

Cointegration Estimators and Changes in the Error Dynamics

Ekaterini Panopoulou* Nikitas Pittis†

January 2006

Abstract

This study investigates the behavior of several widely used cointegration estimators in cases that allow for changes in the error generating mechanism dynamics, such as those induced by a VAR(1) process with VAR(1) coefficients. Our results, based on Monte Carlo simulations, suggest that the best performing estimators are the Autoregressive Distributed Lag (ADL), the Phillips and Loretan (PL) and to a lesser extent the Johansen (JOH) and the Fully Modified PreWhitened (FM-PW) estimators, whose robustness is not contingent upon changes in the error generating mechanism.

JEL classification: *C12, C13, C22*

Keywords: cointegration estimators, random coefficient models, autoregressive coefficient models.

Word Count: 5800 words

Acknowledgments: Financial support from the Greek Ministry of Education and the European Union under “Hrakteitos” grant is greatly appreciated. We are grateful to A. Antzoulatos, S. Kalyvitis, G.Hardouvelis, D. Malliaropoulos, E. Tzavalis and seminar participants at various seminars and conferences for helpful suggestions and comments. The usual disclaimer applies.

**(corresponding author) Department of Economics, National University of Ireland Maynooth, Co.Kildare, Republic of Ireland. E-mail: apano@nuim.ie. Tel: 00353 1 7083793. Fax: 00353 1 7083934.*

†*Department of Banking and Financial Management, University of Piraeus, Karaoli & Dimitriou 80, Piraeus, 18534, Greece. e-mail: npittis@unipi.gr*

1 Introduction

Asymptotic optimality of the common single-equation cointegration estimators is contingent upon certain conditions imposed on the cointegration error and the error that drives the regressor. These conditions ensure the applicability of the Functional Central Limit Theorem (FCLT), thus allowing for asymptotics of integrated processes to be derived.¹ A large class of linear processes, including those generated by the usual stationary and invertible VARMA models, meets these conditions. In the context of the triangular DGP, put forward by Phillips (1988), the vector of the cointegrating error and the error that drives the regressor is usually modelled as a VAR(1) process. The need to parameterise the vector of errors stems from the necessity to evaluate the various estimators in finite samples typically used in empirical applications and/or to derive sample specific critical values for testing a cointegrating relationship of interest. The popularity of the VAR(1) parameterisation of the errors lies on the fact that such a model accommodates the empirical regularities often met in practice. In particular, it is often observed that the cointegration error exhibits a high degree of persistence. This feature can be captured, quite easily, by a VAR(1) model, provided that certain parameters of the model are different from zero. In this context, several Monte Carlo studies of the small sample properties of cointegration estimators have been conducted (see, for example, Phillips and Hansen 1990, Stock and Watson 1993, Christou and Pittis 2002 and Panopoulou and Pittis 2004).

However, an important question naturally arises at this point: Is the typical VAR(1) model versatile enough to accommodate all the empirical regularities exhibited by real data? In other words, are there any other regularities in the data that cannot be captured by a VAR(1) process? If for example, the time heterogeneity properties of real data are more general than those implied by a VAR(1) model, then the simulation studies, performed so far, might have failed to tell the ‘whole story’.

Let us have a look at a popular application of cointegration theory, namely the estimation of the Purchasing Power Parity (PPP) relationship for the UK pound - US dollar. A close inspection of the residual series from a regression of nominal exchange rate on price

¹See Phillips and Durlauf, 1986.

differential reveals that a constant coefficient model can not accommodate the dynamics of this error generating mechanism. Specifically, Figure 1A shows the cointegration residual from the regression mentioned above, for a period of 200 years. Figures 1C and 1D plot artificial data produced by a constant coefficient AR(1) model displaying medium and high degrees of persistence, respectively. Surprisingly, even the highly persistent process (Figure 1D) cannot come close to replicating the empirical observations. On the other hand, visual examination of Figures 1A and 1B, which show artificial data produced by an AR(1) model with an AR(1) coefficient, suggests that the real and artificial data are quite similar. Quite importantly, this picture emerges with very low degrees of persistence of the autoregressive coefficient.

[INSERT FIGURE 1]

Next, we provide visual evidence on the ability of the time-varying coefficient models to accommodate relatively abrupt changes, which are usually referred to as structural breaks in the econometric literature. Consider Figure 2A that plots the evolution of the estimate of the autoregressive coefficient of the cointegration error of the UK/US PPP relationship using a rolling window of 80 observations. A structural break for the autoregressive coefficient of the cointegration error around 1912 is apparent. However, such a behavior cannot be replicated through a constant coefficient model as shown in Figure 2C that plots the respective estimates from artificial data generated by constant coefficients. On the other hand, a VAR(1) process with VAR(1) coefficients seems able to accommodate at least one “break” as shown in Figure 2B and as a result such a parameterisation is likely to imitate real data.

[INSERT FIGURE 2]

The discussion, so far, has shown that a constant coefficient parameterisation of the error generating mechanism is likely to be unrealistic. In this context, we put forward a wide class of processes, i.e. the VAR(1) model with VAR(1) coefficients, which allow for more general heterogeneity properties in the data similar to those encountered in empirical applications. Both the Random Coefficient and the Constant Coefficient VAR(1) models are nested within this class. Interestingly, these models are likely to satisfy the conditions

for the applicability of the FCLT and consequently they are quite appealing to the applied researcher.

To our knowledge, no other study has investigated the behavior of the cointegration estimators in the presence of time varying coefficients in the Data Generation Process (DGP). This issue is of interest given that many financial time series, such as spot and forward exchange rates, short and long term interest rates are often analyzed in a cointegration framework while displaying a considerable degree of time heterogeneity. Specifically, it is our intention to investigate, by means of Monte Carlo simulations, the finite sample performance of several cointegration estimators in cases that the cointegration error and the error that drives the regressor follow a bivariate VAR(1) process with either VAR(1) (AC, henceforth), Random (RC, henceforth) or Constant Coefficients (CC, henceforth). Our analysis reveals a class of estimators, namely the ADL and PL estimators and to a lesser extent the JOH and the FM-PW estimators that are robust to alternative DGPs in the RC or AC framework. However, the applied researcher should be cautious in interpreting these findings, as a safer strategy involves calculating the DGP-specific critical values via Monte Carlo simulations. The developments in the estimation techniques of state space models can ensure the relatively easy calculation of the parameter values that can be utilized in such simulations.

The paper is structured as follows: Section 2 introduces the data generation process (DGP) and discusses the stationarity conditions for the processes under consideration. The parameters of the DGP are related to the “long-run correlation” and “endogeneity” nuisance parameters present in the distribution of the OLS estimator and the finite-sample comparison of a variety of asymptotically efficient cointegration estimator is motivated. A Monte Carlo comparison of these estimators for a plethora of DGPs is presented in Section 3 and conclusions are drawn in Section 4.

2 The DGP, Nuisance parameters and Estimators

2.1 A Bivariate Cointegrated System and Stationarity Conditions

We consider the following bivariate DGP for the I(1) vector $\mathbf{z}_t = [y_t, x_t]^\top$:

$$y_t = \theta x_t + u_{1t} \quad (1)$$

$$\Delta x_t = u_{2t} \quad (2)$$

We further assume that $\mathbf{u}_t = [u_{1t}, u_{2t}]^\top$ is an I(0) process generated as follows:

$$u_t = (M + P_t)u_{t-1} + \varepsilon_t \quad (3)$$

where $\varepsilon_t \sim IID(0, \Sigma_{12})$, M is a 2×2 constant parameter matrix and P_t is a 2×2 time varying matrix. Specifically,

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \left[\begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} + \begin{pmatrix} \rho_{11,t} & \rho_{12,t} \\ \rho_{21,t} & \rho_{22,t} \end{pmatrix} \right] \begin{pmatrix} u_{1t-1} \\ u_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \quad (4)$$

and

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim NIID \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right] \quad (5)$$

Our interest lies on the generating mechanism of the matrix P_t since the properties of the DGP under consideration depend on its structure. Starting from the more general (and uncertain) specification, we can have P_t being a VAR(1) process itself, or a random process or even a constant. More analytically, the following three cases are of interest:

2.1.1 CASE I: u_t follows a VAR(1) process with VAR(1) coefficients

In this case, P_t is a Vector Autoregressive process of order 1, i.e. $P_t = \Phi P_{t-1} + v_t$, where $v_t \sim IID(0, \Sigma_{34})$ and Φ is a 2×2 constant parameter matrix. Setting the off-diagonal elements of matrix P_t equal to zero and dropping the second subscript, we have the following model for P_t :

$$\begin{pmatrix} \rho_{1t} \\ \rho_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} \rho_{1t-1} \\ \rho_{2t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \quad (6)$$

$$\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \sim NIID \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{33} & \sigma_{34} \\ \sigma_{34} & \sigma_{44} \end{pmatrix} \right] \quad (7)$$

