Summability of Stochastic Processes

A Generalization of Integration and Co-Integration valid for Non-linear Processes

by Vanessa Berenguer Rico and Jesús Gonzalo

Universidad Carlos III de Madrid

Very preliminary version

(Please do not distribute without permission yet)

December 10, 2009

Abstract

The order of integration is valid to characterize linear processes; but it is not appropriated 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 regressions; (iv) co-summability represents a generalization of co-integration for non-linear processes. To make this concept empirically applicable asymptotic properties of estimation and inference methods for the degree of summability, $\delta$, are provided.

Keywords: Integrated Processes; Non-linear Balanced Regressions; Non-linear Processes; Summability.

JEL classification: C01; C22.
1 Introduction

The concept of integrability has been widely used during the last decades in the time series literature. In the seventies, after Box and Jenkins (1970), it was a common practice to differentiate the time series until make them stationary. The possible existence of stochastic trends in the data generating processes of macroeconomic variables was one of the major area of research. To this respect, the Dickey-Fuller (1979) test statistic became quite popular being usually applied to test for unit roots. Nelson and Plosser (1982) has been one of the most influential works reporting results on the presence of stochastic trends or unit roots behavior in almost fourteen of the most important U.S. macroeconomic time series.

Linear co-integration was the multivariate counterpart of the integrability concept reconciling the unit roots evidence with the existence of equilibrium relationships advocated by the economic theory. Introduced by Granger (1983) and Engle and Granger (1987), it generated a huge amount of research, being highlighted, among others, the works by Phillips (1986) –giving theoretical and asymptotic explanations to some unexplained and related facts–, and Johansen (1991) –formalizing the system approach to co-integration.

On the other hand, in economic theory terms, it is difficult to justify that some economic variables, like unemployment rates or interest rates, are driven by unit roots. Hence, fractional roots were also putted into play. It has been proved that fractional orders of integration capture the persistence of long memory processes –see for instance, Granger and Joyeux (1980). Moreover, the aggregation process was a theoretical justification for fractional orders of integration to be used. Not only in an univariate framework fractional integration was considered, also fractional co-integration was introduced –see Granger (1986). After fractional integration and co-integration appeared, lot of work has been devoted to this area.

In parallel, non-linear time series models from a stationary perspective were introduced in the literature – see Granger and Terasvirta (1993) or Franses and van Dijk (2003) for some overviews. More recently, the next step has been to study non-linear transformations of integrated processes, see, for instance, Park and Phillips (1999), de Jong (2001), de Jong and Wang (2005) or Pötscher (2004). Natural queries like the order of integration of these non-linear transformations appear in this context. However, such a question does not have a clear answer since the existing definitions of integrability do not properly apply. 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, like the order of integration, is not available to characterize non-linear time series. This does not only affect econometricians, but also economic theorists who
cannot neglect important properties of actual economic variables when choosing functional forms to construct their theories. Second, from a multivariate perspective, it becomes troublesome to determine whether a non-linear regression is or not balanced. Unbalanced equations are related to the familiar problems of spurious relations and misspecification, which are greatly enhanced when managing non-linear functions of variables having a persistency property. In linear setups, the concept of integrability did a good job dealing with balanced/unbalanced relations. However, in non-linear frameworks, the nonexistence of a synoptic quantitative measure makes it difficult, for a set of related variables, to estimate and test this relation with a balanced equation, i.e. with a well specified regression 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 \( y_t = f(x_t, \theta_0) + u_t \), where \( x_t \sim I(1) \), \( u_t \sim I(0) \). For \( f(\cdot) \) non-linear, the order of integration of \( y_t \) is not properly defined implying that the standard concept of co-integration is difficult to be applied. In fact, it was already stated in Granger and Hallman (1991) that a generalization of linear co-integration to a non-linear setup goes through proper extensions of the linear concepts of I(0) and I(1). 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 on hand. 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 and the order of summability is the same as the order of integration. Hence, summability can be understood as 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 regression – linear or not. And maybe more important, non-linear long run equilibrium relationships between non-stationary time series can be properly defined. In particular, we show how the concept of co-summability can be applied to extend co-integration to non-linear setups.

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 study two estimators proposed in McElroy and Politis (2007). Given their asymptotic properties, we finally work only with one of these two estimators. The inference on the true order of summability is based on the subsampling methodology developed in Politis, Romano and Wolf (1999). Although a particular mixing condition required for the use of subsampling is difficult to verify in this context –and right now is beyond the scope of this paper–, we show, by simulations, that the subsampling machinery works quite well when trying to determine the order of summability of an observed time series. We would like to remark that since integrated time series are particular cases of summable stochastic processes, these econometric tools can also be seen as new procedures to estimate and test for the order of integration, integer or fractional. In addition, we also show that this machinery can be used to determine whether a non-linear regression involving non-stationary time series is spurious or specifies a non-linear long run relationship. Finally, an empirical application illustrates how to use in practice the proposed methodology.

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 to empirically deal with summable processes in applications. In addition, we show, in Section 5, that these tools can also be used to determine whether a non-linear regression is spurious or specifies a non-linear long run relationship. In Section 6, the use of the proposed tools is shown with an empirical application. Finally, Section 7 is devoted to some concluding remarks.

2 Order of Integration and Non-linear Processes

In this section, we highlight the applicability problems of the concept of order integration to non-linear models. First, we start recalling some of the definitions of I(0) that the literature has used emphasizing the complications that set in. Second, we show that these definitions cannot be used to determine the order of integration of some relevant univariate time series. And third, and maybe more important, the multivariate implications of such lack of a proper definition for non-linear models are addressed.

2.1 Definitions

Definition 1: A time series $y_t$ is called an integrated process of order $d$ (in short, an I($d$) process) if the time series of $d$th order differences $\Delta^d y_t$ is stationary (an I(0) process).
A natural question that arises after reading this definition is: and what is an \textit{I}(0) process? Attempts to give a definition for \textit{I}(0) processes exists in the literature. Engle and Granger (1987) give the following characterization.

\textbf{Characterization 1:} 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 series between crossing of \( x = 0 \) is finite; (v) the autocorrelations, \( \rho_k \), decrease steadily in magnitude for large enough \( k \), so that their sum is finite.

Trying to model non-linear relationships between extended-memory variables, Granger (1995) gives two different definitions for an \textit{I}(0) process, the theoretical and the practical:

\textbf{Characterization 2:} \textit{Theoretical Definition of \textit{I}(0):} A process is \textit{I}(0) if it is strictly stationary and has a spectrum bounded above and below away from zero at all frequencies.

\textbf{Characterization 3:} \textit{Practical Definition of \textit{I}(0):} \( x_t \) is \textit{I}(0) if it is generated by a stationary autoregressive model \( a(B)x_t = e_t \), where \( e_t \) is zero mean white noise and the roots of the autoregressive polynomial \( a(B) \) are outside the unit circle.

Johansen (1995) defined an \textit{I}(0) as follows.

\textbf{Characterization 4:} A stochastic process \( y_t \) which satisfies \( y_t - E(y_t) = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i} \), with \( \varepsilon_t \sim i.i.d.(0,\sigma^2_{\varepsilon}) \), is called \textit{I}(0) if \( \sum_{i=0}^{\infty} C_i z^i \) converges for \( |z| < 1 \) and \( \sum_{i=0}^{\infty} C_i \neq 0 \).

Therefore, in practical terms, an \textit{I}(0) process can be understood as a second order linear process.

\textbf{Definition 2:} A stochastic process \( y_t \) which satisfies

\[
x_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad C(L) = \sum_{j=0}^{\infty} c_j L^j,
\]

is called \textit{I}(0) if

\[
\sum_{j=0}^{\infty} c_j^2 < \infty,
\]

\( \varepsilon_t \) is i.i.d. with zero mean and \( \sigma^2_{\varepsilon} = E(\varepsilon^2_0) < \infty \).

As stated in Davidson (1999) "it is clear that \textit{I}(0), as commonly understood, is a property of linear models. Let’s state this observation more forcefully: \textit{I}(0), in this framework, is not a property of a time series, but a property of a model. This characterization must give increasing difficulties in view of the numerous generalizations of co-integration now being investigated, which embrace long memory, non-linear and nonparametric approaches to time series modelling. [...] There is a need for a definition that is not model dependent, but describes an objective property of a time series".

With these arguments, Davidson (1999) uses the idea that an \textit{I}(0) process is the first difference of an \textit{I}(1) and gives the following definition.
Definition 3: A time series $y_t$ is $I(0)$ if the process $Y_n$ defined on the unit interval by $Y_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{[n\xi]}(y_t - E(y_t))$, $0 < \xi \leq 1$ where $\sigma_n^2 = \text{Var}(\sum_{t=1}^{n} y_t)$, converges weakly to standard Brownian motion $B$ as $n \to \infty$.

In other words, the standardized partial sums of the series must satisfy a functional central limit theorem (FCLT). As commented in Davidson (1999), "naturally enough, there is plenty of scope for disagreement about Definition 3. For one thing, many people would expect the $I(0)$ class to include any i.i.d. sequence. An i.i.d. sequence of Cauchy variates, for example, fails the weak convergence test. [...] Similarly, we note that Brownian motion is only one member of a class of Gaussian limit processes to which the partial sums can converge under different assumptions".

Although researchers have devoted many efforts in defining an integrated process, still problems remain when trying to apply the existing definitions to some models. We consider the following examples.

2.2 Examples

Example 1: Alpha stable distributed processes

An equally alpha stable distributed process is strictly stationary. However, its first and second moments do not exist. The fact that such a process is identically distributed could incline us to think that this process is $I(0)$. However, this example does not satisfy any of the characterizations or definitions of $I(0)$ given above because of the inexistence of moments.

Example 2: An i.i.d. plus a random variable

Consider the following process

$$y_t = z + e_t,$$

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

(i) $E[y_t] = 0$

(ii) $V[y_t] = \sigma_e^2 + \sigma_z^2$

(iii) $\gamma(k) = \text{Cov}(y_t, y_{t-k}) = \sigma_z^2$ 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 above
characterizations of an $I(0)$ process. Moreover, it cannot be $I(0)$ as described in Definitions 2 and 3.

If $y_t$ is not $I(0)$, to attach any other order of integration to this stochastic process is not obvious. It cannot be an $I(1)$ process since its first difference is not $I(0)$, in fact, it is $I(-1)$. And it becomes difficult to choose any other number with the above given definitions of integrability.

Dealing with non-linear processes we face similar problems. We consider the following examples.

**Example 3 : Product of i.i.d. and random walk**

Let us consider the following process

$$w_t = x_t \eta_t,$$

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

$$x_t = x_{t-1} + \varepsilon_t,$$

with $\varepsilon_t \sim i.i.d. (0, \sigma^2)$. Some properties of $w_t$ are

(i) $E[w_t] = 0$

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

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

It should be not obvious to attach an order of integration to this process. On one hand, the uncorrelation property (iii) could incline us to think that $w_t$ is $I(0)$. However, an $I(0)$ cannot have a trend in the variance according to the above characterizations. On the other hand, this unbounded

1 The autocovariance of the processes in this example can be expressed as

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

Hence, the spectral density is

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

which diverges for all $\lambda$.

2 Assume that $y_t$ is $I(0)$ as described in Definition 2. Then, $y_t = c(L)\varepsilon_t$, where $\varepsilon_t$ is iid. Moreover, the following alternative autoregressive representation exists, $a(L)y_t = \varepsilon_t$, with $a(L) = c(L)^{-1}$. Equivalently, $\varepsilon_t = a(L)z + a(L)\varepsilon_t$, which is a correlated process. But this is a contradiction, therefore, the initial assumption that the process is $I(0)$ must not be true.

Moreover, it cannot be $I(0)$ as described in Definition 3, since

$$Y_n(\xi) = \sigma^{-1}_n \sum_{t=1}^{[n\xi]} (y_t - E(y_t)) = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{(n\sigma_z^2 + \sigma^2)}} \sum_{t=1}^{[n\xi]} (z + \varepsilon_t) \sim B.$$
variance could induce to suspect that the process is $I(1)$. However, its first difference

$$\Delta w_t = x_t \eta_t - x_{t-1} \eta_{t-1},$$

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

$$V[\Delta w_t] = E[(x_t \eta_t)^2] + E[(x_{t-1} \eta_{t-1})^2] - 2E[x_t x_{t-1} \eta_t \eta_{t-1}] = (2t - 1) \sigma^2.$$

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

$$V[\Delta^2 w_t] = E[(x_t \eta_t)^2] + 4E[(x_{t-1} \eta_{t-1})^2] + E[(x_{t-2} \eta_{t-2})^2] = 6(t-1) \sigma^2.$$

In fact, this process can be thought as having an infinite order of integration, in the sense that, the variance of $\Delta^d w_t$ depends on $t$ regardless of the values of $d$ – see, for instance, Yoon (2005).

Therefore, although,

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

any of the definitions above can be strictly used to determine the order of integration of $w_t$, given the behavior of its variance along time. Usually, bounded second moments are required to speak about $I(0)$ time series. And, dependence, although very important, is not the only property describing the behavior of a time series. Heterogeneous distributions – specially when the heterogeneity is prominent – are also important to characterize the evolution of a stochastic process. Volatility along time is fundamental, particularly, in economic time series. In some way, a concept like the order of integration should measure such trending evolution of the variance differencing it from the one of an $I(0)$ process.

In addition, non-linear transformations of highly heterogeneous or volatile processes, although uncorrelated, induce high correlations, as we show with the following example.

**Example 4 : Product of i.i.d. squared and random walk**

Consider the following process

$$q_t = x_t \eta_t^2,$$

where $x_t$ and $\eta_t$ were described in the previous example. The only difference with Example 3 is that now the i.i.d. sequence follows a chi-squared distribution. However, in this case,

$$E[q_t] = E[x_t \eta_t^2] = 0,$$
\[ V[\eta_t] = E[\eta_t^2] = E[x_t^2 \eta_t^4] = E[x_t^2]E[\eta_t^4] = t \sigma_x^2 \mu_4, \]

and

\[ \gamma_q(h) = E[\eta_t \eta_{t-h}] = E[x_t x_{t-h} \eta_t^2 \eta_{t-h}^2] = E[x_t x_{t-h}]E[\eta_t^2 \eta_{t-h}^2] = (t-h) \sigma_x^2 \sigma_\eta^4, \]

where \( E[\eta_t^2] = \mu_4 \). This means that not only the variances if not also the covariances depend on time. Hence, we can see how non-linear transformations of highly heterogenous processes can have an important impact on its stochastic properties. And this impact will be hardly contemplated by the order of integration.

**Example 5 : Square of a random walk**

Consider now the square of the random walk defined in equation (2), that is,

\[ x_t^2 = \varepsilon_t^2 + 2x_{t-1}\varepsilon_t + x_{t-1}^2. \]

To establish the order of integration of this process is again not an obvious task. Granger (1995) showed that \( x_t^2 \) can be seen as a random walk with drift, hence, one could think that \( x_t^2 \) is also \( I(1) \). However, although its first difference

\[ x_t^2 - x_{t-1}^2 = \varepsilon_t^2 + 2x_{t-1}\varepsilon_t, \]

is not correlated,

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

Again any of the above characterizations or definitions of \( I(0) \) can be applied.

