UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 13-12
Economic Series
June, 2013

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Co-summability
From Linear to Non-linear Co-integration[*]

Vanessa Berenguer-Rico
University of Oxford and Institute for New Economic Thinking

Jesús Gonzalo
Universidad Carlos III de Madrid

June 22, 2013

ABSTRACT

While co-integration theory is an ideal framework to study linear relationships among persistent economic time series, the intrinsic linearity in the concepts of integration and co-integration makes it unsuitable to study non-linear long run relations among persistent processes. This drawback hinders the empirical analysis of modern macroeconomics, which often addresses asymmetric responses to policy interventions, multiplicity of equilibria, transitions between regimes or polynomial approximations to unknown functions.

In this paper, to cope with non-linear relations and consequently to generalise co-integration, we formalise the idea of *co-summability*. It is built upon the concept *order of summability* developed by Berenguer-Rico and Gonzalo (2013), which, in turn, was conceived to address non-linear transformations of persistent processes. Theoretically, a co-summable relationship is *balanced* -in terms of the variables involved having the same order of summability- and describes a *long run equilibrium* that can be non-linear -in the sense that the errors have a lower order of summability. To test for these types of equilibria, inference tools for balancedness and cosummability are designed and their asymptotic properties are analysed. Their finite sample performance is studied via Monte Carlo experiments.

The practical strength of co-summability theory is shown through two empirical applications. Specifically, asymmetric preferences of central bankers and the environmental Kuznets curve hypothesis are studied through the lens of co-summability.

Key words: Balancedness, Co-integration, Co-summability, Non-linear Co-integration, Non-linear Processes, Persistence

JEL Codes: C01, C22

[*] The authors acknowledge comments from J.Ll. Carrion-i-Silvestre, J.J. Dolado, B. Nielsen, G. Perez-Quiros, P.C.B. Phillips, J. Pitarakis, C. Velasco, and seminar and conference participants at Nu¢ eld/INET, NES/CEFIR, ECARES, UCM, XXXVI SAE, 5th CFE, 2012 Tsinghua Econometrics, SETA 2012, Southampton Spring Econometrics Eventand ASSA Meetings 2013 for their helpful insights. Financial support from the University of Oxford, Institute for New Economic Thinking, Spanish Ministerio de Ciencia e Innovación (grants ECO2010-19357 and Consolider-2010), Comunidad de Madrid (grant Excelecon) and Bank of Spain (grant ER program) is gratefully acknowledged.

1 Introduction

Co-integration theory has received a great deal of attention from economists and econometricians. From a theoretical perspective, co-integration has played the role of properly combining persistent economic time series with linear long run equilibrium relationships hypothesised by economic theorists. In this sense, co-integration meant a positive step towards consensus in the historical "*measurement without theory*" vs "*theory without measurement*" debate. Economic theories implying co-integrating relationships among economic time series contributed to this step. From an empiricist perspective, co-integration resulted in a clear and precise applied methodology to estimate and test these economic hypotheses.

To provide richer descriptions of economic phenomena, researchers have ventured into the non-linear world. However, the ideas of integration and co-integration cannot be directly used to analyse non-linear equilibrium relationships among persistent variables as these concepts do not properly apply. To be more precise, consider the following non-linear relationship: $y_t = f(x_t, \theta) + u_t$. If it were known that $f(x_t, \theta)$ is $I(d)$, then the standard framework of co-integration would fit perfectly. However, when x_t is persistent, say $I(1)$, then for many interesting non-linear transformations f the order of integration of $f(x_t, \theta)$ may not be well defined. This failure of applicability of the definition of order of integration has two important drawbacks. First, it is not possible to know whether a postulated relationship is balanced –a necessary, although not sufficient, condition for having correctly specified a model. Second, the concept of co-integration cannot be directly extended to non-linear long run relationships. These two consequences originate a clear need for theoretically valid and empirically useful concepts that generalise those of integration and co-integration.

This paper proposes to use the idea of order of summability formalised by Berenguer-Rico and Gonzalo (2013). It was conceived to deal both theoretically and empirically with non-linear transformations of heterogeneous and persistent processes. By making use of this new concept, co-integration theory can be generalised by defining (i) *balancedness* –the order of summability of an explained variable in a postulated hypothesis being equal to that of the, possibly non-linear, more persistent and heterogeneous explanatory variables– and (ii) *co-summability* –the error term of the postulated hypothesis being of a lower order of summability. These two factors are relevant for both econometricians and economic theorists: for the former, when specifying, estimating, and testing econometric models; for the latter when choosing functional forms to construct their theories.

By taking advantage of the order of summability estimator, balancedness and co-summability can be empirically studied. To infer if a postulated relationship is balanced, the rate of convergence estimator in McElroy and Politis (2007) and subsampling techniques can be used. Once balancedness is achieved, researchers must distinguish between spurious and co-summable regressions. This paper proposes a residual based test to disentangle that question; therefore, an estimate of the errors is needed. Parametric and non-parametric approaches to estimate non-linear long run relationships are available in the literature. Park and Phillips (1999, 2001) and Wang and Phillips (2009) develop parametric and non-parametric methods, respectively, from an integrated processes perspective. Alternatively, Karlsen, Myklebust and Tjϕstheim (2007) and Schienle (2011) analyse non-parametric estimation in a recurrent Markov chains setup. Notwithstanding, all these studies assume that the regression model specifies a co-integrating relation, something that should be tested in practice. There have been some –rather limited– proposals in this direction –see, for example, Kasparis (2008) or Choi and Saikkonen (2010).

In this paper, parametric regression models that are non-linear in variables but linear in parameters will be taken into consideration. A more general setting where the model is not only non-linear in variables but also in parameters requires a different empirical, and therefore theoretical, strategy. The non-linear in variables but linear in parameters model considered here, although simple at first sight, is rich enough for empirical purposes, while, at the same time, enclosing its own theoretical features to be analysed by itself. In this scenario, the asymptotic properties of the ordinary least squares estimator under co-summability and no co-summability are studied. These properties guarantee being able to discriminate between spurious and co-summable regressions through a residual based test, which can also be seen as a specification testing procedure.

A natural question arising after finding or defining a non-linear co-summable relationship is whether an error correction representation does exist in a non-linear world. The question is natural given that error correction mechanisms in this framework involve first differences of non-linear processes, which are not properly defined in terms of order of integration. Nevertheless, whether the world is linear or not, modelling the reaction of endogenous variables in a system to deviations of its equilibrium is an important issue. This paper also addresses this question –in a single equation framework– emphasising the fact that to study non-linear error correction models, the ideas of summability and co-summability become a key aspect. Indeed, while balancedness of the error correction representation of a non-linear equilibrium cannot be addressed using the linear concepts of integration and co-integration, it can be analysed using co-summability.

To show the empirical strength of the co-summability theory, the proposed tools are put into practice with two different empirical applications where non-linear transformations of persistent processes occur. Specifically, asymmetric preferences of central bankers and the environmental Kuznets curve are analysed. The former hypothesis is translated in the literature into non-linear Taylor rules when conducting monetary policy –see, for instance, Clarida and Gertler (1997) or Dolado, María-Dolores and Naveira (2005). These non-linearities and the fact that the variables involved in this type of rules are found to be persistent make co-summability appropriate in this context. The latter hypothesis, the environmental Kuznets curve, postulates an inverted U-shaped relationship between pollution and economic development, usually measured by CO_2 emissions and GDP, respectively. Again, this non-linear relationship, typically approximated by a polynomial function, jointly with the well documented persistence of these two measures makes this hypothesis another natural economic context where co-summability theory rightly fits. The empirical findings provide new insights for the econometric treatment of these two hypotheses. In the Taylor rule case, the linear specification does not define a long run relationship –co-summability does not hold– thus suggesting a possible misspecification. Following the asymmetric preferences of central bankers literature, we find that a threshold Taylor rule is not rejected –co-summability holds. Specifically, it is found that the Federal Reserve reacts very asymmetrically to recessions and expansions. With respect to the environmental Kuznets curve, favourable evidence is found when variables are included in logarithms and the polynomial function is of third degree.

The paper is organised as follows. In Section 2, balancedness and co-summability are formally defined and discussed through some economic examples. Section 3 develops an empirical strategy to test for co-summability. First, a test for balancedness is designed. Then, a test for co-summability is proposed. The finite sample performance of these procedures is studied via simulations. Section 4 discusses the error correction representation of a non-linear co-summable relationship, highlighting the fundamental role of the ideas of summability and co-summability when studying this type of representations. In Section 5, the proposed tools are applied to test for the asymmetric preferences of central bankers and the environmental Kuznets curve hypothesis. Section 6 finishes with some concluding remarks. All the proofs are collected in the Appendix.

A word on notation. We use the symbol "$\Longrightarrow$" to signify convergence in distribution and weak convergence indistinctly and "$\xrightarrow{p}$" to signify convergence in probability. By the D-space analogue of a process $y_{nt} = y_t/n^{\alpha_y}$ it is meant $y_n(r) = y_{[nr]}/n^{\alpha_y}$ for $0 \leq r \leq 1$ and where $[.]$ denotes the greatest integer part. Stochastic processes such as $D_y(r)$ or the standard Brownian motion $W(r)$ are defined on $[0,1]$. Finally, all limits given in this paper are taken as the sample size $n \to \infty$.

2 Balancedness and Co-summability

2.1 Order of Summability

The subsequent theory relies on the idea of order of summability of stochastic processes. It was first introduced in a heuristic way by Gonzalo and Pitarakis (2006) and subsequently formalised in Berenguer-Rico and Gonzalo (2013) –BG hereafter.

Definition 1 *: A stochastic process $\{y_t : t \in \mathbb{N}\}$ is said to be summable of order δ, or $S(\delta)$, if there* *exist a slowly varying function*[1] $L(n)$ *and a deterministic sequence* m_t *such that*

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1) \qquad \text{as } n \to \infty, \tag{1}$$

where δ *is the minimum real number that makes* S_n *bounded in probability.*

The order of summability, δ, gives a summary measure of the stochastic properties –persistence and evolution of the variance– of y_t without relying on a particular data generating process. The following examples show the usefulness of this new concept and how to calculate δ.

Let

$$\pi_t = \pi_{t-1} + \varepsilon_t, \tag{2}$$

with $\pi_0 = 0$ and $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$.

Example 1 *: Square of a random walk*

Let us consider the order of integration of

$$\pi_t^2 = \pi_{t-1}^2 + 2\pi_{t-1}\varepsilon_t + \varepsilon_t^2. \tag{3}$$

Granger (1995) points out that π_t^2 can be seen as a random walk with drift; hence, one could think that π_t^2 is $I(1)$. However,

$$V[\pi_t^2 - \pi_{t-1}^2] = 4(t-1)\sigma_\varepsilon^4 + E[\varepsilon_t^4].$$

In fact, the variance of $\Delta^d \pi_t^2$ depends on t regardless of the values of d, i.e. $\pi_t^2 \sim I(\infty)$; but this is not a useful characterisation in practice. Instead, the order of summability can be easily obtained. Given that

$$S_n = \frac{1}{n^2 \sigma_\varepsilon^2} \sum_{t=1}^{n} \pi_t^2 \Longrightarrow \int_0^1 W^2(r) dr,$$

π_t^2 is $S(1.5)$.

Example 2 *: Product of Indicator Function and Random Walk*

Let

$$h_t = 1(v_t \leq \gamma)\pi_t, \tag{4}$$

[1] A positive, Lebesgue measurable function L, on $(0, \infty)$ is slowly varying –in the Karamata's sense– at ∞ if

$$\frac{L(\lambda n)}{L(n)} \to 1 \quad (n \to \infty)\ \forall \lambda > 0.$$

(See Embrechts, Klüppelberg and Mikosh, 1999, p.564).

where $v_t \sim i.i.d.\,(0,1)$ is independent of ε_t, $1(\cdot)$ is the indicator function and γ is a constant. Strictly speaking, $h_t \sim I(\infty)$ as the variance of $\Delta^d h_t$ depends on t regardless of the values of d. Nevertheless,

$$S_n = \frac{1}{n^{\frac{3}{2}} p \sigma_\varepsilon} \sum_{t=1}^{n} h_t \Longrightarrow \int_0^1 W(r)dr,$$

where $p = \Pr(v_t \leq \gamma)$, which implies that h_t is $S(1)$.

Table 1 summarises many other univariate examples considered by BG.