To our knowledge, no stationarity or stability conditions have been derived for a VAR(1) process with VAR(1) coefficients. The only attempt to derive conditions for the stability of such a process is by Weiss (1985) and is limited to the univariate AR(1) process with an AR(1) coefficient.² In this respect, we further simplify our VAR model, so as to be equivalent to two univariate processes, by assuming that the matrices Φ and M are diagonal and $\sigma_{34} = 0$. In such a case, the cointegration error and the error that drives the regressor reduce to:

$$\begin{aligned} u_{1t} &= \mu_{11} + \rho_{1t} u_{1t-1} + e_{1t}, \quad \rho_{1t} = \phi_{11} \rho_{1t-1} + v_{1t} \\ u_{2t} &= \mu_{22} + \rho_{2t} u_{2t-1} + e_{2t}, \quad \rho_{2t} = \phi_{22} \rho_{2t-1} + v_{2t} \end{aligned}$$

The stability condition then (dropping the subscripts) reduces to $R + S^2(\infty) < 1$, with

$$\begin{aligned} R + S^2(\infty) &= \mu^2 + \frac{q}{1 - \phi^2} (1 + 4\mu^2 + 8\mu^2 \lim_{n \rightarrow \infty} \sum_{j=1}^{n-1} \frac{n-j}{n} \phi^j) + \\ &\quad \frac{2q^2}{(1 - \phi^2)^2} (1 + \lim_{n \rightarrow \infty} \sum_{j=1}^{n-1} \frac{n-j}{n} \phi^{2j}) \end{aligned}$$

for u_{1t} and u_{2t} , with q equal σ_{33} and σ_{44} , respectively.

²Specifically, Weiss derived a sufficient condition for the stability of an AR(1) process with an AR(1) coefficient, which might be too strong.

In general, stability does not imply stationarity. However, implicit in the derivation of the stability condition by Weiss is the existence of a finite mean and the condition itself is built on the requirement of a finite variance, which associated with an identical distribution for the errors ε_t provide us with a sufficient second-order stationarity condition, albeit a strong one.

2.1.2 CASE II: u_t follows a VAR(1) process with Random Coefficients

This case is naturally a subset of the previous one. By setting $\Phi = 0$, we have P_t a mean zero process with $E[P_t \otimes P_t] = C$. It is easy to show that the elements of the matrix C can be obtained from those of the matrix Σ_{34} for a diagonal matrix P_t . In deriving the necessary and sufficient conditions for the stationarity and stability of this process, we further assume that P_t is independent of $\{\varepsilon_t\}$. Analytically, the generating mechanism for the elements of the bivariate process P_t are given by the following equation:

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \left[\begin{pmatrix} \mu_{11} & 0 \\ 0 & \mu_{22} \end{pmatrix} + \begin{pmatrix} \rho_{11,t} & 0 \\ 0 & \rho_{22,t} \end{pmatrix} \right] \begin{pmatrix} u_{1t-1} \\ u_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \quad (8)$$

$$\begin{pmatrix} e_{1t} \\ e_{2t} \\ \rho_{11,t} \\ \rho_{22,t} \end{pmatrix} \sim NIID \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{pmatrix} \right] \quad (9)$$

Nicholls and Quinn (1982) derived the stationarity conditions for the multivariate autoregressive model of order n . In our case, their conditions (necessary and sufficient) reduce to the following for our parameter space:

1. The eigenvalues of the matrix M are less than unity³, and

2. The matrix $H = \begin{pmatrix} \sigma_{11} \frac{1-\mu_{11}^2}{1-\mu_{11}^2-\sigma_{33}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \frac{1-\mu_{22}^2}{1-\mu_{22}^2-\sigma_{44}} \end{pmatrix}$ is positive definite.

³Given that M is diagonal, its eigenvalues are less than unity in modulus $\iff |\mu_{11}| < 1$ and $|\mu_{22}| < 1$.

When conditions 1 and 2 are satisfied, the variance-covariance V of \mathbf{u}_t is given by

$$V = \begin{pmatrix} \frac{\sigma_{11}}{1-\mu_{11}^2-\sigma_{33}} & \frac{\sigma_{12}}{1-\mu_{11}\mu_{22}} \\ \frac{\sigma_{12}}{1-\mu_{11}\mu_{22}} & \frac{\sigma_{22}}{1-\mu_{22}^2-\sigma_{44}} \end{pmatrix}.$$

After some algebra, we show that the stationarity of the RC model is ensured by the following condition.

Proposition 1A *VAR(1) RC process generated by (8) and (9) is stationary if $1 - \mu_{11}^2 - \sigma_{33} > 0$ and $1 - \mu_{22}^2 - \sigma_{44} > 0$.*

Proof. See Appendix A ■

2.1.3 CASE III: u_t follows a Constant Coefficients VAR(1) process

Further simplifying our general process by setting Σ_{34} equal to zero, we get the commonly used VAR(1) process with constant coefficients.

$$u_t = Mu_{t-1} + \varepsilon_t \tag{10}$$

In this case, the conditions for stationarity are trivial, i.e. it suffices that the eigenvalues of the matrix M are less than unity in modulus. Given the diagonality of the matrix M , the conditions reduce to $|\mu_{11}| < 1$ and $|\mu_{22}| < 1$.⁴

Next, we relate the parameters of the DGPs introduced so far to the nuisance parameters present in the distribution of the OLS estimator.

2.2 Nuisance parameters

The presence of nuisance parameters in the distribution of the OLS estimator renders standard asymptotic theory useless in the case of cointegration. The reason for the presence of these non-standard asymptotics is that when the elements of \mathbf{u}_t are contemporaneously and/or temporally correlated, two types of second-order asymptotic effects are present in the limiting distribution of the OLS estimator (see Phillips and Loretan 1991): The first is the nuisance parameter, ω_{12}/ω_{22} that describes the “long-run correlation”

⁴It is easy to show that these conditions coincide with the ones for CASE II for $\sigma_{33} = \sigma_{44} = 0$.

effect, due to non-diagonality of the long run covariance matrix $\Omega = [\omega_{ij}]$, $i, j = 1, 2$. The second is the nuisance parameter $\delta_{21} = \sum_{k=0}^{\infty} E(u_{20}u_{1k})$ that describes the “endogeneity” effect. However, the design of our DGPs rules out any feedbacks from the cointegration error to the error that drives the regressor or from the regressor to the error that drives the cointegration error ($\mu_{21} = 0$ and $\mu_{12} = 0$), both nuisance parameters have the same source, namely the contemporaneous correlation between u_{1t} and u_{2t} .

Starting from the latter and more tractable case of the constant coefficient VAR(1) model, we derive the nuisance parameters associated with each of these processes. In general, three types of variances are associated with the process u_t as defined by equations (3) to (5), namely the conditional or contemporaneous covariance matrix, the unconditional covariance matrix and the long-run covariance matrix. In the case that u_t is a stationary constant coefficient VAR(1) process, the long-run covariance matrix Ω and the one-sided covariance matrix Δ , needed to define the asymptotic nuisance parameters, are given by equations (11) and (12), respectively

$$\Omega = (I - M)^{-1}\Sigma_{12}(I - M^{\top})^{-1} \quad (11)$$

$$\Delta = V(I - M^{\top})^{-1} \quad (12)$$

where Σ denotes the innovations covariance matrix of the VAR and V is the unconditional covariance matrix of \mathbf{u}_t given by,

$$vecV = (I - M \otimes M)^{-1}vec\Sigma_{12} \quad (13)$$

With some algebra, we relate the parameters of our DGP (in the CC case) with the following nuisance parameters:

$$\frac{\omega_{12}}{\omega_{22}} = \frac{\sigma_{12}(1 - \mu_{22})}{(1 - \mu_{11})\sigma_{22}} \quad (14)$$

$$\delta_{21} = \frac{\sigma_{12}}{(1 - \mu_{11})(1 - \mu_{11}\mu_{22})} \quad (15)$$

