Journal of Econometrics 178 (2014) 331–341

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Summability of stochastic processes—A generalization of integration for non-linear processes

Vanessa Berenguer-Rico [a,*], Jesús Gonzalo [b,*]

[a] Department of Economics, University of Oxford and Institute for New Economic Thinking, United Kingdom

[b] Department of Economics, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Available online 5 September 2013

JEL classification:
C01
C22

Keywords:
Co-integration
Co-summability
Integrated processes
Non-linear balanced relationships
Non-linear processes
Summability

ABSTRACT

The order of integration is valid to characterize linear processes; but it is not appropriate for non-linear worlds. We propose the concept of summability (a re-scaled partial sum of the process being $O_p(1)$) to handle non-linearities. The paper shows that this new concept, $S(\delta)$: (i) generalizes $I(\delta)$; (ii) measures the degree of persistence as well as of the evolution of the variance; (iii) controls the balancedness of non-linear relationships; (iv) opens the door to the concept of co-summability which represents a generalization of co-integration for non-linear processes. To make this concept empirically applicable, an estimator for δ and its asymptotic properties are provided. The finite sample performance of subsampling confidence intervals is analyzed via a Monte Carlo experiment. The paper finishes with the estimation of the degree of summability of the macroeconomic variables in an extended version of the Nelson–Plosser database.

1. Introduction

No one doubts that the concepts of integration and co-integration have been and still are very useful in time series econometrics. The former by producing a single parameter that was able to summarize the long-memory properties of a given time series. The latter by linking the existence of common trends to long-run linear equilibrium relationships. Thanks, amongst others, to the work by Dickey and Fuller (1979), Nelson and Plosser (1982), Phillips (1986), Engle and Granger (1987) and Johansen (1991), these two concepts are easily handled theoretically as well as empirically.

In parallel, non-linear time series models from a stationary perspective were introduced in the literature—see Granger and Teräsvirta (1993), Franses and van Dijk (2000), Fan and Yao (2003), and Teräsvirta et al. (2010) for some overviews. The introduction of persistent variables into non-linear models – see Park and Phillips (1999, 2001), de Jong and Wang (2005) or Pötscher (2004) for the study of transformations of integrated processes – produced a natural query: Which is the order of integration of these non-linear transformations? Such a question does not have a clear answer since the existing definitions of integrability do not properly apply. Integration is a linear concept.

This lack of definition has at least two important worrying consequences. First, in univariate terms, it implies that an equivalent synthetic measure of the stochastic properties of the time series, such as the order of integration, is not available to characterize non-linear time series. As pointed out by Granger (1995), this does not only affect econometricians but also economic theorists who need to account for important characteristics of economic variables to construct their theories. Second, from a multivariate perspective, it becomes troublesome to determine whether a non-linear model is balanced or not. Unbalancedness is a symptom of a misspecified model, a feature that is easily likely to occur when managing non-linear transformations of persistent variables. In linear setups, the concept of integrability did a good job dealing with balanced/unbalanced relations. However, in non-linear frameworks, the non-existence of a synoptic quantitative measure makes it difficult to check the balancedness of a postulated model.

Additionally, this implies that a definition for non-linear co-integration is difficult to be obtained from the usual concept of integrability. To clarify this point, suppose $y_t = f(x_t, \theta) + u_t$, where $x_t \sim I(1)$ and $u_t \sim I(0)$. For $f(\cdot)$ non-linear, the order of integration of $f(x_t, \theta)$, and hence that of y_t, may not be properly defined implying that the standard concept of co-integration is difficult to be applied. In fact, the literature on non-linear co-integration – see Park and Phillips (2001), Karlsen et al. (2007), Wang and Phillips (2009) – undertakes the whole analysis assuming the existence of a long-run relationship; something that should be tested in practice.

* Corresponding authors.
E-mail addresses: vanessa.berenguer-rico@nuffield.ox.ac.uk (V. Berenguer-Rico), jesus.gonzalo@uc3m.es (J. Gonzalo).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.031

It was already stated in Granger and Hallman (1991) that extensions of the linear concepts $I(0)$ and $I(1)$ are needed to generalize co-integration to non-linear frameworks. This has led some authors to introduce alternative definitions. For instance, Granger (1995) proposed the concepts of Extended and Short Memory in Mean. However, these concepts are neither easy to calculate nor general enough to handle some types of non-linear long run relationships. And, furthermore, a measure of the order of the extended memory is not available. Dealing with threshold effects in co-integrating regressions, Gonzalo and Pitarakis (2006) faced these problems and proposed, in a very heuristic way, the concept of summability (a re-scaled partial sum of the process being $Op(1)$). However, they did not emphasize the avail of such an idea.

In this paper, we define summability properly and show its usefulness and generality. Specifically, we put forward several relevant examples in which the order of integrability is difficult to be established, but the order of summability can be easily determined. Moreover, we show that integrated time series are particular cases of summable processes, in the sense that the order of summability is the same as the order of integration. Hence, summability is a generalization of integrability. Furthermore, summability does not only characterize some properties of univariate time series, but also allows to easily study the balancedness of a postulated relationship —linear or not. And even more important, non-linear long run equilibrium relationships between non-stationary time series can be properly defined. In particular, the concept of co-summability, which can be applied to extend co-integration to non-linear frameworks, is developed by the authors in Berenguer-Rico and Gonzalo (2013).

To make this concept empirically operational, we propose a statistical procedure to estimate and carry out inferences on the order of summability of an observed time series. This makes useful the concept of summability not only in theory but also in practice. To estimate the order of summability, we use an estimator introduced by McElroy and Politis (2007) to analyze the rate of convergence of a statistic which is obtained from a simple least squares regression. The inference on the true order of summability is based on the subsampling methodology developed in Politis et al. (1999). It is shown, by simulations, that the subsampling machinery works reasonably well in finite samples given the generality of the approach. Finally, the proposed methodology is used to estimate the order of summability of the macroeconomic time series in an extended version of the Nelson–Plosser database.

The paper is organized as follows. In the next section, the problems of using the order of integration to characterize non-linear processes are highlighted. In Section 3, our proposed solution based on summability is described and its simple applicability showed. Section 4 describes the statistical tools – estimation and inference – to deal empirically with summable processes. In Section 5, an empirical application shows how to determine the order of summability in practice. Finally, Section 6 is devoted to some concluding remarks. All proofs are collected in the Appendix.

A word on notation. We use the symbol "$\Longrightarrow$" to signify convergence in distribution and weak convergence indistinctly, "$\xrightarrow{p}$" to signify convergence in probability. Stochastic processes such as the standard Brownian motion $W(r)$ are defined on [0, 1]. Finally, all limits given in the paper are taken as the sample size $n \to \infty$.

2. Order of integration and non-linear processes

2.1. Order of integration

Definition 1. A stochastic process $\{y_t : t \in \mathbb{N}\}$ is said to be an integrated process of order d (in short, an $I(d)$ process) if the process of dth order differences $\Delta^d y_t$ is $I(0)$.

A natural question that arises after reading this definition is: And what is an $I(0)$ process? Attempts to give a formal description of $I(0)$ processes exist in the literature. Engle and Granger (1987) give the following characterization.

Engle and Granger (EG) Characterization. *If $y_t \sim I(0)$ with zero mean then* (i) *the variance of y_t is finite;* (ii) *an innovation has only a temporary effect on the value of y_t;* (iii) *the spectrum of y_t, $f(\omega)$, has the property $0 < f(0) < \infty$;* (iv) *the expected length of time between crossing of $x = 0$ is finite;* (v) *the autocorrelations, ρ_k, decrease steadily in magnitude for large enough k, so that their sum is finite.*

Other characterizations have been used as well. Granger (1995) and Johansen (1995) used autoregressive and moving average representations, respectively. Müller (2008) and Davidson (2009) – among others – define an $I(0)$ as a process that satisfies the functional central limit theorem (FCLT). These latter definitions share the same spirit of our summability definition in Section 3. Nevertheless, in all cases, differences must be taken to discover the order of integration and the intrinsic linearity of the difference operator makes it difficult, if not impossible, to characterize – among others – non-linear processes. Integration is a linear concept.