Table 1: Examples: $I(d)$ vs $S(\delta)$

DGP	$I(d)$	$S(\delta)$
$y_{1t} \sim i.i.d.F \in D(\alpha)$	$I(?)$	$S((2-\alpha)/2\alpha)$
$y_{2t} = z + \varepsilon_t$	$I(?)$	$S(1/2)$
$y_{3t} \sim I(d)$	$I(d)$	$S(d)$
$y_{4t} = \pi_t \eta_t$	$I(\infty)$	$S(1/2)$
$y_{5t} = \pi_t \eta_t^2$	$I(\infty)$	$S(1)$
$y_{6t} = \pi_t^2$	$I(\infty)$	$S(3/2)$
$y_{7t} = 1(v_t \leq \gamma)\pi_t$	$I(\infty)$	$S(1)$
$y_{8t} = e^{-\pi_t^2}$	$I(?)$	$S(1/2)$
$y_{9t} = 1/(1+\pi_t^2)$	$I(?)$	$S(1/2)$
$y_{10t} = \log(\lvert\pi_t\rvert)$	$I(?)$	$S(1/2)$
$y_{11t} = (1+e^{-\pi_t})^{-1}$	$I(?)$	$S(1/2)$
$y_{12t} = \rho_t y_{12,t-1} + \varepsilon_t$	$I(\infty)$	$S(\infty)$
$y_{13t} = \phi y_{13,t-1} + \varepsilon_t;\quad \phi > 1$	$I(\infty)$	$S(\infty)$

$D(\alpha)$ denotes the domain of attraction of an α-stable law with $\alpha \in (0,2]$; $z \sim N(0,1)$; $\varepsilon_t \sim i.i.d.\,(0,1)$; $\pi_t = \pi_{t-1} + \varepsilon_t$ and $\pi_0 = 0$; $\eta_t \sim i.i.d.\,(0,1)$; $v_t \sim i.i.d.\,(0,1)$; $\rho_t \sim i.i.d.\,(1,1)$. z, ε_t, η_t, v_t, and ρ_t are independent of each other. In all the DGPs but y_{3t} with $d = 0.5$ and y_{10t} the slowly varying function $L(n)$ is a constant; for y_{3t} with $d = 0.5$ and y_{10t}, $L(n) = 1/log(n)$.

From a multivariate perspective, an applied economist often starts the analysis from a postulated economic relationship, say $y_t = g(x_t, \theta)$. Then, recognising that it is just an approximation to reality and θ is typically unknown, the difference $u_t = y_t - g(x_t, \theta)$ is statistically analysed.

Assumption 0.

$$S_{yn} = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} y_t \Longrightarrow D_y \quad \text{and} \quad S_{gn} = \frac{1}{n^{1/2+\delta_g}} \sum_{t=1}^{n} g(x_t, \theta) \Longrightarrow D_g,$$

where D_y and D_g are two random variables with positive variance.

Under Assumption 0, $y_t \sim S(\delta_y)$ and $g(x_t, \theta) \sim S(\delta_g)$. This weak assumption will be particularly convenient to put forward the balancedness of a theoretical hypothesis as well as to develop the inference theory.

2.2 Balancedness

Definition 2 *: A postulated relationship*

$$y_t = g\left(x_t, \theta\right),$$

is said to be balanced if $y_t \sim S\left(\delta_y\right)$, $g\left(x_t, \theta\right) \sim S\left(\delta_g\right)$*, and* $\delta_y = \delta_g$.

Given a theoretical hypothesis

$$y_t = g\left(x_t, \theta\right), \tag{5}$$

the order of summability of x_t, δ_x, could differ from that of $g\left(x_t, \theta\right)$, δ_g. This means that given δ_y and δ_x, there will be only some appropriate functions g that will generate balanced relationships[2], i.e., $\delta_y = \delta_g$. This is not only important for econometricians but also for economic theorists when choosing functional forms to construct their theories.

Indeed, under Assumption 0, an unbalanced postulated model is clearly misspecified –in a wide sense. When $\delta_y > \delta_g$,

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} g(x_t, \theta) = o_p\left(1\right),$$

which contradicts Assumption 0. If $\delta_y < \delta_g$,

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} g(x_t, \theta),$$

with the right-hand side being unbounded. Again, a contradiction with Assumption 0. Hence, balancedness becomes a necessary, although not sufficient, condition for a correct specification. Particular economic examples will show the relevance of balancedness in practice.

Example 3 *: Endogenous Growth Models (Jones, 1995)*

Endogenous growth theory implies that permanent changes in policy variables, such as the investment rate in physical capital, have permanent effects on the rate of economic growth. The equation of interest is

$$g_{yt} = -\ddot{\delta} + \tilde{A} i_{kt}, \tag{6}$$

where g_{yt} is the growth rate of the economy, $\ddot{\delta}$ is the rate of depreciation, $\tilde{A}$ measures the total factor productivity, and i_{kt} is the investment rate in physical capital. If this equation is balanced, then the persistence of the growth rate would be similar to that of the investment rate. Nevertheless, using time series techniques, it is found that US growth rates exhibit no large persistent changes, while large and permanent movements are found in investment rates. Hence, Jones (1995) argues

[2] In fact, the order of summability of $g\left(x_t, \theta\right)$ could even depend on θ –let, for instance, x_t be a standard random walk and $g\left(x_t, \theta\right) = x_t^{\theta}$, then $\delta_g = \theta/2$.

that endogenous growth models are rejected by this criterion.

Balancedness will be particularly important in non-linear models involving persistent variables. As stated in Granger (1995), non-linear transformations of heterogeneous and persistent processes can have an important impact on their stochastic properties. This impact could be hardly contemplated by the order of integration but can be asserted by the order of summability. The next examples illustrate this point.

Example 4 *: Central Bankers with Asymmetric Preferences*

Consider a central bank with asymmetric preferences with respect to deviations of inflation or output from some particular target level. Under such preferences, the central bank would react more or less aggressively when inflation or output deviates from above, rather than from below, the target. Different modelisations of this hypothesis based on Taylor rules can be found in the literature. For instance, Clarida and Gertler (1997) study the following threshold type of Taylor rule for the Bundesbank

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t 1\left(\tilde{\pi}_t > 0\right) + \theta_2 \tilde{\pi}_t 1\left(\tilde{\pi}_t \leq 0\right) + \theta_3 \tilde{y}_t 1\left(\tilde{\pi}_t > 0\right) + \theta_4 \tilde{y}_t 1\left(\tilde{\pi}_t \leq 0\right), \quad (7)$$

where i_t denotes interest rates, $\tilde{\pi}_t$ are deviations from the inflation target, and $\tilde{y}_t$ is the output gap. On the other hand, Dolado, María-Dolores and Naveira (2005), allowing for a non-linear Phillips curve, derive the following type of optimal monetary policy rule

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t + \theta_2 \tilde{y}_t + \theta_3 \tilde{\pi}_t \tilde{y}_t. \quad (8)$$

In both cases, studying balancedness of these equations will be troublesome using the $I(d)$ framework. Even if it can be said that i_t, $\tilde{\pi}_t$, and $\tilde{y}_t$ are $I(d_i)$, $I(d_{\tilde{\pi}})$, and $I(d_{\tilde{y}})$, respectively, the order of integration of $\tilde{\pi}_t 1\left(\tilde{\pi}_t \geq 0\right)$ or $\tilde{\pi}_t \tilde{y}_t$ would not be well defined. Nevertheless, the generality of the order of summability makes it suitable to be used in both situations. See the empirical application section.

Example 5 *: Environmental Kuznets Curve*

The environmental Kuznets curve indicates an inverted-U relationship between pollution and economic development –see Dasgupta et al. (2001) or Brock and Taylor (2005) for an overview. The usual shape given to this relationship is of a polynomial type. Consider the simplest

$$p_t = \theta_0 + \theta_1 y_t + \theta_2 y_t^2,$$

where p_t is a measure of pollution and y_t is a measure of income, typically CO_2 and GDP, respectively. To check whether this equation is balanced will be troublesome if the order of integration is used. Even if it is known that y_t is $I(d_y)$, the order of integration of y_t^2 could not be well defined. As has been emphasised herein, the order of summability can help to overcome this pitfall. As it does not rely on any particular structure of the data generating process, it is suitable to be generally used. See the empirical application section.

Example 6 *: Predictive regressions*

Predictive regressions are linear specifications linking a noisy variable such as stock returns to past values of a very persistent regressor with the aim of assessing the presence of predictability. Most of the literature parameterises the regressor as a near-unit root process (Lewellen (2004), Campbell and Yogo (2006), Gonzalo and Pitarakis (2012), Lettau and Van Nieuwerburgh (2008)):

$$y_{t+1} = \gamma + \beta x_t + u_{t+1},$$

with $x_t = \rho_T x_{t-1} + v_t$, $\rho_T = 1 - c/T$ and $c > 0$. In the $I(d)$ framework, this type of regression is clearly asymptotically unbalanced (we are regressing an $I(0)$ on a near $I(1)$). Phillips and Lee (2012) propose to solve this problem by setting $\beta = \beta_T$ and letting it to go to zero asymptotically. Alternative balanced solutions can be found under our summability framework. One possible solution in a linear setup is to consider that y_t follows a stable distribution. The near unit root regressor is summable of order one, $\delta_x = 1$, while a stable *i.i.d.* process of parameter α is summable of order $(2-\alpha)/2\alpha$. Balancedness is achieved by an $\alpha = 2/3$. Another possibility consists of specifying a non-linear in variables predictive regression

$$y_{t+1} = \gamma + \beta f(x_t) + u_{t+1},$$

in which x_t is still a near-unit root process but $f(x_t) \sim S(\delta_f)$, $y_t \sim S(\delta_y)$, with $\delta_y = \delta_f$.

2.3 Co-summability

Definition 3 *: Two summable stochastic processes, $y_t \sim S(\delta_y)$ and $x_t \sim S(\delta_x)$, are said to be co-summable if there exists $f(x_t, \theta_f) \sim S(\delta_y)$ such that $u_t = y_t - f(x_t, \theta_f)$ is $S(\delta_u)$, with $\delta_u = \delta_y - \delta$ and $\delta > 0$. In short, $(y_t, x_t) \sim CS(\delta_y, \delta)$.*

Some aspects of this definition are worth noting. First, as stated, Definition 3 is concerned with a bivariate relationship. The extension to a vector of regressors is straightforward. Of particular interest for the subsequent analysis is the case in which $\delta_u = 0$ –*strong co-summability.*

Second, even when x_t is $S(\delta_x)$ with $\delta_x > 0$, some functions f can make $f(x_t, \theta) \sim S(0)$. As in co-integration theory, relations in which y_t and $f(x_t, \theta)$ are $S(0)$ will be excluded from the current co-summability analysis. Notwithstanding, the relevance of these relationships should be emphasised as they allow for the relation between persistent and non-persistent time series in the long run – such as growth rates and levels or returns and persistent macroeconomic variables in predictive regressions– although in a non-linear way. These relations deserve further research outside the present co-summability framework.

Third, a co-summable relationship is balanced. As already stated, balancedness is a necessary, although not sufficient, condition for a correct specification. In fact, when $\delta_y = \delta_g$, a postulated relationship $y_t = g(x_t, \theta)$ could be balanced spuriously. As in standard co-integration theory, spuriousness and co-summability can be distinguished through the fact that only under co-summability $\delta_u < \delta_y$, thus highlighting the existence of an attractor to the equilibrium relationship.

Finally, it is important to emphasise that co-summability mimics the idea of co-integration. This fact facilitates the development of an empirical strategy to test for co-summability that inherits the steps of testing for co-integration, although it uses new econometric tools.

Table 2 summarises the situations that can arise from the different configurations of orders of summability under both co-summability and no co-summability. Two types of unbalancedness are possible independent of whether a long run relationship exists: unbalancedness of type 1 (U1) if $\delta_y > \delta_g$ and unbalancedness of type 2 (U2) if $\delta_y < \delta_g$. A postulated model could be balanced spuriously, both when there is and when there is not co-summability; this is the spurious (S) case. Finally, under co-summability a postulated hypothesis could be correctly specified, such that $g(x_t, \theta) - f(x_t, \theta) = 0$, or misspecified in an admissible sense, such that $f(x_t, \theta) - g(x_t, \theta) \sim S(0)$, case (C).

Table 2: Balancedness and Co-summability

	No Co-summability		**Co-summability**	
Unbalancedness	**U1**	$\delta_y > \delta_g$	**U1**	$\delta_y > \delta_g$
	U2	$\delta_y < \delta_g$	**U2**	$\delta_y < \delta_g$
Balancedness	**S**	$\delta_y = \delta_g$	**S**	$\delta_y = \delta_g$ but $f(x_t, \theta) - g(x_t, \theta) \sim S(\delta_y)$
			C	$\delta_y = \delta_g$ with $f(x_t, \theta) - g(x_t, \theta) \sim S(0)$

$U1$: Unbalancedness of type 1; $U2$: Unbalancedness of type 2; S: Spuriousness; and C: Co-summability.

3 Estimation and Inference

3.1 The model

The co-summable relationship to be analysed in this section is the one described by the following model, linear in parameters but possibly non-linear in variables,

$$y_t = \theta_0 f(x_t) + u_t, \tag{9}$$

where $f : \mathbb{R} \to \mathbb{R}$, θ_0 is unknown, $f(x_t) \sim S(\delta_f)$, $\delta_f > 0$ and $u_t \sim S(0)$ –*strong co-summability*. Relationship (9) can be considered to be an approximation to a more general co-summable relationship $y_t = f(x_t, \theta_0) + u_t$, which will always be better than the standard approximation considered in co-integration theory –linear in parameters and variables. Indeed, stopping at a linear approximation could unbalance the model if, for instance, a higher order polynomial were a better approximation. Moreover, model (9) is empirically very rich, and at the same time, it encloses its own particular theoretical features to be studied by itself. As mentioned in the Introduction, non-linear in parameters models require a different empirical, and therefore theoretical, strategy. Notice that in that case the order of summability of $f(x_t, \theta_0)$ could depend on θ_0. Because it is unknown, balancedness cannot be directly studied as it can be done in the setup considered in this section where $f(x_t, \theta_0) = \theta_0 f(x_t)$ and, hence, the order of summability of $f(x_t, \theta_0)$ does not depend on θ_0. To facilitate the exposition, only the bivariate case (y_t, x_t) will be considered but the extension to a multivariate x_t or to additively separable multiple regression models can be easily adapted.