In what follows, we derive the relevant nuisance parameters for the VAR(1) process with Random Coefficients, u_t . To keep the analysis clear, the derivation of the variance matrices is given in Appendix B. Specifically, the conditional/ contemporaneous covariance matrix is Σ_{12} , the unconditional covariance matrix V is shown to be:

$$vecV = (I - M \otimes M - C)^{-1}vec\Sigma_{12}$$

and the long-run covariance matrix Ω is given by

$$\Omega = V((I - M)^{-1} + (I - M')^{-1}M')$$

In our DGP,

$$\Omega = \begin{pmatrix} \frac{(1+\mu_{11})\sigma_{11}}{(1-\mu_{11})(1-\mu_{11}^2-\sigma_{33})} & \frac{\sigma_{12}}{(1-\mu_{11})(1-\mu_{22})} \\ \frac{\sigma_{12}}{(1-\mu_{22})(1-\mu_{22})} & \frac{(1+\mu_{22})\sigma_{22}}{(1-\mu_{22})(1-\mu_{22}^2-\sigma_{44})} \end{pmatrix}$$

and

$$\Delta = \begin{pmatrix} \frac{\sigma_{11}}{(1-\mu_{11})(1-\mu_{11}^2-\sigma_{33})} & \frac{\sigma_{12}}{(1-\mu_{22})(1-\mu_{11}\mu_{22})} \\ \frac{\sigma_{12}}{(1-\mu_{11})(1-\mu_{11}\mu_{22})} & \frac{\sigma_{22}}{(1-\mu_{22})(1-\mu_{22}^2-\sigma_{44})} \end{pmatrix}$$

Hence, we have:

$$\frac{\omega_{12}}{\omega_{22}} = \frac{\sigma_{12}(1 - \mu_{22}^2 - \sigma_{44})}{(1 - \mu_{11})(1 + \mu_{22})\sigma_{22}} \quad (16)$$

$$\delta_{21} = \frac{\sigma_{12}}{(1 - \mu_{11})(1 - \mu_{11}\mu_{22})} \quad (17)$$

Similar results for the CASE I of a VAR(1) process with VAR(1) coefficients are not easy to derive, due to the algebraic intractability of the process.

2.3 Cointegration estimators

We consider both parametric and semi-parametric cointegration estimators, the majority of which are asymptotically efficient provided that the conditions of the FCLT are satisfied.

Next, we provide a description of these estimators:

•**Simple OLS:** This is the ordinary least square estimator of θ , applied to equation (1).

Dynamic OLS (DOLS(p,t)): This estimator was suggested by Saikonen (1991), Phillips and Loretan (1991) and Stock and Watson (1993). The term DOLS was first used by Stock and Watson who generalised it to systems with higher orders of integration. It utilises the static equation (1), augmented by lags and leads of the first difference of the regressor, i.e.:

$$y_t = \theta x_t + \sum_{i=1}^{p-1} \gamma_i \Delta x_{t-i} + \sum_{j=1}^{t-1} d_j \Delta x_{t+j} + v_t \quad (18)$$

This model provides a direct way to estimate the cointegrating relationship and asymptotically leads to valid test statistics. When there are no feedbacks from the cointegration error to the error that drives the regressor, it is sufficient to augment (1) only by lags of Δx_t . Any serial correlation of v_t does not raise any serious problems in the estimation of θ and can be dealt with by consistently estimating the long-run variance of v_t as proposed by Newey and West (1987).

Dynamic GLS (DGLS(p,t)): This estimator (see Stock and Watson, 1993) is the feasible GLS estimator applied to equation (18), in which the regression error is modelled parametrically, usually as an AR(1) process.

Autoregressive Distributed Lag (ADL(q,r)): This estimator is based on the following ADL(q,r) model:

$$y_t = \sum_{i=0}^q a_i x_{t-i} + \sum_{j=1}^r b_j y_{t-j} + \nu_t$$

The parameter of interest is equal to the long-run multiplier of y_t with respect to x_t . A direct estimate of the parameter of interest θ along with its standard error may be obtained by transforming the ADL model into the Bewley form (see Bewley 1979, Wickens and

Breusch 1988, and Banerjee et. al. 1993):

$$y_t = \theta x_t + \sum_{i=1}^{q-1} a_i \Delta x_{t-i} + \sum_{j=1}^{r-1} b_j \Delta y_{t-j} + \epsilon_t$$

Estimates of the coefficients and their standard errors can be obtained by using the Instrumental Variables (IV) estimator, with the original matrix of regressors being the instrumental variables (see Wickens and Breusch 1988). This means that the ADL estimator of θ is very easy to apply since it involves only IV estimation techniques.

PL(s,l): This is the non-linear-in-parameters estimator of Phillips and Loretan (1991), which utilises the same dynamic structure with that of ADL model and is based on the following structure:

$$y_t = \theta x_t + \sum_{i=1}^s d_i (y_{t-i} - \theta x_{t-i}) + \sum_{j=1}^s f_j \Delta x_{t-j} + \sum_{k=1}^l g_k \Delta x_{t+k} + \epsilon_t$$

This procedure is in the spirit of the Error Correction Model (ECM) with the addition of lagged values of the equilibrium error. The parameter of interest, θ , is estimated by non-linear least squares. Phillips and Loretan (1991) report that in their simulation study, the non-linear formulation did create some problems in estimation. The orders s and l refer to the lags and leads of Δx_t , respectively.

Fully Modified Least Squares (FMLS): Phillips and Hansen (1990) employ semi-parametric corrections for the long run correlation and endogeneity effects, which fully modify the OLS estimator and its attendant standard error, thus obtaining the so-called Fully-Modified Least Squares (FM-S) estimation method. The FMLS estimator is based on consistent estimation of the matrices Ω and Δ , which in turn requires the selection of a kernel and the determination of the bandwidth. We employ the Quadratic Spectral kernel, since it is the best with respect to an asymptotic truncated mean square error criterion in the class of kernels that necessarily generate positive semi-definite estimators of the long-run variance covariance matrix in finite samples. The bandwidth parameter, S_T , has been selected by applying the Andrews (1991) data-dependent procedure. Specifically,

the optimal bandwidth parameter S_T^+ for the Quadratic Spectral kernel is

$$S_T^+ = 1.3221[a(2)T]^{1/5}$$

where $a(2)$ is a function of the unknown spectral density matrix of u_t at frequency zero, its second generalized derivative and a 4×4 weighting matrix of known constants. This means that $a(2)$ and hence S_T^+ are also unknown in practice. Estimates of $a(2)$ may be obtained either by estimating simple parametric models, as suggested by Andrews (1991), or non-parametrically following Newey and West (1994). In our study, we determine the bandwidth by means of the Andrews (1991) data-dependent procedure. Moreover, the “prewhitened” version of FMLS (FM-PW) which filters the error vector \hat{u}_t prior to estimating Ω and Δ is also employed (see Andrews and Monahan, 1992 and Christou and Pittis 2002, for a discussion on the performance of the various versions of the FMLS estimator).

Johansen’s Maximum Likelihood (JOH): Finally, apart from the single-equation estimators discussed above, we also consider the system-based maximum likelihood estimator of θ , suggested by Johansen (1988, 1991). The order of the JOH estimator corresponds to the lag-order of the Vector Autoregressive Model on which this estimator is based. This estimator differs from all the above mentioned estimators in an important respect: It has been developed and proved to be asymptotically optimal in the context of a Gaussian Vector Autoregression which accommodates a rather narrow class of DGPs.

3 Monte Carlo Study

Our aim is to evaluate the performance of the commonly used cointegration estimators, both parametric and non-parametric in the two frameworks under consideration, i.e. the VAR(1) Coefficients (AC) framework and the Random Coefficients (RC) one. The performance of the estimators is assessed against deviations from their performance in the respective Constant Coefficients (CC) models. We specifically employ the following estimators: OLS, DOLS(1), DOLS(4), DGLS(1), DGLS(4), ADL(1,2), ADL(4,4), PL(1,0),

PL(4,0), JOH, FM-S and FM-PW.⁵ As already mentioned, the source of both nuisance parameters is the contemporaneous correlation between u_{1t} and u_{2t} , i.e. σ_{12} in both the CC and RC framework.⁶ Both parameters, namely $\frac{\omega_{12}}{\omega_{22}}$ and δ_{21} are an increasing function of σ_{12} . This is true for the persistence of the cointegration error, as well. On the other hand, the persistence of the error that drives the regressor drives the nuisance parameters to the opposite direction, i.e. as μ_{22} increases, $\frac{\omega_{12}}{\omega_{22}}$ decreases. What is really interesting in this case is that the effect of the random coefficients does not amplify the magnitude of the nuisance parameters. Specifically, as σ_{44} increases, the “long-run correlation” effect decreases, due to increased variation in the error that drives the regressor.