2.2. Examples

Example 1. Alpha Stable i.i.d. Distributed Processes

Let y_t be *i.i.d.* from some distribution $F \in D(\alpha)$, where $D(\alpha)$ denotes the domain of attraction of an α-stable law with $\alpha \in (0, 2]$. y_t is strictly stationary; however, its second moments may not exist. The fact that such a process is *i.i.d.* could incline to think that this process is $I(0)$. However, if second moments do not exist, EG characterization does not apply. Characterizations based on the FCLT could not be used either since they assume a standard Brownian motion in the limit. Hence, it becomes troublesome to establish the order of integration of y_t.

Example 2. An i.i.d. plus a Random Variable

Consider the following process

$$y_t = z + e_t, \tag{1}$$

where $z \sim N(0, \sigma_z^2)$ and $e_t \sim i.i.d.(0, \sigma_e^2)$ are independent of each other. This process has the following properties:

(i) $E[y_t] = 0$
(ii) $V[y_t] = \sigma_z^2 + \sigma_e^2$
(iii) $\gamma_y(k) = Cov(y_t, y_{t-k}) = \sigma_z^2$ for all $k > 0$.

Since it is a strictly stationary process, one could think that it is $I(0)$. However, the autocovariance function is not absolutely summable and its spectrum does not satisfy the required condition in EG characterization.[1] If y_t is not $I(0)$, to attach any other order of integration to this stochastic process is not obvious. It is controversial to say y_t is $I(1)$ since $\Delta y_t = \Delta e_t$ is generally understood as an $I(-1)$; and it becomes difficult to choose any other number using the above definition of order of integration.

Dealing with non-linear processes similar problems are faced.

[1] The autocovariance of the process in this example can be expressed as

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} \left[\frac{\sigma_z^2 + \sigma_e^2}{2\pi} + \frac{\sigma_z^2}{\pi} \sum_{h=1}^{\infty} \cos(\lambda h) \right] d\lambda.$$

Then, the spectral density is

$$f(\lambda) = \frac{\sigma_z^2 + \sigma_e^2}{2\pi} + \frac{\sigma_z^2}{\pi} \sum_{h=1}^{\infty} \cos(\lambda h),$$

which diverges for all λ.

Example 3. Product of i.i.d. and Random Walk

Let

$$w_t = \pi_t \eta_t, \tag{2}$$

where $\eta_t \sim i.i.d.\,(0, 1)$ and

$$\pi_t = \pi_{t-1} + \varepsilon_t, \tag{3}$$

with $\pi_0 = 0$ and $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$ independent of η_t. Some properties of w_t are:

(i) $E[w_t] = 0$

(ii) $V[w_t] = \sigma_\varepsilon^2 t$

(iii) $\gamma_w(h) = E[w_t w_{t-h}] = 0$.

It is not obvious to attach an order of integration to this process. On the one hand, the uncorrelation property (iii) could incline to think that w_t is $I(0)$. However, an $I(0)$ cannot have a trend in the variance according to EG characterization. On the other hand, this unbounded variance could induce to suspect that the process is $I(1)$. Nevertheless, its first difference

$$\Delta w_t = \pi_t \eta_t - \pi_{t-1}\eta_{t-1},$$

cannot be $I(0)$ since, again,

$$\begin{aligned} V[\Delta w_t] &= E[(\pi_t\eta_t)^2] + E[(\pi_{t-1}\eta_{t-1})^2] - 2E[\pi_t\pi_{t-1}\eta_t\eta_{t-1}] \\ &= (2t-1)\sigma_\varepsilon^2. \end{aligned}$$

This means that w_t cannot be $I(1)$. It cannot be $I(2)$ either, since the variance of the second difference is

$$\begin{aligned} V[\Delta^2 w_t] &= E[(\pi_t\eta_t)^2] + 4E[(\pi_{t-1}\eta_{t-1})^2] + E[(\pi_{t-2}\eta_{t-2})^2] \\ &= 6(t-1)\sigma_\varepsilon^2. \end{aligned}$$

In fact, this process can be considered to be $I(\infty)$, in the sense that the variance of $\Delta^d w_t$ depends on t regardless of the value of d—see Yoon (2005).

As pointed out by Granger (1995), non-linear transformations of highly heterogeneous or volatile processes, although uncorrelated, can induce high correlations. This can be seen by analyzing

$$q_t = \pi_t \eta_t^2, \tag{4}$$

where π_t and η_t are defined as before. The only difference is that now the *i.i.d.* sequence, η_t^2, is always positive. However, in this case,

$$E[q_t] = E[\pi_t \eta_t^2] = 0,$$

$$V[q_t] = E[q_t^2] = E[\pi_t^2\eta_t^4] = E[\pi_t^2]E[\eta_t^4] = t\sigma_\varepsilon^2\mu_4,$$

and

$$\begin{aligned} \gamma_q(h) &= E[q_t q_{t-h}] = E[\pi_t\pi_{t-h}\eta_t^2\eta_{t-h}^2] \\ &= E[\pi_t\pi_{t-h}]E[\eta_t^2\eta_{t-h}^2] = (t-h)\sigma_\varepsilon^2\sigma_\eta^4, \end{aligned}$$

where $\mu_4 = E[\eta_t^4]$. Now both variance and covariance depend on time. Hence, it can be seen how non-linear transformations of highly heterogeneous processes can have an important impact on their stochastic properties. This impact will be hardly accounted by the order of integration.

Example 4. Square of a Random Walk

Consider now the square of the random walk defined in (3),

$$\pi_t^2 = \pi_{t-1}^2 + 2\pi_{t-1}\varepsilon_t + \varepsilon_t^2. \tag{5}$$

To establish the order of integration of this process is again not an obvious task. Granger (1995) considers 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,

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

Again, EG characterization cannot be applied to $\Delta\pi_t^2$ or $\Delta^d\pi_t^2$.

Example 5. Product of Indicator Function and Random Walk

Let

$$h_t = 1(v_t \le \gamma)\pi_t, \tag{6}$$

where v_t is *i.i.d.* (0, 1), $1(\cdot)$ is the indicator function, and π_t is the random walk defined in (3). v_t and ε_t are independent of each other. The variance and autocovariances of h_t depend on time, hence, one would think that it is $I(1)$. However, again, the variance of the first difference

$$\begin{aligned} V[\Delta h_t] &= V[1(v_t \le \gamma)\pi_t - 1(v_{t-1} \le \gamma)\pi_{t-1}] \\ &= [2p(1-p)\sigma_\varepsilon^2]t + p(2p-1)\sigma_\varepsilon^2, \end{aligned}$$

where $p = \Pr(v_t \le \gamma)$. In fact, it can be considered, once again, that $h_t \sim I(\infty)$.

Example 6. Park and Phillips (1999, 2001)

Similar incongruities to those encountered in previous examples appear when dealing with the non-linear transformations of $I(1)$ processes studied in Park and Phillips (1999, 2001); for instance, $e^{-\pi_t^2}$, $1/(1+\pi_t^2)$, $\log(|\pi_t|)$, or $(1+e^{-\pi_t})^{-1}$.

Example 7. Stochastic Unit Root and Explosive Processes

Consider, on the one hand, a stochastic unit root process

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \tag{7}$$

where $y_0 = 0$ and $\rho_t \sim i.i.d.(\rho, \omega^2)$ is independent of $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. On the other hand, let z_t be the following explosive process

$$z_t = \phi z_{t-1} + \xi_t, \tag{8}$$

with $z_0 = 0$, $\phi > 1$ and $\xi_t \sim i.i.d.(0, \sigma_\xi^2)$. As in previous examples, to determine the order of integration of y_t and z_t is troublesome.