Because f is unknown, consider that the following least squares regression is carried out

$$y_t = \hat{\theta}_n g(x_t) + \hat{e}_t, \tag{10}$$

where $g : \mathbb{R} \to \mathbb{R}$, x_t and y_t are known by the researcher, and $\hat{\theta}_n$ is the parameter estimate.

Following exactly the same logic of co-integration theory, a two steps empirical strategy is devised. Consider equation (10) and let $y_t \sim S(\delta_y)$, $g(x_t) \sim S(\delta_g)$ and $\hat{e}_t \sim S(\delta_{\hat{e}})$.

Step 1. **Balancedness:** Test $H_o : \delta_y = \delta_g$. If it is not rejected, then go to Step 2.

Step 2. **Strong Co-summability**: Test $H_o : \delta_{\hat{e}} = 0$.

3.2 Testing for Balancedness

To establish balancedness in practice, we propose to start estimating δ_y and δ_g. To carry out this task, the order of summability estimator developed by BG is used. It is based on the convergence rate estimation procedure in McElroy and Politis (2007) and involves a simple least squares regression. The procedure requires the following assumption.

Assumption 1. $P(S_{yn} = 0) = P(S_{gn} = 0) = 0$ for all $n = 1, 2, 3, ...$

Consider the transformation

$$U_{yk} = \log S_{yk}^2 = \log \left[\left(\frac{1}{k^{\frac{1}{2}+\delta_y}} \sum_{t=1}^{k} y_t \right)^2 \right],$$

from which the following regression model can be derived

$$Y_{yk} = \beta_y \log k + U_{yk}, \tag{11}$$

where $Y_{yk} = \log\left[\left(\sum_{t=1}^{k} y_t\right)^2\right]$ and $\beta_y = 1 + 2\delta_y$. BG show that the OLS estimator of $\beta_y = 1 + 2\delta_y$ is $\log n$-consistent. For expository purposes, we include the formal statement.

Proposition 1 *: Let $\hat{\beta}_{yn}$ be the ordinary least squares estimator of β_y in (11). Under Assumption 1, if*

$$\frac{1}{n}\sum_{k=1}^{n} U_{yk} \Longrightarrow D_{yU} \quad \textit{and} \quad \frac{1}{n}\sum_{k=1}^{n} |U_{yk}|^p = O_p(1), \tag{12}$$

for some $1 < p < \infty$ and D_{yU} a random variable, then

$$\log n\left(\hat{\beta}_{yn} - \beta_y\right) \Longrightarrow D_{yU}.$$

Remark: As shown in McElory and Politis (2007) boundedness in probability of U_k suffices to get a consistent estimate of β_y. Nevertheless, to perform inferences on β_y, extra distributional assumptions, such as those in (12), need to be imposed. Notice that

$$\frac{1}{n}\sum_{k=1}^{n} U_{yk} = \frac{1}{n}\sum_{k=1}^{n} \log S_{yk}^2 = -\frac{(1+2\delta)}{n}\sum_{k=1}^{n}\log\left(\frac{k}{n}\right) + \frac{1}{n}\sum_{k=1}^{n}\log\left[\left(\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{k} y_t\right)^2\right].$$

Hence, for the case when y_t is $i.i.d.(0,1)$, following Pötscher (2004), de Jong (2004) or Berkes and Horváth (2006),

$$\frac{1}{n}\sum_{k=1}^{n} U_{yk} \Longrightarrow 1 + \int_0^1 \log\left(W^2(r)\right) dr \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n} |U_{yk}|^p = O_p(1).$$

Similarly, if y_t is a standard random walk, then from Berkes and Horváth (2006),

$$\frac{1}{n}\sum_{k=1}^{n} U_{yk} \Longrightarrow 3 + \int_0^1 \log\left(\left(\int_0^r W(r)dr\right)^2\right) dr \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n} |U_{yk}|^p = O_p(1).$$

Therefore, the asymptotic distribution of $\hat{\beta}_{yn}$ is not invariant to the data generating process of y_t. In BG, it is shown –through simulations– that subsampling confidence intervals can be constructed to undertake inferences on the true δ_y. It is important to remark that the presence of deterministic components in the data generating process biases the order of summability estimator, at least in finite samples. In BG, valid demeaning and detrending procedures are developed. Nevertheless, to facilitate the exposition, no deterministic components will be considered in this section.

Let the regression to estimate the order of summability of $g(x_t)$, that is,

$$Y_{gk} = \beta_g \log k + U_{gk}, \tag{13}$$

where $Y_{gk} = \log\left[\left(\sum_{t=1}^{k} g(x_t)\right)^2\right]$ and $\beta_g = 1 + 2\delta_g$.

To test for balancedness, an auxiliary equation that subtracts (13) from (11) will be used, that is,

$$Y_{yk} - Y_{gk} = (\beta_y - \beta_g) \log k + U_{yk} - U_{gk}.$$

Let $Y_k = Y_{yk} - Y_{gk}$, $\beta = \beta_y - \beta_g$, and $U_k = U_{yk} - U_{gk}$. Then, testing $H_o : \delta_y = \delta_g$ is equivalent to testing $H_o : \beta = 0$ in

$$Y_k = \beta \log k + U_k. \tag{14}$$

Proposition 2 *: Let $\hat{\beta}_n$ be the ordinary least squares estimator of β in (14). Under Assumption 1, if*

$$\frac{1}{n}\sum_{k=1}^{n} U_k \Longrightarrow D_U \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n} |U_k|^p = O_p(1),$$

for some $1 < p < \infty$ and D_U a random variable, then

$$\log n \left(\hat{\beta}_n - \beta\right) \Longrightarrow D_U.$$

Remark: Proposition 2 shows that $\hat{\beta}_n$ is a consistent estimator of the difference $\beta_y - \beta_g$. In particular, under balancedness $\hat{\beta}_n \xrightarrow{p} 0$. Nevertheless, as before, the asymptotic distribution cannot be tabulated in general. As in BG, we propose to use subsampling confidence intervals to undertake inferences. Next, their finite sample performance is analysed via Monte Carlo experiments.

3.2.1 Finite Sample Performance

Let $x_{yt} = x_{y,t-1} + \varepsilon_{yt}$ with $\varepsilon_{yt} \sim i.i.d.N(0,1)$ and $x_{y0} = 0$. $x_{gt} = x_{g,t-1} + \varepsilon_{gt}$ with $\varepsilon_{gt} \sim i.i.d.N(0,1)$ and $x_{g0} = 0$. In addition, let $u_t \sim i.i.d.N(0,1)$ and $v_t \sim i.i.d.N(0,1)$. ε_{yt}, ε_{gt}, u_t, and v_t are independent of each other. We consider the set of data generating processes –DGPs– collected in Table 3.

In all cases, $\hat{\beta}_n$ is calculated. Then, a subsampling confidence interval is computed and the null hypothesis of balancedness, $H_o : \beta = 0 \equiv \delta_y - \delta_g = 0$, is tested. Performance is measured by the coverage probability of two-sided nominal 95% symmetric intervals. Hence, size and power are measured as one minus the coverage probability that zero belongs to the corresponding subsampling confidence interval. The experiment is based on 1000 replicas and three different sample sizes, $n = \{100, 500, 1000\}$. A subsample size $b = \sqrt{n}$ has been chosen. The results are displayed in Table 4.

On the one hand, under the null hypothesis –cases S and C– the test is slightly undersized, leading to an under rejection of the null hypothesis. The implication is a high probability to jump to Step 2 –testing for co-summability– in the proposed empirical strategy. Along the lines of Andrews and Guggenberger (2009), a size-correction procedure could be used to account for these observed size distortions. Nevertheless, in this case, the size corrections could be more involved given the general nature of the problem being treated.

On the other hand, under the alternative hypothesis –cases U1 and U2–, for a given sample size, results show that the greater the difference $\delta_y - \delta_g$ in absolute value, the higher the power of the test. Furthermore, under the alternative hypothesis, for a given DGP, the greater the sample size the higher the power of the test. In other words, by consistency of the test, power increases as we move far away from the null hypothesis and the sample size grows.

Overall, the performance of the test is adequate given its generality and the agnostic assumptions upon which it is built.

Table 3: DGPs: Data Generating Processes

Under No Co-summability					
*	y_t	$g(x_t)$	*	y_t	$g(x_t)$
S	$v_{yt}x_{yt}$	$v_{gt}x_{gt}$	U1	x_{yt}	$v_{gt}x_{gt}$
S	x_{yt}	x_{gt}	U1	x_{yt}^2	$v_{gt}x_{gt}$
S	$1(v_{yt} \leq 0)\, x_{yt}$	$1(v_{gt} \leq 0)\, x_{gt}$	U1	$\sum_{j=1}^{t} x_{yj}$	$v_{gt}x_{gt}$
S	x_{yt}^2	x_{gt}^2	U2	x_{yt}	x_{gt}^2
S	$\sum_{j=1}^{t} x_{yj}$	$\sum_{j=1}^{t} x_{gj}$	U2	x_{yt}	x_{gt}^3
S	$\sum_{j=1}^{t} x_{yj}$	x_{gt}^3	U2	x_{yt}	$\left(\sum_{j=1}^{t} x_{gj}\right)^2$
Under Co-summability					
*	y_t	$g(x_t)$	*	y_t	$g(x_t)$
C	$\ln(\lvert x_{gt}\rvert) + u_t$	$\ln(\lvert x_{gt}\rvert)$	U1	$x_{gt}+u_t$	$v_{gt}x_{gt}$
C	$v_{gt}x_{gt}+u_t$	$v_{gt}x_{gt}$	U1	$x_{gt}^2+u_t$	$v_{gt}x_{gt}$
C	$x_{gt} + v_{gt}+u_t$	x_{gt}	S	$x_{gt}+u_t$	$1(v_{gt} \leq 0)\, x_{gt}$
C	$1(v_{gt} \leq 0)\, x_{gt}+u_t$	$1(v_{gt} \leq 0)\, x_{gt}$	S	$x_{g1t}x_{g2t}+u_t$	x_{g1t}^2
C	$x_{gt}^2+u_t$	x_{gt}^2	U2	$x_{gt}+u_t$	x_{gt}^2
C	$\sum_{j=1}^{t} x_{gj}+u_t$	$\sum_{j=1}^{t} x_{gj}$	U2	$x_{gt}+u_t$	x_{gt}^3

S, C, $U1$, and $U2$ denote spuriousness, co-summability, unbalancedness of type 1, and unbalancedness of type 2, respectively –see Table 2. $x_{yt} = x_{y,t-1} + \varepsilon_{yt}$ with $\varepsilon_{yt} \sim i.i.d.N(0,1)$ and $x_{y0} = 0$. $x_{gt} = x_{g,t-1} + \varepsilon_{gt}$ with $\varepsilon_{gt} \sim i.i.d.N(0,1)$ and $x_{g0} = 0$. In addition, $u_t \sim i.i.d.N(0,1)$ and $v_t \sim i.i.d.N(0,1)$. ε_{yt}, ε_{gt}, u_t, and v_t are independent of each other. x_{g1t} and x_{g2t} are defined as x_{gt} and are independent of each other.

Table 4: Testing for Balancedness: Size and Power

Under No Co-summability											
H_o: $\delta_y = \delta_g$			n			H_o: $\delta_y = \delta_g$			n		
*	δ_y	δ_g	100	500	1000	*	δ_y	δ_g	100	500	1000
S	1/2	1/2	0.013	0.006	0.004	U1	1	1/2	0.315	0.439	0.545
S	1	1	0.056	0.028	0.033	U1	3/2	1/2	0.616	0.838	0.914
S	1	1	0.021	0.006	0.003	U1	2	1/2	0.712	0.861	0.938
S	3/2	3/2	0.004	0.001	0.000	U2	1	3/2	0.174	0.189	0.189
S	2	2	0.046	0.037	0.025	U2	1	2	0.276	0.361	0.401
S	2	2	0.053	0.034	0.032	U2	1	5/2	0.627	0.811	0.893
Under Co-summability											
H_o: $\delta_y = \delta_g$			n			H_o: $\delta_y = \delta_g$			n		
*	δ_y	δ_g	100	500	1000	*	δ_y	δ_g	100	500	1000
C	1/2	1/2	0.000	0.000	0.000	U1	1	1/2	0.550	0.821	0.896
C	1/2	1/2	0.003	0.000	0.000	U1	3/2	1/2	0.882	0.998	0.998
C	1	1	0.000	0.000	0.000	S	1	1	0.007	0.002	0.001
C	1	1	0.000	0.000	0.000	S	3/2	3/2	0.076	0.030	0.033
C	3/2	3/2	0.002	0.010	0.007	U2	1	3/2	0.838	0.950	0.964
C	2	2	0.007	0.052	0.081	U2	1	2	0.743	0.900	0.918

S, C, $U1$, and $U2$ denote spuriousness, co-summability, unbalancedness of type 1, and unbalancedness of type 2, respectively –see Table 2. Hence, S and C represent size while $U1$ and $U2$ correspond to power. See Table 3 for specific details about the DGPs. Performance is measured from coverage probability of two-sided nominal 95% symmetric intervals.