The estimators are evaluated based on their estimation precision and accuracy of statistical inference. As a measure of estimation precision, we calculate the mean bias and for the statistical inference, we calculate the empirical size of the t-test for testing the hypothesis $\theta = 1$ for nominal sizes of 5%. We generate 3000 series of length 150, starting with $u_{10} = u_{20} = 0$, and then discard the initial 50 observations, thus generating a sample of 100 observations.

3.1 DGP1: The effect of randomness of the error that drives the regressor (σ_{44})

In the first DGP, we attempt to isolate the effect of the variance of the process driving the regressor. Specifically, the parameter values are as follows:

$$\text{DGP1: } \left\{ M = \begin{pmatrix} 0.7 & 0 \\ 0 & 0 \end{pmatrix}, \Phi = \begin{pmatrix} 0.3 & 0 \\ 0 & 0 \end{pmatrix}, \Sigma_{12} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \Sigma_{34} = \begin{pmatrix} 0.1 & 0 \\ 0 & \sigma_{44} \end{pmatrix}, \theta = 1, \text{ and } \sigma_{44} = 0.1, 0.5, 0.9 \right\}.$$

The full parameterisation corresponds to the AC case, while when setting $\Phi = \mathbf{0}$ we have the RC case and finally when both Φ and Σ_{34} are set zero, we have the CC case with a relatively high degree of persistence of the cointegration error ($\mu_{11} = 0.7$). The results, concerning the performance of the estimators are reported in Tables 1A-1C.

[INSERT TABLES 1A-1C]

⁵Since our designs do not allow for temporal dependence between the cointegration error and the error that drives the regressor, the augmented versions of the estimators are not considered.

⁶The exact form of the nuisance parameters in the autoregressive coefficients case remains an open issue.

The information content of these tables may be summarised as follows:

(i) The bias and the size of the OLS estimator is a decreasing function of σ_{44} , in line with the decrease in the nuisance parameters. For example, the bias of OLS reduces from 0.141 in the CC case to 0.039 and 0.028 for the RC and AC case, respectively when $\sigma_{44} = 0.9$. Naturally, these bias reductions are followed by size reductions. The OLS size, although significant, reduces to 47% from 76%.

(ii) Irrespective of the degree of randomness of the process, DOLS(1) and especially DGLS(1) are heavily biased even in the CC case. The bias and size of DGLS(1) often exceed the OLS one. Significant size distortions, which reach almost 100% for DGLS(1) are prevalent for all the cases under consideration especially for σ_{44} as low as 0.1. Adding lags of the first difference of the regressor to both DOLS and DGLS seems to work satisfactorily since the respective bias are significantly lower, albeit they remain at around 30%.

(iii) The semi-parametric corrections made to OLS by the FM-S estimator work quite well only in terms of inference. Regarding the estimation of the parameter of interest, FM-S is even more biased than the simple OLS estimator in all cases except for the CC one. For example, in the AC case and for σ_{44} equal to 0.1, the bias of FM-S is 0.171 compared with 0.134 for the OLS estimator. Prewhitening the errors leads to substantial improvement of the performance of the FM estimator. Irrespective of the case considered, the bias and empirical size of FM-PW stays conformably below 0.098 and 15%.

(iv) Surprisingly, JOH, despite its Gaussian origins behaves remarkably well in all the cases, displaying minimal bias and divergence from the standard normal. Irrespective of the process, its empirical size hardly exceeds 11%.⁷

(v) The superiority of ADL and PL estimators, both at estimation and inference, is quite clear. The ADL(4,4) and PL(4,0) perform slightly worse than the less parameterised ones, that is ADL(1,2) and PL(1,0). For these estimators, the empirical sizes range from 7.4% to 10.1%.

Our simulations so far suggest that the best performing estimators both at estimation and inference are the ADL and PL estimators and to a lesser extent, the JOH and FM-PW

⁷In a recent study, Bruggemann and Lutkepohl (2005) provide evidence of JOH producing occasional extreme outliers. However, our DGPs with constant coefficients do not support such a conjecture.

estimators. These estimators form Class A of the best performing ones, while Class B comprises of DOLS(4), DGLS(4) and FM-S and Class C of OLS, DOLS(1) and DGLS(1).

Next, we investigate whether this ranking of the estimators remains unaffected when alternative parameterisations are considered.

3.2 DGP2: The effect of persistence of the cointegration error (μ_{11})

In the second DGP, we attempt to isolate the effect of the persistence of the cointegration error. Specifically, the parameter values are as follows:

DGP2: $\{M = \begin{pmatrix} \mu_{11} & 0 \\ 0 & 0 \end{pmatrix}, \Phi = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}, \Sigma_{12} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \Sigma_{34} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \theta = 1, \text{ and } \mu_{11} = 0.1, 0.5, 0.7\}$. We consider the three cases by appropriately setting Φ and Σ_{34} equal to zero. The results, concerning the performance of the estimators are reported in Tables 2A-2C. We only consider stationary error processes and as a result, for $\mu_{11} = 0.7$, the AC case is not considered.

[INSERT TABLES 2A-2C]

The results are summarised as follows:

(i) The bias of all the estimators is negligible when the persistence of the cointegration error is as low as 0.1. Increasing μ_{11} to 0.5, we get serious biases for the AC case and the OLS, DOLS and DGLS estimators. The same is true for the CC case with μ_{11} equal to 0.7. Moreover, when μ_{11} equals 0.9, all the estimators are severely biased except for the ADL, PL and JOH estimators.

(ii) Naturally, the bias distortions are followed by severe size distortions in the cases aforementioned. Moreover, the estimators of Class B and C display sizable mislocations from the standard normal in all the cases under consideration and not only the persistent cases. For example, for μ_{11} as low as 0.5, the size of the DOLS(1) estimator is equal to 14%, 13%, and 35% for the AC, the RC and the CC case, respectively. On the other hand, the respective values for the FM-PW estimator are 11%, 10% and 8%.

(iii) It is worth noting that OLS displays the largest distributional divergence in the CC case followed by the AC and the RC case. This finding can help us deduct the structure of the nuisance parameters in the AC case that is unknown to us. It seems that

the nuisance parameters in the AC case are decreasing with the variance of the coefficient process, while they increase with the persistence of the coefficient process. For example, for μ_{11} as low as 0.1, the empirical size of the OLS estimator is 17%, 15% and 33% for the AC, RC and the CC case, respectively.

(iv) For a value of μ_{11} greater than 0.5, even adding four lags of the first difference of the regressor to the DOLS/DGLS-type estimators is not sufficient to bring the empirical size close to the 5% nominal size regardless of the framework considered.

(v) The prewhitened version of the FMLS clearly dominates the standard one, both in terms of estimation precision and hypothesis testing. For the highly persistent CC case, the bias and the size of FM-PW is half the one of the FM-S estimator. However, in this case the performance of the FM-PW estimator is ambiguous since its bias is equal to 0.11 and the size is equal to 32%.

(vi) On the other hand, ADL, PL and JOH retain their desirable properties with empirical sizes in the range of 5.5% to 9.9%.

3.3 DGP3: The effect of persistence of the autoregressive coefficient (ϕ_{11})

In the third DGP, we isolate the effect of the persistence of the autoregressive coefficient of the cointegration error. Specifically, the parameter values are as follows:

DGP3: $\{M = \begin{pmatrix} 0.3 & 0 \\ 0 & 0 \end{pmatrix}, \Phi = \begin{pmatrix} \phi_{11} & 0 \\ 0 & 0 \end{pmatrix}, \Sigma_{12} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \Sigma_{34} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}, \theta = 1, \text{ and } \phi_{11} = 0.1, 0.3, 0.5, 0.7\}$. By setting Φ and Σ_{34} equal to zero, we compare

the performance of the cointegration estimators in the autoregressive coefficient environment against the constant coefficient one. The results – reported in Tables 3A-3B are summarised as follows:

[INSERT TABLES 3A-3B]

(i) The performance of all estimators under consideration deteriorates as the degree of persistence of the autoregressive coefficient of the cointegration error increases. This deterioration is more prevalent in terms of the inference quality of the estimators. For example, the empirical size of the OLS estimator increases from 41% to 50% as ϕ_{11} increases from 0.1 to 0.7. Similarly, the size of FM-PW and ADL(1,2) increases from 9%

and 7% to 15% and 9%, respectively.

(ii) The semi-parametric corrections to OLS (i.e. the FM estimators) reduce the distributional divergence of the t-statistic significantly. For example, in the case of ϕ_{11} equal to 0.7, the OLS size is 49.7%, while the respective figures for FM-S and FM-PW are 17.2% and 14.8%. The dominance of the prewhitened version of the FM estimator over the standard one is verified in this case, too.