In all these examples the order of integrability is difficult to be calculated. The standard $I(d)$ classification is not sufficient to handle many stochastic processes.

3. A solution based on summability

3.1. Order of summability

The idea of order of summability of a stochastic process was initially introduced in a heuristic way in Gonzalo and Pitarakis (2006) when dealing with threshold effects in co-integrating regressions. In this section, the concept of summability is formalized and its generality, usefulness, and simplicity are asserted.

Definition 2. 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[2] $L(n)$ and a deterministic sequence m_t such that

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1), \tag{9}$$

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

Remark. When $E[y_t]$ exists, $m_t = E[y_t]$; however, when this is not the case m_t will be case dependent. For instance, in Example 1, $m_t = 0$ if $\alpha \in (0, 1)$ or $\alpha = 1$ and F is symmetric. With respect to δ, when possible, it will be determined by some Central Limit result. For instance, a standard Central Limit Theorem – CLT – produces

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

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

(See Embrechts et al., 1999, p. 564).

$\delta = 0$ and $L(n)$ just a constant. If the process is a random walk, by the Functional Central Limit Theorem – FCLT – and the Continuous Mapping Theorem – CMT –, $\delta = 1$ and $L(n)$ is again a constant term. Although, in many circumstances $L(n)$ will be constant, in some situations[3] the asymptotic theory will force us to use an L function varying with n but slowly in Karamata's sense.

From a more general perspective, the relationship between integrability and summability is discussed in the following two propositions.

Assumption 1. Let y_t be the $I(d)$ process $\Delta^d y_t = C(L) u_t$, where $u_t = \varepsilon_t 1(t > 0)$. ε_t has zero mean, is *i.i.d.*, and $E|\varepsilon_t|^r < \infty$ for $r \geq 4$. In addition, $C(L) = \sum_{j=0}^{\infty} c_j L^j$ satisfies $C(1) \neq 0$ and $\sum_{j=1}^{\infty} j|c_j| < \infty$.

Proposition 1. *Under Assumption 1 if the time series y_t is $I(d)$ with $d \geq 0$, then it is $S(d)$.*

Next proposition deals with processes with negative orders of integration.

Proposition 2. *Under Assumption 1 if the time series y_t is $I(-d)$ with $d = 1, 2, \ldots < \infty$, then it is $S(-0.5)$.*

Since negative orders of integration are not very relevant, only $d \geq 0$ will be considered. Hence, $I(d)$ processes are $S(d)$.

3.2. Examples

For all processes considered in Examples 1–7 the order of integration was not possible to be established. Next, for these examples, it is shown that the order of summability can be easily obtained.

Summability in Example 1 (*α-Stable i.i.d. Process*)**.** Let y_t be symmetric around zero. By the Generalized Central Limit Theorem

$$S_n = \frac{1}{n^{\frac{1}{\alpha}}} L(n) \sum_{t=1}^{n} y_t \Longrightarrow S_\alpha,$$

where $S_\alpha \sim F \in D(\alpha)$. Hence, in this case the time series is said to be summable of order $\delta = (2-\alpha)/2\alpha$. For instance, a Cauchy distributed process ($\alpha = 1$) is $S(0.5)$.

Summability in Example 2 (*An i.i.d. Plus a Random Variable*)**.** From (1)

$$S_n = \frac{1}{n}\sum_{t=1}^{n} y_t = \frac{1}{n}\sum_{t=1}^{n}(z + e_t) = z + \frac{1}{n}\sum_{t=1}^{n} e_t \Longrightarrow z.$$

Therefore, y_t is $S(0.5)$.

Summability in Example 3 (*Product of i.i.d. and Random Walk*)**.** It can be shown – see for instance Park and Phillips (1988) – that

$$S_n = \frac{1}{\sigma_\varepsilon n}\sum_{t=1}^{n} \pi_t \eta_t \Longrightarrow \int_0^1 W_1(r) dW_2(r).$$

This means $\pi_t \eta_t$ is $S(0.5)$ with, for instance, $L(n) = 1/\sigma_\varepsilon$.

For $\pi_t \eta_t^2$ note that,

$$Var\left[\sum_{t=1}^{n} \pi_t \eta_t^2\right] = O(n^3).$$

Then, by Chebyshev's inequality,

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} \pi_t \eta_t^2 = O_p(1),$$

which implies that $\pi_t \eta_t^2$ is $S(1)$.

These two cases show that summability takes into account persistence as well as the variance behavior through time.

Summability in Example 4 (*Square of a Random Walk*)**.** It is well known that

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

Hence, π_t^2 is $S(1.5)$ with, for instance, $L(n) = 1/\sigma_\varepsilon^2$.

Summability in Example 5 (*Product of Indicator Function and Random Walk*)**.** In this case,

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

implying that $1(v_t \leq \gamma)\pi_t$ is $S(1)$ with, for instance, $L(n) = 1/p\sigma_\varepsilon$.

Summability in Example 6 (*Park and Phillips, 1999, 2001*)**.** The order of summability of the processes considered in this example can be obtained by using the asymptotic theory developed in Park and Phillips (1999). Specifically, it can be shown that $e^{-\pi_t^2} \sim S(0)$, $1/(1+\pi_t^2) \sim S(0)$, $\log(|\pi_t|) \sim S(0.5)$, and $(1+e^{-\pi_t})^{-1} \sim S(0.5)$.

Summability in Example 7 (*STUR and Explosive Processes*)**.** Consider the STUR process defined in (7). For simplicity, let $\rho_t \sim i.i.d.(1, 1)$, i.e. set $\rho = \omega^2 = 1$. From Leybourne et al. (1996), it can be shown that

$$S_n = \frac{1}{2^{n/2}}\sum_{t=1}^{n} y_t = O_p(1).$$

With respect to the explosive process (8), from White (1958)

$$S_n = \frac{1}{\phi^n}\sum_{t=1}^{n} z_t = O_p(1).$$

Strictly speaking, the order of summability of y_t and z_t will be ∞. These are cases of non-summable processes.

3.3. Some uses of summability

In the same way integration constitutes the first step to check balancedness of a linear relationship and to analyze co-integration, summability can be used to study non-linear long run relationships.

Definition 3. A postulated relationship

$$y_t = f(x_t, \theta),$$

is said to be balanced if $y_t \sim S(\delta_y)$, $f(x_t, \theta) \sim S(\delta_f)$, and $\delta_y = \delta_f$.

Once balancedness of a non-linear model is established, the analysis of non-linear long run relationships can be done using the concept of co-summability.

[3] Consider the case where the process y_t has density $f(x) = 1/|x|^3$ for $|x| > 1$. In that case, it is known (e.g. Romano and Siegel, 1986, Example 5.47) that

$$\frac{1}{[n \log n]^{1/2}}\sum_{t=1}^{n} y_t \Longrightarrow N(0, 1).$$

Then, $L(n) = (1/\log n)^{1/2}$.

Definition 4. 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 $f(x_t, \theta) \sim S(\delta_y)$ such that $u_t = y_t - f(x_t, \theta)$ is $S(\delta_u)$, with $\delta_u = \delta_y - \delta$ and $\delta > 0$. In short, $(y_t, x_t) \sim CS(\delta_y, \delta)$.

Co-summable processes will share an equilibrium relationship in the long run, i.e. an attractor $y_t = f(x_t, \theta)$ that can be linear or not. This type of equilibrium relationships will be usually established by the economic theory and have significant econometric applications such as, for instance, transition between regimes, multiplicity of equilibria, polynomial approximations to unknown functions or non-linear responses to policy interventions. Applied researchers are often interested on estimating and testing for these type of equilibria. A full treatment of co-summability in a regression framework can be found in Berenguer-Rico and Gonzalo (2013) and Berenguer-Rico (2013).

