by Robinson (1994)) for different parameter regions. Nielsen (2007) suggests a quasi-maximum likelihood estimator in order to estimate jointly β, d and δ. This local Whittle estimator operates in frequency domain and is consistent and asymptotically normal for the entire stationary region of δ and d (so it is restricted to the stationary fractional cointegration case). Finally, Hualde and Robinson (2007) present a time domain maximum likelihood estimator of β, d and δ. Hualde and Robinson (2007) demonstrate consistency and asymptotic normality of this estimator under assumptions of weak fractional cointegration. However their methodology operates in two non-linear optimizations.

The aim of this paper is to analyse various estimation strategies of bivariate fractional cointegration models under the restriction $\delta = 1$. This specific restriction is relevant to a wide range of financial and macroeconomics applications. For instance, Cheung and Lai (1993) bring out some evidence of cointegration at orders $C(1, 0.4)$ exploring the purchasing power parity relationship between United States and United Kingdom (see Henry and Zaffaroni (2003) for a survey of applications in macroeconomics and finance in presence of long-memory). Marinucci and Robinson (2001) analyze the consumption and the GNP in the U.S. case and find them integrated at order 1. However, they get inconclusive results concerning the integration order of cointegrating errors. Caporale and Gil-Alana (2002) test the cointegration hypothesis between unemployment and input prices and find some evidence of cointegration at order $C(1, \lambda)$, where λ depends on the residuals specification (AR(1), AR(2) or white noise). Baillie and Bollerslev (1994) investigate dynamic relationships between different exchange rate series and conclude to cointegration at orders $C(1, 0.11)$. Using the same times series, Nielsen (2004) tests the $C(1, 1)$ cointegration hypothesis against the alternative of fractional cointegration. This paper leads up to the crucial question of testing cointegration that constitutes our second point of interest. Cheung and Lai (1993) provide a finite sample analysis of a fractional cointegration test based upon the log-periodogram estimator. The empirical distribution of their test statistic under the null appears to be negatively skewed. As a consequence this cointegration test rejects the null hypothesis of no cointegration too often. Lobato and Velasco (2007) develop a fractional Wald test of fractional integration and suggest that this test can also be applied for testing fractional cointegration. Since they provide neither asymptotic nor finite sample properties of their test in the case of fractional cointegration, we include it in our simulations.

Under the assumption $\delta = 1$, there are two free parameter of interest: d and β. In the two-step methodology initiated by Cheung and Lai (1993), there is a wide range of candidate to estimate d. The main drawback of the semi-parametric frequency approach of Geweke and Porter-Hudak (1983) used by Cheung and Lai (1993) is a negative bias in small sample. Andrews and Guggenberger (2003) suggest a Log-Periodogram (LPM) Modified estimator in order to correct the bias. In comparison with the LPE and the LPM, the Local Whittle semi-parametric Estimator (LWE) of Künsch (1987) and Robinson (1995) has convenient asymptotic properties, but requires a numerical optimization. In a different spirit, a parametric frequency maximum likelihood (FML) approach is proposed by Fox and Taqqu (1986). Finally, we can can also mention the time domain maximum likelihood estimator of Sowell (1992). An exhaustive list of fractional integration estimator is considered in the Monte-Carlo experiment performed by Nielsen and Frederiksen (2005). Estimators we have mentioned above exhibit best finite sample properties according to their study (except Sowell (1992)).

The second parameter of interest is β. In the standard cointegration case, Stock (1987) demonstrates that the LSE is still consistent but super-convergent at the rate $O(n)$ rather than the optimal rate $O(n^{1/2})$. It implies that the LSE will be biased in small sample. This result is extended by Robinson and Marinucci (2001) to the fractional cointegration case: when $\delta = 1$, it can be shown that the LSE converges at the rate n^{λ} for $\lambda > 1/2$, $(n/\log(n))^{1/2}$ for $\lambda = 1/2$, and $n^{1/2}$ for $0 < \lambda < 1/2$. Hence, it easy to recover the standard results, for instance the spurious regression case when $\lambda = 0$, and the traditional cointegration case when $\lambda = 1$. A second issue is the inconsistency of the LSE estimator of β when cointegrating errors and regressors both have long memory and have non-zero coherence at the zero frequency. Consider the equation $y_t = \beta x_t + \upsilon_t$. One sufficient condition of consistency of the LSE estimator of β is that $\sum_{t=1}^{n} \upsilon_t^2 / \sum_{t=1}^{n} x_t^2$ converges stochastically to zero when $n \to \infty$. This condition is satisfied when $\delta > 1/2$ and $d < 1/2$, ore more generally when the residuals are "less nonstationary" than y_t and x_t (Robinson and Marinucci (2001)). However, the LSE are generally inconsistent when x_t and υ_t are correlated and $d < \delta < 1/2$: in this case, $\sum_{t=1}^{n} x_t^2$ does not dominate $\sum_{t=1}^{n} \upsilon_t^2$. Focusing on the stationary case, Robinson (1994) has developed a semi-parametric Narrow-Band Least Squares estimator (NBLS) in the frequency domain. NBLS exploits the dominance of the spectral density of x_t on υ_t at low frequencies (since cointegration implies $\delta > d$). However the NBLS estimator is confined to the case $\delta < 1/2$ (where y_t, x_t and z_t have finite variances) and in practice it is difficult to ensure that we are in the stationary regions of δ. In order to solve this drawback, Nielsen and Frederiksen (2011) suggest a Fully Modified Narrow-Band Least Squares (FMNBLS) to take into account bias that appears in the limiting distribution of the NBLS in presence of a non-zero long run coherence between regressors and cointegrating errors, in the weak fractional cointegration case ($\lambda < 1/2$). Their approach is equivalent to the so called time domain Fully-Modified Least Squares (FMLS) estimator of Phillips and Hansen (1990). Panopoulou and Pittis (2004) propose to exploit an Autoregressive Distributed Lags structure of the equation (1) to estimate β. They demonstrate the good finite sample properties of the LSE in the ADL form throughout simulations. They also investigate the Dynamic Ordinary Least Squares of Stock and Watson (1993) and show this estimator suffer from truncation bias in finite sample. An other candidate to estimate β is the time domain Generalized Least Squares (GLS) estimator. A frequency domain version of this parametric estimator is given by Robinson and Hidalgo (1997): they demonstrate root-n-consistency and asymptotic normality of this estimator when the true specification of errors is known. More recently, they adapt it to unknown specification of disturbance (see Hidalgo and Robinson (2002)). Finally, we can mention the well-known Maximum Likelihood Estimator (MLE) of Johansen (1988): since this estimator is not designed to estimate fractional cointegrated models, we can expect mixed results in non-stationary regions of cointegrating errors.

All estimators presented above are relevant either in estimation of β or in estimation of d. In the spirit of Nielsen (2007) and Hualde and Robinson (2007) we are interested in the joint estimation of parameters of interest, which are limited to β and d since we assume $\delta = 1$. We are also interested in estimating those parameters for a wide range of integration orders. Since we adopt a parametric approach, we cannot avoid issues of serial correlation and the lack of orthogonality between errors and regressors. In consequence, we suggest an adaptation of the model of Hualde and Robinson (2007) that is still relevant with the Maximum Likelihood Estimator (MLE). We also advise a selection procedure of the parametric form of this model. We investigate the finite sample properties of the MLE of β and d. The question of testing cointegration is also performed using the asymptotic variance, derived by Tanaka (1999), in the framework of a Wald test. We compare our one step methodology with most of estimators mentioned previously. The outline of the paper is as follows. The section 2 presents some technical aspects of the fractional cointegration framework and the model we consider, and then details various estimators used in our simulation. In the section 3 we comment our simulations and the results. We finally lead an empirical study on the present value model in the section 4. The section 5 concludes.

2 SIMULATION DESIGN

2.1 The fractional cointegration model

We consider now a generalized form of equation (1) that allows for serial correlation of errors and weak exogeneity of regressors. Let ω_t and ζ_t be two bivariate processes where $\omega_t = [y_t, x_t]^\top$ is a vector of $I(1)$ variables and $\zeta_t = [\upsilon_t, z_t]^\top$ a vector of their residuals. We assume that $[\Delta^d \upsilon_t, z_t]^\top$ is a two dimensional vector of $I(0)$ variables with a VAR(1) structure driven by $\epsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}]^\top$ and the data generating process (DGP) for y_t is given by the following system,

$$y_t = \beta x_t + \upsilon_t \tag{3}$$

$$x_t = x_{t-1} + z_t \tag{4}$$

$$\begin{pmatrix} (1-L)^d \upsilon_t \\ z_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} (1-L)^d \upsilon_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{5}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim i.i.d\ N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right] \tag{6}$$

Our analysis will be restricted to the case where $\alpha_{21} = \alpha_{12} = \alpha_{22} = 0$: if the two last restrictions are only simplifying assumptions, the first one (α_{21}=0) is crucial as it insures that x_t is weakly exogenous for β, since x_t does not adjust to past cointegration errors. In the standard $I(1)/I(0)$ framework, Johansen (1992) shows that under weak exogeneity of x_t, the single equation modelling is equivalent to the ML estimation of the full system. We assume further that eigenvalues of the matrix A=$[\alpha_{ij}]$, $i, j = 1, 2$ are less than one in modulus: under the restriction $\alpha_{21} = 0$, this hypothesis implies $\alpha_{11} < 1$ (and $\alpha_{22} < 1$ if one would relax our simplifying assumption $\alpha_{22} = 0$).

The contemporaneous correlation between the elements of e_t leads us to suggest a triangular representation in which we assume

$$\varepsilon_{1t} = \alpha \varepsilon_{2t} + \varepsilon_t \tag{7}$$

where,

$$\alpha = \sigma_{12}/\sigma_{22}, \qquad V(\varepsilon) = \sigma_{11} - (\sigma_{12}^2/\sigma_{22}) = \sigma_\varepsilon^2, \qquad \varepsilon_t \sim i.i.d\ N(0, \sigma_\varepsilon^2)$$

The process for the cointegration errors can thus be written more compactly as it follows

$$\Delta^d \upsilon_t = \alpha_{11}\Delta^d \upsilon_{t-1} + \alpha\varepsilon_{2t} + \varepsilon_t \tag{8}$$

Under our set of assumptions, it results that:

- x_t is a weakly exogenous I(1) process for β and follows a simple random walk;

- assuming $\beta \neq 0$, x_t and y_t are two I(1) fractional cointegrated processes provided that $d < 1$. These two processes are not cointegrated if $d = 1$, and are cointegrated under the usual I(1)/I(0) framework if $d = 0$;

- if x_t and y_t are two I(1) processes and are cointegrated under the I(1)/I(0) framework, the long-run multiplier between x_t and y_t is given by the parameter β while the short-run multiplier is given by the sum $\beta + \alpha$. It is thus important to consider the case where $\alpha \neq 0$ since it corresponds to a short-run undershooting ($\alpha < 0$) or overshooting ($\alpha > 0$) following a shock on x_t. It is also important to consider the case where $\alpha_{11} \neq 0$: if the two series are cointegrated under the I(1)/I(0) framework, the cointegrating errors will follow a simple and unrealistic white noise if $\alpha_{11} = 0$.