(iii) The robustness of the JOH and the FM-PW estimators is questionable in this setting. These estimators seem to be more affected than the ADL and PL ones in this more random setting. For example, for a highly persistent autoregressive process, the size of JOH and FM-PW is around 15%, almost double the one of the ADL estimators.

(iv) Turning to the CC case, the best-performing estimators are the ADL, PL, JOH and the FM-PW estimators with an empirical size ranging between 6% and 8%. This class of estimators behave as expected intuitively. Their performance deteriorates as we move on to the AC case. This is not true, however, for the OLS and the DOLS(1) estimators whose performance is significantly worse in the CC case.

4 Conclusions

This study investigates the behavior of several widely used cointegration estimators in cases that allow for changes in the dynamics of the error generating mechanism. In particular, we consider cases where the cointegration error and the first difference of the regressor follow a VAR(1) process with VAR(1) coefficients. The Random Coefficient and the Constant Coefficient VAR(1) models can be thought of as special cases of the aforementioned wide class of models. Within this framework, we define error generating mechanisms that satisfy the conditions for the applicability of the FCLT.

The results from our simulation experiments may be summarised as follows:

First, the ADL and PL estimators and to a lesser extent the JOH and FM-PW estimators are robust to alternative DGPs involving changes in the error dynamics, such as those induced by the wide class of the VAR(1) models with VAR(1) coefficients. These estimators form a class of estimators, say Class A, which exhibits minimal bias and diver-

gence from the standard Normal distribution with empirical sizes that hardly exceed 10% in all the cases under consideration. The parametric estimators, DOLS(4) and DGLS(4) along with the semi-parametric Standard FM estimator seem to form a second class, say Class B, dominated by Class A in every aspect of statistical inference. Finally, the OLS, DOLS(1) and DGLS(1) estimator form Class C, a class of the worst performing estimators. Estimates and hypothesis testing based on these estimators can be totally misleading. Second, this ranking of the estimators remains unaltered when increasing the persistence of the VAR(1) coefficient process or the variance of the Random Coefficient process. Third, the performance of all estimators under consideration deteriorates as the degree of persistence of the autoregressive coefficient increases. This deterioration is more prevalent in terms of the inference quality of the estimators. Fourth, the bias and the size of the OLS estimator is a decreasing function of the variance of the coefficient of the first difference of the regressor (σ_{44}), in line with a decrease in the nuisance parameters. This does not hold for the Class A estimators whose performance worsens slightly with the degree of randomness.

All in all, our analysis draws some interesting conclusions for the applied researcher. He/she should at least employ the best performing estimators of Class A, whose robustness is not contingent upon changes in the error generating mechanism. Given that JOH and FM-PW are more susceptible in such an environment than the ADL and PL estimators, we highly recommend the employment of the latter. In any case, the safer strategy seems to be resorting to Monte Carlo simulations to reveal the small sample performance of the estimators employed. Provided that an adequate parameterisation of the real data can be obtained, this technique can always lead to valid estimation and inference on cointegration errors.

References

- [1] Andrews, D.W.K. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817-858.
- [2] Andrews, D.W.K. and J.C. Monahan (1992), An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, 60, 953-966.
- [3] Banerjee, A., Dolado, J.J., Galbraith, J.W. and D.F. Hendry (1993), *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*, Oxford, Oxford University Press.
- [4] Bewley, R.A (1979), The direct estimation of the equilibrium response in a linear model, *Economics Letters*, 3, 357-61.
- [5] Brüggemann, R. and H. Lutkepohl (2005), Practical problems with reduced-rank ML estimators for cointegration parameters and a simple alternative. *Oxford Bulletin of Economics and Statistics*, 67(5), 673-690.
- [6] Christou, C. and N. Pittis (2002), Kernel and bandwidth selection, prewhitening, and the performance of the fully modified least squares estimation method. *Econometric Theory*, 18, 948-961.
- [7] Johansen, S. (1988), Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.
- [8] Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59, 1551-1580.
- [9] Judge, G., R.C. Hill, W. Griffiths, H Lutkepohl and T.C. Lee (1988), *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons.
- [10] Lutkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, New York.

- [11]Newey, W.K. and K.D. West (1987), A simple positive, semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.
- [12]Nichols Des F. and B.G. Quinn (1982), *Random Coefficient Autoregressive Models: An Introduction*, Lecture Notes in Statistics, *Springer-Verlag*.
- [13]Panopoulou E. and N. Pittis (2004), A comparison of autoregressive distributed lag and dynamic OLS cointegration estimators in the case of a serially correlated cointegration error, *The Econometrics Journal*, 7(2), 585-617.
- [14]Phillips, P.C.B. (1988), Reflections on econometric methodology, *Economic Record*, 64, 344-359.
- [15]Phillips, P.C.B. and E.J. Hansen (1990), Statistical inference in instrumental regressions with I(1) processes, *Review of Economic Studies*, 57, 99-125.
- [16]Phillips, P.C.B. and S.N. Durlauf, (1986), Multiple time series regression with integrated processes, *Review of Economic Studies* 53, 473-496.
- [17]Phillips, P.C.B. and E.J. Hansen (1990), Statistical inference in instrumental variables regressions with I(1) processes, *Review of Economic Studies*, 57, 99-125.
- [18]Phillips, P.C.B. and M. Loretan (1991), Estimating long-run economic equilibria, *Review of Economic Studies*, 58, 407-436.
- [19]Saikkonen, P. (1991), Asymptotically efficient estimation of the cointegrating regressions, *Econometric Theory*, 7,1, 1-27.
- [20]Stock, J.H. and M.W. Watson (1993), A simple estimator of cointegrating vectors in higher-order integrated systems, *Econometrica*, 61, 783-820.
- [21]Wickens, M.R. and T.S. Breusch (1988), Dynamic specification, the long run and the estimation of transformed regression models, *Economic Journal*, 98, (Conference 1988), 189-205.
- [22]Weiss A. (1985) The stability of the AR(1) process with an AR(1) coefficient, *Journal of Time Series Analysis*, 6(3), 181-186.

Appendix A:

Stationarity Conditions for the RC case.

Proof to proposition 1: The starting point for the derivation of the stationarity conditions for our parameter space is the following theorem by Nicholls and Quinn (1982):

Theorem: *A unique F_t -measurable stationary solution $\{u_t\}$ exists to (8) if and only if the eigenvalues of M are less than unity in modulus and the matrix H given by $\text{vec}H = (I - CA)^{-1}\text{vec}\Sigma_{12}$ is positive definite, where $A = (I - M \otimes M)^{-1}$ and Σ_{12} is the covariance matrix of $\{e_t\}$. The covariance matrix V of u_t is then given by $\text{vec}V = A\text{vec}H$.*

The first condition reduces to $|\mu_{11}| < 1$ and $|\mu_{22}| < 1$ given the diagonality of M in our setup. Next, we derive matrix H and then the conditions that ensure that it is positive definite.

As stated in the theorem matrix H is given by $\text{vec}H = (I - CA)^{-1}\text{vec}\Sigma_{12}$, with C being equal to

$$C = E[P_t \otimes P_t] = E \left[\begin{pmatrix} \rho_{11t}\rho_{11t} & 0 & 0 & 0 \\ 0 & \rho_{11t}\rho_{22t} & 0 & 0 \\ 0 & 0 & \rho_{11t}\rho_{22t} & 0 \\ 0 & 0 & 0 & \rho_{11t}\rho_{11t} \end{pmatrix} \right] = \begin{pmatrix} \sigma_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{pmatrix},$$

$$A = (I - M \otimes M)^{-1} = \begin{pmatrix} \frac{1}{1-\mu_{11}^2} & 0 & 0 & 0 \\ 0 & \frac{1}{1-\mu_{11}\mu_{22}} & 0 & 0 \\ 0 & 0 & \frac{1}{1-\mu_{11}\mu_{22}} & 0 \\ 0 & 0 & 0 & \frac{1}{1-\mu_{22}^2} \end{pmatrix}, \text{ since}$$

$$I - M \otimes M = \begin{pmatrix} 1 - \mu_{11}^2 & 0 & 0 & 0 \\ 0 & 1 - \mu_{11}\mu_{22} & 0 & 0 \\ 0 & 0 & 1 - \mu_{11}\mu_{22} & 0 \\ 0 & 0 & 0 & 1 - \mu_{22}^2 \end{pmatrix}$$

$$\text{So, } (I - CA)^{-1} = \begin{pmatrix} \frac{1-\mu_{11}^2}{1-\mu_{11}^2-\sigma_{33}} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1-\mu_{22}^2}{1-\mu_{22}^2-\sigma_{44}} \end{pmatrix}, \text{vec}\Sigma_{12} = \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{12} \\ \sigma_{22} \end{pmatrix}$$

$$\text{Thus, } \text{vec}H = (I - CA)^{-1}\text{vec}\Sigma_{12} = \begin{pmatrix} \sigma_{11} \frac{1-\mu_{11}^2}{1-\mu_{11}^2-\sigma_{33}} \\ \sigma_{12} \\ \sigma_{12} \\ \sigma_{22} \frac{1-\mu_{22}^2}{1-\mu_{22}^2-\sigma_{44}} \end{pmatrix} \Rightarrow$$

$$H = \begin{pmatrix} \sigma_{11} \frac{1-\mu_{11}^2}{1-\mu_{11}^2-\sigma_{33}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \frac{1-\mu_{22}^2}{1-\mu_{22}^2-\sigma_{44}} \end{pmatrix}$$