4. Summability in practice: estimation and inference

Following the same logic as in the integrated world, before any multivariate analysis, i.e. balancedness and co-summability, it is necessary to develop the estimation and inference tools for the order of summability, δ, of univariate processes.

4.1. Estimation of δ

In this section, for simplicity reasons, it will be assumed $L(n) = 1$ in Definition 2. Therefore, the summability condition (9) becomes

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - m_t) = O_p(1). \tag{10}$$

In addition, the next assumption is needed to implement our proposed estimation method of δ.

Assumption 2. $P(S_n = 0) = 0$ for all $n = 1, 2, 3, \ldots$.

Under Assumption 2 and following McElroy and Politis (2007), for a stochastic process y_t satisfying (10),

$$U_n = \log S_n^2 = \log \left(n^{-(1+2\delta)} \left(\sum_{t=1}^{n} (y_t - m_t) \right)^2 \right) = O_p(1). \tag{11}$$

Expression (11) can be written in regression model form as follows

$$Y_k = \beta \log k + U_k, \quad k = 1, 2, \ldots, n, \tag{12}$$

where $\beta = 1 + 2\delta$, $Y_k = \log \left(\sum_{t=1}^{k} (y_t - m_t) \right)^2$, and $U_k = O_p(1)$.

We propose to estimate β by

$$\hat{\beta} = \frac{\sum_{k=1}^{n} Y_k \log k}{\sum_{k=1}^{n} \log^2 k}. \tag{13}$$

Given that $\beta = 1 + 2\delta$, the OLS estimator of δ is

$$\hat{\delta} = \frac{\hat{\beta} - 1}{2}.$$

4.2. Asymptotic properties

From (12) and (13)

$$\hat{\beta} - \beta = \frac{\sum_{k=1}^{n} U_k \log k}{\sum_{k=1}^{n} \log^2 k}. \tag{14}$$

Proposition 3 (*McElroy and Politis, 2007*). *Under Assumption 2, $\hat{\beta} - \beta = o_p(1)$.*

Remark. McElroy and Politis (2007) show that $\hat{\beta}$ is consistent under minimal assumptions. In our context, these assumptions are satisfied by definition of summable processes. Nonetheless, to the best of our knowledge, an asymptotic distribution for $\hat{\beta}$ has not yet been derived. The following proposition addresses this issue.

Proposition 4. *Under Assumption 2, if*

$$\frac{1}{n} \sum_{k=1}^{n} U_k \Longrightarrow D_U \quad \text{and} \quad \frac{1}{n} \sum_{k=1}^{n} |U_k|^p = O_p(1), \tag{15}$$

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

$$\log n(\hat{\beta} - \beta) \Longrightarrow D_U. \tag{16}$$

Remark. As shown in McElroy and Politis (2007) boundedness in probability of U_k suffices to get a consistent estimate of β. Nevertheless, to perform inferences on β, extra distributional assumptions, such as those in (15), need to be imposed. Let $x_t = y_t - m_t$ and notice that

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n} U_k &= \frac{1}{n} \sum_{k=1}^{n} \log S_k^2 \\ &= -\frac{(1+2\delta)}{n} \sum_{k=1}^{n} \log \left(\frac{k}{n} \right) \\ &\quad + \frac{1}{n} \sum_{k=1}^{n} \log \left[\left(\frac{1}{n^{1/2+\delta}} \sum_{t=1}^{k} x_t \right)^2 \right]. \end{aligned}$$

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

$$\frac{1}{n} \sum_{k=1}^{n} U_k \Longrightarrow 1 + \int_0^1 \log \left(W^2(r) \right) dr \quad \text{and}$$

$$\frac{1}{n} \sum_{k=1}^{n} |U_k|^p = O_p(1).$$

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

$$\frac{1}{n} \sum_{k=1}^{n} U_k \Longrightarrow 3 + \int_0^1 \log \left(\left(\int_0^r W(r) dr \right)^2 \right) dr \quad \text{and}$$

$$\frac{1}{n} \sum_{k=1}^{n} |U_k|^p = O_p(1).$$

Remark. In many cases, $L(n) \neq 1$ but still $L(n) = c$, a constant different from zero. In such a case, regression (12) becomes

$$Y_k = \alpha + \beta \log k + U_k, \tag{17}$$

with $\alpha = -2 \log c$. Notice that any c satisfies Definition 2. Therefore, α is not identified. Nevertheless, it is straightforward to get rid of it by subtracting the first observation in regression (17) and estimating the model

$$Y_k^* = \beta \log k + U_k^*, \tag{18}$$

Table 1
Data generating processes: $y_t = m_t + x_t$.

$y_{1t} = m_t + \varepsilon_t, \varepsilon_t \sim iid\, N(0,1)$	$y_{7t} = m_t + \Delta^{0.3}\pi_t$
$y_{2t} = m_t + \pi_t, \pi_t = \sum_{j=1}^{t}\varepsilon_j$	$y_{8t} = m_t + z + \varepsilon_t, z \sim N(0,1)\perp\varepsilon_t$
$y_{3t} = m_t + \sum_{j=1}^{t}\pi_j$	$y_{9t} = m_t + \eta_t\pi_t, \eta_t \sim iid\, N(0,1)\perp\varepsilon_t$
$y_{4t} = m_t + \xi_t, \xi_t \sim iid$ Cauchy	$y_{10t} = m_t + \eta_t^2\pi_t, \eta_t \sim iid\, N(0,1)\perp\varepsilon_t$
$y_{5t} = m_t + \pi_t^2$	$y_{11t} = m_t + 1(v_t \le 0)\pi_t$, $v_t \sim iid\, N(0,1)\perp\varepsilon_t$
$y_{6t} = m_t + t\varepsilon_t$	$y_{12t} = m_t + \log(\lvert\pi_t\rvert)$

where $Y_k^* = Y_k - Y_1$ and $U_k^* = U_k - U_1$. The modified OLS estimator

$$\hat{\beta}^* = \frac{\sum_{k=1}^{n} Y_k^* \log k}{\sum_{k=1}^{n} \log^2 k},$$

satisfies the same asymptotic properties than those of $\hat{\beta}$.

An alternative way to take into account α could be using

$$\tilde{\beta} = \frac{\sum_{k=1}^{n}(Y_k - \bar{Y})(\log k - \overline{\log n})}{\sum_{k=1}^{n}(\log k - \overline{\log n})^2}. \tag{19}$$

The lack of identification of α complicates the properties of $\tilde{\beta}$. For this reason, in this paper only $\hat{\beta}^*$ is considered and consequently $\hat{\delta}^* = (\hat{\beta}^* - 1)/2$.

4.3. Subsampling confidence intervals

In general, the asymptotic distribution of $\hat{\beta}^*$ cannot be tabulated. Nevertheless, subsampling methods can be used to undertake inferences on the order of summability independently of its true value.

Subsampling is consistent under minimal assumptions. The most general result shown in Politis et al. (1999) requires that:

(i) the estimator, properly normalized, has a limiting distribution,

(ii) the distribution functions of the normalized estimator based on the subsamples (of size b) have to be on average close to the distribution function of the normalized estimator based on the entire sample with $\log b/\log n \to 0$, $b/n \to 0$, $b \to \infty$,

(iii) the sequence of the subsampling statistics $Z_{n,b,k} = \log b(\hat{\beta}^*_{n,b,k} - \beta)$, where $\hat{\beta}^*_{n,b,k}$ is the subsample estimator version of $\hat{\beta}^*$, has α-mixing coefficients, $\alpha_{n,b}(h)$, such that $n^{-1}\sum_{h=1}^{n}\alpha_{n,b}(h) \to 0$ as $n \to \infty$.

Conditions (i) and (ii) are guaranteed by Proposition 4. To show that the α-mixing condition (iii) holds in this context is beyond the scope of this paper. The adequacy of the subsampling approach is analyzed via simulations using the twelve data generating processes – DGP – in Table 1.