3.3 Asymptotic Properties of $\hat{\theta}_n$

The test for strong co-summability to be proposed in section 3.4. is a residual based test. Hence, an estimate of the error term is needed. To this end, consider the OLS estimator

$$\hat{\theta}_n = \frac{\sum_{t=1}^{n} g(x_t)\, y_t}{\sum_{t=1}^{n} g^2(x_t)}.$$

3.3.1 Under No Co-summability

To study the asymptotic properties of $\hat{\theta}_n$ under no co-summability the following assumption will be made.

Assumption NC *(No Co-summability): Let y_t be independent of x_t and $(y_{nt}, g_{nt}) = (y_t/n^{\alpha_y}, g(x_t)/n^{\alpha_g})$. The D-space analog of (y_{nt}, g_{nt}) satisfies*

$$(y_n(r), g_n(r)) = \left(\frac{y_{[nr]}}{n^{\alpha_y}}, \frac{g\left(x_{[nr]}\right)}{n^{\alpha_g}} \right) \Longrightarrow (D_y(r), D_g(r)).$$

Assumption NC, for no co-summability, is similar to the assumptions in Granger and Newbold (1974) and Phillips (1986), where linear spurious regressions were analysed. Given that α_y can differ from α_g unbalanced regressions can be studied in the present framework. The relationship between Assumption NC and the order of summability of y_t and $g(x_t)$ follows directly from the continuous mapping theorem –CMT–, that is,

$$\left(\frac{1}{n}\sum_{t=1}^{n}\frac{y_t}{n^{\alpha_y}}, \frac{1}{n}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\right) = \left(\int_0^1 y_n(r)\,dr, \int_0^1 g_n(r)\,dr\right) \Longrightarrow \left(\int_0^1 D_y(r)\,dr, \int_0^1 D_g(r)\,dr\right),$$

which implies $\delta_y = 1/2 + \alpha_y$ and $\delta_g = 1/2 + \alpha_g$.

Proposition 3 *: Under Assumption NC,*

$$n^{(\delta_g - \delta_y)}\hat{\theta}_n \Longrightarrow \frac{\int_0^1 D_y(r)\,D_g(r)\,dr}{\int_0^1 D_g^2(r)\,dr}.$$

Remark: Under unbalancedness of type 1, $\delta_g - \delta_y < 0$. Hence, $\hat{\theta}_n$ diverges. In a spurious relationship, $\delta_g - \delta_y = 0$. Therefore, $\hat{\theta}_n$, without any rescaling, converges to a random variable. Finally, under unbalancedness of type 2, $\delta_g - \delta_y > 0$. Hence, $\hat{\theta}_n$ converges to zero.

3.3.2 Under Co-summability

Let

$$y_t = \theta_0 f(x_t) + u_t. \tag{15}$$

Define

$$v_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t,$$

and $\mathcal{F}_{nt}$ to be the natural filtration of (u_t, x_{t+1}). Finally, denote $E(X|\mathcal{F}_i)$ by E_iX.

Assumption SC *(Strong Co-summability):*

(a)

$$\sup_{t\le n}\|u_t\|_2 < \infty.$$

(b)

$$\sup_{t\le n}\sum_{k=1}^{\infty}\|(E_t u_{t+k} - E_{t-1}u_{t-1+k})\|_2 < \infty \quad \text{and} \quad \sup_{t\le n}\sum_{k=1}^{\infty}|E_t u_{t+k}| < \infty.$$

(c) For some Λ with $Var(\Lambda) \ge 0$,

$$\Lambda_n^* = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(g_{nt} - g_{n,t-1})\sum_{k=1}^{\infty}E_{t-1}u_{t-1+k} \Longrightarrow \Lambda.$$

(d) The D-space analog of (f_{nt}, g_{nt}, v_{nt}) satisfies,

$$\left(f_n\left(r\right), g_n\left(r\right), v_n\left(r\right)\right) = \left(\frac{f\left(x_{[nr]}\right)}{n^{\alpha_f}}, \frac{g\left(x_{[nr]}\right)}{n^{\alpha_g}}, \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t\right) \Longrightarrow \left(D_f\left(r\right), D_g\left(r\right), D_u\left(r\right)\right).$$

Assumption SC describes the stochastic structure of the processes involved in the long run relationship (15). Specifically, conditions (a) and (b) limit the heterogeneity and dependence, respectively, of the error term –the so called Gordin conditions (Gordin, 1969). On the other hand, condition (c) limits the dependence among the regressor and the error term. The limiting Λ describes their long run dependence, which, in this non-linear framework, can be stochastic $-Var\left(\Lambda\right) \geq 0$. Finally, condition (d) is the non-linear counterpart of the usual assumption typically imposed to analyse linear models with integrated time series. Notice that most of the asymptotically homogeneous functions studied in Park and Phillips (1999, 2001) satisfy this condition. Nevertheless, as stated, condition (d) does not require to work under the random walk hypothesis.

Proposition 4 *: Under Assumption SC, if $g\left(x_t\right) = f\left(x_t\right)$ a.s., then*

$$n^{\delta_g}\left(\hat{\theta}_n - \theta_0\right) \Longrightarrow \frac{\int_0^1 D_g\left(r\right) dD_u\left(r\right)}{\int_0^1 D_g^2\left(r\right) dr} + \Lambda.$$

Remark: Because co-summability mimics co-integration theory, it is not surprising that $\hat{\theta}_n$ is consistent, and its rate of convergence depends on δ_g. The asymptotic distribution has been derived following Hansen (1992). It resembles the asymptotic distribution of the OLS estimator of the co-integrating parameter in a linear model. Nevertheless, the non-linearity in variables, as well as Assumption SC, generates a more general asymptotic distribution. The limiting integrals coincide with those in Theorem 3.3. of Park and Phillips (2001) when their theoretical framework is used.

Assumption AM *(Admissible Misspecifications):* $\left(f\left(x_t\right) - g\left(x_t\right)\right) = z_t$ is a martingale difference sequence with $E\left(z_t^2 | \mathcal{F}_{n,t-1}\right) = \sigma_z^2$ *a.s.* for all $t = 1, ..., n$, and $\sup_{1 \leq t \leq n} E\left(\left|z_t\right|^q \left|\mathcal{F}_{n,t-1}\right.\right) < \infty$ *a.s.* for some $q > 2$. Moreover,

$$\left(f_n\left(r\right), g_n\left(r\right), z_n\left(r\right), v_n\left(r\right)\right) \Longrightarrow \left(D_f\left(r\right), D_g\left(r\right), D_z\left(r\right), D_u\left(r\right)\right),$$

where

$$z_n\left(r\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} z_t.$$

Assumption AM gives conditions that ensure that the OLS estimator approaches θ_0 even when the functional form is incorrect or when relevant variables are measured with error or omitted entirely. In this sense, this assumption shares the same spirit as Assumption 3 in White (1981) where, in an i.i.d. setup, conditions under which the least squares estimator has desirable properties as an approximation are provided. In summability terms, notice that under Assumption AM, $f(x_t)$ and $g(x_t)$ are strongly co-summable, i.e., $z_t \sim S(0)$.

Proposition 5 *: Under Assumptions SC and AM,*

$$n^{\delta_g}\left(\hat{\theta}_n - \theta_0\right) \Longrightarrow \theta_0 \frac{\int_0^1 D_g(r)\, dD_z(r)}{\int_0^1 D_g^2(r)\, dr} + \frac{\int_0^1 D_g(r)\, dD_u(r)}{\int_0^1 D_g^2(r)\, dr} + \Lambda.$$

Remark: $\hat{\theta}_n$ is consistent under admissible misspecifications and the rate of convergence depends on δ_g, as in the correct specification case. It seems worth mentioning that the implications of Proposition 5 do not change under a more general characterization of z_t as an $S(0)$ process, for instance a linear dependent process.

Assumption IM *(Inadmissible Misspecifications): Let* $\alpha_m = \max\{\alpha_f, \alpha_g\}$ *and* $z_{nt} = (f(x_t) - g(x_t))/n^{\alpha_m}$ *such that*

$$(f_n(r), g_n(r), z_n(r), v_n(r)) \Longrightarrow (D_f(r), D_g(r), D_z(r), D_u(r)),$$

with

$$z_n(r) = \frac{z_{[nr]}}{n^{\alpha_m}}.$$

Assumption IM considers cases in which the specified function $g(x_t)$ is so different from $f(x_t)$ that the OLS estimator $\hat{\theta}_n$ does not consistently estimate the unknown parameter θ_0. Notice that under Assumption IM, by the CMT, the difference $(f(x_t) - g(x_t)) \sim S(\delta_m)$ where $\delta_m = \max\{\delta_f, \delta_g\}$, such that $f(x_t)$ and $g(x_t)$ are not co-summable.

Proposition 6 *: Under Assumptions SC and IM,*

(i) If $\alpha_g \geq \alpha_f$, then $\hat{\theta}_n$ converges.

(ii) If $\alpha_g < \alpha_f$, then $n^{\delta_g - \delta_f}\left(\hat{\theta}_n - \theta_0\right)$ converges.

Remark: When a model is inadmissibly misspecified and $\alpha_g < \alpha_f$, then $\hat{\theta}_n$ diverges while if $\alpha_g \geq \alpha_f$, then $\hat{\theta}_n$ converges without any rescaling.

3.4 Testing for Strong Co-summability

The asymptotic properties of the OLS estimator derived in the previous section allow us to use OLS residuals to construct a residual-based test for co-summability. The following proposition formalises this fact.

Proposition 7 *: Let $\hat{e}_t$ be the OLS residuals in $y_t = \hat{\theta}_n g(x_t) + \hat{e}_t$.*

(a) Under Assumption NC,

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} \hat{e}_t = O_p(1).$$

(b) Under Assumptions SC and AM,

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} \hat{e}_t = O_p(1).$$

(c) Under Assumptions SC and IM,

(i) if $\alpha_g \geq \alpha_f$,

$$\frac{1}{n^{1/2+\delta_g}} \sum_{t=1}^{n} \hat{e}_t = O_p(1).$$

(ii) if $\alpha_g < \alpha_f$,

$$\frac{1}{n^{1/2+\delta_f}} \sum_{t=1}^{n} \hat{e}_t = O_p(1).$$

Remark: Notice that only under Assumptions SC and AM, $\hat{e}_t \sim S(0)$. In all the other situations, $\hat{e}_t \sim S(\delta_{\hat{e}})$ with $\delta_{\hat{e}} > 0$. Hence, given Proposition 7 a test for strong co-summability, $H_o : \delta_{\hat{e}} = 0$, can be easily constructed because under this null hypothesis the conditions of Proposition 1 are satisfied. The testing procedure can be implemented as follows. First, estimate the order of summability of the residuals. Second, compute the corresponding subsampling confidence interval and check whether zero belongs to this interval or not.

As before, the finite sample performance of the test will be studied via simulations. The considered data generating processes are those in Table 3. Again, performance has been measured by a coverage probability of two-sided nominal 95% symmetric intervals. Size and power are measured as one minus the coverage probability that zero belongs to the corresponding subsampling confidence interval. The experiment is based on 1000 replicas and three different sample sizes, $n = \{100, 500, 1000\}$. A subsample size, $b = \sqrt{n}$, has been chosen. The results are shown in Table 5.

Table 5: Testing for Strong Co-summability: Size and Power

Under No Co-summability											
H_o: $\delta_{\hat{e}}=0$			n			H_o: $\delta_{\hat{e}}=0$			n		
*	δ_y	δ_g	100	500	1000	*	δ_y	δ_g	100	500	1000
S	1/2	1/2	0.139	0.218	0.289	U1	1	1/2	0.835	0.993	0.998
S	1	1	0.634	0.954	0.993	U1	3/2	1/2	0.950	1.000	1.000
S	1	1	0.360	0.844	0.925	U1	2	1/2	0.992	1.000	1.000
S	3/2	3/2	0.853	0.998	1.000	U2	1	3/2	0.669	0.977	0.994
S	2	2	0.998	1.000	1.000	U2	1	2	0.740	0.975	0.995
S	2	2	0.997	1.000	1.000	U2	1	5/2	0.745	0.977	0.992
Under Co-summability											
H_o: $\delta_{\hat{e}}=0$			n			H_o: $\delta_{\hat{e}}=0$			n		
*	δ_y	δ_g	100	500	1000	*	δ_y	δ_g	100	500	1000
C	1/2	1/2	0.009	0.007	0.007	U1	1	1/2	0.745	0.981	0.998
C	1/2	1/2	0.013	0.011	0.015	U1	3/2	1/2	0.956	1.000	1.000
C	1	1	0.015	0.011	0.012	S	1	1	0.237	0.501	0.594
C	1	1	0.008	0.012	0.010	S	3/2	3/2	0.715	0.995	1.000
C	3/2	3/2	0.005	0.010	0.005	U2	1	3/2	0.356	0.842	0.960
C	2	2	0.007	0.011	0.007	U2	1	2	0.450	0.882	0.968

S, C, $U1$, and $U2$ denote spuriousness, co-summability, unbalancedness of type 1, and unbalancedness of type 2, respectively –see Table 2. See Table 3 for specific details about the DGPs. Hence, C represent size while S, $U1$ and $U2$ correspond to power. Performance is measured from coverage probability of two-sided nominal 95% symmetric intervals.