Given that a symmetric matrix is positive definite if and only if all of its eigenvalues are positive, we next prove the condition based on its eigenvalues.⁸

If $\sigma_{12} = 0$, the eigenvalues of H are $\lambda_1 = \sigma_{11} \frac{1-\mu_{11}^2-\sigma_{33}}{1-\mu_{11}^2}$ and $\lambda_2 = \sigma_{22} \frac{1-\mu_{22}^2-\sigma_{44}}{1-\mu_{22}^2}$. Given that $\sigma_{11}, \sigma_{22}, 1-\mu_{11}^2, 1-\mu_{22}^2$ are all positive, H is positive definite $\iff 1-\mu_{11}^2-\sigma_{33} > 0$ and $1-\mu_{22}^2-\sigma_{44} > 0$. These conditions are equivalent to the ones for scalar RC(1) models, which is natural since when $\sigma_{12} = 0$, both u_{1t} and u_{2t} reduce to univariate RC(1) models.

Interestingly, in the case of $\sigma_{12} \neq 0$, the conditions $1-\mu_{11}^2-\sigma_{33} > 0$ and $1-\mu_{22}^2-\sigma_{44} > 0$ are sufficient for the positive definiteness of H .

By setting $1-\mu_{11}^2 = a, 1-\mu_{11}^2-\sigma_{33} = b, 1-\mu_{22}^2 = c$ and $1-\mu_{22}^2-\sigma_{44} = d$, the eigenvalues of H are

$$\lambda_1 = \frac{1}{2bd} \left(ad\sigma_{11} + bc\sigma_{22} - \sqrt{(ad\sigma_{11} + bc\sigma_{22})^2 + 4bd(bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{12})} \right)$$

and

$$\lambda_2 = \frac{1}{2bd} \left(ad\sigma_{11} + bc\sigma_{22} + \sqrt{(ad\sigma_{11} + bc\sigma_{22})^2 + 4bd(bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{12})} \right)$$

⁸See Judge et al. 1988, pp 960-961 for a proof.

In order to ensure that both eigenvalues are positive, it is sufficient to show that $\lambda_1 > 0$, since $\lambda_1 < \lambda_2$ for all parameter values.^{9,10} The following constraints need to be taken into account: $a, b, c, d, \sigma_{11}, \sigma_{22} > 0, |\sigma_{12}| < \sigma_{11}, |\sigma_{12}| < \sigma_{22}, b < a$ and $d < c$.¹¹ We need to show that

$$ad\sigma_{11} + bc\sigma_{22} > \sqrt{(ad\sigma_{11} + bc\sigma_{22})^2 + 4bd(bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{12})} \implies$$

$$4bd(bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{12}) < 0 \implies bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{12} < 0$$

which holds for any parameter configurations subject to the constraints set earlier.¹² Since $|\sigma_{12}| < \sigma_{11}, |\sigma_{12}| < \sigma_{22} \implies \sigma_{12}^2 < \sigma_{11}\sigma_{22}$ and $b < a, d < c \implies bd < ac$.

The covariance matrix V of u_t is given by $vecV = A vecH$, so $V = \begin{pmatrix} \frac{\sigma_{11}}{1-\mu_{11}^2-\sigma_{33}} & \frac{\sigma_{12}}{1-\mu_{11}\mu_{22}} \\ \frac{\sigma_{12}}{1-\mu_{11}\mu_{22}} & \frac{\sigma_{22}}{1-\mu_{22}^2-\sigma_{44}} \end{pmatrix}$

⁹The case of complex eigenvalues is ruled out by the symmetry of the matrix H (see Lutkepohl,1993, pp.456).

¹⁰ $(ad\sigma_{11} + bc\sigma_{22})^2 + 4bd(bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{12}) = (ad\sigma_{11} - bc\sigma_{22})^2 + 4b^2d^2\sigma_{12}^2 > 0$, for every parameter setting.

¹¹Since $\sigma_{33}, \sigma_{44} > 0 \implies 1 - \mu_{11}^2 < 1 - \mu_{11}^2 - \sigma_{33}$ and $1 - \mu_{22}^2 < 1 - \mu_{22}^2 - \sigma_{44}$

¹²Since $|\sigma_{12}| < \sigma_{11}, |\sigma_{12}| < \sigma_{22} \implies \sigma_{12}^2 < \sigma_{11}\sigma_{22}$ and $b < a, d < c \implies bd < ac$. Multiplying the inequalities, we have $bd\sigma_{12}^2 < ac\sigma_{11}\sigma_{22} \implies bd\sigma_{12}^2 - ac\sigma_{11}\sigma_{22} < 0$.

Appendix B: Nuisance parameters for the RC case

Three types of variances are associated with the process u_t as defined by equations (3) to (5), namely the conditional or contemporaneous covariance matrix, the unconditional covariance matrix and the long-run covariance matrix. In what follows, we derive these variances for the RC case.

Conditional/ contemporaneous covariance matrix:

$$E\{[u_t - E(u_t/u_{t-1})][u_t - E(u_t/u_{t-1})]'/u_{t-1}\} = E(\varepsilon_t \varepsilon_t') = \Sigma_{12}$$

Unconditional covariance matrix (V):

$$\begin{aligned} V &= E[u_t u_t'] = E[((M + P_t)u_{t-1} + \varepsilon_t)((M + P_t)u_{t-1} + \varepsilon_t)'] \Rightarrow \\ \text{vec}V &= \text{vec}[E[(M + P_t)u_{t-1}u_{t-1}'(M + P_t)' + \varepsilon_t \varepsilon_t']] \Rightarrow \\ \text{vec}V &= E[(M + P_t) \otimes (M + P_t)\text{vec}(u_{t-1}u_{t-1}') + \text{vec}\Sigma_{12}] \Rightarrow \\ \text{vec}V &= E[(M \otimes M + P_t \otimes P_t)]\text{vec}V + \text{vec}\Sigma_{12} \Rightarrow \\ \text{vec}V &= (M \otimes M + C)\text{vec}V + \text{vec}\Sigma_{12} \Rightarrow \\ \text{vec}V - (M \otimes M + C)\text{vec}V &= \text{vec}\Sigma_{12} \Rightarrow \\ \text{vec}V(I - M \otimes M - C) &= \text{vec}\Sigma_{12} \Rightarrow \end{aligned}$$

$$\text{vec}V = (I - M \otimes M - C)^{-1}\text{vec}\Sigma_{12}$$

Long-run covariance matrix (Ω):

The long-run covariance matrix Ω can be decomposed into three matrices: the unconditional covariance matrix V , and two temporal matrices Λ and Λ' , i.e. $\Omega = V + \Lambda + \Lambda'$, where $\Lambda' = E[T^{-1} \sum_{t=1}^T \sum_{s=t+1}^T u_t u_s'] = \sum_{s=t+1}^T E(u_t u_{t-s}')$.

For $s = 1$, we have $E(u_t u_{t-1}') = E[((M + P_t)u_{t-1} + \varepsilon_t)u_{t-1}'] = E[(M + P_t)]E[u_{t-1}u_{t-1}'] = MV$.

For $s = 2$, we have $E(u_t u_{t-2}') = E[((M + P_t)(M + P_{t-1})u_{t-2} + (M + P_t)\varepsilon_{t-1} + \varepsilon_t)u_{t-2}'] = E[(M + P_t)(M + P_{t-1})]E[u_{t-2}u_{t-2}'] = M^2V$.

For $s = T$, we have $E(u_t u_{t-T}') = M^T V$.