Performance of subsampling is mainly measured by coverage probability, denoted by *CP*, of two-sided nominal 95% symmetric intervals for δ. We also present the mean of the estimated $\delta's$ and the median lower and upper bounds of the estimated confidence intervals. These measures are denoted by $\bar{\delta}^*$, I_{low}, and I_{up}, respectively. The experiment is based on 1000 replicas and three different sample sizes $n = \{100, 200, 500\}$. Subsample size is $b = \sqrt{n}$. Results are collected in Table 2.

The performance of the subsampling method is adequate in general.[4] The coverage probability is around its nominal level and the mean estimated order of summability close to its true value. The subsampling confidence intervals, although wide, get narrower as the sample size increases. The amplitude of the intervals in small samples is basically a direct consequence of not assuming any knowledge about the DGP of the analyzed time series.

4.4. Deterministic components

Until now m_t has been assumed to be known, but this is not the case in practice. As in the integrated world, the presence of deterministic components can affect the estimation of the order of summability.

Let

$$y_t = m_t + x_t,$$

where

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} x_t \Longrightarrow D_x \quad \text{and} \quad \frac{1}{n^{1/2+\gamma}}\sum_{t=1}^{n} m_t \to \mu,$$

with D_x being a random variable with positive variance and μ a constant different from zero.

Consider the following two situations:

a. If $\delta > \gamma$, then

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} x_t + o(1) \Longrightarrow D_x.$$

b. If $\delta < \gamma$, then

$$\frac{1}{n^{1/2+\gamma}}\sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\gamma}}\sum_{t=1}^{n} m_t + o_p(1) \xrightarrow{p} \mu.$$

When $\delta < \gamma$, the order of the deterministic component dominates and it will be confused with the order of summability. Admittedly, even when $\delta > \gamma$, the deterministic components, if not properly considered, can affect the order of summability estimation in finite samples. Although not reported here, for space reasons, Monte Carlo experiments reveal the existence of an important bias effect when deterministic components are present and not properly taken into consideration. Therefore, in order to analyze the order of summability a proper technique to deal with these elements is needed.

Essentially, what is required is an estimator $\hat{m}_t$ such that

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}\left(y_t - \hat{m}_t\right) = O_p(1). \tag{20}$$

In other words, the order of summability of y_t is not affected by subtracting $\hat{m}_t$.

Three usual parametric forms for m_t will be considered: $m_t = m_0$, $m_t = m_0 + m_1 t$, and $m_t = m_0 + m_1 t + m_2 t^2$. For these three cases, a proper treatment of the deterministic components is derived.

Constant term case: Let

$$y_t = m_0 + x_t,$$

where m_0 is a constant and $x_t \sim S(\delta)$ such that

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n} x_t = O_p(1).$$

Assume that only y_t is observed. The standard proposal of demeaning y_t by its arithmetic mean is problematic in this context because

$$\sum_{t=1}^{n}(y_t - \bar{y}) = 0. \tag{21}$$

[4] Notice that the coverage probability for cases 11 and 12 is poor. Nonetheless, the consideration of deterministic components improves dramatically the coverage probability, as it can be seen in Tables 3 and 4.

Table 2
Performance of subsampling intervals for δ. No deterministic components: $m_t = 0$.

DGP	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}
$S(\delta)$	$n = 100$				$n = 200$				$n = 500$			
1 – $S(0)$	0.991	−0.004	−0.699	0.659	0.995	0.005	−0.607	0.566	0.991	0.000	−0.521	0.470
2 – $S(1)$	0.832	0.863	0.383	1.307	0.804	0.880	0.455	1.258	0.807	0.900	0.541	1.220
3 – $S(2)$	0.747	1.634	0.982	2.262	0.797	1.673	1.034	2.292	0.863	1.723	1.076	2.348
4 – $S(0.5)$	0.986	0.496	−0.414	1.387	0.992	0.521	−0.261	1.309	0.994	0.519	−0.185	1.187
5 – $S(1.5)$	0.905	1.516	0.701	2.192	0.900	1.519	0.771	2.107	0.904	1.510	0.828	2.049
6 – $S(1)$	0.990	0.862	−0.052	1.694	0.997	0.891	0.028	1.675	1.000	0.899	0.096	1.635
7 – $S(0.7)$	0.939	0.613	0.038	1.135	0.954	0.627	0.141	1.054	0.949	0.639	0.223	0.998
8 – $S(0.5)$	0.942	0.430	−0.213	1.007	0.929	0.401	−0.149	0.915	0.930	0.447	−0.024	0.875
9 – $S(0.5)$	0.988	0.507	−0.330	1.255	0.984	0.516	−0.206	1.164	0.983	0.501	−0.144	1.063
10 – $S(1)$	0.947	1.171	−0.106	2.311	0.952	1.167	0.099	2.127	0.954	1.124	0.220	1.894
11 – $S(1)$	0.598	0.689	0.220	1.104	0.644	0.743	0.325	1.140	0.650	0.767	0.389	1.105
12 – $S(0.5)$	0.844	0.557	0.041	0.977	0.801	0.630	0.196	0.988	0.705	0.694	0.353	0.982

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. $\bar{\delta}^*$ represents the mean of the estimated orders of summability. I_{low} and I_{up} are the median of the lower and upper bounds of the intervals, respectively. 1000 replicas are used. Subsample size is $b = \sqrt{n}$.

Therefore, the true order of summability cannot be recovered. Next proposition shows that the partial mean

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j,$$

is an alternative operational choice in the sense of satisfying (20).

Proposition 5. *Consider the* constant term case *DGP*

$$y_t = m_0 + x_t, \tag{22}$$

where m_0 is an unknown constant and

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n} |x_t| = O_p(1).$$

If

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j, \tag{23}$$

then $(y_t - \hat{m}_t) \sim S(\delta)$.

Table 3 reports the performance of subsampling confidence intervals after partially demeaning the processes described in Table 1 when $m_t = m_0 = 10$. Results do not depend on the value of m_0.

Results are similar or even better than those obtained without deterministic components. For this reason, we recommend to always partially demean the processes.

Linear trend case: Let

$$y_t = m_0 + m_1 t + x_t,$$

where $x_t \sim S(\delta)$. Next proposition shows how to deal with the deterministic components in this case.

Proposition 6. *Consider the* linear trend case *DGP*

$$y_t = m_0 + m_1 t + x_t, \tag{24}$$

where m_0 and m_1 are two unknown constants and

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n} |x_t| = O_p(1).$$

If

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j - \frac{2}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_j\right), \tag{25}$$

then $(y_t - \hat{m}_t) \sim S(\delta)$.

Note that in the linear trend case, the appropriate $\hat{m}_t$ consists, basically, in a double partial demeaning procedure.[5] Table 4 summarizes the performance of subsampling confidence intervals after properly detrending the DGPs in Table 1 when $m_t = m_0 + m_1 t = 10 + 2t$. As in the previous case, results do not depend on the particular choices of m_0 and m_1.

Results in Table 4 show that the proposed detrending method $\hat{m}_t$ performs adequately in finite samples.

Quadratic trend case: Let

$$y_t = m_0 + m_1 t + m_2 t^2 + x_t,$$

where $x_t \sim S(\delta)$ and m_0, m_1, and m_2 are three unknown constants. The proposed $\hat{m}_t$ in this case is

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j - \frac{2}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i\right) - \frac{3}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i - \frac{2}{j}\sum_{i=1}^{j}\left(y_i - \frac{1}{i}\sum_{h=1}^{i} y_h\right)\right).$$

Essentially, this transformation implies a triple partial demeaning procedure. It can be shown that the use of this $\hat{m}_t$ does not alter the order of summability of $y_t - \hat{m}_t$ and the finite sample performance is adequate (these results are available from the authors upon request).