Restricting some coefficients of this model, we define four specifications which will be used later in our Monte Carlo study.

Table 1: Some parameters combinations

Model	α	α_{11}	α_{12}	α_{21}	α_{22}
A	0	0	0	0	0
B_1	0	> 0	0	0	0
B_2	< 0	0	0	0	0
B_3	< 0	> 0	0	0	0

The first model, denoted A, describes a bivariate relationship between y_t and x_t in line with Cheung and Lai (1993). Indeed, innovations are a simple fractional white noise integrated at order d while x_t is defined as a random walk with drift. The model B_1 corresponds to the case where $\alpha = 0$ and $\alpha_{11} > 0$: the cointegration residuals follows thus an ARFIMA(1,d,0) process with a positive AR coefficient (since $\alpha_{11} > 0$). Conversely, the model B_2 corresponds to the case of $\alpha_{11} = 0$ and $\alpha < 0$. In other terms, we introduce here the second order bias since $\alpha < 0$ breaks the orthogonality condition: $E[\upsilon_t|x_t] \neq 0$. In the specifications B_3, the model faces simultaneously serial correlation and engoneity in the cointegration residuals ($\alpha_{11} > 0$ and $\alpha < 0$) : innovations follow an ARFIMA(1,d,0) process with a positive AR coefficient and are adversely influenced by current shocks on x_t (since $\alpha < 0$).

2.2 Estimators

We will examine eight different estimators of the long run coefficient β: the ordinary least squares (LSE), the generalized least squares (GLS), the fully modified least squares (FMLS), the dynamic LSE (DOLS), the autoregressive distributed lags (ADL), the maximum likelihood estimator (MLE) of Johansen (1988) (although the methodology of Johansen (1988) is clearly limited to the standard I(1)/I(0) case, we considered that it was interesting to evaluate its performances in the case of fractional cointegration), the narrow band least squares (NBLS) and the fully modified narrow band least squares (FMNBLS).

Concerning the fractional differencing parameter (d) of the cointegrating errors, we will consider four estimators: the log-periodogram regression estimator (LPE), the modified log-periodogram estimator (LPM), the gaussian semi-parametric estimator (LWE) and the approximate Whittle Frequency domain Maximum Likelihood estimator (FML). Notice that these estimators of d have not been developed to be applied to estimated residuals. Moreover, some of them are semi-parametric and thus we can expect some biases in presence of short run dynamics in residuals. Alternatively, some of them are parametric and they potentially outperform semi-parametric estimators, provided that we are able to select the true underlying data generation process. In addition we suggest a parametric model, based on Tanaka (1999) and Hualde and Robinson (2007), that allows to estimate simultaneously β and d. It clearly appears as an alternative convenient way to address issues raised by the two-step methods. We denote it FCMLE.

2.2.1 Estimators of β

In the fractional cointegration case, the least squares estimates of the cointegrating parameter is consistent and converge in probability at the rate $O(n^{min(2\delta-1,\lambda)})$ (Robinson and Marinucci (2001) and Robinson and Hualde (2003)). Thus, the case where $\delta = 1$ and $\lambda = 1$ corresponds to the $I(1)/I(0)$ framework and we find back the convergence results demonstrated by Stock (1987). When $\delta = 1$ and $\lambda = 0$, least squares estimates does not converge and the cointegrating regression is spurious. Since we focus on the case $\delta = 1$, we don't cover the stationary case considered by Robinson (1994).

The fractional cointegration framework we consider involves serially dependent errors and thus non-spherical disturbances. Therefore, least squares estimator is no longer efficient and a convenient way to deal with this issue is to use GLS estimators. Since the covariance matrix Ω_υ is unknown, we can apply feasible GLS (FGLS). For instance, the Prais and Winsten (1954) method can be used to build the unknown covariance $\Omega(\varsigma)_\upsilon$, which is assumed to be dependent upon several unobservables parameters ς:

$$\hat{\Omega}_\upsilon^{-1}(\varsigma) = \frac{1}{\sigma_\upsilon^2}\begin{pmatrix} 1 & -\hat{\alpha}_{11} & 0 & 0 & \cdots & 0 \\ -\hat{\alpha}_{11} & 1+\hat{\alpha}_{11}^2 & -\hat{\alpha}_{11} & 0 & \cdots & 0 \\ 0 & -\hat{\alpha}_{11} & 1+\hat{\alpha}_{11}^2 & -\hat{\alpha}_{11} & \cdots & 0 \\ \vdots & \ddots & & & \ddots & \vdots \\ 0 & \cdots & 0 & -\hat{\alpha}_{11} & 1+\hat{\alpha}_{11}^2 & -\hat{\alpha}_{11} \\ 0 & \cdots & 0 & 0 & -\hat{\alpha}_{11} & 1 \end{pmatrix}$$

The FGLS estimator of β is thus:

$$\hat{\beta}_{GLS} = \left(\sum_{t=1}^{n} x_t \hat{\Omega}(\varsigma)_\upsilon^{-1} x_t^\top\right)^{-1} \left(\sum_{t=1}^{n} x_t \hat{\Omega}(\varsigma)_\upsilon^{-1} y_t\right) \tag{9}$$

When $\alpha \neq 0$, the orthogonality condition is no longer satisfied since $Cov(\upsilon_t|x_t) \neq 0$. Phillips and Hansen (1990) suggest a fully modify least square estimator (FMLS), applying non-parametric corrections to the covariance matrix to take into account endogeneity and serial correlation. The FMLS estimator of β is given by

$$\hat{\beta}_{FMLS} = \left(\sum_{t=1}^{n} x_t x_t^\top\right)^{-1} \left(\sum_{t=1}^{n} x_t(y_t - \hat{\lambda}^+) - T\hat{\delta}^+\right), \tag{10}$$

where $\hat{\lambda}^+$ is the correction term associated with the correction for the endogeneity bias while $\hat{\delta}^+$ eliminates the non-centrality bias. Assuming that the long-run variance of residuals is positive definite and can be expressed as $\Omega = \Lambda + \Xi^\top = \Sigma + \Xi + \Xi^\top$, we can partition Ω and Λ such as

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \Lambda_{22} \end{pmatrix},$$

where $\Sigma = \lim_{n\to\infty} 1/n \sum_{t=1}^{n} E(\upsilon_t \upsilon_t^\top)$ and $\Xi = \lim_{n\to\infty} 1/n \sum_{j=1}^{n-1} \sum_{t=1}^{n-j} E(\upsilon_t \upsilon_{t+j}^\top)$. It results that $\hat{\lambda}^+ = y_t - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1} z_t$ and $\hat{\delta}^+ = [0, \hat{\lambda}_{21} - \hat{\Lambda}_{22}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21}]^\top$, where z_t refers to the error term of x_t.

Alternatively, Stock and Watson (1993) consider a parametric correction based upon an augmented regression equation, in order to provide an estimator robust to endogeneity. The dynamic LSE (DOLS) estimator of β is

$$\hat{\beta}_{DOLS} = \left(\sum_{t=k+1}^{n-k} \tilde{u}_t \tilde{u}_t^\top\right)^{-1} \left(\sum_{t=k+1}^{n-k} \tilde{u}_t \tilde{y}_t\right), \tag{11}$$

where, $u_t = [1, x_t^\top]^\top$ and $\tilde{u}_t$ and $\tilde{y}_t$ are regression residuals of u_t and y_t on $\zeta_{2t} = (z_{t+k}^\top, ..., z_{t-k}^\top)^\top$, respectively (for a detailed review of these estimators, see Kurozumi and Hayakawa (2009)).

Johansen (1988) adopts a vectorial approach of the cointegration theory in a I(1)/I(0) framework. Starting from the long run equation, the vectorial error correction model is

$$(1 - L)\mathbf{y}_t = \mu + \Pi \mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \Theta_j \Delta \mathbf{y}_{t-j} + \mathbf{v}_t$$

where $\mathbf{y}_t = [y_t, x_t]^\top$ and $\mathbf{v}_t$ is an $k = 2 \times 1$ vector of innovations. This equation is clearly a combination of $I(1)$ and $I(0)$ variables. However, if the coefficient matrix Π has reduced rank $r < k$, then we can decompose Π as $k \times r$ matrix α and ξ such that $\Pi = \alpha\xi^\top$. It appears now that $\xi^\top \mathbf{y}_t$ is stationary since ξ is the matrix of cointegrating vectors, while α is the matrix of adjustment coefficients.

Let $\mathbf{u}_{1t}$ be the residuals in the regressions of $\Delta \mathbf{y}_t$ on $\mathbf{z}_t = \{\Delta \mathbf{y}_{t-1}, \Delta \mathbf{y}_{t-2}, ..., \Delta \mathbf{y}_{t-p+1}\}$ and $\mathbf{u}_{2t}$ the residuals in the regressions of $\mathbf{y}_{t-p}$ on $\mathbf{z}_t$. With respect to parameters $\Theta_1, ..., \Theta_{p-1}, \mu$, the maximum likelihood function is the following,

$$\Rightarrow L_{max}^{-2/n}(r) = |S_{00}| \prod_{j=1}^{r} (1 - \hat{\lambda}_j) \tag{12}$$

where the eigenvalue $\hat{\lambda}_j$ is the jth largest canonical correlation of $(1 - L)\mathbf{y}_t$ with $\mathbf{y}_{t-1}$ after correcting for lag differences and S_{ij} the cross-correlation matrix between variables in set i and set

j for $i, j = \mathbf{u}_{1t}, \mathbf{u}_{2t}$[1]. Notice that in a bivariate cointegration model, the Π matrix contains at most one cointegrating vector.

When $\lambda = 1$, the model defined by equations (3), (4), (5), and (6), can also be parametrized as an autoregressive distributed lag (ADL) model (see Panopoulou and Pittis (2004)). Hence the conditional density of y_t, with an ADL(1,1), can be expressed as

$$D(y_t|x_t, \omega_{t-1}, \theta) = N(\tilde{\beta}x_t + c_1 y_{t-1} + c_2 x_{t-1}, \sigma_e^2), \qquad \theta \equiv (c_0, c_1, c_2, \sigma_e^2),$$

with

$$\begin{aligned}
\tilde{\beta} &= \beta + \frac{\sigma_{12}}{\sigma_{22}} \\
c_1 &= \alpha_{11} - \alpha_{21}\frac{\sigma_{12}}{\sigma_{22}} = \alpha_{11} \\
c_2 &= \alpha_{12} - (\alpha_{22} + 1 - \alpha_{21}\beta)\frac{\sigma_{12}}{\sigma_{22}} - \alpha_{11}\beta = -\alpha_{11}\beta \\
\sigma_u^2 &= \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}
\end{aligned}$$

Hence, the ADL equation is given by

$$y_t = \tilde{\beta}x_t + c_1 y_{t-1} + c_2 x_{t-1} + e_t, \tag{13}$$

and the equation of β by

$$\beta = \frac{\tilde{\beta} + c_2}{1 - c_1}, \tag{14}$$

where error term e_t in this ADL representation is orthogonal to z_t. Panopoulou and Pittis (2004) show that the LSE of (13) can be employed to obtain an efficient estimator of β using (14).