**Example 6 : A stochastic unit root process**

A stochastic unit root process, in short STUR, is a simple non-linear time series model defined as follows

\[ y_t = (1 + \eta_t)y_{t-1} + \varepsilon_t, \]

where \( \eta_t \sim i.i.d.(0, \sigma_\eta^2) \) and \( \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2) \). Assume that \( \eta_t \) and \( \varepsilon_t \) are independent each other. Given that \( E[\eta_t] = 0, y_t \) has a unit root only on average. Yoon (2006) showed that a STUR process is strictly stationary and has no finite moments. Taking a characterization of long memory based on the variance of the partial sum, Yoon (2003) shows that STUR processes can be confused with an \( I(1.5) \) process, although they are strictly stationary. Again the order of integration of such a process is not obvious.

**Example 7 : Product of indicator function and random walk**

\[ V[\eta_t] = E[\eta_t^2] = E[x_t^2 \eta_t^4] = E[x_t^2]E[\eta_t^4] = t \sigma_x^2 \mu_4, \]

and

\[ \gamma_q(h) = E[\eta_t \eta_{t-h}] = E[x_t x_{t-h} \eta_t^2 \eta_{t-h}^2] = E[x_t x_{t-h}]E[\eta_t^2 \eta_{t-h}^2] = (t-h) \sigma_x^2 \sigma_\eta^4, \]

where \( E[\eta_t^2] = \mu_4 \). This means that not only the variances if not also the covariances depend on time. Hence, we can see how non-linear transformations of highly heterogenous processes can have an important impact on its stochastic properties. And this impact will be hardly contemplated by the order of integration.

**Example 5 : Square of a random walk**

Consider now the square of the random walk defined in equation (2), that is,

\[ x_t^2 = \varepsilon_t^2 + 2x_{t-1}\varepsilon_t + x_{t-1}^2. \]

To establish the order of integration of this process is again not an obvious task. Granger (1995) showed that \( x_t^2 \) can be seen as a random walk with drift, hence, one could think that \( x_t^2 \) is also \( I(1) \). However, although its first difference

\[ x_t^2 - x_{t-1}^2 = \varepsilon_t^2 + 2x_{t-1}\varepsilon_t, \]

is not correlated,

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

Again any of the above characterizations or definitions of \( I(0) \) can be applied.

**Example 6 : A stochastic unit root process**

A stochastic unit root process, in short STUR, is a simple non-linear time series model defined as follows

\[ y_t = (1 + \eta_t)y_{t-1} + \varepsilon_t, \]

where \( \eta_t \sim i.i.d.(0, \sigma_\eta^2) \) and \( \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2) \). Assume that \( \eta_t \) and \( \varepsilon_t \) are independent each other. Given that \( E[\eta_t] = 0, y_t \) has a unit root only on average. Yoon (2006) showed that a STUR process is strictly stationary and has no finite moments. Taking a characterization of long memory based on the variance of the partial sum, Yoon (2003) shows that STUR processes can be confused with an \( I(1.5) \) process, although they are strictly stationary. Again the order of integration of such a process is not obvious.

**Example 7 : Product of indicator function and random walk**
Consider the following process

\[ h_t = 1(v_t \leq \gamma)x_t, \]

where \( v_t \) is i.i.d., \( 1(\cdot) \) is the usual indicator function, and \( x_t \) is the random walk defined in (2). It is another example where the concept of integrability is difficult to apply. Its variance and covariances depend on time, hence, one would think that \( h_t \) is \( I(1) \). However, the first difference of this process is not \( I(0) \) as described in the definitions given above since

\[
V[\Delta h_t] = V[1(v_t \leq \gamma)x_t - 1(v_{t-1} \leq \gamma)x_{t-1}]
= E[(1(v_t \leq \gamma)x_t)^2 + 2(1(v_t \leq \gamma)1(v_{t-1} \leq \gamma)x_t x_{t-1}) + (1(v_{t-1} \leq \gamma)x_{t-1})^2]
= [2p(1-p)\sigma_x^2]t + p(2p - 1)\sigma_x^2.
\]

In fact, it can be considered, once again, that \( h_t \) has an infinite order of integration.

In all these examples the concept of integrability is difficult to use. And a conclusion from these considerations is that the standard \( I(d) \) classifications are not sufficient to handle several situations.

\section*{2.3 Multivariate Implications}

This lack of a proper definition for non-linear univariate time series translates to multivariate relationships. First, it cannot be determined whether a non-linear regression is balanced or not. And second, a generalization of the standard concept of co-integration to non-linear relationships is not straightforward.

To clarify these two issues, consider the following model

\[ y_t = \theta f(x_t) + u_t, \]

where \( x_t \sim I(1), u_t \sim I(0) \). As we have highlighted above, the order of integration of \( f(x_t) \) and, hence, of \( y_t \) cannot be characterized.

With respect the first implication, note that if the order of integration of \( f(x_t) \) is not properly defined, then, it is not possible to use the order of integration to determined whether this regression model is balanced. As stated in Granger (1995), an equation will be called balanced if the major properties of the endogenous variable are available amongst the right-hand side explanatory variables and there are no unwanted strong properties on that side. Balanced regressions are a necessary—although not sufficient—condition for a good specification. Hence, the question of balance is related to the familiar concept of misspecification. Moreover, non-linear functions of variables with a persistency property will enhance the opportunities for unbalanced regression, as Granger (1995) showed. Therefore, a first step in the estimation of a regression model—linear or not—should be devoted to determine the balancedness of the corresponding regression.
Balancedness of a regression opens the door to long run equilibrium relationships. However, as the second implication states, the standard concept of co-integration cannot be applied for many interesting functions \( f \). Even assuming that the order of integration of the errors is zero, the order of the observable variables in the model cannot be characterized. This invalidates a direct extension of the linear concept of co-integration to non-linear relationships.

All these issues makes necessary to extend the concept of integratedness to allow for more general types of processes and some authors have proposed new concepts. Among others, Granger (1995) proposed to use the concepts of Extended and Short Memory in Mean defined as follows.

**Definition 4**: \( y_t \) will be called short memory in mean (abbreviated as SMM) if the conditional \( h \)-step forecast in mean

\[
f_{t,h} = E[y_{t+h} | I_t], \quad h > 0,
\]


tends to a constant \( m \) as \( h \) becomes large. Thus, as one forecasts into the distant future, the information available in \( I_t \) comes progressively less relevant. More formally, using a squared norm, \( y_t \) is SMM if

\[
E \left[ |f_{t,h} - m|^2 \right] < c_h,
\]

where \( c_h \) is some sequence that tends to zero as \( h \) increases.

**Definition 5**: If \( y_t \) is not SMM, so that \( f_{t,h} \) is a function of \( I_t \) for all \( h \), it will be called "extended memory in mean", denoted EMM. Thus, as one forecasts into the distant future, values known at present will generally continue to be helpful.

Granger (1995) gave a way to quantify the order of memory of a SMM process as follows. If \( c_h \) in Definition 4 is such that \( c_h = O(h^{-\theta}) \), \( \theta > 0 \), then the process under study can be said to be SMM of order \( \theta \). Nevertheless, a way to establish the order of EMM is not available. Even so, other authors have used the SMM and EMM concepts. For instance, Gourieroux and Jasiak (1999) denoted SMM and EMM by non-linear integrated (NLI) and non-linear integrated of order zero (NLI(0)), respectively. And Escanciano and Escribano (2008) proposed the pairwise equivalent measures of the previous concepts. However, for some DGPs the conditional forecast could be difficult to obtain. And hence, SMM and EMM are neither easy to calculate nor general enough to handle some types of non-linear long run relationships.
3 A Solution Based on Summability

In this section, we propose to use the concept of order of summability. Specifically, we start by giving a formal definition. Then, we show that the order of summability of the processes considered in Examples 1-7 can be determined without difficulty. Additionally, we prove that any $I(d)$ stochastic process is $S(d)$. Finally, summability is applied to solve the multivariate problems of balanced regressions and co-integration in non-linear frameworks.

3.1 Order of Summability Definition

Dealing with threshold effects in co-integrating regressions, Gonzalo and Pitarakis (2006) faced the applicability problems of the order of integration, SMM, and EMM concepts. To solve these problems, they defined the order of summability without exploiting its potential scope. As it will be seen, it is a simple, useful, and general idea that gives more insights with respect to the degree of memory and variance structure than the previous concepts of Short and Extended Memory. Besides, summability is closely related to the limiting properties of the sum of the process under study. Hence, it is very advantageous in knowing the type of statistics that can be used to estimate and carry out inferences on some population peculiarities of the process. In addition, and maybe eventually more important, summability allows to deal with more general balanced/unbalanced regressions given the measurability of the degree of summability. This, in turn, implies to be able to easily extend the study of linear long run equilibrium relationships to non-linear environments, preserving the main features of the original concept of co-integration. Knowing the statistical properties of the stochastic processes and those of the models that relate them, the appropriate statistical procedures can be chosen in order to estimate and test postulated theoretical relationships. As it will be shown, summability allows to carry out this task without relying on linear or unchanging long run relationships.

The definition of summability given by Gonzalo and Pitarakis (2006) was as follows.

**Definition 6**: A time series $y_t$ is said to be summable of order $\delta$, symbolically represented as $S(\delta)$, if the sum $S_n = \sum_{t=1}^{n} (y_t - E[y_t])$ is such that $S_n/n^{1+\delta} = O_p(1)$ as $n \to \infty$.

Since any $o_p(1)$ process is also $O_p(1)$, given Definition 6, a time series can have an infinite number of orders of summability. Additionally, if $E[y_t]$ does not exist, as it is the case, for instance, of the Cauchy distribution, then Definition 6 cannot be applied.

To skip these pitfalls but still with the same spirit, we propose the following slightly different definition.
Definition 7: Summability of order \( \delta \): A time series \( y_t \) is said to be summable of order \( \delta \), symbolically represented as \( S(\delta) \), if there exist a nonrandom sequence \( m_t \) such that

\[
S_n = \frac{1}{n^{1+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1) \quad \text{as } n \to \infty,
\]

where \( \delta \) is the minimum number that makes \( S_n \) bounded in probability\(^3\) and \( L(n) \) is a slowly-varying function\(^4\).

Definition 7 is just a correction and a generalization of Definition 6 for summability of stochastic processes. By taking \( \delta \) to be the minimum number that makes \( S_n \) bounded in probability, we avoid the problem of an infinite number of orders of summability. By considering a general \( m_t \), we can also allow for processes without first moments. And, introducing the slowly varying function \( L(n) \), we can consider more general normalizing factors.

Note that, when possible, the order of summability will be determined by some Central Limit result. In the i.i.d. CLT, for instance, \( \delta = 0 \) and \( L(n) \) is just a constant, the inverse of the standard deviation of the time series. When the time series is a standard random walk, the FCLT will establish that \( \delta = 1 \) and \( L(n) \) is again a constant term, the inverse of the standard deviation of the innovations. Although, in many circumstances \( L(n) \) will be constant, in some situations the asymptotic theory will enforce us to use an \( L \) function varying with \( n \) but slowly in the Karatama’s sense.

Summability allows for a huge variety of processes. For instance, an \( I(0) \) process is always \( S(0) \). And, the usual \( I(1) \) process, a random walk, is \( S(1) \). These are common processes studied in the literature for which the concept of integratedness works well. But, as we remarked above, the concept of integrability does not apply to some models. And it is here where the concept of summability starts to become a useful device.

3.2 Examples

From an univariate perspective, in all processes considered in Examples 1-7 the order of integration was difficult to establish. Next, we show that the order of summability can be directly obtained for all of the above examples.

\(^3\) \( S_n \) is said to be bounded in probability if, for every \( \epsilon > 0 \), there exists a positive real number \( M_\epsilon \) such that \( P(|S_n| \geq M_\epsilon) \leq \epsilon , \) for all \( n \).

\(^4\) A positive measurable function \( L \), defined on some neighbourhood \( [0, \infty) \) of infinity, and satisfying

\[
\frac{L(\lambda n)}{L(n)} \to 1 \quad (n \to \infty) \quad \forall \lambda > 0,
\]

is said to be slowly varying (in Karatama’s sense).
**Summability in Example 1** (\(\alpha\)-stable distributed process): For an \(\alpha\)-stable Levy distributed process, \(y_t\), it can be shown that when the Levy distribution is symmetric with \(0 < \alpha \leq 2\) the normalized sum
\[
S_n = \frac{1}{n^\delta} \sum_{t=1}^{n} y_t,
\]
converges to a Levy distribution. Hence, in this case the time series is said to be summable of order \(\delta = (2 - \alpha)/2\alpha\) with \(L(n) = 1\). For a Cauchy distribution \(\alpha = 1\), which implies that a Cauchy distributed process is \(S(0.5)\). When \(\alpha = 2\), the \(y_t\) are normally distributed, hence,
\[
S_n = \frac{1}{n^2} \sum_{t=1}^{n} y_t = O_p(1).
\]
That is, \(y_t\) is \(S(0)\) in this case\(^5\).

**Summability in Example 2** (A white noise plus a random variable): It is easy to see that
\[
S_n = \frac{1}{n^{1+\delta}} \sum_{t=1}^{n} y_t = \frac{1}{n^{1+\delta}} \sum_{t=1}^{n} (z + e_t) = \frac{1}{n^{-\frac{1}{2}+\delta}} z + \frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} e_t,
\]
which converges for \(\delta = 0.5\), that is,
\[
S_n = z + \frac{1}{n} \sum_{t=1}^{n} e_t \implies z.
\]
Therefore, \(y_t\) is \(S(0.5)\) and \(L(n) = 1\) in this case.

In Examples 3-7, we considered several non-linear models. Next, we show that the order of summability of those processes can be easily determined.

**Summability in Example 3** (Product of i.i.d. and random walk): It can be shown –see for instance, Park and Phillips (1988)– that
\[
S_n = \frac{1}{\sigma_x n} \sum_{t=1}^{n} x_t \eta_t \implies \int_0^1 W_1(r) dW_2(r).
\]
This means that \(w_t\) is \(S(0.5)\) with \(L(n) = 1/\sigma_x\).

**Summability in Example 4** (Product of i.i.d. squared and random walk): For \(q_t\) we have,
\[
Var \left[ \sum_{t=1}^{n} x_t \eta_t^2 \right] = O(n^3).
\]
\(^5\)Consider the case where the process \(y_t\) have 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
\[
\frac{1}{[n \log n]^{1/2}} \sum_{t=1}^{n} y_t \implies N(0, 1).
\]
This is a case where \(L(n) = (1/ \log n)^{1/2}\), and not just a constant.
Then, by the Chebyshev’s inequality, we know that
\[ \frac{1}{n^{3/2}} \sum_{t=1}^{n} x_t \eta_t^2 = O_p(1). \]
Hence, \( q_t \) is \( S(1) \). Comparing Examples (3) and (4), we can see that summability is taken into account not only the covariance structure if not also the variance behavior along time.