Remark. It can be shown that if the order of the trend that is subtracted is higher than the true one, then the order of summability of the detrended process, $y_t - \hat{m}_t$, is preserved; that is, it has the same order of summability that y_t. However, because of inefficiency issues, in general, it is not recommended to subtract a very high polynomial trend.

Overall, the methodology proposed in this section to estimate the order of summability works reasonably well in finite samples. It is important to notice that our method does not assume any knowledge about the model generating the data. The trade off is that the confidence intervals are not very narrow.

[5] Other proper detrending procedures work too. We thank Franco Peracchi for pointing out the alternative methodology of applying a partial OLS detrending, i.e. $\hat{m}_t = \hat{\alpha}_t + \hat{\beta}_t t$ where $\hat{\alpha}_t = (1/t)\sum_{j=1}^{t} y_j - \hat{\beta}_t (1/t)\sum_{j=1}^{t} j$ and $\hat{\beta}_t = \sum_{j=1}^{t}\left(y_j - (1/t)\sum_{j=1}^{t} y_j\right)\left(j - (1/t)\sum_{j=1}^{t} j\right) / \sum_{j=1}^{t}\left(j - (1/t)\sum_{j=1}^{t} j\right)^2$. This choice will be particularly interesting when fractional deterministic trends are present.

Table 3
Performance of subsampling intervals for δ. Constant term: $m_t = 10$.

DGP	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}
$S(\delta)$	$n = 100$				$n = 200$				$n = 500$			
1 – $S(0)$	0.982	0.085	−0.613	0.720	0.984	0.072	−0.523	0.618	0.987	0.061	−0.443	0.515
2 – $S(1)$	0.896	0.838	0.232	1.339	0.885	0.878	0.346	1.322	0.882	0.894	0.453	1.286
3 – $S(2)$	0.698	1.608	0.971	2.208	0.792	1.655	0.996	2.262	0.860	1.715	1.065	2.337
4 – $S(0.5)$	0.970	0.420	−0.424	1.185	0.969	0.443	−0.329	1.132	0.967	0.455	−0.171	1.039
5 – $S(1.5)$	0.752	1.208	0.378	1.956	0.788	1.266	0.506	1.957	0.814	1.305	0.624	1.920
6 – $S(1)$	0.981	0.775	−0.108	1.542	0.992	0.805	−0.020	1.555	0.999	0.822	0.049	1.515
7 – $S(0.7)$	0.970	0.582	−0.092	1.160	0.976	0.609	0.041	1.099	0.979	0.608	0.145	1.021
8 – $S(0.5)$	0.825	0.091	−0.594	0.736	0.707	0.071	−0.540	0.606	0.544	0.059	−0.442	0.524
9 – $S(0.5)$	0.985	0.398	−0.365	1.102	0.986	0.420	−0.259	1.041	0.986	0.443	−0.167	0.964
10 – $S(1)$	0.910	0.856	0.018	1.568	0.911	0.897	0.146	1.594	0.900	0.915	0.242	1.513
11 – $S(1)$	0.812	0.602	−0.134	1.291	0.831	0.667	0.008	1.278	0.841	0.711	0.123	1.271
12 – $S(0.5)$	0.943	0.525	−0.032	1.019	0.923	0.538	0.075	0.934	0.922	0.539	0.182	0.853

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. $\bar{\delta}^*$ represents the mean of the estimated orders of summability. I_{low} and I_{up} are the median of the lower and upper bounds of the intervals, respectively. 1000 replicas are used. Subsample size is $b = \sqrt{n}$.

Table 4
Performance of subsampling intervals for δ. Linear trend: $m_t = 10 + 2t$.

DGP	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}
$S(\delta)$	$n = 100$				$n = 200$				$n = 500$			
1 – $S(0)$	0.933	0.282	−0.428	0.927	0.949	0.264	−0.359	0.831	0.953	0.228	−0.292	0.703
2 – $S(1)$	0.918	0.817	0.176	1.380	0.907	0.834	0.271	1.327	0.900	0.872	0.391	1.289
3 – $S(2)$	0.788	1.581	0.811	2.285	0.854	1.637	0.889	2.328	0.931	1.705	0.989	2.363
4 – $S(0.5)$	0.958	0.504	−0.274	1.174	0.965	0.501	−0.194	1.106	0.956	0.499	−0.098	1.028
5 – $S(1.5)$	0.726	1.096	0.329	1.816	0.755	1.144	0.433	1.818	0.799	1.198	0.539	1.790
6 – $S(1)$	0.973	0.727	−0.151	1.477	0.982	0.750	−0.058	1.464	0.997	0.795	0.033	1.473
7 – $S(0.7)$	0.978	0.616	−0.057	1.214	0.986	0.613	0.032	1.123	0.989	0.642	0.152	1.052
8 – $S(0.5)$	0.928	0.283	−0.429	0.929	0.912	0.273	−0.336	0.846	0.814	0.233	−0.280	0.726
9 – $S(0.5)$	0.985	0.456	−0.312	1.131	0.988	0.451	−0.220	1.080	0.991	0.467	−0.141	1.023
10 – $S(1)$	0.849	0.748	−0.047	1.436	0.858	0.770	0.055	1.411	0.865	0.805	0.150	1.393
11 – $S(1)$	0.794	0.621	−0.113	1.279	0.803	0.654	−0.030	1.254	0.832	0.707	0.076	1.281
12 – $S(0.5)$	0.928	0.559	−0.008	1.065	0.929	0.554	0.093	0.972	0.900	0.574	0.209	0.885

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. $\bar{\delta}^*$ represents the mean of the estimated orders of summability. I_{low} and I_{up} are the median of the lower and upper bounds of the intervals, respectively. 1000 replicas are used. Subsample size is $b = \sqrt{n}$.

5. Empirical application

After Nelson and Plosser (1982) accounted for unit root behavior in almost all the fourteen US macroeconomic time series in their database, many researchers have used the same dataset to confirm or refuse their conclusions with alternative approaches. In what follows, we contribute to this literature by applying the above developed methodology to estimate and infer the order of summability of the time series included in an extended version of the Nelson and Plosser (1982) database.[6] As a novelty, we do not impose any linearity assumption.

More precisely, we estimate the order of summability of the fourteen macroeconomic aggregates with $\hat{\delta}^* = (\hat{\beta}^* - 1)/2$ and derive the subsampling confidence intervals denoted by (I_L^*, I_U^*). It is well known in the literature that deterministic components are an important issue for these time series. Since the order of the deterministic trend is unknown, we propose to use in practice a traditional graphical device. If a trending behavior is observed, include *at least* a linear trend. If the time series evolve around a constant, consider *at least* a constant term. Using this device and knowing that it is always better to subtract a higher order trend than a lower one with respect to the true order, a quadratic trend has been considered for all the variables but interest and unemployment rates. Results are shown in Table 5.

Observe that the variable with the lowest order of summability is the unemployment rate and the one with the highest the consumer price index. On the other hand, variables like nominal and real GNP, stock of money, wages, industrial production or S&P share similar orders of summability, around one. The amplitude of the confidence intervals is in line with the wide confidence intervals reported in Stock (1991) for the largest autoregressive root and in Arteche and Orbe (2005) for the fractional order of integration. Notice that our methodology does not assume any model for the data.

Table 5
Order of summability. Estimation and inference.

log(variable)	Order of summability		
Quadratic trend	$\hat{\delta}^*$	I_L^*	I_U^*
Consumer price index	2.369	1.112	3.625
Employment	0.579	0.185	0.973
GNP deflator	0.900	0.168	1.631
Nominal GNP	1.031	0.557	1.505
Industrial production	0.738	0.082	1.393
GNP per capita	0.938	0.278	1.599
Real GNP	0.898	0.287	1.510
Wages	0.961	0.341	1.580
Real wages	1.070	0.320	1.821
S&P	0.702	0.121	1.283
Money	0.913	0.279	1.548
Velocity	0.576	−0.010	1.163
Linear trend	$\hat{\delta}^*$	I_L^*	I_U^*
Interest	0.934	0.359	1.509
Unemployment	0.162	−0.603	0.928

$\hat{\delta}^*$ denotes the estimated order of summability. I_L^* and I_U^* denote the lower and upper bounds of the corresponding subsampling intervals.