As can be evidenced, the testing procedure is undersized –case C– while power increases as we move away from the null hypothesis and the sample size grows –cases S, U1, and U2.

Remark: When a constant term is introduced in the proposed model, that is,

$$y_t = \hat{m}_n + \hat{\theta}_n g(x_t) + \hat{e}_t,$$

where $\hat{m}_n$ is the OLS estimator of a constant term, then the OLS residuals satisfy

$$\sum_{t=1}^{n} \hat{e}_t = \sum_{t=1}^{n} (y_t - \hat{m}_n - \hat{\theta}_n g(x_t)) = 0,$$

which implies that $\hat{e}_t$ cannot be used to infer $\delta_{\hat{e}}$. Partially demeaned residuals

$$\tilde{e}_t = \hat{e}_t - \frac{1}{t}\sum_{j=1}^{t} \hat{e}_j,$$

can be used instead in that case.

4 Error Correction Model Representation

Error Correction Models –ECMs– have a long tradition in econometrics. From the study of its own history, two main approaches can be distinguished: the LSE tradition, on the one hand, and the Engle-Granger-Johansen standpoint, on the other.

The LSE tradition –with Phillips (1954, 1957), Sargan (1964), Davidson, Hendry, Srba and Yeo (1978) or Nickell (1985)– conceived ECMs as models deriving from dynamic decision rules of economic agents. As noted by Alogoskoufis and Smith (1991), for the LSE tradition, ECMs are structural representations of dynamic adjustments towards an equilibrium about which economic theory can be informative. There is another particular feature of their approach, namely, the assumptions regarding exogeneity, under which single equation ECMs are mainly considered. The single equation modelling contrasts with the Engle-Granger-Johansen approach. This second tradition considers ECMs as statistical representations of co-integrated systems, which do not distinguish a priori between endogenous and exogenous variables. In any case, both approaches share the error correction principle, that is, the idea that a proportion of the disequilibrium in one period is corrected through changes in the variables of the system such that it tends to return to the equilibrium.

Given the nature of the model studied in this paper, we follow the LSE tradition as only single equation error correction models are considered in this section. A full treatment of the Engle-Granger-Johansen approach is beyond the scope of this paper and is under current investigation.

Consider an economic agent that tries to minimise the following quadratic loss function

$$Q = \frac{1}{2}\left(y_t - y_t^*\right)^2 + \frac{\vartheta}{2}\left(\Delta y_t\right)^2, \tag{16}$$

where ϑ is the ratio of the marginal cost of adjustment relative to the marginal cost of being away from equilibrium. As shown in Alogoskoufis and Smith (1991), the optimal solution of (16) can be written as

$$\Delta y_t = \lambda \Delta y_t^* - \lambda\left(y_{t-1} - y_{t-1}^*\right), \tag{17}$$

where $\lambda = (1+\vartheta)^{-1}$. If the target level depends in a non-linear manner upon an observable variable, x_t, for instance $y_t^* = \theta f(x_t)$, (17) becomes

$$\Delta y_t = \lambda\theta\Delta f(x_t) - \lambda\left(y_{t-1} - \theta f(x_{t-1})\right), \tag{18}$$

which can be understood as a non-linear partial adjustment model. As an example of the above minimisation problem, consider that the economic agent is a policy maker that tries to achieve a particular targeted level of pollution y_t^*, which may depend non-linearly on the level of GDP per capita x_t, as predicted by the environmental Kuznets curve.

By adding an $S(0)$ disturbance term ϵ_t to (18), that is,

$$\Delta y_t = \lambda\theta\Delta f(x_t) - \lambda(y_{t-1} - \theta f(x_{t-1})) + \epsilon_t,$$

the ECM representation is obtained in which the term $\Delta f(x_t)$, in contrast to the linear co-integration case where it can be ignored without serious harm, becomes a key element to obtain a balanced representation, $\delta_{\Delta y} = \delta_{\Delta f}$. As emphasised in section 2, the order of integration of $\Delta f(x_t)$, and hence of Δy_t, could not be well defined. Therefore, summability and co-summability become essential concepts to study ECMs associated with non-linear equilibrium relationships.

The econometric analysis and statistical treatment of the ECM representation of nonlinear equilibrium relationships require a more general setup that the one intended in this paper. Nevertheless, the derivation of the ECM in this section allows to catch sight of the interesting but open question about the existence of a Granger Representation Theorem for nonlinear equilibrium relationships; a subject that has been rarely studied in the literature. As it can be seen from the above discussion, ECM representations of nonlinear equilibrium relations will present distinctive features that will require additional econometric tools.

5 Empirical Application

5.1 Asymmetric preferences of central bankers

There seems to be nowadays certain consensus about the superiority of rules versus discretion in the practice of monetary policy. As noted by Taylor (1993), the advantage of rules over discretion is like the advantage of a cooperative over a non-cooperative solution in game theory. Optimal rules have been traditionally derived in a linear-quadratic framework in which policy makers have a quadratic objective function and operate in an economy that is described by a linear dynamic system –see, for instance, Svensson (1997). Linear Taylor rules are obtained in this framework when interest rates are taken to be the policy instrument, implying that central banks adjust interest rates according to output and inflation deviations from their targets. A traditional representative Taylor rule looks like

$$i_t = \theta_0 + \theta_1\tilde{\pi}_t + \theta_3\tilde{y}_t, \tag{19}$$

where i_t denotes nominal interest rates and $\tilde{\pi}_t$ and $\tilde{y}_t$ are deviations of inflation and output from their targets, respectively. Using equation (19), or some slightly modified version of it, several authors have tried to quantify the parameters that define the practice of monetary policy in different countries –see, for instance, Clarida, Galí and Gertler (1998, 2000).

It is somehow surprising that little attention has been paid to the fact that the variables involved in the Taylor rule are known to be highly persistent, something that should be taken into account when long time periods are analysed. There are, however, several works that address this issue, for instance, Siklos and Wohar (2005), Österholm (2005), and Christensen and Nielsen (2009). The fact that traditional Taylor rules do not appear to be congruent with the data once persistence is taken into consideration –usually through integration and co-integration theory– seems to be a common feature of these studies. This conclusion points to the possibility of an incorrect specification of the traditional Taylor rule.

On the other hand, although consistent with this conclusion, a stream of the literature has emphasised the hypothesis of asymmetric preferences of central bankers, which is often translated into non-linear Taylor rules. Next, the two cases described in Example 4 will be considered. Recall that, Clarida and Gertler (1997) consider a threshold type of Taylor rule in which the reaction of the monetary authority is different when inflation or output deviates from above, rather than from below, the target. Specifically,

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t^{(k)} 1\left(v_t > 0\right) + \theta_2 \tilde{\pi}_t^{(k)} 1\left(v_t \leq 0\right) + \theta_3 \tilde{y}_t 1\left(v_t > 0\right) + \theta_4 \tilde{y}_t 1\left(v_t \leq 0\right), \tag{20}$$

where $\tilde{\pi}_t^{(k)}$ are deviations of the rate of inflation between periods t and $t-k$, and v_t can be either $\tilde{\pi}_t^{(k)}$ or $\tilde{y}_t$. Alternatively, Dolado, María-Dolores and Naveira (2005) derive a non-linear optimal rule when non-linearities in the Phillips curve are allowed. The main prediction of this model is that the optimal rule should contain the interaction between inflation and output gaps, that is,

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t^{(k)} + \theta_2 \tilde{y}_t + \theta_3 \tilde{\pi}_t^{(k)} \tilde{y}_t. \tag{21}$$

Note that if i_t, $\tilde{\pi}_t^{(k)}$, or $\tilde{y}_t$ are highly persistent, the non-linear nature of these two specifications invalidate the use of standard co-integration theory to analyse the relevance of these models. Nevertheless, co-summability can be used given its generality when allowing for persistence and non-linearities at the same time. Moreover, the linearity in parameters of both equations makes suitable the application of the tools to test for co-summability developed in section 3.

To this end, we use US monthly time series covering the period 1954:07-2013:03, which are obtained from the Federal Reserve Bank of St. Louis. Specifically, we use (i) federal funds rate as interest rates, (ii) annual ($t/t-12$ basis; $k=12$) percentage rate in the CPI for inflation, (iii) (logged) industrial production index for output. Following the usual practice in the literature, to measure the output gap, we detrend (logged) industrial production using the HP filter with a coefficient of 14.800. For the inflation target, we use a fixed 2% level. Figure 1 shows the temporal evolution of these three measures –i_t, $\tilde{\pi}_t^{(k)}$, and $\tilde{y}_t$.

Figure 1: Optimal Rules of Monetary Policy

Variables: Taylor Rule

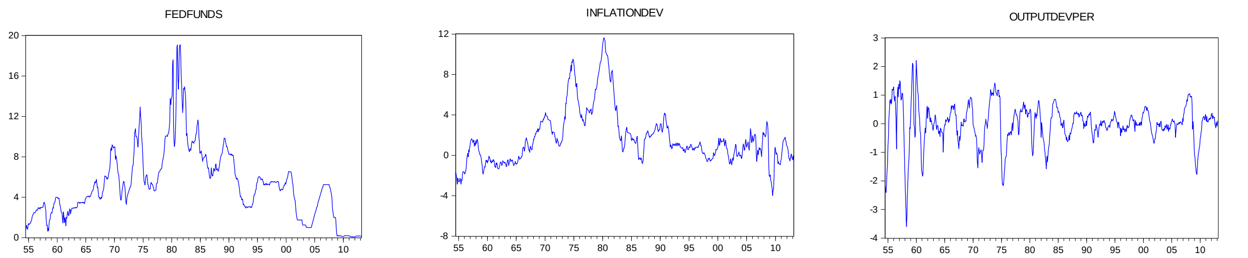


Table 6 reports the estimated orders of summability of all the variables contained in equations (20) and (21) as well as their corresponding subsampling confidence interval. All the variables have been partially demeaned to compute their orders of summability. Moreover, to control for a possible constant term in regression model (11) the first observation is subtracted –see BG for these technical details.

Table 6: Order of Summability: Estimation and Inference

Variables	$\hat{\delta}$	I_L	I_U
i_t	0.813	0.419	1.207
$\tilde{\pi}_t^{(k)}$	0.862	0.404	1.321
$\tilde{y}_t$	0.490	0.055	0.925
$\tilde{\pi}_t^{(k)}\tilde{y}_t$	0.198	-0.381	0.778
$\tilde{\pi}_t^{(k)}1\left(\tilde{\pi}_t^{(k)} > 0\right)$	0.814	0.459	1.169
$\tilde{\pi}_t^{(k)}1\left(\tilde{\pi}_t^{(k)} \leq 0\right)$	0.697	0.271	1.122
$\tilde{y}_t1\left(\tilde{\pi}_t^{(k)} > 0\right)$	0.155	-0.502	0.813
$\tilde{y}_t1\left(\tilde{\pi}_t^{(k)} \leq 0\right)$	0.398	-0.232	1.029
$\tilde{\pi}_t^{(k)}1\left(\tilde{y}_t > 0\right)$	0.805	0.301	1.309
$\tilde{\pi}_t^{(k)}1\left(\tilde{y}_t \leq 0\right)$	0.725	0.240	1.210
$\tilde{y}_t1\left(\tilde{y}_t > 0\right)$	0.496	0.129	0.862
$\tilde{y}_t1\left(\tilde{y}_t \leq 0\right)$	0.626	0.186	1.065

$\hat{\delta}$ denotes the estimated order of summability calculated from regression (11) after subtracting the first observation. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals. All the variables have been partially demeaned.

Results in Table 6 indicate that interest rates, i_t, and inflation gap, $\tilde{\pi}_t^{(k)}$, have a similar order of summability of approximately 0.8, while the estimated order of summability for the output gap, $\tilde{y}_t$, is approximately 0.5. It is worth emphasising that zero does not belong to any of the subsampling confidence intervals of these three time series. This confirms that persistence has to be properly addressed when using this database. With respect to the non-linear variables, different results are found. While the subsampling confidence intervals for $\tilde{\pi}_t^{(k)}\tilde{y}_t$, $\tilde{y}_t 1\left(\tilde{\pi}_t^{(k)} > 0\right)$, and $\tilde{y}_t 1\left(\tilde{\pi}_t^{(k)} \leq 0\right)$ do contain zero, all the other confidence intervals do not.