**Summability in Example 5** (*Squared of a random walk*): For the square root of a random walk, it is also well known that
\[ S_n = \frac{1}{n^{2+\delta}} \sum_{t=1}^{n} x_t^2, \]
converges when \( \delta = 1.5 \). Specifically,
\[ S_n = \frac{1}{n^2 \sigma^2} \sum_{t=1}^{n} x_t^2 \rightarrow \int_0^1 W^2(r)dr. \]
Hence, we can conclude that \( x_t^2 \) is \( S(1.5) \).

**Summability in Example 6** (*A STUR process*): As demonstrated in Yoon (2003), the variance of the partial sums of the STUR models considered in Example 6 grows at a rate corresponding to an \( I(1.5) \) process. Therefore,
\[ S_n = \frac{1}{n^2} \sum_{t=1}^{n} y_t = O_p(1). \]
Hence, STUR processes are \( S(1.5) \).

**Summability in Example 7** (*Product of indicator function and random walk*): In this case,
\[ S_n = \frac{1}{n^2 \sigma^2} \sum_{t=1}^{n} h_t \Rightarrow \int_0^1 W(r)dr, \]
meaning that \( h_t \) is \( S(1) \) with \( L(n) = 1/\sigma^2 \).

These examples show that some processes have not well defined its order of integration, but they have a well established order of summability. The later concept allows studying the stochastic properties of non-stationary time series without imposing linear structures. Moreover, the order of summability keeps the original idea of measuring the memory, dependence, or persistence but giving a richer characterization of its degree than other existing concepts in the literature. Additionally, integrated time series are particular cases of summable processes as we show next.

### 3.3 Integrability implies Summability

In this subsection, we discuss the relationship between integrability and summability. For the former, we will use Definition 2 and, for the latter, Definition 7. Definition 2 can be considered as
the most general practical definition of an $I(0)$ process. Furthermore, assuming that,

$$\sum_{j=0}^{\infty} j^2 c_j^2 < \infty,$$

Definition 2 allows us to easily show the relation between integrated and summable processes as follows.

**Proposition 1**: Let $d \geq 0$. If a time series is $I(d)$, then it is $S(d)$.

**Proof**: We will divide the proof in four steps.

(i) $d = 0$. Using the Beveridge-Nelson decomposition as in Phillips and Solo (1992) the linear process described in Definition 2 can be expressed as

$$x_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t,$$

where

$$\tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t = \sum_{j=0}^{\infty} \tilde{c}_j \varepsilon_{t-j}, \quad \tilde{c}_j = \sum_{k=j+1}^{\infty} c_k.$$

Now, the sum of $x_t$ can be computed as

$$\sum_{t=1}^{n} x_t = C(1) \sum_{t=1}^{n} \varepsilon_t + \tilde{\varepsilon}_0 - \tilde{\varepsilon}_n,$$

and the following CLT holds

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t \overset{d}{\rightarrow} N(0, \sigma^2 C(1)^2),$$

-see Phillips and Solo (1992). This implies that $x_t$ is $S(0)$. And hence, every $I(0)$ process is $S(0)$.

(ii) $d = d_0 \in (0,1/2)$. It was shown in Hosking (1996) that if a time series $x_t \sim I(d)$ with $d \in (0,1/2)$, then it satisfies the following CLT

$$n^{\alpha/2}(\bar{x} - \mu) \quad \Rightarrow \quad N\left(0, \frac{2\lambda}{(1 - \alpha)(2 - \alpha)}\right),$$

where $\alpha = 1 - 2d$ and $\lambda$ is determined by the powerlaw decay of the autocovariance function of $x_t$.

(iii) For the case $d = 1/2$, we can use Theorem 2.2. of Liu (1998). In such theorem, it is shown that

$$\frac{\sqrt{2}}{K^{1/2} n \log^{1/2} n} S_T(s) \overset{d}{\rightarrow} sB(1),$$

where $S_T(s)$ is the $D$-space analog of the partial sum process of an $I(1/2)$ time series. Note that in this case $L(n) = \sqrt{2}/K^{1/2} \log^{1/2} n$.

(iv) For $d > 1/2$ we can apply Theorem 2.3 in Liu (1998). **Q.E.D.**
Remark 1: Proposition 1 tells us that $I(d)$ processes are $S(d)$ when $d$ is a positive real number. The same is not true when the order of integration is negative as we show in the following proposition.

Proposition 2: If a process is $I(-d)$, $d = 1, 2, ..., < \infty$, then it is $S(-0.5)$.

Proof: Let $x_t$ be $I(0)$. The sum of its $d$ difference is

$$\sum_{t=1}^{n} \Delta^d x_t = C(1) \sum_{t=1}^{n} \Delta^d \varepsilon_t + \Delta^d \varepsilon_0 - \Delta^d \varepsilon_n.$$

Now,

$$\Delta^d \varepsilon_0 - \Delta^d \varepsilon_n = O_p(1),$$

by definition of $\varepsilon_t$. With respect the component

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

note that,

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

By Definition 2, and

$$\sum_{t=1}^{n} \Delta^d \varepsilon_t = \Delta^{d-1} \sum_{t=1}^{n} \Delta \varepsilon_t = \Delta^{d-1} (\varepsilon_n - \varepsilon_0) = O_p(1),$$

for all $d = 1, 2, ..., < \infty$. Q.E.D.

Remark 2: Since negative integers of integration are not the most important and/or interesting ones, we will consider only $d \geq 0$ and conclude that integrated processes are particular cases of summable processes.

3.4 Balancedness and Co-summability

As we have seen, the concept of summability overcomes the pitfalls that appear when trying to establish the order of integration of some non-linear transformations of integrated processes. It gives a measure of the degree of persistence as well as of the evolution of the variance of stochastic processes along time. Furthermore, the order of summability generalizes the idea of order of integration in the sense that integrated time series can be seen as particular cases of summable processes. But, maybe more important, the concept of summability controls the balancedness of non-linear regressions and generalizes the standard concept of co-integration to non-linear long run relationships.
Definition 8: A regression model of the form

\[ y_t = \theta f(x_t) + u_t, \]

will be said to be balanced if the order of summability of \( y_t \) is the same as the order of summability of \( z_t = f(x_t) \).

Once the balancedness of a non-linear regression is established, to speak about non-linear long run relationships can be done using the concept of co-summability.

Definition 9: Two \( S(\delta) \) stochastic processes, \( y_t \) and \( z_t = f(x_t) \), with \( \delta > 0 \), will be said to be co-summable if there exists a constant \( \theta \) such that \( u_t = y_t - \theta f(x_t) \) is \( S(\delta_0) \), with \( \delta > \delta_0 \). In short, \((y_t, z_t) \sim CS(\delta, \delta - \delta_0)\).

Co-summable processes will share an equilibrium relationship in the long run, i.e. an attractor \( y_t = \theta f(x_t) \) that can be linear or not. This type of equilibrium relationships will be usually established by the economic theory and have interesting econometric applications that include, for instance, transition behavior between regimes, multiplicity of equilibria, or non-linear responses to intervention policies. Applied researchers will be interested in estimating and testing for those type of equilibria.

4 Summability in practice

In this section, we propose econometric tools to empirically estimate and infer unknown the order of summability of observed time series in applications. Firstly, two estimators of the order of summability are studied. Given its asymptotic properties, we advice to use only one of them. Secondly, a subsampling inference methodology is analyzed showing, by means of simulations, that it behaves reasonably well.

4.1 Order of summability estimation

If a stochastic process, \( y_t \), satisfies

\[ S_n = \frac{1}{n^{1+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1), \]

we say that \( y_t \) is summable of order \( \delta \). For a \( \delta \)-summable stochastic process, it should be true that

\[ S_n^2 = \left( \frac{1}{n^{1+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) \right)^2 = n^{-1-2\delta} L^2(n) \left( \sum_{t=1}^{n} (y_t - m_t) \right)^2 = O_p(1). \]
Taking logs

\[ U_n = \log S_n^2 = \log \left( n^{-(1+2\delta)} L^2(n) \left( \sum_{t=1}^{n} (y_t - \mu_t) \right)^2 \right) = O_p(1). \] (5)

Equation (5) can be written as

\[ U_n = -(1 + 2\delta) \log n + 2 \log L(n) + \log \left( \sum_{t=1}^{n} (y_t - \mu_t) \right)^2 \]

\[ = -\beta \log n + 2 \log L(n) + \log T_k \]

\[ = -\beta \log n - \alpha + Y_n, \]

where \( \beta = 1 + 2\delta, \alpha = -2 \log L(n), \) and \( Y_n = \log T_n. \) In regression model form,

\[ Y_n = \alpha + \beta \log n + U_n, \] (6)

with \( U_n = O_p(1). \)

Following McElroy and Politis (2007), we propose to estimate \( \beta \) with

\[ \hat{\beta}_1 = \frac{\sum_{t=1}^{n} Y_t \log t}{\sum_{t=1}^{n} \log^2 t} = \beta + \frac{\sum_{t=1}^{n} (\alpha + U_t) \log t}{\sum_{t=1}^{n} \log^2 t}, \] (7)

or

\[ \hat{\beta}_2 = \frac{\sum_{t=1}^{n} (Y_t - \bar{Y})(\log t - \log n)}{\sum_{t=1}^{n} (\log t - \log n)^2} = \beta + \frac{\sum_{t=1}^{n} (U_t - \bar{U})(\log t - \log n)}{\sum_{t=1}^{n} (\log t - \log n)^2}, \] (8)

and \( \delta \) with

\[ \hat{\delta} = \frac{\hat{\beta}_t - 1}{2}. \]

Since \( \log L(t) = o(\log t), \) \( \alpha = -2 \log L(n) \) can be treat as "approximately constant" – see McElroy and Politis (2007) for details. In fact, in order to keep things simple, we will assume in the following that \( \alpha \) is constant. In fact, this is the case in all the examples we considered.

4.2 Asymptotic Properties

In this section, we study the asymptotic properties of \( \hat{\beta}_1 \) and \( \hat{\beta}_2. \) Let \( x_t = y_t - m_t. \) We start focusing on

\[ \hat{\beta}_1 - \beta = \frac{\sum_{t=1}^{n} V_t \log t}{\sum_{t=1}^{n} \log^2 t}, \]

with \( V_t = \alpha + U_t. \)

**Proposition 3**: \( \hat{\beta}_1 - \beta = o_p(1). \)

**Proof**. \( U_t \) is \( O_p(1) \) by definition of summable processes. Hence, Theorem 3.1. in McElroy and Politis (2007) applies. Q.E.D.
Remark 3. McElroy and Politis (2007) show that $\hat{\beta}_1$ is consistent under minimal assumptions. In our context, these assumptions are satisfied by construction and by definition of summable processes. Nonetheless, an asymptotic distribution for $\hat{\beta}_1$ and $\hat{\beta}_2$ has not been derived. We address this issue in the following, in the context of estimating the order of summability.

Proposition 4: If

$$S_n = \frac{1}{n^{1/2+\delta}} L(n) \sum_{t=1}^{n} x_t \implies D(r,\delta),$$

where $D(r,\delta)$ is some random variable or process with positive variance, then

$$\log n(\hat{\beta}_1 - \beta) \implies \alpha + \int_{0}^{1} U(r,\delta) dr,$$

where $U(r,\delta) = \log \left( D(r,\delta)^2 \right)$.

Proof:

The denominator satisfies

$$\frac{1}{n \log^2 n} \sum_{t=1}^{n} \log^2 t \to 1,$$

hence, we write

$$\hat{\beta}_1 - \beta = \frac{\frac{1}{n \log^2 n} \sum_{t=1}^{n} V_t \log t}{\frac{1}{n \log^2 n} \sum_{t=1}^{n} \log^2 t} = \frac{\frac{1}{n \log^2 n} \sum_{t=1}^{n} (\alpha + U_t) \log t}{\frac{1}{n \log^2 n} \sum_{t=1}^{n} \log^2 t}$$

$$= \frac{1}{n \log^2 n} \sum_{t=1}^{n} \log t + \frac{1}{n \log^2 n} \sum_{t=1}^{n} U_t \log t.$$

In order to derive an asymptotic distribution we concentrate on

$$U_n = \log \left( n^{-\beta} \left( \sum_{t=1}^{n} x_t \right)^2 \right)$$

$$= \log \left( S_n^2 \right),$$

where

$$S_n = \frac{1}{n^{1/2+\delta}} \sum_{t=1}^{n} x_t.$$

By assumption it is true that

$$S_n = \frac{1}{n^{1/2+\delta}} \sum_{t=1}^{n} x_t \implies D(r,\delta),$$

Hence,

$$U_n = \log \left( S_n^2 \right) \implies \log(D(r,\delta)^2).$$
Now, consider the following processes

\[
S_{[nr]}(\delta) = \frac{1}{[nr]^{1/2+\delta}} \sum_{t=1}^{[nr]} x_t = \begin{cases} 
0 & \text{for } 0 \leq r < \frac{1}{n} \\
\frac{x_1}{1/2+\delta} = S_1 & \text{for } \frac{1}{n} \leq r < \frac{2}{n} \\
\frac{x_1+x_2}{2/2+\delta} = S_2 & \text{for } \frac{2}{n} \leq r < \frac{3}{n} , \\
\vdots & \\
\frac{x_1+x_2+\ldots+x_n}{n/2+\delta} = S_n & \text{for } r = 1
\end{cases}
\]

and

\[
U_{[nr]}(\delta) = \log S_{[nr]}^2(\delta) = \begin{cases} 
0 & \text{for } 0 \leq r < \frac{1}{n} \\
\log \left( \frac{x_1}{1/2+\delta} \right)^2 = U_1 & \text{for } \frac{1}{n} \leq r < \frac{2}{n} \\
\log \left( \frac{x_1+x_2}{2/2+\delta} \right)^2 = U_2 & \text{for } \frac{2}{n} \leq r < \frac{3}{n} , \\
\vdots & \\
\log \left( \frac{x_1+x_2+\ldots+x_n}{n/2+\delta} \right)^2 = U_n & \text{for } r = 1
\end{cases}
\]

where \([\cdot]\) denotes the integer part and \(r = t/n\). Note that

\[
S_{[nr]}(\delta) = \frac{1}{[nr]^{1/2+\delta}} \sum_{t=1}^{[nr]} x_t \Rightarrow D(r, \delta),
\]

and

\[
U_{[nr]}(\delta) \Rightarrow \log(D(r, \delta)^2) \equiv U(r, \delta).
\]

Hence, by the Continuous Mapping Theorem

\[
\int_0^1 S_{[nr]}(\delta) d\delta = \frac{1}{n} \sum_{t=1}^n S_t \Rightarrow \int_0^1 D(r, \delta) d\delta,
\]

\[
\int_0^1 U_{[nr]}(\delta) d\delta = \frac{1}{n} \sum_{t=1}^n U_t \Rightarrow \int_0^1 U(r, \delta) d\delta.
\]

Now, we concentrate on the numerator of \(\log n(\beta_1 - \beta)\), specifically in,

\[
\frac{1}{\log n} \sum_{t=1}^n U_t \log t.
\]

It can be written as

\[
\frac{1}{\log n} \sum_{t=1}^n U_t \log t = \frac{1}{\log n} \sum_{t=1}^n U_t \left( \log \left( \frac{t}{n} \right) + \log n \right)
= \frac{1}{\log n} \left( \frac{1}{n} \sum_{t=1}^n U_t \log r \right) + \frac{1}{n} \sum_{t=1}^n U_t.
\]

Now,

\[
\frac{1}{n} \sum_{t=1}^n U_t \log r = \int_0^1 \log r U_{[nr]}(\delta) d\delta \Rightarrow \int_0^1 \log r U(r, \delta) d\delta,
\]
hence, \[ \frac{1}{\log n} \left( \frac{1}{n} \sum_{t=1}^{n} U_t \log r \right) = o_p(1). \]

Therefore, \[ \frac{1}{n \log n} \sum_{t=1}^{n} U_t \log t = \frac{1}{n} \sum_{t=1}^{n} U_t + o_p(1) \implies \int_0^1 U(r, \delta) dr. \]

This implies that

\[
\log n(\hat{\beta}_1 - \beta) = \frac{\frac{1}{n \log n} \sum_{t=1}^{n} (\alpha + U_t) \log t}{\frac{1}{n \log^2 n} \sum_{t=1}^{n} \log^2 t} = \frac{\alpha \cdot \frac{1}{n \log n} \sum_{t=1}^{n} \log t}{\frac{1}{n \log^2 n} \sum_{t=1}^{n} \log^2 t} + \frac{1}{n \log n} \sum_{t=1}^{n} U_t \log t \implies \alpha + \int_0^1 U(r, \delta) dr.
\]

Q.E.D.