In a seminal paper, Robinson (1994) provides a semi-parametric frequency domain narrow-band least squares (NBLS) estimator of β. Robinson (1994) restricts his analysis to the stationary region of δ and d, and deals with long-run non-zero coherence between υ_t and x_t. Since we impose $\delta = 1$, our simulation is not theoretically relevant with this framework. However, we are still interested in the finite sample performance of the NBLS when $\delta \geq 1/2$. Assume any non-negative real γ which transforms the following hypothetical model,

$$(1 - L)^{\gamma} y_t = \beta(1 - L)^{\gamma} x_t + (1 - L)^{\gamma}\upsilon_t,$$

into a stationary model. In the special case investigated by Robinson (1994), $\gamma = 0$ since it is assumed that variables of the model are covariance stationary. Then, the NBLS estimator of β is defined by:

$$\hat{\beta}_{NBLS}(m, \gamma) = \hat{F}_{xx}^{-1}(\gamma, 1, m)\hat{F}_{xy}(\gamma, 1, m), \tag{15}$$

[1] Lasak (2010) suggests to extend the Johansen MLE estimator to fractional processes. However, we do not consider this estimator since the asymptotic theory is still a work in progress.

where the average co-periodogram is

$$\hat{F}_{q,r}(\gamma,k,l) = \frac{2\pi}{n}\sum_{j=k}^{l} Re(I_{qr}(\gamma,\lambda_j)), \qquad 0 \leq k \leq l \leq n-1, \tag{16}$$

and the co-periodogram is

$$\hat{I}_{q,r}(\gamma,\lambda) = \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n}((1-L)^{\gamma}q_t)((1-L)^{\gamma}r_s)'e^{-i(t-s)\lambda} \tag{17}$$

Nielsen and Frederiksen (2011) extend the NBLS estimator to the case where $\gamma > 0$ (in this case, the time series are covariance non-stationary) and propose a fully modified NBLS (FMNBLS) estimator in order to correct the bias that appears in presence of non-zero coherence between errors and regressors. The FMNBLS estimator of β is expressed as

$$\tilde{\beta}_{FMNBLS}(m_3,\gamma) = \hat{\beta}_{NBLS}(m_3,\gamma) - \lambda_{m_3}^{-\hat{d}}\lambda_{m_3}^{\hat{\delta}}\lambda_{m_2}^{\hat{d}}\lambda_{m_2}^{-\hat{\delta}}\tilde{\Gamma}_{m_2}(\gamma), \tag{18}$$

where λ_j is the angular frequency. The estimators of the bias is given by

$$\tilde{\Gamma}(m_2,\gamma) = \tilde{F}_{xx}^{-1}(\gamma,m_0+1,m_2)\tilde{F}_{x\hat{\upsilon}}(\gamma,m_0+1,m_2), \tag{19}$$

where the modified average co-periodogram $\tilde{F}_{q,r}$ is

$$\tilde{F}_{q,r}(\gamma,k,l) = \frac{2\pi}{n}\sum_{j=k}^{l} Re(e^{i\lambda_j(d_q-d_r)/2}I_{qr}(\gamma,\lambda_j)), \qquad 0 \leq k \leq l \leq n-1 \tag{20}$$

Bandwidths can be chosen such as $m_0 = m_3 = 0.3$, $m_1 > 0.675$ and $m_2 \in (0.675, 0.85)$, when $\lambda_{min} = 0.4$ (λ refers here to the reduction parameter).
Nielsen and Frederiksen (2011) use the local Whittle estimator of Robinson (1995) to estimate δ with bandwidth m_1. Nielsen and Frederiksen (2011) demonstrate the asymptotic normality of this estimator, albeit δ is estimated beforehand.

2.2.2 Estimators of d

In this section we review some estimators of d that we consider in the Monte Carlo study[2]. Fist we recall some useful properties of long memory process. Let υ_t an autoregressive fractionally integrated moving average model, ARFIMA(p,d,q), defined by

$$\Phi(L)(1-L)^d\upsilon_t = \Psi(L)e_t, \qquad d \in (0, 1/2) \tag{21}$$

[2]The time domain exact maximum likelihood estimator (EML) of Sowell (1992) is not considered here. According to the Monte-Carlo study of Nielsen and Frederiksen (2005), finite sample performances of EML estimator strongly depend on specifications and has poor finite sample properties relatively to other parametric estimators like FML. Notice also that Dubois et al. (2004) investigate the EML-based fractional cointegration test and point out that the power of this test decreases when the sample size increases. They also bring out that the size of this test strongly depends on the short run coefficients and perform badly in some cases.

where

$$\Phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j, \qquad \Psi(L) = 1 + \sum_{j=1}^{q} \psi_j L^j,$$

are polynomials of orders p and q in the lag operator, with roots outside the unit circle. The autocorrelation function of the process in (21) decays at an hyperbolic rate $(2d-1)$ and thus satisfies

$$\rho_k \sim c_\rho k^{2d-1}, \qquad 0 < c_\rho < \infty, \qquad \text{as } k \to \infty \tag{22}$$

The autocovariance function of υ_t is defined by,

$$\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_\upsilon(\lambda) e^{-ik\lambda} d\lambda, \qquad \pi \geq \lambda \geq -\pi \tag{23}$$

where $f_\upsilon(\lambda)$ denotes the spectral density function of υ_t:

$$f_\upsilon(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{i\lambda}|^{-2d} \left| \frac{\Psi(e^{i\lambda})}{\Phi(e^{i\lambda})} \right|^2, \tag{24}$$

with $\lambda_j = \frac{2\pi j}{n}$ the angular frequencies. Under some additional conditions it can be shown that equation (22) is equivalent to the following approximation of (24), near zero frequency,

$$f_\upsilon(\lambda) \sim g|\lambda|^{-2d}, \qquad 0 < g < \infty, \qquad \text{as } k \to 0 \tag{25}$$

Exploiting those properties, Fox and Taqqu (1986) suggest a convenient parametric frequency maximum likelihood estimator (FML) based upon the spectral density of υ_t and the log-periodogram $I(\lambda)$,

$$I(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^{n} \upsilon_t e^{it\lambda} \right|^2, \tag{26}$$

The Whittle form of the log-likelihood function they obtain is:

$$\mathcal{L}(d, \Phi, \Psi, \sigma^2) = - \sum_{j=1}^{n/2} \left[\ln f_y(\lambda_j) + \frac{I(\lambda_j)}{f_y(\lambda_j)} \right], \tag{27}$$

Assuming gaussianity and $d \in \{-1/2, 1/2\}$, the FML estimator verifies asymptotically

$$\sqrt{n}(\hat{d}_{FML} - d) \to N\left(0, \frac{4\pi}{\Omega(\varsigma)}\right),$$

where $\Omega(\varsigma)$ is a $p \times p$ matrix, depending on unknown parameters ς and with j, kth entry such that,

$$\omega_{j,k}(\varsigma) = \int_{-\pi}^{\pi} f(\upsilon, \varsigma) \frac{\partial^2}{\partial \varsigma_j \partial \varsigma_k} f^{-1}(\upsilon, \varsigma) d\upsilon$$

The so called log-periodogram least square regression (LPE) of Geweke and Porter-Hudak (1983) is one of the most commonly applied estimator. While such estimator does not depend on short-run dynamics parameters, it is less efficient since only $\sqrt{m}$-consistency is achieved. From equation (25), Geweke and Porter-Hudak (1983) suggest to estimate the following equation using the LSE:

$$\ln I(\lambda_j) = c - d \ln(4\sin^2(\lambda_j/2)) + e(\lambda_j), \qquad j = 1, ..., m \tag{28}$$

The bandwidth number for the log-periodogram is arbitrarily fixed to $m = \sqrt{n}$. The LSE of $d \in (-1/2, 1/2)$ is asymptotically normally distributed such as:

$$\sqrt{m}(\hat{d}_{LPE} - d) \to N\left(0, \frac{\pi^2}{24}\right)$$

Andrews and Guggenberger (2003) investigate the bias source of the LPE estimator and suggest to replace the constant term c in (28) by the polynomial $\sum_{r=0}^{R} c_\tau \lambda_j^{2r}$, leading to the new regression model:

$$\ln I(\lambda_j) = \sum_{r=0}^{R} c_\tau \lambda_j^{2r} - d \ln(4\sin^2(\lambda_j/2)) + \varepsilon, \tag{29}$$

Thus, assuming that $d \in \{-1/2, 1/2\}$, the asymptotic distribution of this modified log-periodogram estimator (LPM) is

$$\sqrt{m}(\hat{d}_{LPM} - d) \to N\left(0, \frac{\pi^2}{24} c_R\right), \qquad c_{R=1} = 2.25, \qquad d \in \{-1/2, 1/2\}$$

Another semi-parametric approach is suggested by Künsch (1987) using a local Whittle approach and revisited by Robinson (1995). The Whittle form log-likelihood function of the local Whittle estimator (LWE) is defined as follows:

$$\mathcal{L}(g, d) = -\frac{1}{m} \sum_{j=1}^{m} \left[\ln(g\lambda_j^{-2d}) + \frac{I(\lambda_j)}{g\lambda_j^{-2d}} \right], \tag{30}$$

where $g\lambda^{-2d}$ corresponds to the approximation of the spectral density in (25). Robinson (1995) also demonstrates that the asymptotic distribution of the LWE is normal:

$$\sqrt{m}(\hat{d}_{LWE} - d) \to N(0, 1/4), \qquad d \in \{-1/2, 1/2\}$$

Various others long memory estimators have been developed during the last three decades (see Nielsen and Frederiksen (2005) for a Monte Carlo comparison of numerous ARFIMA(p,d,q) estimators). However, we limit ourselves to the most diffused and efficient.