Following the steps of the proposed empirical strategy to test for co-summability, balancedness of (20) and (21) is next analysed. Given that these equations contain more than one regressor, the test for balancedness considers as explanatory variable the sum of all the regressors. Notice that, if they are not co-summable, then their sum must have an order of summability equal to the highest order of summability of the regressors. When the test is carried out separately for each regressor, the same conclusions are obtained. The corresponding results are collected in Table 7. As can be seen, the null hypothesis of balancedness, $H_o : \delta_y = \delta_g$, is not rejected in either case –zero belongs to the two subsampling intervals. Therefore, Step 2 in the proposed empirical strategy –testing for co-summability– is conducted.

Table 7: Testing for Balancedness

Balancedness	$\hat{\beta}_n = \hat{\delta}_y - \hat{\delta}_g$	I_L	I_U
threshold	0.010	-0.377	0.397
cross product	0.096	-0.337	0.531

$\hat{\delta}_y$ and $\hat{\delta}_g$ denote the estimated order of summability of the endogenous variable and the sum of the explanatory variables, respectively. The variables have been partially demeaned. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals.

Table 8 collects the parameter estimates of equations (20) and (21) jointly with the results of the test for co-summability associated with each regression. Some aspects are worth emphasising. First, the traditional Taylor rule does not specify a strong co-summable relationship –zero does not belong to the corresponding subsampling confidence interval. Second, focusing on the non-linear specifications, it can be seen that only a threshold type of Taylor rule in which the Federal Reserve reacts asymmetrically to output deviations is not rejected –zero belongs to the interval in this case. Finally, the difference between the parameters associated to $\tilde{y}_t 1\left(\tilde{y}_t > 0\right)$ and $\tilde{y}_t 1\left(\tilde{y}_t \leq 0\right)$ is remarkable. This fact clearly reflects a greater aversion to recessions than to expansions of the monetary authorities in the US.

Table 8: Testing for Co-summability

Taylor Rules	i_t	i_t	i_t	i_t
1	3.641	3.606	3.603	3.781
$\tilde{\pi}_t^{(k)}$	0.957	0.959		
$\tilde{y}_t$	0.744	0.445		
$\tilde{\pi}_t^{(k)}\tilde{y}_t$		0.173		
$\tilde{\pi}_t^{(k)}1\left(\tilde{\pi}_t^{(k)} > 0\right)$			0.965	
$\tilde{\pi}_t^{(k)}1\left(\tilde{\pi}_t^{(k)} \leq 0\right)$			0.916	
$\tilde{y}_t1\left(\tilde{\pi}_t^{(k)} > 0\right)$			0.974	
$\tilde{y}_t1\left(\tilde{\pi}_t^{(k)} \leq 0\right)$			0.290	
$\tilde{\pi}_t^{(k)}1\left(\tilde{y}_t > 0\right)$				1.052
$\tilde{\pi}_t^{(k)}1\left(\tilde{y}_t \leq 0\right)$				0.820
$\tilde{y}_t1\left(\tilde{y}_t > 0\right)$				0.119
$\tilde{y}_t1\left(\tilde{y}_t \leq 0\right)$				0.807
$\hat{\delta}_{\hat{e}}$	0.428	0.471	0.437	0.403
I_L	0.036	0.087	0.011	-0.005
I_U	0.819	0.854	0.863	0.811

$\hat{\delta}_{\hat{e}}$ denotes the estimated order of summability of the residuals calculated from regression (11) after subtracting the first observation. Pseudo residuals have been partially demeaned. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals.

5.2 Environmental Kuznets Curve

The environmental Kuznets curve –EKC– suggests an inverted U-shaped relationship between pollution and economic development. The argument is as follows. Agents living in poor economies are more concerned with labor and income than with the environment; consequently, environmental regulation is limited at poorer stages. As economies gain in wealth, agents value more the environment, production becomes cleaner, and more efficient regulatory institutions are formed –see Dasgupta et al. (2001).

This hypothesis has been controversial, prompting conflicting views from researchers and policymakers. The literature –see Grossmann and Krueger (1995) or Brock and Taylor (2005)– identifies, mainly, three different channels linking pollution and economic activity: scale, composition, and technique effects. Ceteris paribus (i) emissions rise when the scale of economic activity, as measured by real GDP, increases; (ii) emissions fall through the composition effect when the goods produced in an economy become cleaner; (iii) emissions fall when the techniques of production are less contaminating. The EKC hypothesis depends on the relative relevance of these three effects. To identify them, a structural modelling should be carefully undertaken. Nevertheless, the empirical literature on the EKC has mainly used a reduced form approach. Typically, polynomial relationships between pollution and income have been considered, that is,

$$p_t = \theta_0 + \theta_1 y_t + \theta_2 y_t^2 + ... + \theta_k y_t^k, \tag{22}$$

where p_t is a measure of pollution and y_t is a measure of income. Several empirical issues arise in this setup. A first issue is concerned with the measures chosen for p_t and y_t. While GDP has been used as a measure of income many measures of pollutants have been used. Commonly used measures for p_t are CO_2, NO_x, and SO_2. Empirical evidence is mixed for different pollutants. A second issue relates to the curvature of the EKC. There seems not to be a clear agreement about the order of the polynomial to be used. Grossman and Krueger (1995) used a cubic specification, while Holtz-Eakin and Selden (1995) preferred the quadratic one. Other authors tend to compare both specifications in practice. A third empirical ambiguity arises as p_t and y_t are sometimes treated in levels (Grossman and Krueger, 1995), other times in logarithms (Hong and Wagner, 2008), and, still at other times, both cases are compared (Holtz-Eakin and Selden, 1995). Finally, it is surprising that only a few authors have taken into consideration persistence of the variables involved in the EKC. Some exceptions include Perman and Stern (2003), Hong and Wagner (2008) and Jalil and Mahmud (2009). When persistence is taken into consideration, the empirical evidence on the EKC is mixed.

As an illustration, we apply co-summability theory to disentangle some of the empirical features on the EKC. We use annual GDP and CO_2 emissions per capita in the US during the period 1870-2007. GDP and population are taken from Angus Maddison and CO_2 emissions from the Carbon Dioxide Information Analysis Centre. Figure 2 shows the evolution of GDP and CO_2 emissions per capita, both in levels –*co*2*pcus*, *gdppcus*– and logarithms –*lco*2*pcus*, *lgdppcus*.

Figure 2: Environmental Kuznets Curve Hypothesis

Variables: Environmental Kuznets Curve

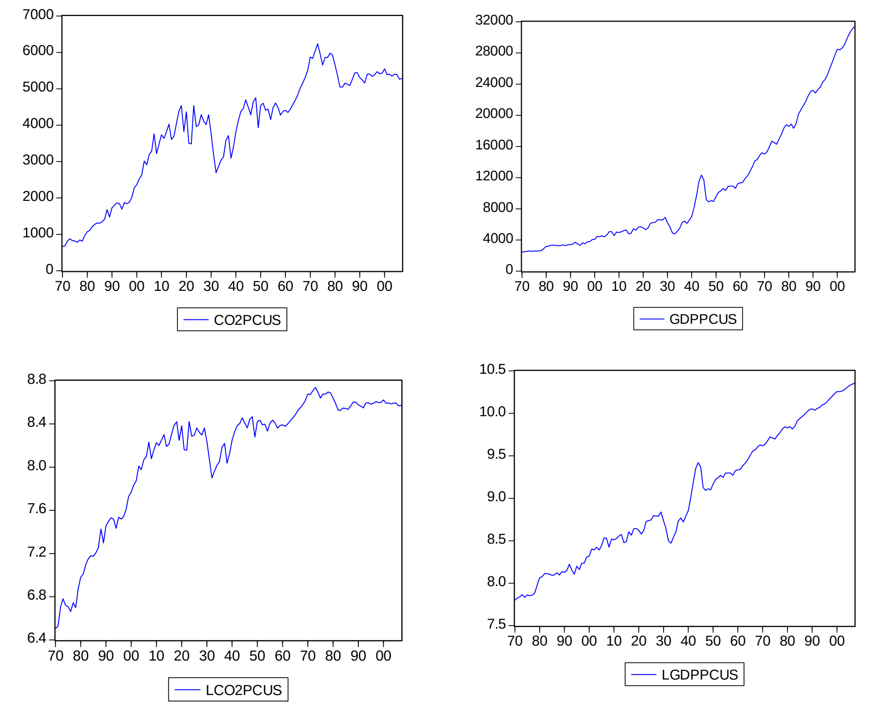


Table 9 reports the estimated orders of summability of all the variables contained in (22) for $k = 4$. The corresponding subsampling confidence intervals are provided as well. As expected, the order of summability of GDP per capita increases as successive powers are taken. In general, these results show that persistence must be taken into account –zero does not belong to any of the confidence intervals.

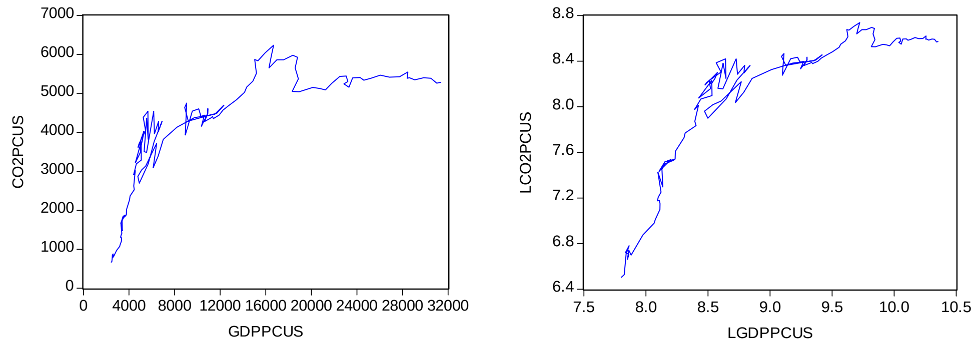
Table 9: Order of Summability: Estimation and Inference

Variables	$\hat{\delta}$	I_L	I_U
$co2pc$	0.893	0.286	1.500
$gdppc$	1.424	0.599	2.249
$gdppc^2$	1.779	0.795	2.763
$gdppc^3$	2.090	0.952	3.229
$gdppc^4$	2.391	1.082	3.699
$lco2pc$	0.705	0.160	1.250
$lgdppc$	0.876	0.195	1.557
$lgdppc^2$	0.950	0.255	1.645
$lgdppc^3$	1.017	0.270	1.764
$lgdppc^4$	1.112	0.260	1.963

$\hat{\delta}$ denotes the estimated order of summability calculated from regression (11) after subtracting the first observation. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals. All the variables have been partially detrended.

Figure 3 plots the relationship between GDP and CO_2 emissions per capita in levels and logarithms. Although it seems there is a diminishing marginal propensity to pollute, the postulated inverted U-shape should be more carefully and formally analysed.

Figure 3: Environmental Kuznets Curve Hypothesis



Results of testing for balancedness are reported in Table 10. Notice that when variables are in levels, balancedness is only achieved under the linear specification. However, when variables are in logarithms, balancedness is obtained under quadratic and cubic polynomials.

Table 10: Testing for Balancedness

	Balancedness	$\hat{\beta}_n = \hat{\delta}_y - \hat{\delta}_g$	I_L	I_U
$co2$				
	gdp	-0.104	-0.995	0.787
	gdp^2	-2.494	-4.809	-0.179
$lco2$				
	$lgdp$	0.530	0.054	1.007
	$lgdp^2$	-0.216	-0.979	0.545
	$lgdp^3$	-0.865	-1.957	0.225
	$lgdp^4$	-1.534	-2.944	-0.124

$\hat{\delta}_y$ and $\hat{\delta}_g$ denote the estimated order of summability of the endogenous variable and the sum of the explanatory variables, respectively. The variables have been partially detrended. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals.

Results to test for co-summability are collected in Tables 11 and 12 for levels and logarithms, respectively. From Table 11, it is clear that co-summability does not hold either for the linear or the quadratic specification. The latter result was expected given the balancedness test. Nevertheless, in Table 12, results are more optimistic. Co-summability is not rejected for the cubic specification, which is compatible with the shape observed in Figure 3. These results are invariant to the inclusion of a deterministic trend. Summarising, from the co-summability results, we recommend to use the logarithmic transformation and polynomials of third degree when empirically studying parametric reduced forms of the EKC in the US.

Table 11: Testing for Co-summability

EKC	$co2$	$co2$	$co2$	$co2$
1	2190.538	1090.780	470.265	749.636
t		54.820		39.204
gdp	0.149	-0.098	0.520	0.112
gdp^2			-1.249e-005	-4.752e-006
$\hat{\delta}_{\hat{e}}$	1.750	0.883	1.447	0.989
I_L	0.945	0.143	0.483	0.322
I_U	2.556	1.623	2.411	1.656

$\hat{\delta}_{\hat{e}}$ denotes the estimated order of summability of the residuals calculated from regression (11) after subtracting the first observation. Pseudo residuals have been partially demeaned. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals.