**Remark 4:** When the series under study is i.i.d.\((0,1)\), for instance, the classical CLT applies, that is

\[ S_n = \frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t \implies N(0,1). \]

which has \(r = 1\) and \(\delta = 0\). Moreover, in this case,

\[ \log n(\hat{\beta}_1 - \beta) \implies \log (\chi^2_1). \]

Similarly if the time series that we consider was a standard random walk, then

\[ S_n = \frac{1}{n^{3/2}} \sum_{t=1}^{n} x_t \implies \int_0^1 W(r) dr, \]

and

\[ \log n(\hat{\beta}_1 - \beta) \implies \int_0^1 U(r, 1) dr, \]

where \(r \in [0,1]\), \(W(r)\) is a Wiener Process, and \(U(r, 1) = \log \left( \int_0^1 W(r) dr \right)^2 \).

**Remark 5:** By Propositions 3 and 4, we know that \(\hat{\beta}_1\) is log \(n\)-consistent. However, the asymptotic distribution will depend on the nuisance parameter \(\alpha\), unless \(\alpha = 0\), i.e. \(L(n) = 1\). In the following, we study the asymptotic properties of

\[ \hat{\beta}_2 = \frac{\sum_{t=1}^{n} (Y_t - \bar{Y})(\log t - \log n)}{\sum_{t=1}^{n} (\log t - \log n)^2}, \]

and

\[ \hat{\alpha} = \bar{Y} - \hat{\beta}_2 \log n. \]
Proposition 5: If
\[ S_n = \frac{1}{n^{1/2+\delta}} L(n) \sum_{t=1}^{n} x_t \Rightarrow D(r, \delta), \]

where \( D(r, \delta) \) is some random variable or process with positive variance, then
\[ (\hat{\beta}_2 - \beta) \Rightarrow \int_0^1 (1 + \log r) U(r, \delta) dr, \]

where \( U(r, \delta) = \log (D(r, \delta)^2) \).

Proof: The denominator in this case satisfies
\[ \frac{1}{n} \sum_{t=1}^{n} (\log t - \log n)^2 \rightarrow 1. \]

With respect the numerator, note that
\[ \frac{1}{n} \sum_{t=1}^{n} U_t (\log t - \log n) = \frac{1}{n} \sum_{t=1}^{n} U_t \log(t) - \frac{1}{n} \log n \sum_{t=1}^{n} U_t \]
\[ = \frac{1}{n} \sum_{t=1}^{n} U_t \left( \log \left( \frac{t}{n} \right) + \log n \right) - \left( \frac{1}{n} \sum_{t=1}^{n} \left( \log \left( \frac{t}{n} \right) + \log n \right) \right) \left( \frac{1}{n} \sum_{t=1}^{n} U_t \right) \]
\[ = \frac{1}{n} \sum_{t=1}^{n} U_t \log \left( \frac{t}{n} \right) - \left( \frac{1}{n} \sum_{t=1}^{n} \log \left( \frac{t}{n} \right) \right) \left( \frac{1}{n} \sum_{t=1}^{n} U_t \right) . \]

Note that
\[ \left( \frac{1}{n} \sum_{t=1}^{n} \log \left( \frac{t}{n} \right) \right) \rightarrow \left( \int_0^1 \log r dr \right) = -1, \]

hence,
\[ \frac{1}{n} \sum_{t=1}^{n} U_t (\log t - \log n) \Rightarrow \int_0^1 \log r U(r, \delta) dr - \left( \int_0^1 \log r dr \right) \int_0^1 U(r, \delta) dr \]
\[ = \int_0^1 (1 + \log r) U(r, \delta) dr. \]

Q.E.D.

Remark 6: When the time series under study is i.i.d.(0,1), for instance,
\[ (\hat{\beta}_2 - \beta) \Rightarrow \int_0^1 (1 + \log r) U(1, 0) dr = U(1, 0) \int_0^1 (1 + \log r) dr = 0, \]

showing that \( \hat{\beta}_2 \) is a consistent estimator of the true \( \beta \). However, if the process we consider was a random walk, then
\[ (\hat{\beta}_2 - \beta) \Rightarrow \int_0^1 (1 + \log r) U(r, 1) dr, \]

where \( U(r, 1) = \log \left( \left( \int_0^1 W(r) dr \right)^2 \right) \) and \( W(r) \) is a Wiener Process, loosing the consistency of \( \hat{\beta}_2 \) in this case.
Proposition 6: If

\[ S_n = \frac{1}{n^{1/2+\delta}} L(n) \sum_{t=1}^{n} x_t \implies D(r, \delta), \]

where \( D(r, \delta) \) is some random variable or process with positive variance, then

\[ \frac{1}{\log n} (\hat{\alpha} - \alpha) \implies -\int_0^1 (1 + \log r) U(r, \delta) dr, \]

where \( U(r, \delta) = \log (D(r, \delta)^2) \).

Proof: The OLS estimator of the constant term, \( \alpha \), is

\[ \hat{\alpha} = \bar{Y} - \hat{\beta}_2 \log n = \frac{1}{n} \sum_{t=1}^{n} Y_t - \hat{\beta}_2 \frac{1}{n} \sum_{t=1}^{n} \log t = \frac{1}{n} \sum_{t=1}^{n} (\alpha + \beta \log t + U_t) - \hat{\beta}_2 \frac{1}{n} \sum_{t=1}^{n} \log t \]

\[ = \alpha + \frac{1}{n} \sum_{t=1}^{n} U_t - \left( \hat{\beta}_2 - \beta \right) \frac{1}{n} \sum_{t=1}^{n} \log t. \]

Hence,

\[ \hat{\alpha} - \alpha = \frac{1}{n} \sum_{t=1}^{n} U_t - \left( \hat{\beta}_2 - \beta \right) \frac{1}{n} \sum_{t=1}^{n} \log t, \]

which satisfies

\[ \frac{1}{\log n} (\hat{\alpha} - \alpha) = \frac{1}{\log n} \frac{1}{n} \sum_{t=1}^{n} U_t - \left( \hat{\beta}_2 - \beta \right) \frac{1}{n} \log n \sum_{t=1}^{n} \log t = - \left( \hat{\beta}_2 - \beta \right) \frac{1}{n} \log n \sum_{t=1}^{n} \log t + o_p(1) \]

\[ \implies -\int_0^1 (1 + \log r) U(r, \delta) dr. \]

Q.E.D.

Remark 7: Proposition 6 shows that the constant term in the regression model (6) cannot be consistently estimated by OLS. Moreover, by Proposition (5), the OLS estimator of the slope will not be consistent for all orders of summability. Therefore, we incline towards the use of \( \hat{\beta}_1 \). Remember, however, that when a constant must be introduced in the regression model, \( \hat{\beta}_1 \) is still consistent but the asymptotic distribution depends on the unknown nuisance parameter \( \alpha \). In order to get rid of the nuisance parameter, we propose to estimate, instead of

\[ Y_t = \alpha + \beta \log t + U_t, \]

the following modified regression model

\[ Y_t^* = \beta \log t + U_t^*, \]

where \( Y_t^* = Y_t - Y_1 \) and \( U_t^* = U_t - U_1 \). Hence, the modified OLS estimator

\[ \hat{\beta}_1^* = \frac{\sum_{t=1}^{n} Y_t^* \log t}{\sum_{t=1}^{n} \log^2 t}, \]

satisfies the same asymptotic properties than \( \hat{\beta}_1 \) when \( L(n) = 1 \). That is, it will be \( \log n \)-consistent with an asymptotic distribution which does not depend on the nuisance parameter \( \alpha \).
Remark 8: Up to now, we have assumed that \( m_t \) in \( x_t = y_t - m_t \) is known. However, in practice, we do not observe \( m_t \). As we next show, a proper estimator \( \hat{m}_t \) could be used instead.

Assumption 1. \( \hat{m}_t \) is an estimator of \( m_t \) such that

\[
\hat{m}_t = m_t + \nu_t,
\]

with

\[
\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} \nu_t = O_p(1).
\]

Proposition 7: Let \( y_t \) to satisfy (4). Under Assumption 1

\[
\hat{S}_n = \frac{1}{\frac{1}{2}+\delta} \sum_{t=1}^{n} (y_t - \hat{m}_t) = O_p(1).
\]

Proof: Under Assumption 1,

\[
\hat{S}_n = \frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - \hat{m}_t) = \frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - m_t - \nu_t) = \frac{1}{\frac{1}{2}+\delta} \sum_{t=1}^{n} (y_t - m_t) - \frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} \nu_t = O_p(1).
\]

Hence, the result holds. Q.E.D.

Remark 9: Proposition 7 says that to estimate \( \delta \) with \( \hat{\delta} \) when \( m_t \) is unknown, we just need to find an estimator \( \hat{m}_t \) satisfying Assumption 1. It is worth mentioning that such assumption is quite weak since it does not even require consistency of \( \hat{m}_t \). The condition

\[
\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} \nu_t = O_p(1),
\]

does not imply in general that \( \nu_t = o_p(1) \).

Remark 10: Because of Propositions 3 and 7, Assumption 1 is enough to guarantee consistency of \( \hat{\beta}_1^* \) when \( m_t \) is substituted by \( \hat{m}_t \). Because of Proposition 4, in order to get an asymptotic distribution of \( \hat{\beta}_1^* \) when \( m_t \) is unknown, we should assume instead of Assumption 1,

\[
\frac{1}{n^{1/2+\delta}} \sum_{t=1}^{n} \nu_t \Rightarrow D_{\nu}(r, \delta),
\]

where \( D_{\nu}(r, \delta) \) is some asymptotic limit. Under this condition, Proposition 4 still holds if we replace \( m_t \) by \( \hat{m}_t \). A deeper analysis of particular choices of \( \hat{m}_t \) for several \( m_t \) will be carry out below.
4.3 Subsampling Confidence Intervals

Although the asymptotic distribution of $\hat{\beta}_1$ does not depend on the variance of the errors driving the stochastic process under consideration, it depends on the true order of summability. Hence, a unique limiting distribution is not available. However, subsampling methods could be an alternative way to carry out inferences on the order of summability independently of its true value.

The subsampling methodology is consistent under minimal assumptions. The most general result shown in Politis, Romano and Wolf (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

(iii) $\log b / \log n \to 0$, $b/n \to 0$, $b \to \infty$, and $n^{-1} \sum_{h=1}^{n} \alpha_{n,b}(h) \to 0$ as $n \to \infty$, where $\alpha_{n,b}(h)$ are the $\alpha$-mixing coefficients sequence of $Z_{n,b,t} = \log b (\hat{\beta}_{n,b,t} - \beta)$.

Since the $\alpha$-mixing conditions required for the subsampling to be consistent are difficult to verify in this framework, we show its validity with a simulation study.

4.3.1 Without Deterministic Components

The DGPs that we consider next are the following ones:

| DGP 1: $x_{1t} = \varepsilon_t \sim i.i.d. N(0, 1)$ | DGP 6: $x_{6t} = \Delta^{0.3} \left( \sum_{j=1}^{t} \varepsilon_j \right)$ |
| DGP 2: $x_{2t} = \sum_{j=1}^{t} \varepsilon_j$ | DGP 7: $x_{7t} = z + \varepsilon_t$, $z \sim N(0, 1)$ |
| DGP 3: $x_{3t} = \sum_{j=1}^{t} \sum_{i=1}^{j} \varepsilon_i$ | DGP 8: $x_{8t} = \eta_t \left( \sum_{j=1}^{t} \varepsilon_j \right)$, $\eta_t \sim i.i.d. N(0, 1) \perp \varepsilon_t$ |
| DGP 4: $x_{4t} = \xi_t \sim Cauchy$ | DGP 9: $x_{9t}$ is a STUR process |
| DGP 5: $x_{5t} = \left( \sum_{j=1}^{t} \varepsilon_j \right)^2$ | DGP 10: $x_{10t} = 1(v_t \leq 0) \left( \sum_{j=1}^{t} \varepsilon_j \right)$, $v_t \sim i.i.d. N(0, 1) \perp \varepsilon_t$ |

Performance of subsampling is mainly measured by coverage probability of two-sided nominal 95% symmetric intervals for the parameter $\delta = (\beta - 1)/2^6$. We also look at the mean and standard deviation length. The coverage probability of the intervals is denoted by $CP$ and the mean and standard deviation length by $IQR_{95\%}$ and $sd(IQR_{95\%})$, respectively. The experiment is based on

---

9In the simulations we ran, two-sided symmetric confidence intervals performed much better than equally tailed confidence intervals.
Table 1: Performance of the subsampling methodology. Without Deterministic Components