2.2.3 A simultaneous estimator of β and d

Most of the time, the two-step methodology (e.g. Cheung and Lai (1993)) uses semi-parametric methods to estimate fractional integration parameters and parametric methods to estimate long-run parameters. This approach involves a loss of efficiency since the estimators are globally less than $\sqrt{n}$-consistent. According to Hualde and Robinson (2007) these parameters ought to be jointly estimated parametrically to achieve $\sqrt{n}$-consistency and they advise to use the Gaussian Maximum Likelihood Estimator in order to estimate δ, β and d. Outside the cointegration context, Tanaka (1999) proves that the time domain Gaussian maximum likelihood estimator of d is asymptotically efficient and normally distributed. He shows that $\sqrt{n}(\hat{d} - d) \to N(0, 6/\pi^2)$ when υ_t is a fractional white noise, and that $\sqrt{n}(\hat{d} - d) \to N(0, \omega^{-2})$, where ω is detailed in (39), when υ_t follows a process similar to (21). Moreover, Tanaka (1999) proposes a two-step parametric cointegration methodology based on the LSE of β and the MLE of d. We suggest to estimate jointly β and d by MLE, assuming $\delta = 1$ and Gaussian errors. First we rewrite the model (8) using the following reparametrization

$$\begin{aligned} \varepsilon_t &= \Delta^{\tilde{d}} y_t + \Delta^d (y_{t-1} - \beta x_{t-1}) - \beta \Delta^{\tilde{d}} x_t \\ &\quad - \sum_{j=1}^{p} \alpha_{11}^{j} (\Delta^d y_{t-j} - \beta \Delta^d x_{t-j}) - \sum_{l=1}^{m} \left[\alpha^l \Delta^{\tilde{d}} x_{t-l+1} + \alpha^l (\Delta x_{t-l+1} - \Delta^{\tilde{d}} x_{t-l+1}) \right], \end{aligned} \tag{31}$$

where $\tilde{d} = d + \delta = d + 1$ and Δ^d is the binomial expansion defined in (2). The Gaussian MLE of (31), denoted FCMLE, is defined as

$$\hat{\theta} = \arg\min_{\theta} f_\theta(x), \qquad d \in \mathscr{D}, \tag{32}$$

where $\theta = (d, \beta)'$ is the vector of parameters of interest, $\mathscr{D} \subseteq (0, 1)$ and the log likelihood function of $f_\theta(x)$ is given by

$$\mathcal{L}(\theta \mid \mathbf{y}) = -\frac{1}{2} \sum_{t=1}^{n} \left(\ln(2\pi) + \ln|\Omega| + \varepsilon_t \Omega^{-1} \varepsilon_t' \right) \tag{33}$$

Proving consistency and limiting distribution of the Gaussian MLE of (31) is outside the scope of this paper. Difficulties in studying asymptotic theory when $\delta > 1/2$ are briefly discussed in Hualde and Robinson (2007) and motivates a Monte-Carlo experiment (the use of Monte Carlo simulations when appropriate asymptotic theory is not sufficiently developed is advocated by da Silva and Robinson (2008)). However, the Gaussian MLE of (31) raises some important issues. For instance, our estimator is subject to the uniform convergence problem described in Saikkonen (1995) : in cointegrated systems, different rates of convergence can apply in different directions of the parameter space. Consequently, we are not sure that the $\sqrt{n}$-consistency is achieved for all parameters of interest. One solution to overcome this problem is to concentrate out β (Saikkonen (1995), Robinson and Hualde (2003)). Nevertheless, we choose to not concentrate β out and the numerical simulations presented in section 3 suggest that, at least empirically, this issue does not result in major consequences. In addition, our simulation results permit the interpretation that the Gaussian MLE of $\hat{\beta}(\delta, \hat{d})$ is root-n-consistent, since Robinson and Hualde (2003) show that the asymptotic distributions of $\hat{\beta}(\delta, d)$ and $\hat{\beta}(\delta, \hat{d})$ are equivalent as long as $\hat{d}$ is root-n-consistent.

This model is fully parametric: in order to select the most appropriate model, we suggest to perform an unconstrained estimation and then to proceed to a sequential model selection of lag structure using the t-statistics. Since the true parametric model is unknown, this procedure can generate some selection bias that will be assessed in our Monte Carlo study.

2.3 Fractional cointegration tests

Given the estimation methods of the key parameters β and d discussed above, it is an interesting point to test for fractional cointegration hypothesis. Since both series, y_t and x_t, are assumed to be $I(1)$, the question is to test for the hypothesis:

H_0: $d = 1$ (y_t and x_t are not cointegrated)

H_1: $d < 1$ (y_t and x_t are cointegrated)

In the standard cointegration framework cointegrating errors are assumed to be $I(1)$ or $I(0)$. Unfortunately, usual unit root test have low power against the fractional alternative (see Dittman (2000) for a comparison of some residuals-based tests for fractional cointegration).

To deal with this issue, Lobato and Velasco (2007)[3] propose a fractional integration Wald test. We apply this test on cointegration residuals in order to test the fractional cointegration hypothesis (Lobato and Velasco (2007) mention this possibility in the conclusion of their article). Considering the simplest case where the cointegrating errors follow an fractional white noise, we can rewrite $(1-L)^d \upsilon_t = \varepsilon_t$ as $(1-L)\upsilon_t = (1-(1-L)^{d-1})(1-L)\upsilon_t + \varepsilon_t$ and thus,

$$(1-L)\upsilon_t = \varphi_2((1-L)^{d-1} - 1)(1-L)\upsilon_t + \varepsilon_t, \tag{34}$$

where $\varphi_2 = 0$ under the null hypothesis and $\varphi_2 = -1$ under the alternative hypothesis. To make the regressor continuous when $d = 1$, the equation (34) must be re-written:

$$(1-L)\upsilon_t = \varphi_2 v_{t-1}(d) + \varepsilon_t, \qquad v_{t-1}(d) = \frac{(1-L)^{d-1} - 1}{1-d}(1-L)\upsilon_t \tag{35}$$

Lobato and Velasco (2007) also consider the case where $(1-L)^d$ is serially correlated. The testing equation is in this case

$$(1-L)\upsilon_t = \varphi_2\left(\Phi(L)v_{t-1}(d)\right) + \sum_{j=1}^{p} \alpha_j (1-L)\upsilon_{t-j} + \varepsilon_t, \tag{36}$$

where $v_{t-1}(d)$ is given in equation (35). Lobato and Velasco (2007) show that estimation of α is consistent and under the null hypothesis, a simple *t-test* can be used.

We also consider LPE, LWE and FML residual-based test of fractional cointegration (for all semi-parametric tests we select a bandwidth equal to $n^{0.5}$). Following the methodology of

[3]Notice that Dolado et al. (2002) suggest a fractional Dickey-Fuller test, but Lobato and Velasco (2007) show that this test is inefficient. Moreover, their tests are locally asymptotically equivalent to the optimal Lagrange multiplier tests of Robinson (1991).

Cheung and Lai (1993), we lead the LPE residual-based test using the following statistic under the null hypothesis of no cointegration:

$$T_{LPE} = \frac{(\hat{d}-1)}{\hat{\sigma}}, \qquad \hat{\sigma}^2 = \frac{m\pi^2}{6 \det XX'}, \qquad X_k = \left(1, -2\ln\left(2\sin\left(\frac{\lambda_k}{2}\right)\right)\right)$$

Since residuals are estimated, critical values are non-standard. We use here critical values tabulated by Sephton (2002). The LWE et FML residual-based tests follow the same methodology, the only difference being that we use critical values from a Student distribution in order to test $(1-\hat{d}_\upsilon) = 0$ under the null hypothesis.

In the case of the the Gaussian MLE, we apply the fractional integration Wald test introduced in Tanaka (1999). The test statistic depends on whether errors are serially dependent or not. When errors follow a fractional white noise (that is, $(1-L)^d \upsilon_t = \varepsilon_t$) the test statistic is,

$$W_{FCMLE} = \sqrt{n} \times \frac{(\hat{d} - d_{H_0:d=1})}{\sqrt{6/\pi^2}} \tag{37}$$

Conversely, in the case when errors are autocorrelated the test statistic is,

$$W_{FCMLE} = \sqrt{n} \times \hat{\omega} \times (\hat{d} - d_{H_0:d=1}), \tag{38}$$

where $\hat{\omega}$ depends on the Fisher information matrix and the ARMA structure of (31). $\hat{\omega}$ can be easily computed in the simplest case of one AR lag:

$$\hat{\omega}^2 = \frac{\pi^2}{6} - \frac{1-\alpha_{11}^2}{\alpha_{11}^2}(\log(1-\alpha_{11}))^2 \tag{39}$$

3 MONTE-CARLO SIMULATIONS

Several Monte Carlo studies concerning estimators of β and/or d and/or δ are given in the literature. For instance, Robinson and Hualde (2003) investigate the finite sample properties of a GLS estimator of β when $\delta - d > 1/2$. Hualde and Robinson (2007) explore the finite sample behavior of a MLE of β, d and δ when $\delta - d < 1/2$. However they restrict their experiment to the comparison with the LSE of β and they do not provide any comparison with other long memory parameters estimators. da Silva and Robinson (2008) focus on nonlinearity and compare two narrow band estimators of β and three frequency estimators of d, without relaxing orthogonality conditions. They find out that optimal choice depends on the specification: for instance they show that the weighted narrow band least squares estimator has generally both lowest bias and lowest root mean squared error (RMSE). However, in presence of nonlinearity, the simple narrow band least squares estimator of Robinson (1994) is shown to be optimal. Concerning the fractional integration parameter of cointegrating residuals, it appears that the modified local Whittle estimator is not systematically the optimal choice, considering the finite sample performance of the simple local Whittle approach. Marinucci and Robinson (2001) compare the NBLS with the LSE and conclude to the superiority of the NBLS in terms of bias and MSE. Another interesting approach is led by Kurozumi and Hayakawa (2009), considering only the $I(1)/I(0)$ case. They analytically explain the poor finite sample performance of three efficient estimators (i.e. Phillips and Hansen (1990), Stock and Watson (1993) and the canonical cointegrating regression method) with a moderate serial correlation. They finally explain why finite samples performance of these estimators are clearly different while they are asymptotically equivalent. Remaining in the $I(1)/I(0)$ paradigm, Panopoulou and Pittis (2004) compare several time domain estimators of β.

3.1 Parametrization

We suggest a Monte Carlo study in order to investigate finite sample properties of both first and second step estimators described above, plus the one step FCMLE. As in Nielsen and Frederiksen (2005), we generate 1000 replications of artificial time series, with 512 observations, from the data generation process (DGP) given by equations (3), (4), (5), and (6) where $\beta = 1.0$. We consider seven values of d which represent different long range dependence behaviors: $d = \{0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 1.0\}$. The different values of the nuisance parameters α and α_{11} are detailed below:

Table 2: The different parametrizations

Model	α	α_{11}	α_{12}	α_{21}	α_{22}
A	0	0	0	0	0
B_1	0	0.8	0	0	0
B_2	−0.5	0	0	0	0
B_3	−0.5	0.8	0	0	0

For each estimators, we compute the root mean squared error (RMSE), defined by RMSE= $\frac{1}{I}\sum_{i=1}^{I}(E[\widehat{d_i} - d]) \equiv var(\widehat{d}) + (Bias[\widehat{d|d}])^2$, and the bias in order to compare finite sample performances[4]. Values in bold indicate the smallest RMSE. Concerning fractional cointegration tests, results are reported in percentage of rejection of the null hypothesis at a threshold of 5%. The critical value at 5% used for LPE residual-based is −2.25 (see Sephton (2002)). The test of Lobato and Velasco (2007) is based on the FML estimator of d. Bandwidths are arbitrary fixed to $n^{0.5}$ for all semi-parametric frequency estimators of d. The optimal lags of ADL, MLE and FML, are selected using the bayesian information criterion (BIC) of Schwarz (1978). Others information criterion have also been envisaged, however, there is no particular changes in the results. Regarding the two-step methodologies, estimation of d are performed using the LSE cointegrating residuals. Concerning the nuisance parameters of in the FCMLE, the model selection is based on sequential *t-tests*. In order to extract an approximation of the performance of our selection procedure, we also perform the FCMLE under the hypothesis that the true DGP is known (FCMLE$_k$) and compare it with the selected model FCMLE$_u$.