So Λ' reduces to $\Lambda' = MV + M^2V + \dots$ and given that all the eigenvalues of Λ are less than unity in modulus, we have:

$$\Lambda = M(I + M + M^2 + \dots)V = M(I - M)^{-1}V.$$

$$\text{In a similar mode, } \Lambda = V(I - M')^{-1}M'.$$

Adding up the three terms we have

$$\Omega = V + \Lambda + \Lambda' \Rightarrow$$

$$\Omega = V + M(I - M)^{-1}V + V(I - M')^{-1}M' \Rightarrow V(I + M(I - M)^{-1} + (I - M')^{-1}M') \Rightarrow$$

$$\Omega = V((I - M)^{-1} + (I - M')^{-1}M')$$

With some algebra, we relate the parameters of our DGP to the nuisance parameters, the first ω_{12}/ω_{22} that describes the “long-run correlation” effect, due to non-diagonality of the long run covariance matrix $\Omega = [\omega_{ij}]$, $i, j = 1, 2$ and the second $\delta_{21} = \sum_{k=0}^{\infty} E(u_{20}u_{1k})$ that describes the “endogeneity” effect. Specifically, we have:

$$\Sigma_{12} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

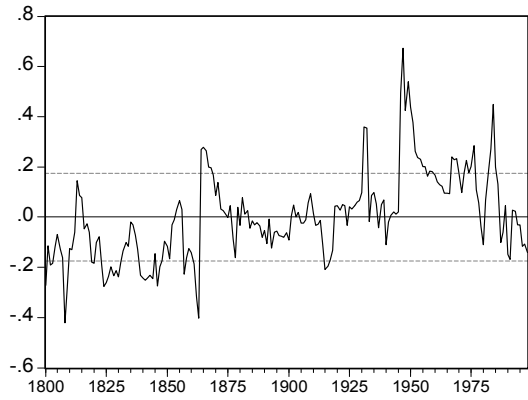
$$V = \begin{pmatrix} \frac{\sigma_{11}}{1 - \mu_{11}^2 - \sigma_{33}} & \frac{\sigma_{12}}{1 - \mu_{11}\mu_{22}} \\ \frac{\sigma_{12}}{1 - \mu_{11}\mu_{22}} & \frac{\sigma_{22}}{1 - \mu_{22}^2 - \sigma_{44}} \end{pmatrix}$$

$$\Omega = V + \Lambda + \Lambda' = \begin{pmatrix} \frac{(1 + \mu_{11})\sigma_{11}}{(1 - \mu_{11})(1 - \mu_{11}^2 - \sigma_{33})} & \frac{\sigma_{12}}{(1 - \mu_{11})(1 - \mu_{22})} \\ \frac{\sigma_{12}}{(1 - \mu_{22})(1 - \mu_{22})} & \frac{(1 + \mu_{22})\sigma_{22}}{(1 - \mu_{22})(1 - \mu_{22}^2 - \sigma_{44})} \end{pmatrix}$$

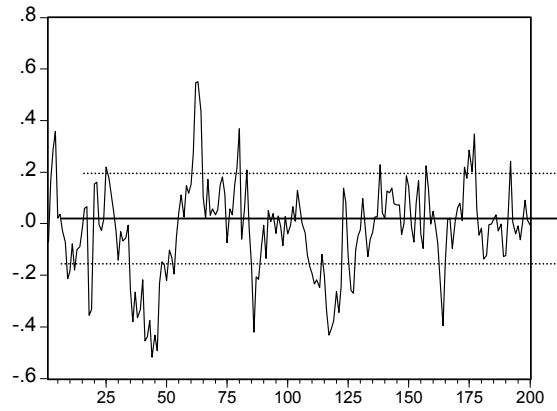
$$\Delta = V + \Lambda = \begin{pmatrix} \frac{\sigma_{11}}{(1 - \mu_{11})(1 - \mu_{11}^2 - \sigma_{33})} & \frac{\sigma_{12}}{(1 - \mu_{22})(1 - \mu_{11}\mu_{22})} \\ \frac{\sigma_{12}}{(1 - \mu_{11})(1 - \mu_{11}\mu_{22})} & \frac{\sigma_{22}}{(1 - \mu_{22})(1 - \mu_{22}^2 - \sigma_{44})} \end{pmatrix}$$

Figure 1

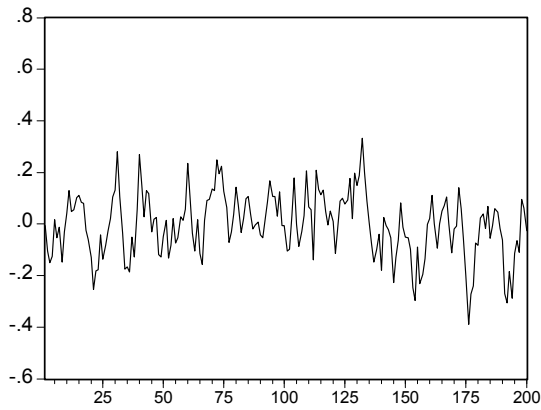
A. UK: PPP (cointegration residual)



B. Artificial data (AR(1) process with AR(1) coefficient)



C. Artificial data (AR(1) process with Constant Autoregressive Coefficient equal to 0.7)



D. Artificial data (AR(1) process with Constant Autoregressive Coefficient equal to 0.9)

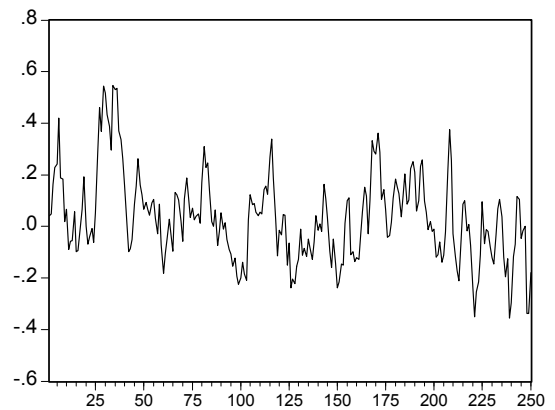
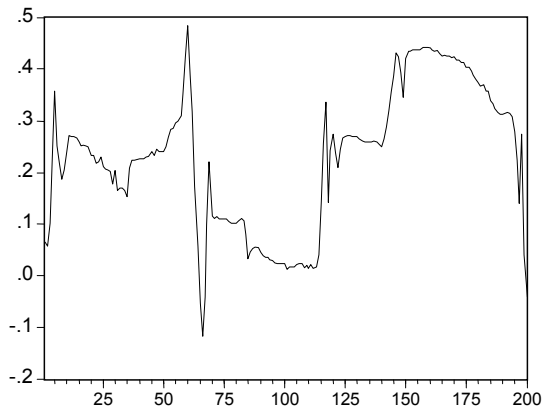


Figure 2

A. UK PPP series: AR coefficient of cointegration residual
(Rolling 80 obs. window)



B. VAR(1) with VAR(1) coefficient process
AR coefficient of cointegration residual
(Rolling 80 obs. window)



C. VAR(1) with Constant Coefficients
AR coefficient of cointegration residual
(Rolling 80 obs. window)

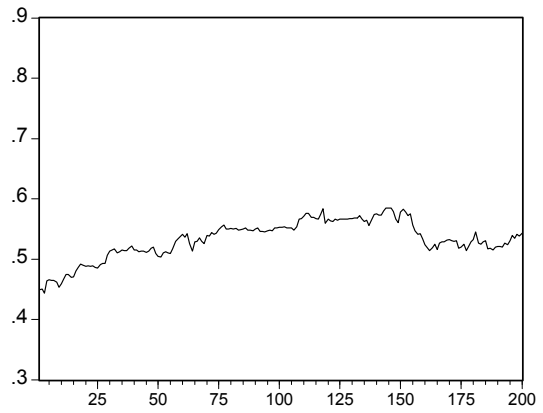


Table 1: Properties of estimators (DGP1)

DGP 1: $M = \begin{pmatrix} 0.7 & 0 \\ 0 & 0 \end{pmatrix}$, $\Phi = \begin{pmatrix} 0.3 & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma_{12} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$, $\Sigma_{34} = \begin{pmatrix} 0.1 & 0 \\ 0 & \sigma_{44} \end{pmatrix}$, $\theta = 1$