Table 12: Testing for Co-summability

EKC	$lco2$	$lco2$	$lco2$	$lco2$	$lco2$	$lco2$
1	2.290	10.023	-42.883	-41.941	-280.718	-290.421
t		0.019		0.001		-0.003
$lgdp$	0.646	-0.359	10.665	10.501	90.013	92.866
$lgdp^2$			-0.551	-0.546	-9.340	-9.623
$lgdp^3$					0.323	0.333
$\hat{\delta}_{\hat{e}}$	1.503	1.351	0.792	0.796	0.240	0.247
I_L	0.724	0.529	0.189	0.172	-0.342	-0.305
I_U	2.281	2.172	1.395	1.419	0.823	0.801

$\hat{\delta}_{\hat{e}}$ denotes the estimated order of summability of the residuals calculated from regression (11) after subtracting the first observation. Pseudo residuals have been partially demeaned. I_L and I_U denote the lower and upper bounds of the corresponding 95% subsampling confidence intervals.

6 Concluding Remarks

Co-integration theory is not designed to deal with situations in which non-linearities and persistence occur simultaneously. Accordingly, there is a clear need for theoretically valid and empirically useful concepts that generalise the concepts of integration and co-integration to non-linear environments.

The order of summability concept has made it possible to define non-linear long run relationships between persistent processes under exactly the same logic as that of co-integration theory. It has easily allowed for (i) the definition of balancedness of a postulated model –a necessary condition for a correct specification; and (ii) the definition of non-linear long run relationships by means of the concept of co-summability –a direct extension of co-integration valid for non-linear equilibria. These two pieces are relevant for both econometricians and economic theorists: for the former when specifying, estimating, and testing econometric models; for the latter when choosing functional forms to construct their theories.

Further research is going in two directions: (i) non-linear in parameters regression models and (ii) vector error correction models.

7 Appendix

Proof of Proposition 1: See Berenguer-Rico and Gonzalo (2013).

Proof of Proposition 2: The estimator of interest is

$$\left(\hat{\beta}_n - \beta\right) = \frac{\sum_{k=1}^{n} U_k \log k}{\sum_{k=1}^{n} \log^2 k},$$

or equivalently

$$\log n \left(\hat{\beta}_n - \beta\right) = \frac{\frac{1}{n \log n} \sum_{k=1}^{n} U_k \log k}{\frac{1}{n \log^2 n} \sum_{k=1}^{n} \log^2 k}.$$

The denominator satisfies

$$\frac{1}{n \log^2 n} \sum_{k=1}^{n} \log^2 k \longrightarrow 1.$$

The numerator can be written as

$$\begin{aligned}
\frac{1}{n \log n} \sum_{k=1}^{n} U_k \log k &= \frac{1}{n \log n} \sum_{k=1}^{n} U_k \log\left(\frac{k}{n} n\right) \\
&= \frac{1}{n \log n} \sum_{k=1}^{n} U_k \left(\log\left(\frac{k}{n}\right) + \log n\right) \\
&= \frac{1}{n} \sum_{k=1}^{n} U_k + \frac{1}{\log n} \left(\frac{1}{n} \sum_{k=1}^{n} U_k \log\left(\frac{k}{n}\right)\right).
\end{aligned}$$

Now, let q be such that $1/p + 1/q = 1$. By Hölder's inequality,

$$\frac{1}{n} \sum_{k=1}^{n} \left| U_k \log\left(\frac{k}{n}\right)\right| \leq \left(\frac{1}{n} \sum_{k=1}^{n} |U_k|^p\right)^{1/p} \left(\frac{1}{n} \sum_{k=1}^{n} \left|\log\left(\frac{k}{n}\right)\right|^q\right)^{1/q},$$

hence,

$$\frac{1}{n}\sum_{k=1}^{n}U_k\log\left(\frac{k}{n}\right)\leq\frac{1}{n}\sum_{k=1}^{n}\left|U_k\log\left(\frac{k}{n}\right)\right|\leq\left(\frac{1}{n}\sum_{k=1}^{n}|U_k|^p\right)^{1/p}\left(\frac{1}{n}\sum_{k=1}^{n}\left|\log\left(\frac{k}{n}\right)\right|^q\right)^{1/q}.$$

Therefore,

$$\frac{1}{n}\sum_{k=1}^{n}U_k\log\left(\frac{k}{n}\right)=O_p(1),$$

which implies that the numerator satisfies

$$\frac{1}{n\log n}\sum_{k=1}^{n}U_k\log k=\frac{1}{n}\sum_{k=1}^{n}U_k+o_p(1)\Longrightarrow D_U.$$

All together gives the stated result

$$\log n(\hat{\beta}-\beta)=\frac{\frac{1}{n\log n}\sum_{k=1}^{n}U_k\log k}{\frac{1}{n\log^2 n}\sum_{k=1}^{n}\log^2 k}\Longrightarrow D_U.$$

Q.E.D.

Proof of Proposition 3*:* The OLS estimator can be rewritten as

$$n^{(\delta_g-\delta_y)}\hat{\theta}_n=\frac{\frac{1}{n}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\frac{y_t}{n^{\alpha_y}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}=\frac{\int_0^1 g_n(r)\,y_n(r)\,dr}{\int_0^1 g_n^2(r)\,dr}.$$

Hence, under Assumption NC, by the CMT,

$$n^{(\delta_g-\delta_y)}\hat{\theta}_n\Longrightarrow\frac{\int_0^1 D_g(r)\,D_y(r)\,dr}{\int_0^1 D_g^2(r)\,dr}.$$

Q.E.D.

Proof of Proposition 4: Let

$$V_{nt}-V_{n,t-1}=\frac{u_t}{\sqrt{n}},$$

and write

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{f(x_t)}{n^{\alpha_f}}u_t=\int_0^1 f_n(r)\,dV_n(r).$$

Following Hansen (1992), define

$$\epsilon_t=\sum_{k=0}^{\infty}\left(E_t u_{t+k}-E_{t-1}u_{t+k}\right),\qquad z_t=\sum_{k=1}^{\infty}E_t u_{t+k},$$

such that

$$u_t=\epsilon_t+z_{t-1}-z_t,\qquad E_{t-1}\epsilon_t=0.$$

In this scenario, a martingale difference approximation can be used, that is,

$$\int_0^1 f_n(r)\, dV_n(r) = \int_0^1 f_n(r)\, dY_n(r) + \Lambda_n^*,$$

with $Y_{nt} = Y_t/\sqrt{n}$, $Y_t = \sum_i^t \epsilon_i$, and

$$\Lambda_n^* = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (f_{nt} - f_{n,t-1})\, z_{t-1} - \frac{1}{\sqrt{n}} f_{nn} z_n.$$

Let $\epsilon_{nt} = \epsilon_t/\sqrt{n}$. To apply Theorem 3.1. in Hansen (1992), that is,

$$\int_0^1 f_n(r)\, dY_n(r) \Longrightarrow \int_0^1 D_f(r)\, dD_u(r),$$

it must be showed that:

(i)

$$\sum_{t=1}^{n} E\epsilon_{nt}^2 < \infty,$$

and (ii)

$$\sup_{t\le n} |Y_{nt} - V_{nt}| \xrightarrow{p} 0.$$

With respect (i), note that

$$\begin{aligned}
\sum_{t=1}^{n} E\epsilon_{nt}^2 \le \sup_{t\le n} E\epsilon_t^2 &= \left(\sup_{t\le n} \|u_t + z_t - z_{t-1}\|_2\right)^2 \\
&\le \left(\sup_{t\le n} \|u_t\|_2 + \sup_{t\le n} \|z_t - z_{t-1}\|_2\right)^2 \\
&= \left(\sup_{t\le n} \|u_t\|_2 + \sup_{t\le n} \left\| \sum_{k=1}^{\infty} (E_t u_{t+k} - E_{t-1} u_{t-1+k}) \right\|_2\right)^2 \\
&\le \left(\sup_{t\le n} \|u_t\|_2 + \sup_{t\le n} \sum_{k=1}^{\infty} \|(E_t u_{t+k} - E_{t-1} u_{t-1+k})\|_2\right)^2 < \infty,
\end{aligned}$$

by conditions (a) and (b) of Assumption SC.

With respect (ii), note that

$$\begin{aligned}
\sup_{t\le n} |Y_{nt} - V_{nt}| &\le 2\frac{1}{\sqrt{n}} \sup_{t\le n} |z_t| \\
&= 2\frac{1}{\sqrt{n}} \sup_{t\le n} \left| \sum_{k=1}^{\infty} E_t u_{t+k} \right| \xrightarrow{p} 0,
\end{aligned}$$

by condition (b).

It remains to analyse

$$\Lambda_n^* = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(f_{nt} - f_{n,t-1}\right) z_{t-1} - \frac{1}{\sqrt{n}} f_{nn} z_n.$$

The second component of the right hand side satisfies

$$\frac{1}{\sqrt{n}} f_{nn} z_n \leq \sup_{t\leq n} \frac{1}{\sqrt{n}} \left|f_{nt} z_t\right| \leq \sup_{t\leq n} \left|f_{nt}\right| \sup_{t\leq n} \frac{1}{\sqrt{n}} \left|z_t\right|.$$

By condition (d) of Assumption SC,

$$\sup_{t\leq n} \left|f_{nt}\right| = O_p\left(1\right),$$

and by condition (b)

$$\frac{1}{\sqrt{n}} \sup_{t\leq n} \left|z_t\right| \xrightarrow{p} 0,$$

implying

$$\frac{1}{\sqrt{n}} f_{nn} z_n \xrightarrow{p} 0.$$

Hence, by condition (c)

$$\Lambda_n^* \Longrightarrow \Lambda.$$

All together gives

$$n^{\delta_g}\left(\hat{\theta}_n - \theta_0\right) = \frac{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{g(x_t)}{n^{\alpha_g}} u_t}{\frac{1}{n} \sum_{t=1}^{n} \frac{g^2(x_t)}{n^{2\alpha_g}}} = \frac{\int_0^1 f_n\left(r\right) dY_n\left(r\right)}{\int_0^1 f_n^2\left(r\right) dr} + \Lambda_n^* \Longrightarrow \frac{\int_0^1 D_f\left(r\right) dD_u\left(r\right)}{\int_0^1 D_f^2\left(r\right) dr} + \Lambda,$$

as stated. **Q.E.D.**

Proof of Proposition 5: The OLS estimator in terms of u_t, $g\left(x_t\right)$, and $f\left(x_t\right)$,

$$\hat{\theta}_n = \theta_0 \frac{\sum_{t=1}^{n} g\left(x_t\right) f\left(x_t\right)}{\sum_{t=1}^{n} g^2\left(x_t\right)} + \frac{\sum_{t=1}^{n} g\left(x_t\right) u_t}{\sum_{t=1}^{n} g^2\left(x_t\right)},$$

can be rewritten as

$$\hat{\theta}_n = \theta_0 \frac{\sum_{t=1}^{n} g\left(x_t\right)\left(f\left(x_t\right) + g\left(x_t\right) - g\left(x_t\right)\right)}{\sum_{t=1}^{n} g^2\left(x_t\right)} + \frac{\sum_{t=1}^{n} g\left(x_t\right) u_t}{\sum_{t=1}^{n} g^2\left(x_t\right)} = \theta_0 + \theta_0 \frac{\sum_{t=1}^{n} g\left(x_t\right)\left(f\left(x_t\right) - g\left(x_t\right)\right)}{\sum_{t=1}^{n} g^2\left(x_t\right)} + \frac{\sum_{t=1}^{n} g\left(x_t\right) u_t}{\sum_{t=1}^{n} g^2\left(x_t\right)}.$$

Equivalently,

$$n^{\delta_g}\left(\hat{\theta}_n-\theta_0\right)=\theta_0\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\left(f\left(x_t\right)-g\left(x_t\right)\right)}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}+\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}u_t}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}. \tag{23}$$

By Assumption AM, convergence of the first summand of the right hand side of (23) follows from results in Kurtz and Protter (1991). Convergence of the second summand follows from Proposition 4. In particular,

$$n^{\delta_g}\left(\hat{\theta}_n-\theta_0\right)\Longrightarrow\theta_0\frac{\int_0^1 D_g\left(r\right)dD_z\left(r\right)}{\int_0^1 D_g^2\left(r\right)dr}+\frac{\int_0^1 D_g\left(r\right)dD_u\left(r\right)}{\int_0^1 D_g^2\left(r\right)dr}+\Lambda.$$

Q.E.D.