[4] All computations are performed using RATS 8.01. RATS codes and unreported results are available upon request.

Table 3: Residuals face neither serial correlation nor endogeneity ($\alpha_{11} = \alpha = 0$)

	Model *A*		I(1)/I(0)		Strong fractional				Weak fractional						Spurious	
	d		0,0		0,2		0,4		0,6		0,8		0,9		1,0	
n	θ	Est.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
512	$\hat{\beta}$	LSE	0,000	**0,006**	0,000	0,013	0,000	0,030	-0,002	0,079	0,005	0,197	-0,006	0,344	-0,008	0,571
		ADL	0,000	**0,006**	0,000	0,013	0,000	0,032	-0,003	0,085	0,004	0,209	-0,010	0,364	-0,013	0,599
		DOLS	0,000	**0,006**	0,000	0,013	0,000	0,031	-0,002	0,080	0,005	0,200	-0,006	0,350	-0,008	0,580
		FMLS	0,000	**0,006**	0,000	0,013	0,000	0,030	-0,002	0,080	0,005	0,199	-0,006	0,348	-0,008	0,577
		GLS	0,000	**0,006**	0,000	**0,012**	0,000	0,028	-0,001	0,055	0,002	**0,049**	0,001	**0,044**	0,001	**0,042**
		MLE	0,000	**0,006**	0,000	0,013	0,000	0,034	0,136	4,766	-0,275	9,573	-0,142	4,945	-0,597	25,959
		NBLS	0,001	0,020	0,000	0,021	0,001	0,032	0,001	0,054	0,001	0,096	0,006	0,167	-0,004	0,244
		FMNBLS	0,001	0,020	0,000	0,019	-0,001	0,020	0,000	0,034	0,003	0,081	-0,005	0,188	-0,002	0,360
		$FCMLE_u$	0,000	**0,006**	0,001	0,014	0,000	**0,019**	0,001	**0,030**	-0,002	0,059	-0,011	0,120	-0,017	0,224
		$FCMLE_k$	0,000	**0,006**	0,000	0,014	0,000	0,018	0,002	0,028	0,002	0,042	0,001	0,044	0,001	0,043
	$\hat{d}_v$	LPE	-0,042	0,184	-0,038	0,180	-0,044	0,191	-0,043	0,194	-0,038	0,190	-0,052	0,190	-0,071	0,191
		LWE	-0,057	0,159	-0,049	0,157	-0,047	0,173	-0,033	0,172	-0,028	0,161	-0,037	0,155	-0,049	0,156
		LPM	-0,076	0,318	-0,073	0,327	-0,087	0,334	-0,097	0,346	-0,102	0,347	-0,112	0,343	-0,133	0,336
		FML	-0,019	**0,057**	-0,019	**0,056**	-0,016	**0,060**	-0,010	0,065	-0,002	0,052	-0,003	**0,053**	-0,022	0,053
		$FCMLE_u$	-0,019	0,061	-0,030	0,075	-0,017	0,061	-0,014	**0,060**	-0,009	**0,051**	-0,002	0,059	-0,011	**0,051**
		$FCMLE_k$	-0,014	0,038	-0,016	0,048	-0,010	0,038	-0,010	0,038	-0,006	0,040	0,004	0,040	-0,007	0,032

LSE residual-based tests	0,0		0,2		0,4		0,6		0,8		0,9		1,0	
	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1
LPE	0,999	0,001	0,987	0,013	0,898	0,102	0,637	0,363	0,245	0,755	0,131	0,869	0,069	0,931
LWE	1,000	0,000	1,000	0,000	0,993	0,007	0,910	0,090	0,521	0,479	0,288	0,712	0,129	0,871
LV	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,971	0,029	0,553	0,447	0,125	0,875
FML	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,815	0,185	0,219	0,781	0,028	0,972
$FCMLE_u$	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,992	0,008	0,855	0,145	0,074	0,926
$FCMLE_k$	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,839	0,161	0,056	0,944

Table 4: Residuals face serial correlation ($\alpha_{11} = 0.8, \alpha = 0$)

		Model B_1	I(1)/I(0)		Strong fractional				Weak fractional						Spurious	
		d	0,0		0,2		0,4		0,6		0,8		0,9		1,0	
n	θ	Est.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
512	$\hat{\beta}$	LSE	0,000	0,027	0,001	0,061	0,001	0,149	-0,008	0,393	0,024	0,983	-0,028	1,713	0,082	2,780
		ADL	-0,001	0,029	0,001	0,065	-0,003	0,160	-0,017	0,419	0,029	1,039	-0,027	1,801	0,106	2,931
		DOLS	0,000	0,028	0,001	0,063	0,001	0,152	-0,009	0,400	0,025	1,000	-0,029	1,743	0,085	2,824
		FMLS	0,000	0,028	0,002	0,062	0,001	0,151	-0,008	0,398	0,025	0,994	-0,028	1,734	0,084	2,809
		GLS	0,000	**0,023**	0,001	**0,035**	0,000	**0,043**	0,001	**0,047**	0,002	0,053	0,001	0,061	0,001	0,069
		MLE	-0,001	0,032	0,003	0,076	0,001	0,371	0,337	12,998	0,667	12,965	2,935	88,529	0,985	75,996
		NBLS	0,004	0,071	0,001	0,088	0,004	0,144	0,006	0,261	0,006	0,472	0,026	0,821	-0,025	1,214
		FMNBLS	0,002	0,075	0,002	0,081	0,003	0,169	0,008	0,520	0,003	0,962	0,027	1,573	-0,017	2,194
		FCMLE$_u$	0,000	0,030	-0,002	0,048	-0,006	0,074	-0,005	0,056	-0,001	**0,043**	-0,003	**0,045**	-0,004	**0,063**
		FCMLE$_k$	0,000	0,030	0,001	0,041	0,000	0,042	0,000	0,040	0,002	0,035	0,000	0,035	0,001	0,032
	$\hat{d}_v$	LPE	0,109	0,209	0,114	0,210	0,109	0,214	0,101	0,214	0,088	0,204	0,059	0,188	0,021	0,184
		LWE	0,104	0,183	0,117	0,196	0,123	0,211	0,127	0,205	0,126	0,198	0,101	0,181	0,084	0,175
		LPM	-0,050	0,308	-0,048	0,320	-0,056	0,326	-0,074	0,338	-0,083	0,340	-0,088	0,329	-0,114	0,343
		FML	0,024	0,147	0,038	**0,177**	0,017	0,180	0,014	0,278	0,249	0,389	0,252	0,319	0,147	0,208
		FCMLE$_u$	0,019	**0,143**	0,013	0,184	0,012	**0,149**	0,042	**0,167**	0,062	**0,164**	0,040	**0,111**	-0,005	**0,078**
		FCMLE$_k$	0,017	0,134	0,010	0,176	0,012	0,148	0,041	0,163	0,058	0,154	0,036	0,105	-0,008	0,070

LSE residual-based tests	0,0		0,2		0,4		0,6		0,8		0,9		1,0	
	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1
LPE	0,993	0,007	0,937	0,063	0,702	0,298	0,323	0,677	0,097	0,903	0,044	0,956	0,025	0,975
LWE	1,000	0,000	0,998	0,002	0,949	0,051	0,656	0,344	0,173	0,827	0,080	0,920	0,040	0,960
LV	1,000	0,000	0,996	0,004	0,952	0,048	0,696	0,304	0,142	0,858	0,036	0,964	0,030	0,970
FML	1,000	0,000	0,972	0,028	0,927	0,073	0,762	0,238	0,226	0,774	0,068	0,932	0,023	0,977
FCMLE$_u$	1,000	0,000	0,994	0,006	0,974	0,026	0,852	0,148	0,449	0,551	0,194	0,806	0,050	0,950
FCMLE$_k$	1,000	0,000	0,999	0,001	0,973	0,027	0,850	0,150	0,450	0,550	0,197	0,803	0,050	0,950

Table 5: Residuals face non-zero coherence with regressors ($\alpha_{11} = 0, \alpha = -0.5$)

	Model B_2		I(1)/I(0)		Strong fractional				Weak fractional						Spurious	
	d		0,0		0,2		0,4		0,6		0,8		0,9		1,0	
n	θ	Est.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
512	$\hat{\beta}$	LSE	-0,005	0,009	-0,011	0,018	-0,028	0,044	-0,070	0,111	-0,168	0,267	-0,270	0,447	-0,414	0,709
		ADL	0,000	**0,006**	-0,004	**0,014**	-0,019	0,038	-0,060	0,108	-0,160	0,272	-0,269	0,462	-0,412	0,732
		DOLS	0,000	**0,006**	-0,006	**0,014**	-0,022	0,039	-0,064	0,107	-0,164	0,266	-0,267	0,450	-0,413	0,716
		FMLS	0,000	**0,006**	-0,006	0,015	-0,023	0,040	-0,065	0,108	-0,165	0,266	-0,268	0,449	-0,413	0,714
		GLS	-0,005	0,009	-0,014	0,021	-0,055	0,086	-0,390	0,413	-0,475	0,478	-0,493	0,495	-0,497	**0,498**
		MLE	0,000	**0,006**	-0,004	**0,014**	-0,015	0,038	-0,049	0,310	-0,176	2,948	-0,085	13,876	-0,238	25,164
		NBLS	0,000	0,022	-0,005	0,025	-0,018	0,041	-0,045	0,086	-0,116	0,193	-0,160	0,294	-0,256	0,551
		FMNBLS	0,001	0,026	-0,001	0,021	-0,003	0,026	-0,008	**0,045**	-0,024	**0,131**	-0,014	**0,260**	-0,180	0,608
		FCMLE$_u$	0,000	**0,006**	-0,001	0,016	-0,001	**0,023**	0,001	0,046	-0,074	0,206	-0,259	0,377	-0,465	0,539
		FCMLE$_k$	0,000	0,006	0,000	0,016	-0,001	0,023	0,001	0,046	0,003	0,132	0,019	0,849	0,362	3,041
	$\hat{d}_v$	LPE	-0,024	0,179	-0,018	0,168	-0,038	0,189	-0,045	0,186	-0,036	0,187	-0,054	0,191	-0,066	0,188
		LWE	-0,038	0,150	-0,033	0,145	-0,042	0,167	-0,041	0,171	-0,029	0,156	-0,040	0,155	-0,042	0,153
		LPM	-0,038	0,304	-0,039	0,304	-0,073	0,327	-0,086	0,328	-0,089	0,336	-0,112	0,347	-0,131	0,352
		FML	-0,014	**0,045**	-0,016	**0,049**	-0,017	**0,059**	-0,016	0,052	-0,009	0,057	-0,006	**0,053**	-0,021	**0,048**
		FCMLE$_u$	-0,015	0,054	-0,023	0,065	-0,014	0,064	-0,011	**0,047**	-0,001	**0,056**	0,019	0,065	-0,009	0,055
		FCMLE$_k$	-0,011	0,032	-0,012	0,041	-0,008	0,035	-0,009	0,036	-0,008	0,046	-0,005	0,047	-0,013	0,055