Panel A	AC ($\sigma_{44} = 0.1$)		RC ($\sigma_{44} = 0.1, \Phi = \mathbf{0}$)		CC ($\Sigma_{34} = \Phi = \mathbf{0}$)	
Estimator	Mean bias	Size	Mean bias	Size	Mean bias	Size
OLS	0.134	66.50	0.128	67.95	0.141	76.15
DOLS(1)	0.100	39.50	0.090	40.10	0.100	52.55
DOLS(4)	0.044	25.65	0.031	23.45	0.036	30.55
DGLS(1)	0.458	91.85	0.450	91.80	0.631	99.85
DGLS(4)	0.106	21.20	0.065	15.95	0.084	32.35
ADL(1,2)	0.001	8.30	0.005	7.55	0.004	7.40
ADL(4,4)	0.009	9.50	0.004	9.10	0.003	8.40
PL(1,0)	0.001	8.30	0.005	7.55	0.004	7.40
PL(4,0)	0.009	10.40	0.003	9.80	0.003	9.35
JOH	-0.004	8.95	-0.001	6.95	0.000	7.15
FM-S	0.171	24.10	0.143	21.40	0.070	31.20
FM-PW	0.098	13.90	0.078	12.95	0.022	12.00
Panel B	AC ($\sigma_{44} = 0.5$)		RC ($\sigma_{44} = 0.5, \Phi = \mathbf{0}$)			
Estimator	Mean bias	Size	Mean bias	Size		
OLS	0.064	55.05	0.077	58.20		
DOLS(1)	0.046	29.75	0.055	30.00		
DOLS(4)	0.019	24.10	0.018	20.95		
DGLS(1)	0.160	50.60	0.186	60.05		
DGLS(4)	0.040	12.85	0.036	11.65		
ADL(1,2)	-0.004	10.05	0.004	8.85		
ADL(4,4)	0.002	11.30	0.002	10.40		
PL(1,0)	-0.004	10.05	0.004	8.85		
PL(4,0)	0.001	11.85	0.001	10.50		
JOH	-0.013	10.65	-0.002	8.75		
FM-S	0.086	23.05	0.089	20.85		
FM-PW	0.055	15.05	0.054	13.60		
Panel C	AC ($\sigma_{44} = 0.9$)		RC ($\sigma_{44} = 0.9, \Phi = \mathbf{0}$)			
Estimator	Mean bias	Size	Mean bias	Size		
OLS	0.028	47.35	0.039	47.40		
DOLS(1)	0.021	25.20	0.028	24.80		
DOLS(4)	0.009	22.90	0.010	21.70		
DGLS(1)	0.054	23.75	0.074	31.95		
DGLS(4)	0.015	10.70	0.018	10.20		
ADL(1,2)	-0.002	9.60	0.003	9.55		
ADL(4,4)	0.000	10.60	0.001	11.05		
PL(1,0)	-0.002	9.60	0.003	9.55		
PL(4,0)	-0.001	10.90	0.000	11.10		
JOH	-0.005	11.70	0.000	9.85		
FM-S	0.039	20.75	0.048	20.25		
FM-PW	0.025	13.05	0.031	13.25		

Table 2: Properties of estimators (DGP2)

DGP 2: $M = \begin{pmatrix} \mu_{11} & 0 \\ 0 & 0 \end{pmatrix}$, $\Phi = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$, $\Sigma_{12} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$, $\Sigma_{34} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$, $\theta = 1$.

Panel A	AC ($\mu_{11} = 0.1$)		RC ($\mu_{11} = 0.1, \Phi = \mathbf{0}$)		CC ($\mu_{11} = 0.1, \Sigma_{34} = \Phi = \mathbf{0}$)	
Estimator	Mean bias	Size	Mean bias	Size	Mean bias	Size
OLS	0.031	16.90	0.029	14.85	0.053	32.75
DOLS(1)	0.004	9.90	0.003	9.75	0.006	10.50
DOLS(4)	-0.001	9.70	-0.001	8.85	0.000	8.85
DGLS(1)	0.006	8.20	0.004	7.65	0.006	7.45
DGLS(4)	-0.001	8.10	-0.001	7.40	0.000	5.75
ADL(1,2)	0.001	8.50	-0.001	7.50	0.000	5.60
ADL(4,4)	0.000	8.85	0.000	8.70	0.000	7.10
PL(1,0)	0.001	8.50	-0.001	7.50	0.000	5.60
PL(4,0)	-0.001	9.10	-0.001	8.75	0.000	8.25
JOH	0.000	6.75	-0.001	6.00	0.000	5.45
FM-S	0.016	9.60	0.016	9.60	0.008	8.45
FM-PW	0.015	9.80	0.015	9.80	0.008	6.50
Panel B	AC ($\mu_{11} = 0.5$)		RC ($\mu_{11} = 0.5, \Phi = \mathbf{0}$)		CC ($\mu_{11} = 0.5, \Sigma_{34} = \Phi = \mathbf{0}$)	
Estimator	Mean bias	Size	Mean bias	Size	Mean bias	Size
OLS	0.201	40.95	0.048	31.90	0.090	63.40
DOLS(1)	0.201	13.85	0.024	12.60	0.046	34.75
DOLS(4)	0.216	12.15	0.001	11.00	0.006	15.00
DGLS(1)	0.268	12.20	0.047	11.70	0.265	80.50
DGLS(4)	0.428	9.95	0.004	6.85	0.009	7.95
ADL(1,2)	0.099	8.80	0.000	6.70	0.001	6.75
ADL(4,4)	0.028	9.70	0.001	8.20	0.001	7.70
PL(1,0)	0.096	8.80	0.000	6.70	0.001	6.75
PL(4,0)	0.025	9.90	0.001	9.00	0.001	8.30
JOH	0.105	9.90	-0.001	6.15	0.000	6.05
FM-S	0.053	13.90	0.047	13.65	0.032	19.00
FM-PW	0.043	11.00	0.035	10.05	0.011	7.90
Panel C	RC ($\mu_{11} = 0.7, \Phi = \mathbf{0}$)		CC ($\mu_{11} = 0.7, \Sigma_{34} = \Phi = \mathbf{0}$)			
Estimator	Mean bias	Size	Mean bias	Size		
OLS	0.071	39.00	0.141	76.15		
DOLS(1)	0.049	16.25	0.100	52.55		
DOLS(4)	0.012	14.50	0.036	30.55		
DGLS(1)	0.112	19.05	0.631	99.85		
DGLS(4)	0.029	7.90	0.084	32.35		
ADL(1,2)	0.007	7.60	0.004	7.40		
ADL(4,4)	0.007	9.00	0.003	8.40		
PL(1,0)	-0.067	7.60	0.004	7.40		
PL(4,0)	-0.076	9.35	0.003	9.35		
JOH	-0.011	7.85	0.000	7.15		
FM-S	0.109	17.40	0.070	31.20		
FM-PW	0.095	10.55	0.022	12.00		

Table 3: Properties of estimators (DGP3)

DGP 3: $M = \begin{pmatrix} 0.3 & 0 \\ 0 & 0 \end{pmatrix}$, $\Phi = \begin{pmatrix} \phi_{11} & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma_{12} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$, $\Sigma_{34} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$, $\theta = 1$.

Panel A	AC ($\phi_{11} = 0.1$)		AC ($\phi_{11} = 0.3$)		CC ($\Sigma_{34} = \Phi = \mathbf{0}$)	
Estimator	Mean bias	Size	Mean bias	Size	Mean bias	Size
OLS	0.061	41.40	0.063	43.15	0.067	48.40
DOLS(1)	0.019	14.75	0.021	15.40	0.021	21.45
DOLS(4)	0.000	11.55	0.001	12.15	0.001	10.60
DGLS(1)	0.030	14.90	0.035	16.35	0.033	22.80
DGLS(4)	0.001	7.20	0.002	7.70	0.001	6.20
ADL(1,2)	0.000	7.30	0.002	7.35	0.001	6.20
ADL(4,4)	0.000	8.55	0.001	8.45	0.001	7.15
PL(1,0)	0.000	7.30	0.002	7.35	0.001	6.20
PL(4,0)	0.000	8.85	0.000	9.10	0.001	8.25
JOH	0.000	6.40	0.000	7.70	0.000	5.60
FM-S	0.037	11.10	0.041	12.30	0.017	12.25
FM-PW	0.025	8.80	0.030	10.05	0.008	6.80

Panel B	AC ($\phi_{11} = 0.5$)		AC ($\phi_{11} = 0.7$)	
Estimator	Mean bias	Size	Mean bias	Size
OLS	0.067	46.10	0.080	49.70
DOLS(1)	0.025	16.25	0.039	17.50
DOLS(4)	0.003	13.05	0.014	14.25
DGLS(1)	0.046	20.80	0.086	28.60
DGLS(4)	0.005	8.15	0.031	10.60
ADL(1,2)	0.004	7.75	0.013	8.50
ADL(4,4)	0.001	8.35	0.007	8.55
PL(1,0)	0.004	7.75	0.013	8.50
PL(4,0)	0.001	9.25	0.103	9.85
JOH	0.000	10.25	-0.001	14.05
FM-S	0.049	13.85	0.088	17.15
FM-PW	0.040	11.65	0.081	14.75