Proof of Proposition 6: The OLS estimator can be written as

$$n^{\alpha_g-\alpha_m}\left(\hat{\theta}_n-\theta_0\right)=\theta_0\frac{\frac{1}{n}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\frac{(f(x_t)-g(x_t))}{n^{\alpha_m}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}+\frac{\frac{1}{n^{1+\alpha_m}}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}u_t}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}.$$

By Proposition 4,

$$n^{\alpha_g-\alpha_m}\left(\hat{\theta}_n-\theta_0\right)=\theta_0\frac{\frac{1}{n}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\frac{(f(x_t)-g(x_t))}{n^{\alpha_m}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}+o_p\left(1\right).$$

Now,

(i) If $\alpha_g\geq\alpha_f$, then $\alpha_g-\alpha_m=0$ and

$$\left(\hat{\theta}_n-\theta_0\right)=\theta_0\frac{\frac{1}{n}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\frac{(f(x_t)-g(x_t))}{n^{\alpha_m}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}+o_p\left(1\right)\Longrightarrow\theta_0\frac{\int_0^1 D_g\left(r\right)D_z\left(r\right)\theta dr}{\int_0^1 D_g^2\left(r\right)dr}.$$

(ii) If $\alpha_g<\alpha_f$, then $\alpha_g-\alpha_m=\alpha_g-\alpha_f=\delta_g-\delta_f<0$ and $\hat{\theta}_n$ diverges since

$$n^{\delta_g-\delta_f}\left(\hat{\theta}_n-\theta_0\right)=\theta_0\frac{\frac{1}{n}\sum_{t=1}^{n}\frac{g(x_t)}{n^{\alpha_g}}\frac{(f(x_t)-g(x_t))}{n^{\alpha_f}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{g^2(x_t)}{n^{2\alpha_g}}}+o_p\left(1\right)\Longrightarrow\theta_0\frac{\int_0^1 D_g\left(r\right)D_z\left(r\right)dr}{\int_0^1 D_g^2\left(r\right)dr}.$$

Q.E.D.

Proof of Proposition 7:

(a) Under NC, $n^{\delta_g-\delta_y}\hat{\theta}_n$ converges, hence

$$\frac{1}{n^{1/2+\delta_y}}\sum_{t=1}^{n}\hat{e}_t = \frac{1}{n^{1/2+\delta_y}}\sum_{t=1}^{n}y_t - n^{\delta_g-\delta_y}\hat{\theta}_n\frac{1}{n^{1/2+\delta_g}}\sum_{t=1}^{n}g\left(x_t\right) = O_p\left(1\right).$$

(b) Under Assumptions SC and AM, $n^{\delta_g}\left(\hat{\theta}_n - \theta_0\right)$ converges, hence

$$\begin{aligned}
\frac{1}{n^{1/2}}\sum_{t=1}^{n}\hat{e}_t &= \frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(y_t - \hat{\theta}_n g\left(x_t\right)\right) \\
&= \frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(\theta_0 f\left(x_t\right) + u_t + \theta_0 g\left(x_t\right) - \theta_0 g\left(x_t\right) - \hat{\theta}_n g\left(x_t\right)\right) \\
&= \frac{1}{n^{1/2}}\sum_{t=1}^{n}u_t - n^{\delta_g}\left(\hat{\theta}_n - \theta_0\right)\frac{1}{n^{1/2+\delta_g}}\sum_{t=1}^{n}g\left(x_t\right) + \theta_0\frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(f\left(x_t\right) - g\left(x_t\right)\right) \\
&= O_p\left(1\right).
\end{aligned}$$

(d) Under Assumptions SC and IM

(i) If $\alpha_g \geq \alpha_f$*, then* $\hat{\theta}_n$ converges. Therefore,

$$\begin{aligned}
\frac{1}{n^{1/2+\delta g}}\sum_{t=1}^{n}\hat{e}_t &= \frac{1}{n^{1/2+\delta g}}\sum_{t=1}^{n}u_t + \left(\hat{\theta}_n - \theta_0\right)\frac{1}{n^{1/2+\delta_g}}\sum_{t=1}^{n}g\left(x_t\right) + \theta_0\frac{1}{n^{1/2+\delta g}}\sum_{t=1}^{n}\left(f\left(x_t\right) - g\left(x_t\right)\right) \\
&= \left(\hat{\theta}_n - \theta_0\right)\frac{1}{n^{1/2+\delta_g}}\sum_{t=1}^{n}g\left(x_t\right) + \theta_0\frac{1}{n^{1/2+\delta g}}\sum_{t=1}^{n}\left(f\left(x_t\right) - g\left(x_t\right)\right) + o_p\left(1\right) \\
&= O_p\left(1\right).
\end{aligned}$$

(ii) If $\alpha_g < \alpha_f$*, then* $n^{\delta_g-\delta_f}\left(\hat{\theta}_n - \theta_0\right)$ converges. Hence,

$$\begin{aligned}
\frac{1}{n^{1/2+\delta_f}}\sum_{t=1}^{n}\hat{u}_t &= \frac{1}{n^{1/2+\delta_f}}\sum_{t=1}^{n}u_t + n^{\delta_g}\left(\hat{\theta}_n - \theta_0\right)\frac{1}{n^{1/2+\delta_f+\delta_g}}\sum_{t=1}^{n}g\left(x_t\right) + \theta_0\frac{1}{n^{1/2+\delta_f}}\sum_{t=1}^{n}\left(f\left(x_t\right) - g\left(x_t\right)\right) \\
&= n^{\delta_g-\delta_f}\left(\hat{\theta}_n - \theta_0\right)\frac{1}{n^{1/2+\delta_g}}\sum_{t=1}^{n}g\left(x_t\right) + \theta_0\frac{1}{n^{1/2+\delta_f}}\sum_{t=1}^{n}\left(f\left(x_t\right) - g\left(x_t\right)\right) + o_p\left(1\right) \\
&= O_p\left(1\right).
\end{aligned}$$

Q.E.D.

References

Alogoskoufis, G., and R. Smith (1991): “On error correction models: Specification, interpretation, estimation,” *Journal of Economic Surveys*, 5, 97-128.

Andrews, D. W. K., and P. Guggenberger (2009): “Hybrid and size-corrected subsampling methods,” *Econometrica*, 77, 721-762.

Berenguer-Rico, V., and J. Gonzalo (2013): “Summability of stochastic processes: A generalization of integration for non-linear processes,” forthcoming *Journal of Econometrics*.

Berkes, I., and L. Horváth (2006): “Convergence of integral functionals of stochastic processes,” *Econometric Theory*, 22, 304-322.

Brock, W. A., and M. S. Taylor (2005): “Economic growth and the environment: A review of theory and empirics,” *Handbook of Economic Growth*, Volume 1, Part 2, 1749-1821.

Campbell, J. Y., and M. Yogo (2006): “Efficient tests of stock return predictability,” *Journal of Financial Economics*, 81, 27-60.

Choi, I., and P. Saikkonen (2010): “Tests for non-linear cointegration,” *Econometric Theory*, 26, 682-709.

Christensen, A. M., and H. B. Nielsen (2009): “Monetary Policy in the Greenspan Era: A Time Series Analysis of Rules vs. Discretion,” *Oxford Bulletin of Economics and Statistics*, 71, 69-89.

Clarida, R., and M. Gertler (1997): “How the Bundesbank conducts monetary policy.” In: Romer, D. (Ed.), *Reducing Inflation*. Chicago: University of Chicago Press.

Clarida, R., J. Galí, and M. Gertler (1998): “Monetary policy rules in practice. Some international evidence,” *European Economic Review*, 42, 1033-1067.

Clarida, R., J. Galí, and M. Gertler (2000): “Monetary policy rules and macroeconomic stability: Evidence and some theory,” *Quarterly Journal of Economics*, 115, 147-180.

Dasgupta, S., B. Laplante, H. Wang, and D. Wheeler (2002): “Confronting the Environmental Kuznets Curve,” *The Journal of Economic Perspectives*, 16, 147-168.

Davidson, J. E. H., D. F. Hendry, F. Srba, and J. S. Yeo (1978): “Econometric modelling of the aggregate time series relationship between consumer’s expenditure and income in the United Kingdom,” *Economic Journal*, 88, 661-692.

de Jong, R. M. (2004): Addendum to “Asymptotics for nonlinear transformations of integrated time series,” *Econometric Theory*, 20, 627-635.

Dolado, J. J., R. María-Dolores, and M. Naveira (2005): “Are monetary-policy reaction functions asymmetric?: The role of non-linearity in the Phillips curve,” *European Economic Review*, 49, 485-503.

Embrechts, P., C. Klüppelberg, and T. Mikosh (1999): *Modelling extremal events for insurance and finance*. Berlin: Springer-Verlag.

Gonzalo, J., and J. Y. Pitarakis (2006): “Threshold effects in co-integrating regressions,” *Oxford*

Bulletin of Economics & Statistics, 68, 813-833.

Gonzalo, J., and J. Y. Pitarakis (2012): "Regime Specific Predictability in Predictive Regressions," *Journal of Business and Economic Statistics*, 30, 229-241.

Gordin M. I. (1969): "The Central Limit Theorem for Stationary Processes," *Soviet Math Dokl.*, 10, 1174-1176.

Granger, C. W. J. (1995): "Modelling non-linear relationships between extended-memory variables," *Econometrica*, 63, 265-279.

Granger, C. W. J., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111-120.

Grossman, G. M., and A. B. Krueger (1995): "Economic growth and environment," *Quarterly Journal of Economics*, 110, 353-377.

Hansen, B. E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory*, 8, 489-500.

Holtz-Eakin, D., and T. M. Selden (1995): "Stoking the fires? CO_2 emissions and economic growth," *Journal of Public Economics*, 57, 85-101.

Hong, S. H., and M. Wagner (2008): "Non-linear cointegration analysis and the environmental Kuznets curve," Mimeo.

Jalil, A., and S. F. Mahmud (2009): "Environment Kuznets curve for CO_2 emissions: A cointegration analysis for China," *Energy Policy*, 37, 5167-5172.

Jones, Ch. I. (1995): "Time series tests of endogenous growth models," *Quarterly Journal of Economics*, 110, 495-525.

Karlsen, H. A., T. Myklebust, and D. Tjϕstheim (2007): "Nonparametric estimation in a nonlinear cointegration type model," *The Annals of Statistics*, 35, 252-299.

Kasparis, I. (2008): "Detection of functional form misspecification in cointegrating relations," *Econometric Theory*, 24, 1373-1403.

Kurtz, T. G., and P. Protter (1991): "Weak limit theorems for stochastic integrals and stochastic differential equations," *The Annals of Probability*, 19, 1035-1070.

Lettau, M., and S. Van Nieuwerburgh (2008): "Reconciling the return predictability evidence," *Review of Financial Studies*, 21, 1607-1652.

Lewellen, J. (2004): "Predicting returns with financial ratios," *Journal of Financial Economics*, 74, 209-235.

McElroy, T., and D. N. Politis (2007): "Computer-intensive rate estimation, diverging statistics, and scanning," *The Annals of Statistics*, 35, 1827-1848.

Nickell, S. (1985): "Error correction, partial adjustment and all that: an expository note," *Oxford Bulletin of Economics and Statistics*, 47, 119-29.

Österholm, P. (2005): "The Taylor Rule: A Spurious Regression?," *Bulletin of Economic Re-*

search, 57, 217-247.

Park, J. Y., and P. C. B. Phillips (1999): “Asymptotics for non-linear transformations of integrated time series,” *Econometric Theory*, 15, 269-298.

Park, J. Y., and P. C. B. Phillips (2001): “Non-linear regressions with integrated time series,” *Econometrica*, 69, 117-161.

Perman, R. J., and D. Stern (2003): “Evidence from panel unit root and cointegration tests that the Environmental Kuznets Curve does not exist,” *Australian Journal of Agricultural and Resource Economics*, 47, 325-347.

Phillips, A. W. (1954): “Stabilization policy in a closed economy,” *Economic Journal*, 64, 290-323.

Phillips, A. W. (1957): “Stabilization policy and the time form of lagged responses,” *Economic Journal*, 67, 265-277.

Phillips, P. C. B. (1986): “Understanding Spurious Regressions in Econometrics,” *Journal of Econometrics*, 33, 311-340.

Phillips, P. C. B., and J. H. Lee (2012): “Long Horizon Predictive Regression,” Mimeo presented at SETA 2012.

Pötscher, B. M. (2004): “Non-linear functions and convergence to Brownian motion: beyond the continuous mapping theorem,” *Econometric Theory,* 20, 1-22.

Sargan, J. D. (1964): “Wages and prices in the United Kingdom: a study in econometric methodology,” In. P. E. Hart, G. Mills, and J. K. Whitaker (eds.), Econometric Analysis for National Economic Planning, Vol. 16 of Colston Papers. London: Butterworth Co., 25-63.

Schienle, M. (2011): “Nonparametric nonstationary regression with many covariates,” Mimeo.

Siklos, P. L., and M. E. Wohar (2005): “Estimating Taylor-type rules: An unbalanced regression?,” *Econometric Analysis of Financial and Economic Time Series*, 20, 239-276.

Svensoon, L. E. O. (1997): “Inflation forecast targeting: Implementing and monitoring inflation targets,” *European Economic Review*, 41, 1111-1146.

Taylor, J. B. (1993): “Discretion versus policy rules in practice,” *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

Wang, Q., and P. C. B. Phillips (2009): “Structural nonparametric cointegrating regressions,” *Econometrica*, 77, 1901-1948.

White, H. (1981): “Consequences and Detection of Misspecified non-linear Regression Models,” *Journal of the American Statistical Association*, 76, 419-433.