LSE residual-based tests	0,0		0,2		0,4		0,6		0,8		0,9		1,0	
	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1
LPE	1,000	0,000	0,987	0,013	0,907	0,093	0,642	0,358	0,242	0,758	0,136	0,864	0,058	0,942
LWE	1,000	0,000	0,999	0,001	0,988	0,012	0,931	0,069	0,533	0,467	0,300	0,700	0,128	0,872
LV	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,972	0,028	0,566	0,434	0,104	0,896
FML	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,850	0,150	0,237	0,763	0,032	0,968
FCMLE$_u$	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,988	0,012	0,611	0,389	0,041	0,959
FCMLE$_k$	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,911	0,089	0,083	0,917

Table 6: Residuals face both serial correlation and endogeneity ($\alpha_{11} = 0.8, \alpha = -0.5$)

	Model B_3		I(1)/I(0)		strong fractional				weak fractional						Spurious	
	d		0,0		0,2		0,4		0,6		0,8		0,9		1,0	
n	θ	Est.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
512	$\hat{\beta}$	LSE	-0,022	0,041	-0,053	0,089	-0,135	0,214	-0,335	0,541	-0,809	1,312	-1,295	2,195	-1,985	3,498
		ADL	-0,008	**0,031**	-0,034	**0,077**	-0,112	0,206	-0,323	0,553	-0,791	1,344	-1,290	2,268	-1,985	3,633
		DOLS	-0,014	0,034	-0,046	0,083	-0,128	0,210	-0,332	0,545	-0,814	1,329	-1,311	2,231	-2,011	3,550
		FMLS	-0,016	0,036	-0,048	0,085	-0,130	0,212	-0,333	0,544	-0,814	1,324	-1,308	2,220	-2,006	3,534
		GLS	-0,458	0,463	-0,475	0,477	-0,489	0,492	-0,496	0,499	-0,499	0,502	-0,498	0,502	-0,501	0,507
		MLE	-0,001	0,032	-0,011	0,080	-0,044	0,239	-0,282	2,749	-0,682	17,404	-3,484	84,141	2,306	82,262
		NBLS	-0,005	0,079	-0,038	0,110	-0,103	**0,205**	-0,240	0,439	-0,582	0,966	-0,795	1,457	-1,250	2,728
		FMNBLS	0,005	0,088	-0,016	0,102	-0,006	0,209	-0,239	0,653	-1,048	1,557	-1,668	2,396	-2,684	3,949
		FCMLE$_u$	0,001	0,042	-0,013	0,120	-0,059	0,211	-0,053	**0,207**	-0,008	**0,103**	-0,008	**0,101**	-0,001	**0,145**
		FCMLE$_k$	0,001	0,042	0,004	0,109	0,005	0,122	0,002	0,133	-0,006	0,111	-0,012	0,129	-0,012	0,154
	$\hat{d}_v$	LPE	0,128	0,217	0,133	0,216	0,117	0,218	0,104	0,207	0,087	0,200	0,047	0,189	0,007	0,185
		LWE	0,121	0,192	0,132	0,200	0,127	0,209	0,126	0,204	0,123	0,194	0,091	0,175	0,063	0,170
		LPM	-0,011	0,296	-0,013	0,302	-0,042	0,321	-0,056	0,327	-0,066	0,330	-0,085	0,332	-0,109	0,344
		FML	0,031	0,154	0,045	0,178	0,028	0,182	0,029	0,270	0,247	0,382	0,223	0,298	0,119	0,186
		FCMLE$_u$	0,016	**0,128**	0,007	**0,156**	0,008	**0,142**	0,030	**0,157**	0,057	**0,149**	0,047	**0,111**	-0,004	**0,084**
		FCMLE$_k$	0,015	0,122	0,004	0,145	0,008	0,130	0,025	0,144	0,056	0,148	0,043	0,104	-0,008	0,065

LSE residual-based tests	0,0		0,2		0,4		0,6		0,8		0,9		1,0	
	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1	H0	H1
LPE	0,990	0,010	0,930	0,070	0,704	0,296	0,339	0,661	0,088	0,912	0,055	0,945	0,039	0,961
LWE	0,999	0,001	0,998	0,002	0,943	0,057	0,640	0,360	0,185	0,815	0,083	0,917	0,048	0,952
LV	1,000	0,000	0,993	0,007	0,949	0,051	0,664	0,336	0,119	0,881	0,062	0,938	0,050	0,950
FML	0,999	0,001	0,973	0,027	0,935	0,065	0,757	0,243	0,208	0,792	0,076	0,924	0,019	0,981
FCMLE$_u$	1,000	0,000	0,991	0,009	0,978	0,022	0,884	0,116	0,442	0,558	0,159	0,841	0,048	0,952
FCMLE$_k$	1,000	0,000	1,000	0,000	0,982	0,018	0,895	0,105	0,443	0,557	0,160	0,840	0,048	0,952

3.2 Simulation Results

3.2.1 The $I(1)/(0)$ case

The first columns of tables 3 to 6 correspond to the traditional cointegration case ($\delta = 1$ and $d = 0$). When cointegrating errors follow a simple white noise (table 3), most estimators of β (LSE, ADL, DOLS, FMLS, GLS, MLE and FCMLE) perform well with similar bias and RMSE. Adding short-run noise (table 4), the RMSE is deteriorated, but these estimation methods seem fairly robust to endogeneity (table 5). The worst performances are observed when residuals face both autocorrelation and endogeneity (table 6).

Concerning semi-parametric estimators of d based on cointegrating residuals, it is well known that in the presence of short-run dynamics, a smaller bandwidth is more appropriate to avoid biased results. Since we have selected a bandwidth equal to $T^{0.5}$, these methods are moderately impacted by autocorrelation (table 4 and 6). Overall, FML estimator and FCMLE exhibit lowest bias and RMSE.

3.2.2 The fractional cointegration case with stationary residuals

In this case, most estimators of β (LSE, ADL, DOLS, FMLS, GLS, MLE and FCMLE) exhibit similar performances, both in terms of bias and RMSE. It may be noted that the RMSE of these estimators increases gradually as d increase. When residuals face endogeneity, $\sum_{t=1}^{n} y_t^2 / \sum_{t=1}^{n} x_t^2$ no longer dominates asymptotically $\sum_{t=1}^{n} \upsilon_t^2 / \sum_{t=1}^{n} x_t^2$ and biases of most of estimators of β increase as long as d goes up. The behavior of the estimators of d is very similar to the $I(1)/I(0)$ case. Figure 1 compares the histograms of two estimators of β (LSE and FCMLE) and d (LPR and FCMLE) in the model A with stationary residuals ($d = 0.2$): the histograms are close to a normal distribution (see also the Figure 3 below for the case of weak fractional cointegration case, $d = 0.6$).

3.2.3 The weak fractional cointegration case

When cointegrating errors are not stationary (that is, $d \geq 0.5$), the MLE estimator of Johansen (1988) is not convergent and clearly inappropriate. GLS estimator outperform other estimators of β in model A and B_1 but is no longer convergent when residuals face endogeneity (models B_2 and B_3): referring to our notation, this estimator cannot distinguish between β and $\beta + \alpha$ when d is high and $\alpha \neq 0$. The problem of endogeneity also affect the behavior of other estimation methods (LSE, ADL, DOLS, FMLS and FCMLE). Without endogeneity, the bias increases moderately with d while the variance increases sharply (table 3 and 4). In the presence of non-zero coherence between errors and regressors both bias and variances are strongly impacted when d increase. In this context, the FMNBLS estimator behaves correctly in the presence of endogeneity (table 5) but its performance deteriorates when $\alpha_{11} \neq 0$ (tables 4 and 6). Concerning the estimators of d, the simulations show that the bias increases slightly when d goes up, while the RMSEs remain broadly similar.

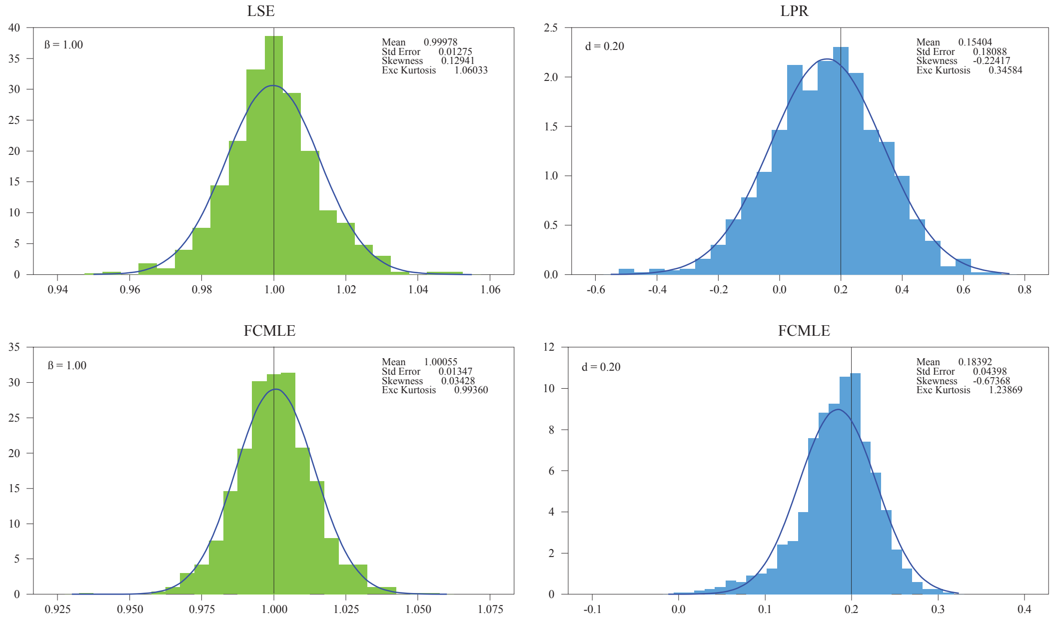

Figure 1: Finite sample distributions of $\hat{\beta}$ and $\hat{d}$ (model A, $d = 0.2$ and $n = 512$)

3.2.4 The spurious regression case

The case where $d = 1$ corresponds to the no-cointegration case, introducing the well-known spurious regression issue. Consequently, estimates of β are highly significant but don't fit with any econometric inference. In our Monte-Carlo experiment, it leads to dramatic increase in variances for all estimators of β (tables 3 - 4). Moreover, in our framework, equation (8) can be reformulated as follows,

$$\Delta y_t = (\alpha + \beta)\Delta x_t + \alpha_{11}\Delta y_{t-1} - \alpha_{11}\beta\Delta x_{t-1} + \varepsilon_t \tag{40}$$

This equation highlights an additional problem: the identification of parameter β. When $\alpha_{11} = 0$ there is no way to identify separately α and β and all estimators are severely biased (table 5). Conversely our fully-parametric FCMLE allows to identify separately β and α_{11} and thus reduces the bias quite considerably.