6 The data have been downloaded from P.C.B. Phillips' webpage.

Overall, the estimated orders of summability of the fourteen macroeconomic variables seem to be quite reasonable in economic and econometric terms. Regarding the latter aspect of the empirical exercise, we would like to highlight the similarities of our results with those found in the fractional literature. With respect to the economic content of the results, as already stated, variables like real and nominal GNP, industrial production, or nominal money have similar orders of summability and higher than those of unemployment or velocity of money. Additionally, in a heuristic way, it can be seen that these results do not go against the quantity theory of money.

6. Conclusion

Time Series Econometrics has not yet been able to properly handle non-linearities with persistent variables. This is mainly due to the fact that the concept of integration, and consequently co-integration, is too linear and not always well defined for non-linear processes. This lack of a proper definition has two important multivariate consequences. First, it is not possible to characterize the balancedness of a non-linear postulated model relating persistent variables. This is a necessary condition for an appropriate model specification. Second, co-integration cannot be directly extended to analyze non-linear long run relationships. The concept of summability is able to solve these problems. This paper shows how to calculate, estimate, and undertake inference on the order of summability, δ.

Acknowledgments

The authors benefited from the helpful comments of J.J. Dolado, G. Elliott, C.W.J. Granger, S. Johansen, S.J. Leybourne, O. Linton, B. Nielsen, A. Pagan, P.C.B. Phillips, D. Politis, P. Robinson, M.H. Seo, W. Stute, R. Taylor, C. Velasco, two anonymous referees and the participants at the 18th Annual Symposium of the SNDE, Nottingham Sir Clive Granger Memorial Conference, 10th World Congress of the Econometric Society, 2010 NBER-NSF Time Series Conference and the 2010 European Winter Meetings of the Econometric Society. Financial support from SEJ-2007-63098, ECO-2010-19357, CONSOLIDER 2010 (CSD 2006-00016), and EXCELECON S-2007/HUM-044 grants is gratefully acknowledged.

Appendix

Proof of Proposition 1. Applying the Beveridge–Nelson decomposition as in Phillips and Solo (1992)

$$\Delta^d y_t = C(1)\, u_t + \tilde{u}_{t-1} - \tilde{u}_t,$$

with

$$\tilde{u}_t = \tilde{C}(L)\, u_t = \sum_{j=0}^{\infty} \tilde{c}_j L^j u_t = \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} c_k u_{t-j}.$$

Then,

$$y_t = C(1)\, \Delta^{-d} u_t + \Delta^{-d}\left(\tilde{u}_{t-1} - \tilde{u}_t\right),$$

and

$$\frac{1}{n^{1/2+d}} \sum_{t=1}^{n} y_t = C(1) \frac{1}{n^{1/2+d}} \sum_{t=1}^{n} \Delta^{-d} u_t - \frac{1}{n^{1/2+d}} \Delta^{-d} \tilde{u}_n.$$

Let $\kappa(n,d)$ be the slowly varying function defined in Liu (1998)

$$\kappa(n,d) = \begin{cases} \dfrac{\sigma_u^2 \Gamma(1-2d)}{(1+2d)\,\Gamma(1+d)\,\Gamma(1-d)} & \text{if } d \geq 0 \text{ and} \\ \quad d \neq \dfrac{2k+1}{2}\ \forall k \in \mathbb{N} & \\ \dfrac{\sigma_u^2}{\pi} \log n \quad \text{if } d = \dfrac{2k+1}{2}\ \forall k \in \mathbb{N} \cup \{0\}, & \end{cases}$$

and consider

$$\frac{1}{n^{1/2+d}} \kappa(n,d)^{-1/2} \sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+d}} \kappa(n,d)^{-1/2} \sum_{t=1}^{n} \Delta^{-d} u_t - \kappa(n,d)^{-1/2} \frac{1}{n^{1/2+d}} \Delta^{-d} \tilde{u}_n. \tag{26}$$

Boundedness in probability of the first component of the right hand side of (26) is shown in Liu (1998). Boundedness in probability of the second term can be shown by noticing that

$$Var\left[\frac{1}{n^{1/2+d}} \Delta^{-d} \tilde{u}_n\right] = \frac{1}{n^{1+2d}} Var\left[\Delta^{-d} \tilde{u}_n\right] = o(1).$$

In particular, for $d = 0$,

$$Var\left[\frac{1}{n^{1/2}} \Delta^{-d} \tilde{u}_n\right] = \frac{1}{n} Var\left[\tilde{u}_n\right] = \frac{1}{n} \sigma_u^2 \sum_{s=0}^{\infty} \tilde{c}_s^2 = o(1),$$

since $\sum_{j=1}^{\infty} j\left|c_j\right| < \infty$. Finally, for $d > 0$, given that $\Delta^{-d} = \sum_{i=0}^{\infty} a_i L^i$ with $a_i = O\left(i^{d-1}\right)$, we have

$$\begin{aligned} Var\left[\frac{1}{n^{1/2+d}} \Delta^{-d} \tilde{u}_n\right] &= \frac{1}{n^{1+2d}} Var\left[\Delta^{-d} \tilde{u}_n\right] \\ &= \frac{1}{n^{1+2d}} Var\left[\sum_{i=0}^{\infty} a_i \tilde{u}_{n-i}\right] \\ &= \frac{1}{n^{1+2d}} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i a_j Cov\left[\tilde{u}_{n-i}, \tilde{u}_{n-j}\right] \\ &= \frac{1}{n^{1+2d}} \sum_{i=0}^{\infty} a_i \sum_{j=0}^{\infty} a_j \sum_{s=0}^{\infty} \tilde{c}_s \sum_{p=0}^{\infty} \tilde{c}_p \sigma_u^2 1(i-j=p-s) = o(1), \end{aligned}$$

where $1(i-j=p-s)$ is the corresponding indicator function. Hence, $y_t \sim S(\delta)$. □

Proof of Proposition 2. The sum of y_t is

$$\sum_{t=1}^{n} y_t = C(1) \sum_{t=1}^{n} \Delta^d u_t - \Delta^d \tilde{u}_n = A_n - B_n,$$

where $A_n = C(1) \sum_{t=1}^{n} \Delta^d u_t$ and $B_n = \Delta^d \tilde{u}_n$. By definition of $\tilde{u}_t$,

$$B_n = \Delta^d \tilde{u}_n = O_p(1),$$

for all $d = 1, 2, \ldots < \infty$. With respect to A_n note that,

$$C(1) < \infty,$$

and

$$\sum_{t=1}^{n} \Delta^d u_t = \Delta^{d-1} \sum_{t=1}^{n} \Delta u_t = \Delta^{d-1} u_n = O_p(1),$$

for all $d = 1, 2, \ldots < \infty$. Therefore,

$$A_n = C(1) \sum_{t=1}^{n} \Delta^d u_t = O_p(1),$$

as well. And, all together implies that

$$\sum_{t=1}^{n} y_t = A_n - B_n = O_p(1),$$

or equivalently that $y_t \sim S(-0.5)$. □

Proof of Proposition 3. By Assumption 2 and definition of summable process, U_k is $O_p(1)$. Hence, Theorem 3.1 in McElroy and Politis (2007) applies. □