<table>
<thead>
<tr>
<th>DGP</th>
<th>CP</th>
<th>IQR(_{95%})</th>
<th>sd(IQR(_{95%}))</th>
<th>CP</th>
<th>IQR(_{95%})</th>
<th>sd(IQR(_{95%}))</th>
<th>CP</th>
<th>IQR(_{95%})</th>
<th>IQR(_{95%})</th>
</tr>
</thead>
<tbody>
<tr>
<td>S((\delta))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td></td>
<td></td>
<td></td>
<td>n = 200</td>
<td></td>
<td></td>
<td>n = 500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – S(0)</td>
<td>0.991</td>
<td>2.933</td>
<td>0.519</td>
<td>0.995</td>
<td>2.520</td>
<td>0.333</td>
<td>0.991</td>
<td>2.125</td>
<td>0.256</td>
</tr>
<tr>
<td>2 – S(1)</td>
<td>0.832</td>
<td>2.063</td>
<td>0.669</td>
<td>0.804</td>
<td>1.760</td>
<td>0.529</td>
<td>0.807</td>
<td>1.498</td>
<td>0.423</td>
</tr>
<tr>
<td>3 – S(2)</td>
<td>0.747</td>
<td>2.730</td>
<td>0.598</td>
<td>0.797</td>
<td>2.659</td>
<td>0.517</td>
<td>0.863</td>
<td>2.612</td>
<td>0.443</td>
</tr>
<tr>
<td>4 – S(0.5)</td>
<td>0.986</td>
<td>3.897</td>
<td>0.680</td>
<td>0.992</td>
<td>3.458</td>
<td>0.539</td>
<td>0.994</td>
<td>2.978</td>
<td>0.396</td>
</tr>
<tr>
<td>5 – S(1.5)</td>
<td>0.905</td>
<td>3.392</td>
<td>1.054</td>
<td>0.900</td>
<td>2.982</td>
<td>0.888</td>
<td>0.904</td>
<td>2.671</td>
<td>0.758</td>
</tr>
<tr>
<td>6 – S(0.7)</td>
<td>0.939</td>
<td>2.390</td>
<td>0.625</td>
<td>0.954</td>
<td>2.019</td>
<td>0.499</td>
<td>0.949</td>
<td>1.693</td>
<td>0.387</td>
</tr>
<tr>
<td>7 – S(0.5)</td>
<td>0.942</td>
<td>2.532</td>
<td>0.673</td>
<td>0.929</td>
<td>2.152</td>
<td>0.495</td>
<td>0.930</td>
<td>1.836</td>
<td>0.420</td>
</tr>
<tr>
<td>8 – S(0.5)</td>
<td>0.988</td>
<td>3.437</td>
<td>0.678</td>
<td>0.984</td>
<td>2.993</td>
<td>0.560</td>
<td>0.983</td>
<td>2.587</td>
<td>0.501</td>
</tr>
<tr>
<td>9 – S(1.5)</td>
<td>0.606</td>
<td>3.485</td>
<td>0.853</td>
<td>0.637</td>
<td>3.324</td>
<td>0.817</td>
<td>0.682</td>
<td>3.190</td>
<td>0.790</td>
</tr>
<tr>
<td>10 – S(1)</td>
<td>0.597</td>
<td>1.965</td>
<td>0.624</td>
<td>0.665</td>
<td>1.772</td>
<td>0.504</td>
<td>0.653</td>
<td>1.536</td>
<td>0.418</td>
</tr>
</tbody>
</table>

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. IQR\(_{95\%}\) denotes the mean length of the intervals and sd(IQR\(_{95\%}\)) its corresponding standard deviation.

1000 replicas and we use three different sample sizes \(n = \{100, 200, 500\}\). Moreover, we choose a subsample size \(b = \sqrt{n^7}\). The results are collected in Table 1.

As it can be seen, the coverage probability is near to the nominal 95% level in almost all the cases we have considered. For DGPs 1-3, we see that as the order of summability increase the coverage probability decreases for a given sample size. However, as expected, the higher the sample size the better the coverage probability. This also can be seen in DGPs 9-10 where the subsampling seems to work worst for small sample sizes. However, although not reported here, for \(n = \{1000, 5000, 10000\}\) the coverage probability for DGPs 9-10 increases considerably. Moreover, for DGP 10 the performance is optimal when, instead of using an indicator function for defining the stochastic process, we use a step function taking values different from zero. The zeros reduce the actual sample size that the subsampling methodology use.

Finally, we want to remark that the mean and standard deviation of the length of the intervals is almost the same for all the DGPs. Moreover, as expected they are better for higher sample sizes.

\(^7\)To keep things simple we choose \(b = \sqrt{n}\), although a choice based in the minimum volatility method should be preferred. Anyway, if \(b \to \infty\) and \(b/n \to 0\) as \(n \to \infty\), any choice of \(b\) will yield the required consistency of subsampling methods.
All together, it seems that the subsampling methodology works quite well when trying to carry out inferences on the order of summability of a time series.

### 4.3.2 With Deterministic Components

When unknown determinist components describe the DGPs, Proposition 7 and Remark 10 say that a proper estimator of those deterministic components can be used to infer the corresponding order of summability.

Next, we study the performance of the subsampling confidence intervals for when the parameters that characterize the deterministic components of the process, $m_t$, are properly estimated. The DGPs to be considered are the following ones.

<table>
<thead>
<tr>
<th>DGP 11: $x_{1t} = m_t + \varepsilon_t \sim i.i.d. N(0, 1)$</th>
<th>DGP 15: $x_{5t} = m_t + \left( \sum_{j=1}^{t} \varepsilon_j \right)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP 12: $x_{2t} = m_t + \sum_{j=1}^{t} \varepsilon_j$</td>
<td>DGP 16: $x_{6t} = m_t + \Delta^{0.3} \left( \sum_{j=1}^{t} \varepsilon_j \right)$</td>
</tr>
<tr>
<td>DGP 13: $x_{3t} = m_t + \sum_{j=1}^{t} \sum_{i=1}^{j} \varepsilon_i$</td>
<td>DGP 17: $x_{7t} = m_t + \eta_t \left( \sum_{j=1}^{t} \varepsilon_j \right), \eta_t \sim i.i.d. N(0, 1) \perp \varepsilon_t$</td>
</tr>
<tr>
<td>DGP 14: $x_{4t} = m_t + \xi_t \sim \text{Cauchy}$</td>
<td>DGP 18: $x_{8t} = m_t + 1(v_t \leq 0) \left( \sum_{j=1}^{t} \varepsilon_j \right), v_t \sim i.i.d. N(0, 1) \perp \varepsilon_t$</td>
</tr>
</tbody>
</table>

Two usual parametric forms for $m_t$ will be considered. First, we will study the case in which $m_t = m$, i.e. an unknown constant term. Second, we will consider a linear deterministic trend, $m_t = m_0 + m_1 t$, with $m_0$ and $m_1$ unknown.

**Example 8**: Let

\[ x_t = m + w_{xt}, \]

where $m$ is a constant and $w_{xt} \sim S(\delta_{wx})$ in the sense that

\[ \frac{1}{n^{0.5 + \delta_{wx}}} \sum_{t=1}^{n} w_{xt} \rightarrow D(r, \delta_{wx}), \tag{9} \]

where $D(r, \delta_{wx})$ is a random variable or process with positive variance and $r \in [0, 1]$ but $p(r = 0) = 0$. The time series of interest is $x_t$. By definition of order of summability, we know that $x_t \sim S(\delta_{wx})$. Let us assume that $\delta_{wx}$ is unknown and we only observe $x_t$. By Proposition 7 and Remark 10, a proper estimator of $m$ can be used to infer $\delta_{wx}$. To demean $x_t$ with the arithmetic mean of $x_t$, say $\bar{x}$, is problematic in this context because

\[ \sum_{t=1}^{n} (x_t - \bar{x}) = 0. \]
An alternative operational choice could be the following weighted partial mean

\[ \hat{m}_t = \sum_{j=1}^{t} j x_j \sum_{j=1}^{t} j = \sum_{j=1}^{t} j (m + w_{xj}) \sum_{j=1}^{t} j = m + \sum_{j=1}^{t} j w_{xj} \]

\[ = m + \frac{2}{t(t+1)} \sum_{j=1}^{t} j w_{xj}. \]

Let

\[ \nu_{xt} = \frac{2}{t(t+1)} \sum_{j=1}^{t} j w_{xj}. \]

By (9) it holds that

\[ \frac{1}{n^{1/2 + \delta_{wx} + 1}} \sum_{t=1}^{n} t w_{xt} = \frac{1}{n^{1/2 + \delta_{wx}}} \sum_{t=1}^{n} \frac{t}{n} w_{xt} \implies D_r(r, \delta_{wx}), \]

where \( D_r(r, \delta_{wx}) \) is the corresponding asymptotic distribution, which is subindexed by \( r = t/n \) to remark the difference with respect \( D(r, \delta_{wx}) \). Hence, the D-sapce analog of

\[ \frac{1}{t^{1/2 + \delta_{wx} + 1}} \sum_{j=1}^{t} j w_{xt} = \frac{1}{t^{1/2 + \delta_{wx}}} \sum_{j=1}^{t} \frac{j}{t} w_{xt}, \]

satisfies

\[ \frac{1}{[nr]^{1/2 + \delta_{wx}}} \sum_{j=1}^{[nr]} sw_{xj} \]

\[ \begin{cases} 0 & \text{for } 0 \leq s < \frac{1}{t} \\ \frac{1}{1/2 + \delta_{wx}} w_{x1} & \text{for } \frac{1}{t} \leq s < \frac{2}{t} \\ \frac{1}{2^{1/2 + \delta_{wx}}} (\frac{1}{2} w_{x1} + w_{x2}) & \text{for } \frac{2}{t} \leq s < \frac{3}{t} \\ \frac{1}{3^{1/2 + \delta_{wx}}} (\frac{1}{3} w_{x1} + \frac{2}{3} w_{x2} + w_{x3}) & \text{for } \frac{3}{t} \leq s < \frac{4}{t} \\ \frac{1}{4^{1/2 + \delta_{wx}}} (\frac{1}{4} w_{x1} + \frac{2}{4} w_{x2} + \frac{3}{4} w_{x3} + w_{x4}) & \text{for } \frac{4}{t} \leq s < \frac{5}{t} \\ \vdots & \vdots \\ \frac{1}{t^{1/2 + \delta_{wx}}} (\frac{1}{t} w_{x1} + \frac{2}{t} w_{x2} + \frac{3}{t} w_{x3} + \frac{4}{t} w_{x4} + \ldots + w_{xt}) & \text{or } s = r \end{cases} \]

\[ \implies D^s_r(s, \delta_{wx}), \]

where \( D^s_r(s, \delta_{wx}), s = j/t, \text{ and } r = t/n \). By the CMT,

\[ \int_0^1 \left( \frac{1}{[nr]^{1/2 + \delta_{wx}}} \sum_{j=1}^{[nr]} sw_{xj} \right) dr = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{t^{1/2 + \delta_{wx}}} \sum_{j=1}^{t} \frac{j}{t} w_{xt} \implies \int_0^1 D^s_r(s, \delta_{wx}) dr. \]
We are interested on $v_{xt}$, for which

$$\frac{1}{n^{1/2+\delta_{wx}}} \sum_{t=1}^{n} v_{xt} = \frac{2}{n^{1/2+\delta_{wx}}} \sum_{t=1}^{n} \frac{2}{t(t+1)} \sum_{j=1}^{t} j w_{xj}$$

$$= \frac{2}{n} \sum_{t=1}^{n} \frac{1}{t^{1/2+\delta_{wx}}} \sum_{j=1}^{t} \frac{j}{t+1} w_{xj}$$

$$= \frac{2}{n^{1/2+\delta_{wx}}} \sum_{t=1}^{n} \frac{t-1/2+\delta_{wx}}{t^{1/2+\delta_{wx}}} \sum_{j=1}^{t} \frac{j}{t+1} w_{xj}.$$  

Therefore,

$$\frac{1}{n^{1/2+\delta_{wx}}} \sum_{t=1}^{n} v_{xt} = \frac{1}{n} \sum_{t=1}^{n} r^{-1/2+\delta_{wx}} \frac{1}{t^{1/2+\delta_{wx}}} \sum_{j=1}^{t} \frac{j}{t+1} w_{xj} \implies 2 \int_{0}^{1} r^{-1/2+\delta_{wx}} D_s(s, \delta_{wx}) \, dr.$$  

Therefore, for all $\delta_{wx} \geq 0$, $\hat{m}_t$ is a proper estimator of $m$ in the sense of Proposition 7 and Remark 10.

Previous example says that to control for constant deterministic elements, we can subtract to the time series of interest, $x_t$, the weighted mean

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

By doing so, the asymptotic properties of the order of summability estimator proposed in this section and described in Propositions 3 and 4 are still true. Table 2 shows the performance of the subsampling confidence intervals for $\delta$ in DGPs 11-18 when $m_t = m = 10$. However, it should be emphasized that, as the above example shows, results are independent of the particular choice of $m$. As before, we compute the coverage coverage probability of two-sided nominal 95% symmetric intervals as well as the mean and standard deviation lengths of the intervals. The experiment is based on 1000 replicas, three different sample sizes $n = \{100, 200, 500\}$, and a subsample size $b = \sqrt{n}$.

As it can be seen in Table 2, the results are similar to those obtained without deterministic components. This signifies that $\hat{m}_t$ demeans the processes in such a way that the order of summability can be inferred without interferences from the unknown constant term.

Example 9: Let

$$x_t = m_0 + m_1 t + w_{xt},$$
Table 2: Performance of the subsampling methodology. With Constant Term

| DGP | CP | IQR
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 100</td>
<td>sd(IQR)</td>
</tr>
<tr>
<td>$S(\delta)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 – $S(0)$</td>
<td>0.983</td>
<td>2.839</td>
</tr>
<tr>
<td>12 – $S(1)$</td>
<td>0.941</td>
<td>2.592</td>
</tr>
<tr>
<td>13 – $S(2)$</td>
<td>0.729</td>
<td>2.756</td>
</tr>
<tr>
<td>14 – $S(0.5)$</td>
<td>0.974</td>
<td>3.432</td>
</tr>
<tr>
<td>15 – $S(1.5)$</td>
<td>0.758</td>
<td>3.367</td>
</tr>
<tr>
<td>16 – $S(0.7)$</td>
<td>0.980</td>
<td>2.756</td>
</tr>
<tr>
<td>17 – $S(0.5)$</td>
<td>0.989</td>
<td>3.167</td>
</tr>
<tr>
<td>18 – $S(1)$</td>
<td>0.812</td>
<td>3.033</td>
</tr>
</tbody>
</table>

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. IQR denotes the mean length of the intervals and sd(IQR) its corresponding standard deviation. Where $m_0$ and $m_1$ are unknown parameters and $w_{xt} \sim S(\delta_{wx})$ in the sense that

$$
\frac{1}{n^{1/2+\delta_{wx}}} \sum_{t=1}^{n} w_{xt} \to D(r, \delta_{wx}),
$$

as before. In this case, the detrending procedure consists of two steps. Let, as in Example 8,

$$
\hat{m}_{0t} = \frac{\sum_{j=1}^{t} j x_t}{\sum_{j=1}^{t} j} = \frac{\sum_{j=1}^{t} j (m_0 + m_1 j + w_{xj})}{\sum_{j=1}^{t} j} = m_0 + m_1 \frac{\sum_{j=1}^{t} j^2}{\sum_{j=1}^{t} j} + \frac{\sum_{j=1}^{t} j w_{xj}}{\sum_{j=1}^{t} j} = m_0 + m_1 \frac{2t + 1}{3} + \frac{\sum_{j=1}^{t} j w_{xj}}{\sum_{j=1}^{t} j}.
$$

The first step in this example is the transformation

$$
x_t - \hat{m}_{0t} = m_1 t + w_{xt} - m_1 \frac{2t + 1}{3} - \frac{\sum_{j=1}^{t} j w_{xj}}{\sum_{j=1}^{t} j} = m_1 \frac{t - 1}{3} + w_{xt} - \frac{\sum_{j=1}^{t} j w_{xj}}{\sum_{j=1}^{t} j}.
$$

Next, consider just the observations $t \geq 2$, or equivalently, take $t \geq 1$ and rewrite

$$
x_{t+1} - \hat{m}_{0t+1} = \frac{m_1}{3} t + w_{xt+1} - \frac{\sum_{j=1}^{t+1} j w_{xj}}{\sum_{j=1}^{t+1} j}.
$$
Now, let
\[ \hat{m}_{1t} = \frac{\sum_{j=1}^{t} j(x_{j+1} - \hat{m}_{0j+1})}{\sum_{j=1}^{t} j^2} = \frac{\sum_{j=1}^{t} j \left( \frac{m_{0j+1}}{3} + w_{xj+1} - \frac{\sum_{j=1}^{t} j w_{xj+1}}{3} \right)}{\sum_{j=1}^{t} j^2} \]
\[ = \frac{m_1}{3} + \frac{\sum_{j=1}^{t} j w_{xj+1}}{\sum_{j=1}^{t} j^2} - \frac{\sum_{j=1}^{t} j \frac{\sum_{j=1}^{t} j w_{xj+1}}{3}}{\sum_{j=1}^{t} j^2}. \]

The second transformation is
\[ x_{t+1} - \hat{m}_{0t+1} - \hat{m}_{1t} = \frac{\sum_{j=1}^{t+1} j w_{xj}}{\sum_{j=1}^{t+1} j} - \frac{t \sum_{j=1}^{t+1} j w_{xj+1}}{\sum_{j=1}^{t+1} j^2} - \frac{\sum_{j=1}^{t} j \frac{\sum_{j=1}^{t} j w_{xj+1}}{3}}{\sum_{j=1}^{t} j^2} \]
\[ = w_{xt+1} - \frac{2}{(t+2)(t+1)} \sum_{j=1}^{t+1} j w_{xj} - \frac{6}{(2t+1)(t+1)} \sum_{j=1}^{t} j w_{xj+1} \]
\[ + \frac{6}{(2t+1)(t+1)} \sum_{j=1}^{t} \frac{2j}{(j+2)(j+1)} \sum_{i=1}^{j+1} w_{xi}. \]

Again, using the CMT, it could be shown that
\[ \frac{1}{n^{1/2+\delta_{ux}}} \sum_{t=1}^{n} (x_{t+1} - \hat{m}_{0t+1} - \hat{m}_{1t}), \]
converge to some functional of $D(r, \delta_{ux})$.