3.2.5 Synthesis

Overall, β estimators can be gathered into four categories.

i) LSE, ADL, DOLS and FMLS perform similarly in all cases. According to our simulations there is no significant benefit to the use of methods ADL, DOLS or FMLS compared to OLS to correct the endogeneity or autocorrelation bias in the fractional cointegration case. These biases appears to be closely related to the long memory parameter value. However, they are negligible as long as $d < 0.4$.

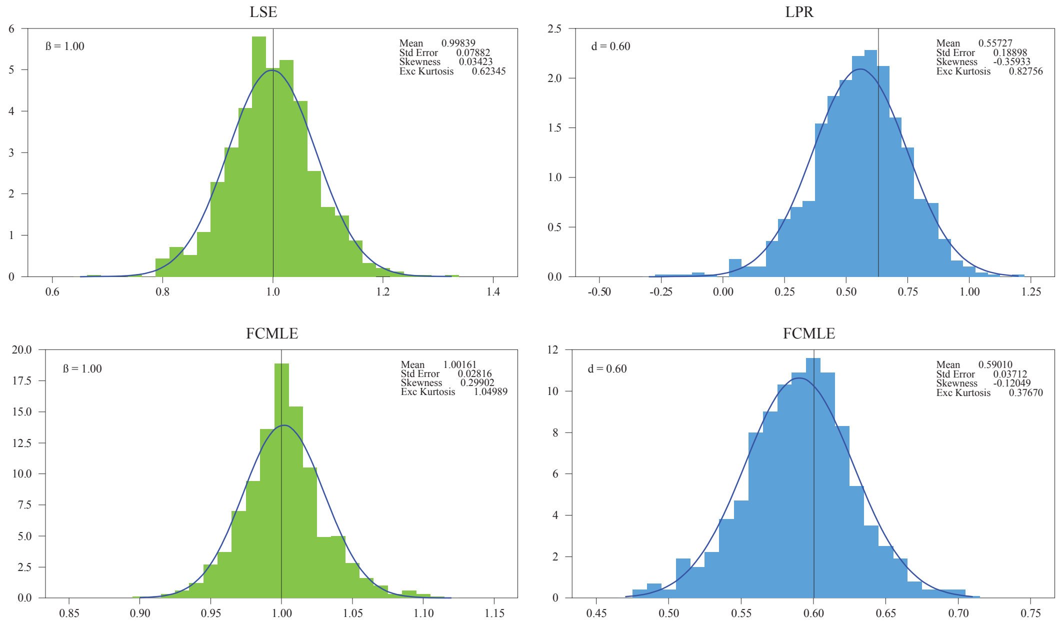

Figure 2: Finite sample distributions of $\hat{\beta}$ and $\hat{d}$ (model A, $d = 0.6$ and $n = 512$)

ii) GLS estimator is heavily biased when residuals face both serial correlation and endogeneity (table 6) and is not robust to endogeneity when residuals are non-stationary (table 5).

iii) NBLS perform generally better than the estimators discussed above, although theoretically, it should not be applied in the parameter space of δ that we consider. FMNBLS appear to be less robust to autocorrelation than NBLS. Nevertheless this problem is confined to weak fractional cointegration and spurious regression cases. In line with the theory, FMNBLS outperform other estimators when residuals are subject to endogeneity bias (table 5).

iv) In comparison with other estimators, the FCMLE is less affected by endogeneity and autocorrelation biases when d increases. Overall, this estimator clearly outperforms other methods both in terms of bias and variance.

Concerning d estimators we can identify two categories.

i) Regarding the non-parametric estimators of d, our simulations do not reveal large differences between LPR and LWE whatever the cases studied. In most cases LWE exhibit a slight advantage in terms of RMSE. In contrast and in line with the theory, the LPM estimator is characterized by a significant bias reduction at the cost of increased variance. However its performance deteriorates in the spurious regression case.

ii) The behavior of FML and FCMLE are closely related and globally these estimators outperforms the non-parametric estimators. The FCMLE is less sensitive to serial correlation bias that affect FML when d is close to 1 (table 4 and 6).

3.2.6 Fractional cointegration tests: powers and sizes

Overall, most tests are satisfactory in terms of power, since they reject H_0 when H_0 is false with a probability of 100%, as long as residuals are stationary. As we can expect, the power of the different tests decreases when d increases. Moreover, all tests are sensitive to autocorrelation, which causes a dramatic deterioration of powers. This behavior is much more pronounced in the case of the LPE test (figure 3) and FCMLE clearly exhibits the best results. Size are globally satisfactory expect in some cases for the test of Lobato and Velasco (2007) and the LWE test (table 3 and 5).

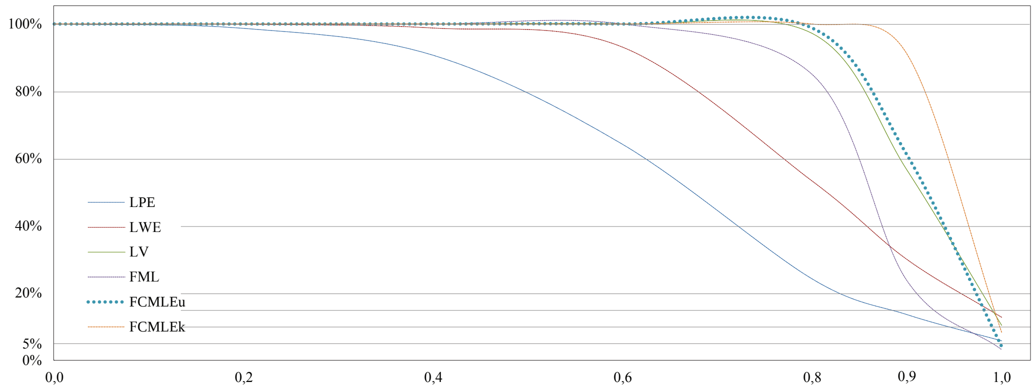

Figure 3: Power and size of various tests at 5%, for the model B_2 with $n = 512$.

4 EMPIRICAL ILLUSTRATION

4.1 The present value model

We present an empirical investigation of the present value model to the S&P 500 stock prices (Campbell and Shiller (1987)). Under the efficient markets hypothesis (EMH), an asset price y_t depends linearly on the present discounted value of expected dividends x_t,

$$y_t = \sum_{j=1}^{\infty} \frac{1}{(1+r)^j} \times E[x_{t+j}|I_t] \tag{41}$$

Solving this equation leads to a cointegration relationship between y_t and x_t. Numerous papers have investigated this relation from the restrictive point of view of the $I(1)/I(0)$ paradigm. Most studies have led to inconclusive results, or have concluded against the EMH because "*spread between stock prices and dividends moves too much and deviations from the present value model are quite persistent*" (see Campbell and Shiller (1987)). Recently, some studies have explored the possibility of fractionally integrated deviations from equilibrium. Marinucci and Robinson (2001) find little evidence in favor of weak cointegration (that is, $\delta - d < 1/2$) and they cannot reject the null hypothesis of no cointegration. In the same spirit, Caporale and Gil-Alana (2004) reveal strong evidence of in favor weak fractional cointegration with $d \in [0.6, 0.7]$ and conclude to the rejection of the null hypothesis.

4.2 Estimations results

In order to test the EMH, we use the updated monthly database of Campbell and Shiller (1987)[5]: y_t refers to the Standard and Poor's composite stock price index from January 1871 to December 2011 and x_t refers to the associated dividends.

Firstly, we lead several integration tests upon y_t and x_t including the fractional Wald test of Lobato and Velasco (2007). Results support the hypothesis of non-stationary $I(1)$ process for both y_t and x_t. The LV test confirms that processes are both $I(1)$, against the alternative hypothesis of $I(d)$ processes.

Table 7: Unit root tests

	S&P500				Dividend			
Sample (T)	1871-2010 (1667)				1871-2010 (1667)			
(μ)	KPSS	ADF	PP	LV	KPSS	ADF	PP	LV
Level	30.14	1.085	0.730	0.898	31.64	3.964	1.709	0.16
Difference	0.266	-30.32	-29.98	-	0.674	-9.903	-9.709	-
Crit. value at 5%	0.463	-2.864	-2.864	-1.96	0.463	-2.864	-2.864	-1.96
($\mu + \delta_t$)	KPSS	ADF	PP	LV	KPSS	ADF	PP	LV
Level	6.559	-1.639	-1.878	-	6.340	-2.153	-1.900	-
Difference	5.789	-74.28	-62.88	-	4.084	-152.6	-88.16	-
Crit. value at 5%	0.146	-3.415	-3.415	-	0.146	-3.415	-3.415	-

Secondly, we lead several estimations corresponding to different sample sizes. We consider first the original dataset of Campbell and Shiller (1987). Next, we consider the complete sample. Finally we consider the complete sample excluding the Cowles (1939)'s data (i.e. from 1871 to 1926), since calculation methods are different. We perform estimations using the FCMLE and we also report the Cheung and Lai (1993) methodology and the combination {LSE,FML}. Results are reported in the table (8).

Results depend heavily on the period and the estimation method. For instance, the methodology of Cheung and Lai (1993) cannot reject at 5% the null hypothesis of no cointegration ($T_{LPE} = -1.43 > -2.25$) as well as our one step methodology, when we consider the complete sample. Conversely, the LSE/FML methodology reject the null hypothesis accepting cointegration ($T_{FML} = -3.79 < -1.645$). Removing the Cowles (1939)'s data, the conclusion is the same for the LSE/LPE methodology ($T_{LPE} = -1.07 > -2.25$), whereas the LSE/FML and the FCMLE both lead to reject the null hypothesis ($T_{FML} = -2.87 < -1.645$ and $W_{FCMLE} = -2.40 < -1.645$). According to the data set used by Campbell and Shiller (1987), the LSE/LPE and the LSE/FML reject H_0 ($T_{FML} = -4.21 < -1.645$ and $T_{LPE} = -3.28 < -1.645$) while

[5]http://www.econ.yale.edu/ shiller/

Table 8: Estimation of β, d and tests

(n)	1871-2011 (1680)			1926-2011 (1032)			1871-1986 (1380)		
θ	FCMLE	LSE/LPE	LSE/FML	FCMLE	LSE/LPE	LSE/FML	FCMLE	LSE/LPE	LSE/FML
d	1.0061	0.7959	0.9086	0.8611	0.8812	0.8510	0.9984	0.6549	0.7913
	(0.0318)	(0.1158)	(0.0424)	(0,10172)	(0.1346)	(0.0517)	(0.0329)	(0.1232)	(0.0495)
β	0.8973	1.1819	1.1819	1.0298	1.2572	1.2572	0.8588	1.0638	1.0638
	(0.1379)	(0.0046)	(0.0046)	(0.1967)	(0.0073)	(0.0073)	(0.1416)	(0.0058)	(0.0058)
α_{11}^1	0.2531	-	0.3618	0.3786	-	0.3899	0.2575	-	0.4309
	(0.0512)	-	(0.0479)	(0.1214)	-	(0.0586)	(0.0557)	-	(0.0543)
α_{11}^2	-0.0567	-	-0.1056	-	-	-	-0.0498	-	-
	(0.0316)	-	(0.0257)	-	-	-	(0.0336)	-	-
$\overline{R}^2$	0.3874	0.9750	0.9750	0.3139	0.9661	0.9661	0.3837	0.9605	0.9605

the FCMLE Wald test cannot reject H_0. However, with respect to power of tests, we cannot eliminate any doubt of type I error and results have to be interpreted carefully.