Proof of Proposition 4. Expression (14) can be rewritten as

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

The denominator satisfies

$$\frac{1}{n\log^2 n}\sum_{k=1}^{n}\log^2 k\to 1\quad\text{as } n\to\infty.$$

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,

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

hence,

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

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

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

Proof of Proposition 5. From (22) and (23)

$$y_t-\hat{m}_t=y_t-\frac{1}{t}\sum_{j=1}^{t}y_j=x_t-\frac{1}{t}\sum_{j=1}^{t}x_j.$$

Then,

$$\begin{aligned}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}(y_t-\hat{m}_t)&=\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(x_t-\frac{1}{t}\sum_{j=1}^{t}x_j\right)\\&=\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_t\left(1-\sum_{j=t}^{n}\frac{1}{j}\right)\\&\approx\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_t\left(1+\log(t/n)\right),\end{aligned}$$

where "≈" means that both terms are of the same asymptotic order in probability. Now,

$$\begin{aligned}&P\left(\left|\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_t(1+\log(t/n))\right|>M_\varepsilon\right)\\&\leq P\left(\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}|x_t|+\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}|x_t||\log(t/n)|>M_\varepsilon\right)\\&\leq P\left(\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}|x_t|>\frac{M_\varepsilon}{2}\right)\\&\quad+P\left(\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}|x_t||\log(t/n)|>\frac{M_\varepsilon}{2}\right)\\&<\varepsilon,\end{aligned}$$

where $M_\varepsilon\in(0,\infty)$ and $\varepsilon>0$. The last inequality is obtained from the stated assumption on $|x_t|$ and the properties of $\log(t/n)$. Hence,

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}(y_t-\hat{m}_t)=O_p(1).$$

By construction of $\hat{m}_t$, last expression is not $o_p(1)$. Therefore, $(y_t-\hat{m}_t)$ is $S(\delta)$. □

Proof of Proposition 6. Let us consider the following five steps:

(i) First, the partial mean is computed

$$\frac{1}{t}\sum_{j=1}^{t}y_j=m_0+m_1\frac{1}{t}\sum_{j=1}^{t}j+\frac{1}{t}\sum_{j=1}^{t}x_j.$$

(ii) Second, the partial mean is subtracted from y_t

$$\begin{aligned}y_t-\frac{1}{t}\sum_{j=1}^{t}y_j&=m_1t+x_t-m_1\frac{1}{t}\sum_{j=1}^{t}j-\frac{1}{t}\sum_{j=1}^{t}x_j\\&=m_1t-m_1\frac{1}{t}\frac{t(t+1)}{2}+x_t-\frac{1}{t}\sum_{j=1}^{t}x_j\\&=\frac{m_1}{2}(t-1)+x_t-\frac{1}{t}\sum_{j=1}^{t}x_j.\end{aligned}$$

(iii) Third, compute

$$\begin{aligned}\frac{2}{t}\sum_{j=1}^{t}\left(y_j-\frac{1}{j}\sum_{i=1}^{j}y_i\right)&=\frac{2}{t}\sum_{j=1}^{t}\left(\frac{m_1}{2}(j-1)+x_j-\frac{1}{j}\sum_{i=1}^{j}x_i\right)\\&=\frac{m_1}{2}(t-1)+\frac{2}{t}\sum_{j=1}^{t}x_j-\frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j}x_i.\end{aligned}$$

(iv) Fourth, subtracting the quantity obtained in step (iii) from that obtained in step (ii)

$$\begin{aligned}&y_t-\frac{1}{t}\sum_{j=1}^{t}y_j-\frac{2}{t}\sum_{j=1}^{t}\left(y_j-\frac{1}{j}\sum_{i=1}^{j}y_i\right)\\&=x_t-\frac{3}{t}\sum_{j=1}^{t}x_j+\frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j}x_i.\end{aligned}$$

(v) Finally,

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(y_t-\frac{1}{t}\sum_{j=1}^{t}y_j-\frac{2}{t}\sum_{j=1}^{t}\left(y_j-\frac{1}{j}\sum_{i=1}^{j}y_i\right)\right)$$

$$=\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(x_t-\frac{3}{t}\sum_{j=1}^{t}x_j+\frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j}x_i\right)$$

$$=\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_t\left(1-3\sum_{j=t}^{n}\frac{1}{j}+2\sum_{j=t}^{n}\frac{1}{j}\sum_{i=t}^{j}\frac{1}{i}\right).$$

The result follows from similar arguments as those in the proof of Proposition 5. □

References

Arteche, J., Orbe, J., 2005. Bootstrapping the log-periodogram regression. Economics Letters 86, 79–85.

Berenguer-Rico, V., 2013. Co-summability and NLLS regressions: generalizing co-integration to nonlinear models. Mimeo.

Berenguer-Rico, V., Gonzalo, J., 2013. Co-summability: from linear to non-linear co-integration. UC3M Working Paper.

Berkes, I., Horváth, L., 2006. Convergence of integral functionals of stochastic processes. Econometric Theory 22, 304–322.

Davidson, J., 2009. When is a time series $I(0)$? In: Jennifer, Castle, Neil, Shepherd (Eds.), The Methodology and Practice of Econometrics, a Festschrift for David F. Hendry. Oxford University Press.

de Jong, R.M., 2004. Addendum to "Asymptotics for nonlinear transformations of integrated time series". Econometric Theory 20, 627–635.

de Jong, R.M., Wang, Ch.H., 2005. Further results on the asymptotics for non-linear transformations of integrated time series. Econometric Theory 21, 413–430.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimator for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Embrechts, P., Klüppelberg, C., Mikosh, T., 1999. Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York.

Franses, P.H., van Dijk, D., 2000. Non-linear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge.

Gonzalo, J., Pitarakis, J.Y., 2006. Threshold effects in co-integrating regressions. Oxford Bulletin of Economics & Statistics 68, 813–833.

Granger, C.W.J., 1995. Modelling non-linear relationships between extended-memory variables. Econometrica 63, 265–279.

Granger, C.W.J., Hallman, J., 1991. Non-linear transformations of integrated time series. Journal of Time Series Analysis 12, 207–224.

Granger, C.W.J., Teräsvirta, T., 1993. Modelling Non-linear Economic Relationships. Oxford University Press, Oxford.

Johansen, S., 1991. Estimation and hypothesis testing of co-integration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

Johansen, S., 1995. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

Karlsen, H.A., Myklebust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. The Annals of Statistics 35, 252–299.

Leybourne, S.J., McCabe, B.P.M., Tremayne, A.R., 1996. Can economic time series be differenced to stationarity? Journal of Business and Economic Statistics 14, 435–446.

Liu, M., 1998. Asymptotics of non-stationary fractional integrated series. Econometric Theory 14, 641–662.

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

Müller, U.K., 2008. The impossibility of consistent discrimination between $I(0)$ and $I(1)$ processes. Econometric Theory 24, 616–630.

Nelson, Ch.R., Plosser, Ch.I., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. Journal of Monetary Economics 10, 139–162.

Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4, 468–497.

Park, J.Y., Phillips, P.C.B., 1999. Asymptotics for non-linear transformations of integrated time series. Econometric Theory 15, 269–298.

Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. Journal of Econometrics 33, 311–340.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. The Annals of Statistics 20, 971–1001.

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.

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

Romano, J.P., Siegel, A.F., 1986. Counterexamples in Probability and Statistics. Wadsworth and Brooks/Cole, Monterey, California.

Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in US macroeconomic time series. Journal of Monetary Economics 28, 435–459.

Teräsvirta, T., Tjøstheim, D., Granger, C.W.J., 2010. Modelling Nonlinear Economic Time Series. Oxford University Press, Oxford.

Wang, Q., Phillips, P.C.B., 2009. Structural nonparametric cointegrating regressions. Econometrica 77, 1901–1948.

White, J.S., 1958. The limiting distribution of the serial correlation coefficient in the explosive case. Annals of Mathematical Statistics 29, 1188–1197.

Yoon, G., 2005. An introduction to $I(\infty)$ processes. Economic Modelling 22, 473–483.