Table 3 shows the performance of the subsampling confidence intervals for $\delta$ in DGPs 11-18 when $m_t = m_0 + m_1 t = 10 + 2t$. However, as the in the previous example, results are independent of the particular choice of $m_0$ and $m_1$. All measures are as in Table 2. One more time, we can see that the coverage probability is close to the nominal level of 95% in almost all the cases. Also, the mean and the standard deviation lengths of the intervals are quite reasonable and improves as the sample size grows.

Overall, we can say that the methodology that we have proposed in this section is quite successful to determine the order of summability of univariate time series.

5 Spurious vs Co-summable Regressions

In this section, we propose a three steps strategy for applied researchers interested in knowing whether an economic hypothesis of the type $y_t = \theta f(x_t)$ is true. It follows the scheme of the standard single equation co-integration literature, but it works with some different statistical tools.
Table 3: Performance of the subsampling methodology. With Linear Trend

<table>
<thead>
<tr>
<th>DGP</th>
<th>CP</th>
<th>IQR$_{95%}$</th>
<th>sd(IQR$_{95%}$)</th>
<th>CP</th>
<th>IQR$_{95%}$</th>
<th>sd(IQR$_{95%}$)</th>
<th>CP</th>
<th>IQR$_{95%}$</th>
<th>IQR$_{95%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S(\delta)$</td>
<td></td>
<td>n = 100</td>
<td></td>
<td></td>
<td>n = 200</td>
<td></td>
<td></td>
<td>n = 500</td>
<td></td>
</tr>
<tr>
<td>11 – $S(0)$</td>
<td>0.976</td>
<td>2.867</td>
<td>0.485</td>
<td>0.977</td>
<td>2.461</td>
<td>0.386</td>
<td>0.979</td>
<td>2.062</td>
<td>0.277</td>
</tr>
<tr>
<td>12 – $S(1)$</td>
<td>0.959</td>
<td>2.825</td>
<td>0.554</td>
<td>0.962</td>
<td>2.444</td>
<td>0.464</td>
<td>0.966</td>
<td>2.108</td>
<td>0.434</td>
</tr>
<tr>
<td>13 – $S(2)$</td>
<td>0.869</td>
<td>3.307</td>
<td>0.624</td>
<td>0.935</td>
<td>3.208</td>
<td>0.541</td>
<td>0.974</td>
<td>3.040</td>
<td>0.442</td>
</tr>
<tr>
<td>14 – $S(0.5)$</td>
<td>0.973</td>
<td>3.279</td>
<td>0.589</td>
<td>0.974</td>
<td>2.924</td>
<td>0.455</td>
<td>0.964</td>
<td>2.476</td>
<td>0.319</td>
</tr>
<tr>
<td>15 – $S(1.5)$</td>
<td>0.793</td>
<td>3.406</td>
<td>0.772</td>
<td>0.844</td>
<td>3.126</td>
<td>0.662</td>
<td>0.881</td>
<td>2.827</td>
<td>0.566</td>
</tr>
<tr>
<td>16 – $S(0.7)$</td>
<td>0.991</td>
<td>2.865</td>
<td>0.524</td>
<td>0.989</td>
<td>2.447</td>
<td>0.443</td>
<td>0.998</td>
<td>2.077</td>
<td>0.361</td>
</tr>
<tr>
<td>17 – $S(0.5)$</td>
<td>0.993</td>
<td>3.214</td>
<td>0.588</td>
<td>0.997</td>
<td>2.859</td>
<td>0.475</td>
<td>0.995</td>
<td>2.542</td>
<td>0.431</td>
</tr>
<tr>
<td>18 – $S(1)$</td>
<td>0.845</td>
<td>3.090</td>
<td>0.637</td>
<td>0.870</td>
<td>2.830</td>
<td>0.542</td>
<td>0.888</td>
<td>2.604</td>
<td>0.482</td>
</tr>
</tbody>
</table>

$CP$ denotes the coverage probability of two-sided nominal 95% symmetric intervals. $IQR_{95\%}$ denotes the mean length of the intervals and $sd(IQR_{95\%})$ its corresponding standard deviation.

### 5.1 A Three Steps Strategy

A regression model

$$y_t = \theta f(x_t) + u_t,$$ (11)

where $y_t$ and $x_t$ are observed, $f$ is known but $\theta$ is not, is the object of interest for the econometrician.

To determine if a regression model like (11) specifies a long run equilibrium relationship, it is not sufficient to carry out inferences on the unknown parameter $\theta$. First, it should be determined whether the regression model is balanced. We propose to achieve this task by using Definition 8, i.e. by determining whether $y_t$ and $f(x_t)$ have the same order of summability, $\delta$. Even in that case, the danger of spurious regressions documented in the linear case can arise here as well. Using Definition 9, we will say that a balanced regression is spurious if $\delta = \delta_0$, where $\delta_0$ is the order of summability of $u_t$. On the contrary, if $\delta > \delta_0$ the regression will specify a co-summable relationship between $y_t$ and $f(x_t)$. In this case, it will be interesting to distinguish between strong and weak co-summable regressions. In the former case, $\delta_0 = 0$, while it is positive in the later. To empirically determine balancedness and to rule out spurious regressions, we propose the following 3-steps empirical strategy.

**Step 1.** To check the balancedness of the regression model, estimate the order of summability of $y_t$ and $z_t = f(x_t)$ using the technique developed above.

**Step 2.** If $y_t$ and $z_t = f(x_t)$ have the same order of summability, $\delta$, estimate $\theta$ in (11) and compute $\hat{u}_t$. 

32
Step 3. To rule out spurious regressions, estimate the order of summability of \( \hat{u}_t \) using the same technique as in Step 1.

Given the fact that model (11) is linear in its parameters, we propose the use of the OLS estimator, \( \hat{\theta} \), of the parameter \( \theta \), in Step 2. Next, we study its asymptotic properties.

**Proposition 8**: If

(i) \[
\frac{1}{n^{1/2+\delta-1}} \sum_{t=1}^{n} \Delta z_t = \frac{1}{n^{1/2+\delta-1}} (z_{n} - z_{0}) = \frac{1}{n^{1/2+\delta-1}} z_{n} + o_p(1) \Rightarrow Z(r, \delta),
\]

(ii) \[
\frac{1}{n^{1/2+\delta_0}} \sum_{t=1}^{n} u_t \Rightarrow D(r, \delta_0),
\]

(iii) \[
\frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n} \Delta z_t u_t \rightarrow \Lambda,
\]

hold, then

\[
n^{\delta-\delta_0} (\hat{\theta} - \theta) \Rightarrow \frac{Z(r, \delta)D(r, \delta_0) + \Lambda}{2 \int_{0}^{1} Z^2(r, \delta) dr}.
\]

**Proof**: The OLS estimator, \( \hat{\theta} \), is

\[
\hat{\theta} = \theta + \frac{\sum_{t=1}^{n} z_t u_t}{\sum_{t=1}^{n} z_t^2}.
\]

Let us focus on the denominator. By condition (i) and the CMT,

\[
\frac{1}{n^{1/2+\delta}} \sum_{t=1}^{n} z_t \Rightarrow \left( \int_{0}^{1} Z(r, \delta) dr \right),
\]

and

\[
\frac{1}{n^{2\delta}} \sum_{t=1}^{n} z_t^2 \Rightarrow \int_{0}^{1} Z^2(r, \delta) dr.
\]

With respect the numerator, note that by conditions (i) and (ii),

\[
\left( \frac{1}{n^{-1/2+\delta}} \sum_{t=1}^{n} \Delta z_t \right) \left( \frac{1}{n^{1/2+\delta_0}} \sum_{t=1}^{n} u_t \right) \Rightarrow Z(r, \delta)D(r, \delta_0).
\]
or equivalently,

\[
\frac{1}{n^{\delta+\delta_0}} \left( \sum_{t=1}^{n} \Delta z_t u_t + 2 \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} u_t \Delta z_s \right) = \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n} \Delta z_t u_t + 2 \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n-1} u_t \sum_{s=t+1}^{n} \Delta z_s \\
= \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n} \Delta z_t u_t + 2 \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n-1} u_t (z_n - z_t) \\
= \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n} \Delta z_t u_t + 2 \frac{z_n}{n^{1/2+\delta}} \frac{1}{n^{1/2+\delta_0}} \sum_{t=1}^{n-1} u_t - 2 \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n-1} u_t z_t \\
= \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n} \Delta z_t u_t + 2 \frac{z_n}{n^{1/2+\delta}} \frac{1}{n^{1/2+\delta_0}} \sum_{t=1}^{n-1} u_t - 2 \frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n-1} u_t z_t 
\]

\[
\implies Z(r, \delta) D(r, \delta_0).
\]

Hence, by conditions (i)-(iii)

\[
\frac{1}{n^{\delta+\delta_0}} \sum_{t=1}^{n-1} u_t z_t \implies 2Z(r, \delta) D(r, \delta_0) - Z(r, \delta) D(r, \delta_0) + \Lambda \\
\quad \implies Z(r, \delta) D(r, \delta_0) + \Lambda.
\]

Therefore,

\[
n^{\delta-\delta_0} (\hat{\theta} - \theta) = \frac{\sum_{t=1}^{n} z_t u_t}{\sum_{t=1}^{n} z_t^2} \implies \frac{Z(r, \delta) D(r, \delta_0) + \Lambda}{\frac{\int_{0}^{1} Z^2(r, \delta) d\delta}{2}}.
\]

as stated. Q.E.D.

As in the linear regression case, Proposition 8 tell us that if \( z_t \sim S(\delta), u_t \sim S(\delta_0) \), and the re-scaled sample covariance between \( \Delta z_t \) and \( u_t \) converge, then the OLS estimator of \( \theta \) is \((\delta - \delta_0)\)-consistent. However, before carrying any inference on \( \theta \), it must be determined whether the regression model specifies a spurious or a long run relationship. The following proposition shows that the machinery developed in previous section can be used to this regard.

**Proposition 9** : Let the conditions (i), (ii) and (iii) in Proposition 8 hold. Then

\[
\frac{1}{n^{1/2+\delta_0}} \sum_{t=1}^{n} \hat{u}_t \implies D(r, \delta_0) - \frac{Z(r, \delta) D(r, \delta_0) + \Lambda}{2 \int_{0}^{1} Z^2(r, \delta) d\delta} \int_{0}^{1} Z(r, \delta) d\delta.
\]

**Proof**: By definition of the OLS residuals,

\[
\sum_{t=1}^{n} \hat{u}_t = \sum_{t=1}^{n} u_t - (\hat{\theta} - \theta) \sum_{t=1}^{n} z_t.
\]

The results immediately holds by conditions (i), (ii), and (iii). Q.E.D.
Proposition 9 means that \( \hat{u}_t \) has the same order of summability that \( u_t \) regardless of whether the regression is spurious or co-summable.

Similarly to the previous section, if a constant term is introduced in the regression model, it must be properly taken into consideration. To be more precise, let

\[
y_t = m + \theta f(x_t) + u_t,
\]

where \( m \) is an unknown constant term. In this case,

\[
\sum_{t=1}^{n} \hat{u}_t = \sum_{t=1}^{n} (y_t - \hat{m} - \hat{\theta} f(x_t)) = \sum_{t=1}^{n} \left( u_t - (\hat{m} - m) - (\hat{\theta} - \theta) f(x_t) \right) = 0,
\]

which implies that \( \hat{u}_t \) cannot be used to infer \( \delta_0 \). Nevertheless, the following pseudo residuals could be used instead. Let

\[
\tilde{u} = y_t - \hat{\theta} f(x_t) = m + u_t - (\hat{\theta} - \theta) f(x_t).
\]

Then,

\[
\sum_{t=1}^{n} \tilde{u} \neq 0,
\]

and the effect of \( m \) on the estimation of the order of summability of \( u_t \) can be taken into account by using Proposition 7 and Remark 10.