5 FINAL COMMENTS

In this paper we have compared through Monte Carlo simulations the finite sample properties of different estimation methods of a fractional cointegration model for a wide range of the integration order of residuals, in the case where the integration order of regressors is known and equal to 1. To improve efficiency, we have proposed to estimate jointly all parameters of interest, using the Gaussian maximum likelihood estimator. Our approach complements the contributions of Hualde and Robinson (2007), which is restricted to weak fractional cointegration (that is, $\delta - d < 1/2$), and Nielsen (2007) solely concerned with the stationary fractional cointegration case (that is, $0 \leq d < \delta < 1/2$). We have studied the finite sample properties of the Gaussian maximum likelihood estimator and several estimators that operate in two steps. The Monte Carlo experiment shows that our one-step parametric time domain approach compares favorably with other estimators of β and d and sometimes performs better, even when cointegrating errors face endogeneity and serial correlation. We have also suggested to use the fractional integration Wald test of Tanaka (1999), which has good power and size according to our simulation. Finally, we have revisited the test of the present value model of Campbell and Shiller (1987) and found little evidence in favor of fractional cointegration.

References

[Andrews and Guggenberger (2003)] Andrews, D. W. K., Guggenberger, P. (2003), "A bias-reduced log-periodogram regression estimator for the long memory parameter," *Econometrica*, 71, 675-712

[Baillie and Bollerslev (1994)] Baillie, R. T., Bollerslev, T. (1994), "Cointegration, fractional cointegration and exchange rate dynamics," *Journal of Finance*, 49, 737-45.

[Campbell and Shiller (1987)] Campbell, J. Y., Shiller, R. J. (1987), "Cointegration and tests of present value models," *Journal of Political Economy*, 95, 1062-1088.

[Caporale and Gil-Alana (2002)] Caporale, G. M., Gil-Alana, L. A. (2002), "Unemployment and input prices: a fractional cointegration approach," *Applied Economics Letters*, 9, 347-351.

[Caporale and Gil-Alana (2004)] Caporale, G. M., Gil-Alana, L. A. (2004), "Fractional cointegration and tests of present value models," *Review of Financial Economics*, 13, 245-58.

[Cheung and Lai (1993)] Cheung, Y. W., Lai K. S. (1993), "A fractional cointegration analysis of purchasing power parity," *Journal of Business & Economic Statistics*, 11, 103-112.

[Cowles (1939)] Cowles, A. (1939), "Common stock indexes," *Bloomington: Principia Press*, (2nd ed.).

[da Silva and Robinson (2008)] da Silva, A. G., Robinson, P. M. (2008), "Finite Sample Performance in Cointegration Analysis of Nonlinear Time Series with Long Memory," *Econometric Theory*, 27, 268-297.

[Dittman (2000)] Dittman, I. (2000), "Residual-based tests for fractional cointegration: a Monte Carlo study," *Journal Time Series Analysis*, 21, 615-47.

[Dolado et al. (2002)] Dolado, J. J., Gonzalo, J., Mayoral, L. (2002), "A Fractional DickeyFuller Test for Unit Roots," *Econometrica*, 70, 1963-2006.

[Dubois et al. (2004)] Dubois, E., Lardic, S., Mignon, V. (2004), "The Exact Maximum Likelihood-Based Test for Fractional Cointegration: Critical Values, Power and Size," *Computational Economics*, 24, 239-255.

[Engle and Granger (1987)] Engle, R. F., Granger, C. W. J. (1987), "Cointegration and error correction: representation, estimation and testing," *Econometrica*, 55, 251-276.

[Fox and Taqqu (1986)] Fox, R., Taqqu, M. S. (1986), "Large-sample properties of parameter estimates for strongly dependent stationary gaussian series," *Annals of Statistics* 14, 517-532.

[Geweke and Porter-Hudak (1983)] Geweke, J., Porter-Hudak, S. (1983), "The estimation and application of long memory time series models," *Journal of Time Series Analysis* 4, 221-238.

[Granger (1986)] Granger, C. W. J. (1986), "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics, 48, 213-228.*

[Granger and Joyeux (1980)] Granger, C. W. J., Joyeux, R. (1980), "An introduction to long memory time series models and fractional differencing," *Journal of Time Series Analysis*, 1, 15-29.

[Henry and Zaffaroni (2003)] Henry, M., Zaffaroni, P. (2003), "The long range dependence paradigm for macroeconomics and finance," *In: Doukhan, P., Oppenheim, G., Taqqu, M. S., eds. Theory and Applications of Long-Range Dependence.* Boston: Birkhuser, pp. 417-438.

[Hidalgo and Robinson (2002)] Hidalgo, F. J., Robinson, P. M. (2002), "Adapting to Unknown Disturbance Autocorrelation in Regression With Long Memory," *Econometrica*, 70, 1545-1581.

[Hosking (1981)] Hosking, J.R.M. (1981), "Fractional differencing," *Biometrika* 68, 165-176.

[Hualde and Robinson (2007)] Hualde, J., Robinson, P. M. (2007), "Root-n-consistent estimation of weak fractional cointegration," *Journal of Econometrics*, 140, 450-484.

[Johansen (1988)] Johansen, S. (1988). "Statistical analysis of cointegrating vectors," *Journal of Economic Dynamics and Control*, 12, 231-54.

[Künsch (1987)] Künsch, H. R. (1987), "Statistical aspects of self-similar processes," In: Prokhorov, Y., Sazanov, V. V., eds. Proceedings of the First World Congress of the Bernoulli Society. Utrecht: VNU Science Press, pp. 67-74.

[Kurozumi and Hayakawa (2009)] Kurozumi, E., Hayakawa, K. (2009), "Asymptotic properties of the efficient estimators for cointegrating regression models with serially dependent errors," *Journal of Econometrics*, 149, 118-135.

[Lasak (2010)] Lasak, K. (2010), "Maximum likelihood estimation of fractionally cointegrated systems," *CREATES Research Paper*, 2008-53.

[Lobato and Velasco (2007)] Lobato, I. N., Velasco, C. (2007), "Efficient Wald tests for fractional unit roots," *Econometrica*, 75, 575-589.

[Marinucci and Robinson (2001)] Marinucci, D., Robinson, P. M. (2001), "Semiparametric fractional cointegration," *Journal of Econometrics*, 105, 225-47.

[Nielsen (2004)] Nielsen, M. Ø. (2004), "Optimal Residual-Based Tests for Fractional Cointegration and Exchange Rate Dynamics," *Journal of Business & Economic Statistics*, 22, 331-345.

[Nielsen (2007)] Nielsen, M. Ø. (2007), "Local Whittle Analysis of Stationary Fractional Cointegration and the Implied-Realized Volatility Relation," *Journal of Business & Economic Statistics*, 25, 427-446.

[Nielsen and Frederiksen (2005)] Nielsen, M. Ø., Frederiksen, P. H. (2005), "Finite Sample Comparison of Parametric, Semi-parametric, And Wavelet Estimators of Fractional Integration," *Econometric Reviews*, 24, 405-443.

[Nielsen and Frederiksen (2011)] Nielsen, M. Ø., Frederiksen, P. H. (2011), "Fully modified narrow-band least squares estimation of weak fractional cointegration," *Econometrics Journal*, 14, 77-120.

[Panopoulou and Pittis (2004)] Panopoulou, E., Pittis, N. (2004), "A comparison of autoregressive distributed lag and dynamic LSE cointegration estimators in the case of a serially correlated cointegration error," *Econometrics Journal*, 7, 585-617.

[Phillips and Hansen (1990)] Phillips, P. C. B., Hansen, B. E. (1990), "Statistical inference in instrumental regressions with I(1) processes," *Review of Economic Studies*, 57, 99-125.

[Prais and Winsten (1954)] Prais, S. J., Winsten, C. B. (1954), "Trend Estimators and Serial Correlation," *Cowles Commission Discussion Paper*, 383, Chicago.

[Robinson (1991)] Robinson, P. M. (1991), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67-84.

[Robinson (1994)] Robinson, P. M. (1994), "Semiparametric analysis of long-memory time series," *Annals of Statistics*, 22, 515-539.

[Robinson (1995)] Robinson, P. M. (1995), "Gaussian semiparametric estimation of long range dependence," *Annals of Statistics*, 23, 1630-1661.

[Robinson and Hidalgo (1997)] Robinson, P. M., Hidalgo, F. J. (1997), "Time series regression with long-range dependence," *The Annals of Statistics*, 25, 77-104.

[Robinson and Hualde (2003)] Robinson, P.M., Hualde, J. (2003), "Cointegration in fractional systems with unkown integration orders," *Econometrica*, 71, 1727-1766

[Robinson and Marinucci (2001)] Robinson, P. M., Marinucci, D. (2001), "Narrow-Band Analysis of Nonstationary Processes," *The Annals of Statistics*, 29, 947-986.

[Saikkonen (1995)] Saikkonen, P. (1995), "Problems with the Asymptotic Theory of Maximum Likelihood in Integrated and Cointegrated Systems," *Econometric Theory*, 11, 888-911.

[Schwarz (1978)] Schwarz, G. E. (1978), "Estimating the dimension of a model," *Annals of Statistics*, 6, 461-464.

[Sephton (2002)] Sephton, P. S. (2002), "Fractional cointegration: Monte Carlo estimates of critical values, with an application," *Applied Financial Economics*, 12, 331-335.

[Stock (1987)] Stock, J. H. (1987), "Asymptotic properties of least squares estimators of cointegrating vector," *Econometrica*, 55, 1035-1056.

[Stock and Watson (1993)] Stock, J. H., Watson, M.W. (1993), "A simple estimator of cointegrating vectors in higher-order integrated systems," *Econometrica*, 61, 783-820.

[Sowell (1992)] Sowell, F. B. (1992), "Maximum likelihood estimation of stationary univariate fractionally integrated time series models," *Journal of Econometrics*, 53, 165-188.

[Tanaka (1999)] Tanaka, K. (1999), "The nonstationary fractional unit root," *Econometric Theory*, 15, 549-582.