### 5.2 Finite Sample Performance

In this section, we run several Monte Carlo experiments to study the performance of the order of summability estimator and the subsampling inference when applied to the estimated residuals from both spurious and co-summable regressions. In the experiments, we use the following driving errors:

\[
\varepsilon_{yt} \sim i.i.d.N(0, 1), \quad \varepsilon_{xt} \sim i.i.d.N(0, 1); \quad z_x \sim N(0, 1); \quad \eta_{yt} \sim i.i.d.N(0, 1) \quad \text{and} \quad \eta_{xt} \sim i.i.d.N(0, 1);
\]

\[
v_{yt} \sim i.i.d.N(0, 1) \quad \text{and} \quad v_{xt} \sim i.i.d.N(0, 1). \quad \text{All of them are independent each other.}
\]

#### 5.2.1 Spurious Regressions

In the spurious regression case, we run \( i = 1, \ldots, 7 \) regressions. The DGPs for \( y_{it} \) and \( z_{it} \), are defined in Table 4. After calculating the OLS estimator of \( \theta \) in

\[
y_{it} = \theta z_{it} + u_{it},
\]

with \( \theta = 0 \), the order of summability of \( \hat{u}_{it} \) is estimated and inferred. Performance of subsampling is measured as in previous experiments. The results are collected in Table 5.
Table 4: DGPs for the Spurious Regression Case

<table>
<thead>
<tr>
<th>DGP</th>
<th>( y_{1t} = \sum_{j=1}^{t} \varepsilon_{yj} )</th>
<th>( z_{1t} = \sum_{j=1}^{t} \varepsilon_{xj} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP 1</td>
<td>( y_{2t} = \sum_{j=1}^{t} \sum_{j=1}^{t} \varepsilon_{yj} )</td>
<td>( z_{2t} = \sum_{j=1}^{t} \sum_{j=1}^{t} \varepsilon_{xj} )</td>
</tr>
<tr>
<td>DGP 2</td>
<td>( y_{3t} = \left( \sum_{j=1}^{t} \varepsilon_{yj} \right)^2 )</td>
<td>( z_{3t} = \left( \sum_{j=1}^{t} \varepsilon_{xj} \right)^2 )</td>
</tr>
<tr>
<td>DGP 3</td>
<td>( y_{4t} = \Delta^{0.3} \left( \sum_{j=1}^{t} \varepsilon_{yj} \right) )</td>
<td>( z_{4t} = \Delta^{0.3} \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
<tr>
<td>DGP 4</td>
<td>( y_{5t} = \eta_{yt} \left( \sum_{j=1}^{t} \varepsilon_{yj} \right) )</td>
<td>( z_{5t} = \eta_{xt} \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
<tr>
<td>DGP 5</td>
<td>( y_{6t} = 1(v_{yt} \leq 0) \left( \sum_{j=1}^{t} \varepsilon_{yj} \right) )</td>
<td>( z_{6t} = 1(v_{xt} \leq 0) \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
<tr>
<td>DGP 6</td>
<td>( y_{7t} = \sum_{j=1}^{t} \varepsilon_{yj} )</td>
<td>( z_{7t} = 1(v_{xt} \leq 0) \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
</tbody>
</table>

Table 5: Performance of the subsampling methodology. Spurious Regression. Without Constant

<table>
<thead>
<tr>
<th>DGP</th>
<th>( CP )</th>
<th>( IQR_{95%} )</th>
<th>( sd(IQR_{95%}) )</th>
<th>( CP )</th>
<th>( IQR_{95%} )</th>
<th>( sd(IQR_{95%}) )</th>
<th>( CP )</th>
<th>( IQR_{95%} )</th>
<th>( IQR_{95%} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( n = 100 )</td>
<td>( n = 200 )</td>
<td>( n = 500 )</td>
<td>( n = 100 )</td>
<td>( n = 200 )</td>
<td>( n = 500 )</td>
</tr>
<tr>
<td>1</td>
<td>0.755</td>
<td>2.415</td>
<td>0.625</td>
<td>0.703</td>
<td>1.984</td>
<td>0.466</td>
<td>0.672</td>
<td>1.634</td>
<td>0.406</td>
</tr>
<tr>
<td>2</td>
<td>0.524</td>
<td>2.881</td>
<td>0.636</td>
<td>0.590</td>
<td>2.683</td>
<td>0.550</td>
<td>0.669</td>
<td>2.527</td>
<td>0.482</td>
</tr>
<tr>
<td>3</td>
<td>0.765</td>
<td>3.374</td>
<td>0.856</td>
<td>0.771</td>
<td>2.983</td>
<td>0.764</td>
<td>0.813</td>
<td>2.673</td>
<td>0.649</td>
</tr>
<tr>
<td>4</td>
<td>0.955</td>
<td>2.612</td>
<td>0.607</td>
<td>0.943</td>
<td>2.217</td>
<td>0.441</td>
<td>0.938</td>
<td>1.808</td>
<td>0.328</td>
</tr>
<tr>
<td>5</td>
<td>0.986</td>
<td>3.324</td>
<td>0.649</td>
<td>0.990</td>
<td>2.930</td>
<td>0.541</td>
<td>0.982</td>
<td>2.524</td>
<td>0.463</td>
</tr>
<tr>
<td>6</td>
<td>0.734</td>
<td>2.818</td>
<td>0.816</td>
<td>0.745</td>
<td>2.502</td>
<td>0.726</td>
<td>0.793</td>
<td>2.202</td>
<td>0.537</td>
</tr>
<tr>
<td>7</td>
<td>0.836</td>
<td>2.540</td>
<td>0.682</td>
<td>0.810</td>
<td>2.145</td>
<td>0.513</td>
<td>0.833</td>
<td>1.896</td>
<td>0.475</td>
</tr>
</tbody>
</table>

\( CP \) denotes the coverage probability of two-sided nominal 95\% symmetric intervals. \( IQR_{95\%} \) denotes the mean length of the intervals and \( sd(IQR_{95\%}) \) its corresponding standard deviation.
The experiments show that the subsample performance on the estimated errors when the regression is spurious is similar to the univariate time series case. It is worthwhile to emphasize that patterns shown in the univariate time series case are even more marked here. For instance, when two I(2) independent processes are spuriously related, the coverage probability for the order of summability of the residuals needs a higher number of observations to work as in the univariate case.

As we have mentioned above, the effect of introducing a constant term into the regression model, i.e.

$$y_{it} = m + \theta z_{it} + u_{it},$$

will enforce us to use the pseudo-residuals

$$\tilde{u}_{it} = y_{it} - \hat{\theta} z_{it}.$$ 

To show the performance of the subsampling in this case, we have run the same experiments as in Table 5 but introducing an intercept into the regression model. The intercept has been set $m = 10$, but, as in the univariate case, results are not depending on this choice. The associated results, reported in Table 6, show that the pseudo-residuals are a useful construction. The performance is quite similar to the previous case in which the constant term was not introduced in the regression.

### 5.2.2 Co-summable Regressions

Next, we focus on co-summable regressions. In particular, we will study strong co-summable regressions in the sense that the error term will be assumed to be summable of order zero. In fact,
Table 7: DGPs for the Strong Co-summable Regression Case

<table>
<thead>
<tr>
<th>DGP</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( z_{1t} = \sum_{j=1}^{t} \varepsilon_{xj} )</td>
</tr>
<tr>
<td>2</td>
<td>( z_{21t} = \sum_{j=1}^{t} \sum_{i=1}^{t} \varepsilon_{xj} )</td>
</tr>
<tr>
<td>3</td>
<td>( z_{3t} = \left( \sum_{j=1}^{t} \varepsilon_{xj} \right)^2 )</td>
</tr>
<tr>
<td>4</td>
<td>( z_{4t} = \Delta^{0.3} \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
<tr>
<td>5</td>
<td>( z_{5t} = z_x + \varepsilon_{xt} )</td>
</tr>
<tr>
<td>6</td>
<td>( z_{6t} = \eta_{xt} \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
<tr>
<td>7</td>
<td>( z_{7t} = 1(\nu_{xt} \leq 0) \left( \sum_{j=1}^{t} \varepsilon_{xj} \right) )</td>
</tr>
</tbody>
</table>

we used

\[ y_{it} = \theta z_{it} + u_{it}, \]

with \( u_{it} \sim i.i.d. N(0, 1), i = 1, ..., 7 \). The explanatory variables to be considered in the experiments are described in Table 7. We fixed \( \theta = 1 \) and the errors driving \( z_{it} \) were defined above.

Results shown in Table 8 indicate that the subsampling methodology is quite successful detecting strong co-summable regressions. Again, this result is in line with those found in the univariate case. As Table 9 reports, no much differences can be found when an intercept, \( m \), is introduced in the regression and it is taken into account through the pseudo-residuals. As before, we set \( m = 10 \) but the results are independent of that choice.

On the whole, we can conclude that the estimator of the order of summability as well as the subsampling inference that we have proposed can be used to distinguish between non-linear spurious and co-summable regressions quite successfully.

6 Empirical Application

In this section, we carry out an empirical application to illustrate how to infer in practice the order of summability of observed time series. Firstly, we carry out an univariate analysis using an extended version of the Nelson and Plosser (1982) database. After inferring the order of summability of the time series included in that database, a multivariate empirical study of the quantitative theory of money shows how to distinguish between spurious or co-summable regressions in practice.
Table 8: Performance of the subsampling methodology. Co-summable Regression. Without Constant

<table>
<thead>
<tr>
<th>DGP</th>
<th>$CP$</th>
<th>$IQR_{95%}$</th>
<th>$sd(IQR_{95%})$</th>
<th>$CP$</th>
<th>$IQR_{95%}$</th>
<th>$sd(IQR_{95%})$</th>
<th>$CP$</th>
<th>$IQR_{95%}$</th>
<th>$IQR_{95%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td></td>
<td></td>
<td></td>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td>n = 500</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.995</td>
<td>2.927</td>
<td>0.505</td>
<td>0.990</td>
<td>2.532</td>
<td>0.369</td>
<td>0.991</td>
<td>2.135</td>
<td>0.267</td>
</tr>
<tr>
<td>2</td>
<td>0.993</td>
<td>2.909</td>
<td>0.515</td>
<td>0.991</td>
<td>2.528</td>
<td>0.364</td>
<td>0.989</td>
<td>2.131</td>
<td>0.273</td>
</tr>
<tr>
<td>3</td>
<td>0.995</td>
<td>2.927</td>
<td>0.511</td>
<td>0.992</td>
<td>2.538</td>
<td>0.356</td>
<td>0.990</td>
<td>2.129</td>
<td>0.261</td>
</tr>
<tr>
<td>4</td>
<td>0.992</td>
<td>2.937</td>
<td>0.525</td>
<td>0.992</td>
<td>2.517</td>
<td>0.355</td>
<td>0.986</td>
<td>2.135</td>
<td>0.274</td>
</tr>
<tr>
<td>5</td>
<td>0.986</td>
<td>2.952</td>
<td>0.519</td>
<td>0.994</td>
<td>2.529</td>
<td>0.370</td>
<td>0.990</td>
<td>2.135</td>
<td>0.263</td>
</tr>
<tr>
<td>6</td>
<td>0.986</td>
<td>2.949</td>
<td>0.531</td>
<td>0.992</td>
<td>2.522</td>
<td>0.354</td>
<td>0.988</td>
<td>2.110</td>
<td>0.263</td>
</tr>
<tr>
<td>7</td>
<td>0.998</td>
<td>2.926</td>
<td>0.497</td>
<td>0.988</td>
<td>2.518</td>
<td>0.363</td>
<td>0.988</td>
<td>2.124</td>
<td>0.269</td>
</tr>
</tbody>
</table>

$CP$ denotes the coverage probability of two-sided nominal 95\% symmetric intervals. $IQR_{95\%}$ denotes the mean length of the intervals and $sd(IQR_{95\%})$ its corresponding standard deviation.

Table 9: Performance of the subsampling methodology. Co-summable Regression. With Constant

<table>
<thead>
<tr>
<th>DGP</th>
<th>$CP$</th>
<th>$IQR_{95%}$</th>
<th>$sd(IQR_{95%})$</th>
<th>$CP$</th>
<th>$IQR_{95%}$</th>
<th>$sd(IQR_{95%})$</th>
<th>$CP$</th>
<th>$IQR_{95%}$</th>
<th>$IQR_{95%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td></td>
<td></td>
<td></td>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td>n = 500</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.988</td>
<td>2.870</td>
<td>0.470</td>
<td>0.988</td>
<td>2.467</td>
<td>0.353</td>
<td>0.993</td>
<td>2.068</td>
<td>0.229</td>
</tr>
<tr>
<td>2</td>
<td>0.990</td>
<td>2.888</td>
<td>0.480</td>
<td>0.988</td>
<td>2.465</td>
<td>0.346</td>
<td>0.993</td>
<td>2.068</td>
<td>0.232</td>
</tr>
<tr>
<td>3</td>
<td>0.987</td>
<td>2.869</td>
<td>0.486</td>
<td>0.986</td>
<td>2.459</td>
<td>0.349</td>
<td>0.994</td>
<td>2.063</td>
<td>0.225</td>
</tr>
<tr>
<td>4</td>
<td>0.992</td>
<td>2.858</td>
<td>0.451</td>
<td>0.988</td>
<td>2.467</td>
<td>0.350</td>
<td>0.990</td>
<td>2.064</td>
<td>0.228</td>
</tr>
<tr>
<td>5</td>
<td>0.983</td>
<td>2.867</td>
<td>0.498</td>
<td>0.983</td>
<td>2.449</td>
<td>0.348</td>
<td>0.986</td>
<td>2.063</td>
<td>0.241</td>
</tr>
<tr>
<td>6</td>
<td>0.988</td>
<td>2.853</td>
<td>0.456</td>
<td>0.991</td>
<td>2.453</td>
<td>0.330</td>
<td>0.982</td>
<td>2.071</td>
<td>0.256</td>
</tr>
<tr>
<td>7</td>
<td>0.986</td>
<td>2.870</td>
<td>0.491</td>
<td>0.990</td>
<td>2.457</td>
<td>0.337</td>
<td>0.991</td>
<td>2.070</td>
<td>0.256</td>
</tr>
</tbody>
</table>

$CP$ denotes the coverage probability of two-sided nominal 95\% symmetric intervals. $IQR_{95\%}$ denotes the mean length of the intervals and $sd(IQR_{95\%})$ its corresponding standard deviation.
6.1 Extended N&P Database

After Nelson and Plosser (1982) accounted for unit root behavior in almost fourteen U.S. macroeconomic time series, many researchers have been using the same database, or some extended version of it, 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 univariate time series included in an extended version of the Nelson and Plosser (1982) database. As a novelty, we do not impose any linearity assumption.

We will divide the univariate study in two different exercises. In the first one, a graphical study of the behavior of the variances of $S_n$ for growing sample sizes and several choices of $\delta$ will give us a first intuition about the true order of summability of the time series in question. Then, in the second and main exercise, we will estimate and infer the order of summability of the fourteen U.S. macroeconomic time series using the techniques we have studied in the previous sections.

With respect the first exercise of analyzing the behavior of the variance of $S_n$ for growing sample sizes and different choices of $\delta$, note that when the order of summability that is imposed in $S_n$ is less than the true one, the variances grow. Conversely, when the imposed $\delta$ is higher than the actual value, the variances tend to zero. Only when the chosen order is the exact one or it is close to it, the variances stabilize.

To save space, we will only report the graphs of $S_n$ for the U.S. real GNP data, which contains annual observations for the period 1909-1988. The temporal evolution of the real GNP and its logarithm is illustrated in Table 10. In addition, we report, in the same table, the evolution of the demeaned and detrended series as described in Examples 8 and 9.

Regarding the graphical study of the variance of $S_n$ for growing sample sizes and different choices of $\delta$, results concerning the real GNP and its demeaned and detrended time series are shown in Table 11. Specifically, we report the graphs in which we impose $\delta_l = \{0.5, 0.55, 0.6\}$, $\delta_{dem} = \{1.5, 1.6, 1.7\}$, and $\delta_{det} = \{1.7, 1.8, 1.9\}$, for real GNP, demeaned real GNP, and detrended real GNP, respectively. While variances of $S_n$ for growing sample sizes in the demeaned and the detrended cases seem to stabilize when choosing similar orders of summability about 1.7, a lower value of 0.55 seems to be enough when no deterministic components are taken into account. This seems to indicate that demeaning the U.S. real GNP could be enough to control for deterministic components in its DGP. As it can be seen in Table 12, similar results are found when the logarithm of the real GNP is considered.

Next, we present results concerning the second exercise of applying the statistical machinery

---

8 The data have been taken from P.C.B. Phillips’ webpage.
9 Although not reported here, results are similar for the other variables in the database.
Table 10: Temporal Evolution of U.S. GNP

<table>
<thead>
<tr>
<th></th>
<th>Real GNP</th>
<th>log (Real GNP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>levels</td>
<td><img src="#" alt="Graph 1" /></td>
<td><img src="#" alt="Graph 2" /></td>
</tr>
<tr>
<td>demeaned levels</td>
<td><img src="#" alt="Graph 3" /></td>
<td><img src="#" alt="Graph 4" /></td>
</tr>
<tr>
<td>detrended levels</td>
<td><img src="#" alt="Graph 5" /></td>
<td><img src="#" alt="Graph 6" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Real GNP</th>
<th>log (Real GNP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>logs</td>
<td><img src="#" alt="Graph 7" /></td>
<td><img src="#" alt="Graph 8" /></td>
</tr>
<tr>
<td>demeaned logs</td>
<td><img src="#" alt="Graph 9" /></td>
<td><img src="#" alt="Graph 10" /></td>
</tr>
<tr>
<td>detrended logs</td>
<td><img src="#" alt="Graph 11" /></td>
<td><img src="#" alt="Graph 12" /></td>
</tr>
</tbody>
</table>

Developed in previous sections to the each of the fourteen U.S. macroeconomic time series of the N&P database. In particular, we estimate its order of summability with $\delta^*_t$ and derive the subsampling confidence intervals, denoted by $(I^*_L, I^*_U)$. The same quantities have been computed for the demeaned and detrended time series, denoted by $\hat{\delta}^*_\text{dem}$ and $\hat{\delta}^*_\text{det}$, respectively. The associated bounds of the confidence intervals are denoted by $(I^*_L, I^*_U)$ as well. The results associated to its levels and logarithms are shown in Tables 13 and 14, respectively.

Focusing on the second column of Table 13, it is immediately seen that the estimated order of summability is similar, around 0.5-0.6, for almost the fourteen macroeconomic variables. In particular, if we look at results for real GNP, the resemblance of the conclusions that can be extracted from the estimation and the first graphical exercise are evident. In the extremes, the index of industrial production is the variable with a higher estimated order of summability, around 1; and the velocity of money is the one with the lower estimated order, being it 0.35. It is particularly bright the narrowness of practically all the confidence intervals shown in columns three and four of the same table. However, these nice results should be taken with caution. Although not reported here for the sake of space, Monte Carlo experiments evidenced us that, at least in finite samples, the estimated order of summability of several $S(1)$ processes with a mean different from zero tends to be around 0.5, and the subsampling intervals are quite narrow. In other words, the
Table 11: Real GNP. Graphs of the variances of $S_n$

<table>
<thead>
<tr>
<th></th>
<th>$\delta_l = 0.5$</th>
<th>$\delta_l = 0.55$</th>
<th>$\delta_l = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real GNP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Demeaned Real GNP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Detrended Real GNP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td><img src="image9" alt="Graph" /></td>
</tr>
</tbody>
</table>
Table 12: log (Real GNP). Graphs of the variances of $S_n$

<table>
<thead>
<tr>
<th>log (Real GNP)</th>
<th>$\delta_l = 0.5$</th>
<th>$\delta_l = 0.55$</th>
<th>$\delta_l = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demeaned log (Real GNP)</td>
<td>$\delta_{dem} = 1.3$</td>
<td>$\delta_{dem} = 1.4$</td>
<td>$\delta_{dem} = 1.5$</td>
</tr>
<tr>
<td>Detrended log (Real GNP)</td>
<td>$\delta_{det} = 1.1$</td>
<td>$\delta_{det} = 1.2$</td>
<td>$\delta_{det} = 1.3$</td>
</tr>
</tbody>
</table>
### Table 13: Order of Summability. Estimation and Inference I

<table>
<thead>
<tr>
<th>Variable</th>
<th>levels</th>
<th>mean</th>
<th>trend</th>
<th>trend squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\delta}_1$</td>
<td>$I_L^*$</td>
<td>$I_U^*$</td>
<td>$\hat{\delta}_m$</td>
</tr>
<tr>
<td>cpi</td>
<td>0.588</td>
<td>0.514</td>
<td>0.663</td>
<td>0.700</td>
</tr>
<tr>
<td>employ</td>
<td>0.638</td>
<td>0.554</td>
<td>0.722</td>
<td>1.609</td>
</tr>
<tr>
<td>gnpdefl</td>
<td>0.623</td>
<td>0.541</td>
<td>0.705</td>
<td>1.437</td>
</tr>
<tr>
<td>nomgnp</td>
<td>0.915</td>
<td>0.681</td>
<td>1.150</td>
<td>1.687</td>
</tr>
<tr>
<td>interest</td>
<td>0.546</td>
<td>0.486</td>
<td>0.605</td>
<td>1.103</td>
</tr>
<tr>
<td>indprod</td>
<td>1.011</td>
<td>0.736</td>
<td>1.286</td>
<td>1.447</td>
</tr>
<tr>
<td>gnppe</td>
<td>0.580</td>
<td>0.512</td>
<td>0.649</td>
<td>1.535</td>
</tr>
<tr>
<td>realgnp</td>
<td>0.681</td>
<td>0.574</td>
<td>0.788</td>
<td>1.476</td>
</tr>
<tr>
<td>wages</td>
<td>0.803</td>
<td>0.628</td>
<td>0.978</td>
<td>1.540</td>
</tr>
<tr>
<td>rwages</td>
<td>0.614</td>
<td>0.545</td>
<td>0.683</td>
<td>1.128</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>0.675</td>
<td>0.534</td>
<td>0.816</td>
<td>1.184</td>
</tr>
<tr>
<td>unemploy</td>
<td>0.660</td>
<td>0.367</td>
<td>0.953</td>
<td>0.493</td>
</tr>
<tr>
<td>velocity</td>
<td>0.345</td>
<td>0.238</td>
<td>0.453</td>
<td>0.799</td>
</tr>
<tr>
<td>money</td>
<td>1.070</td>
<td>0.760</td>
<td>1.380</td>
<td>1.752</td>
</tr>
</tbody>
</table>

$\hat{\delta}_i$, $\hat{\delta}_{dem}$, and $\hat{\delta}_{det}$ are the values of the estimator of the order of summability of the time series in levels, demeaned levels and detrended levels, respectively. $I_L^*$ and $I_U^*$ denotes the lower and upper bounds of the corresponding subsampling intervals.
Table 14: Order of Summability. Estimation and Inference II

<table>
<thead>
<tr>
<th>Variable</th>
<th>log</th>
<th>log mean</th>
<th>log trend</th>
<th>log trend squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\delta}_U^*$</td>
<td>$I_L^*$</td>
<td>$I_U^*$</td>
<td>$\hat{\delta}_{lm}^*$</td>
</tr>
<tr>
<td>cpi</td>
<td>0.521</td>
<td>0.502</td>
<td>0.540</td>
<td>0.599</td>
</tr>
<tr>
<td>employ</td>
<td>0.512</td>
<td>0.504</td>
<td>0.519</td>
<td>1.481</td>
</tr>
<tr>
<td>gnpdefl</td>
<td>0.527</td>
<td>0.507</td>
<td>0.547</td>
<td>1.307</td>
</tr>
<tr>
<td>nomgnp</td>
<td>0.528</td>
<td>0.511</td>
<td>0.545</td>
<td>1.294</td>
</tr>
<tr>
<td>interest</td>
<td>0.546</td>
<td>0.486</td>
<td>0.605</td>
<td>1.047</td>
</tr>
<tr>
<td>indprod</td>
<td>1.097</td>
<td>0.783</td>
<td>1.411</td>
<td>0.990</td>
</tr>
<tr>
<td>gnpce</td>
<td>0.509</td>
<td>0.501</td>
<td>0.517</td>
<td>1.443</td>
</tr>
<tr>
<td>realgnp</td>
<td>0.530</td>
<td>0.511</td>
<td>0.548</td>
<td>1.296</td>
</tr>
<tr>
<td>wages</td>
<td>0.536</td>
<td>0.514</td>
<td>0.557</td>
<td>1.253</td>
</tr>
<tr>
<td>rwages</td>
<td>0.531</td>
<td>0.513</td>
<td>0.550</td>
<td>1.018</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>0.561</td>
<td>0.504</td>
<td>0.618</td>
<td>1.006</td>
</tr>
<tr>
<td>unemploy</td>
<td>0.563</td>
<td>0.323</td>
<td>0.802</td>
<td>0.275</td>
</tr>
<tr>
<td>velocity</td>
<td>0.366</td>
<td>0.227</td>
<td>0.505</td>
<td>0.933</td>
</tr>
<tr>
<td>money</td>
<td>0.705</td>
<td>0.594</td>
<td>0.816</td>
<td>1.236</td>
</tr>
</tbody>
</table>

$\hat{\delta}_i^*$ , $\hat{\delta}_{dem}^*$ , and $\hat{\delta}_{det}^*$ are the values of the estimator of the order of summability of the time series in logs, demeaned logs and detrended logs, respectively. $I_L^*$ and $I_U^*$ denotes the lower and upper bounds of the corresponding subsampling intervals.
Table 15: Order of Summability. Estimation and Inference II

<table>
<thead>
<tr>
<th>Variable</th>
<th>log</th>
<th>log mean</th>
<th>log trend</th>
<th>log trend squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\delta}_{I_U}^*$</td>
<td>$I_L^*$</td>
<td>$I_U^*$</td>
<td>$\hat{\delta}_{lm}^*$</td>
</tr>
<tr>
<td>Δcpi</td>
<td>0.521</td>
<td>0.502</td>
<td>0.540</td>
<td>0.019</td>
</tr>
<tr>
<td>Δemploy</td>
<td>0.512</td>
<td>0.504</td>
<td>0.520</td>
<td>-0.105</td>
</tr>
<tr>
<td>Δgnpdefl</td>
<td>0.529</td>
<td>0.509</td>
<td>0.549</td>
<td>0.377</td>
</tr>
<tr>
<td>Δnomgnp</td>
<td>0.528</td>
<td>0.511</td>
<td>0.545</td>
<td>0.306</td>
</tr>
<tr>
<td>Δinterest</td>
<td>0.535</td>
<td>0.490</td>
<td>0.580</td>
<td>0.273</td>
</tr>
<tr>
<td>Δindprod</td>
<td>1.107</td>
<td>0.789</td>
<td>1.425</td>
<td>-0.302</td>
</tr>
<tr>
<td>Δgnppc</td>
<td>0.509</td>
<td>0.501</td>
<td>0.518</td>
<td>0.933</td>
</tr>
<tr>
<td>Δrealgnp</td>
<td>0.529</td>
<td>0.511</td>
<td>0.547</td>
<td>0.902</td>
</tr>
<tr>
<td>Δwages</td>
<td>0.535</td>
<td>0.514</td>
<td>0.555</td>
<td>1.086</td>
</tr>
<tr>
<td>Δrwages</td>
<td>0.528</td>
<td>0.511</td>
<td>0.545</td>
<td>-0.018</td>
</tr>
<tr>
<td>ΔS&amp;P</td>
<td>0.554</td>
<td>0.504</td>
<td>0.604</td>
<td>0.016</td>
</tr>
<tr>
<td>Δunemploy</td>
<td>0.514</td>
<td>0.249</td>
<td>0.778</td>
<td>-0.397</td>
</tr>
<tr>
<td>Δvelocity</td>
<td>0.375</td>
<td>0.240</td>
<td>0.510</td>
<td>0.305</td>
</tr>
<tr>
<td>Δmoney</td>
<td>0.693</td>
<td>0.593</td>
<td>0.794</td>
<td>0.044</td>
</tr>
</tbody>
</table>

$\hat{\delta}_{li}$, $\hat{\delta}_{ldem}$, and $\hat{\delta}_{ldet}$ are the values of the estimator of the order of summability of the time series in logs, demeaned logs and detrended logs, respectively. $I_L^*$ and $I_U^*$ denotes the lower and upper bounds of the corresponding subsampling intervals.
untreated deterministic components introduce biases in the estimation process of the true order of summability. Comparing this experimental evidence with the empirical results in columns two to four of Table 13, we really believe that some attention must be paid to the deterministic components of the time series, mainly to the mean. With this objective, columns five to ten of Table 13 show the point and interval estimates of $\delta$ for the demeaned and detrended cases. It is remarkable the fact that the variable with a lower order of summability is the unemployment rate in all cases; while several variables, like nominal and real GNP, industrial production, wages or stock of money, share the highest orders of summability. Moreover, it is noteworthy that for almost all time series, the estimated orders of summability in the demeaned and detrended cases are similar. Some differences are found, however, for employment, real GNP, and nominal wages. Even so, because of its associated $\delta^*$ around 0.6, we prefer to focus on the demeaned case. To opt for the trend case an initial estimated order, $\delta_1^*$, about 1.5 would be expected. Finally, we can see, in Table 14, that similar conclusions are extracted when analyzing the logarithms to the macroeconomic time series.

Overall, the estimated orders of summability for the fourteen macroeconomic variables seem to be quite reasonable in economic and econometric terms. Regarding the later aspect of the empirical exercice, we would like to highlight the similarities of our results with those found in the fractional literature. With respect the economic content of the results, note that variables like employment, real and nominal GNP, industrial production, or nominal money have similar orders of summability and higher than those of unemployment or velocity of money. Particularly interesting, for the coming multivariate study, will be the fact that the logarithms of demeaned nominal GNP and nominal money have the same estimated order of summability.

### 6.2 Quantitative Theory of Money

Previous univariate results incite to verify the quantitative theory of money for the U.S. economy. The logarithms of nominal GNP and nominal money when demeaned and detrended have a highly similar order of summability, which implies to have a balanced regression between them. The joint temporal evolution of the nominal GNP and nominal money, as well as its demeaned and detrended versions, are illustrated in Table 16. As it can be seen, both time series move together along time, although the relationship, if any, seems to be less strengthen after the Second World War.

Next, we proceed to determine, using the techniques proposed in previous sections, whether the closely joint evolution of the time series of interest is a long run relationship or, by the contrary, it is a spurious finding. Specifically, we run three different sets of regressions. The first one deals with the time series when no control for the deterministic components has been undertaken.
However, as shown in the univariate analysis, it seems to be necessary to demean the univariate time series. Anyway, in a regression exercise, we can, alternatively, introduce an intercept to account for constant terms in the DGPs of the variables involved. The conclusions from this first set of regressions can be elicited from Table 17. In columns 2 and 4, results obtained when no deterministic components are taken into consideration are reported. In this case, the estimates point at a strong co-summable regression since $\hat{\delta}^* = 0.275$, when log(gnp) explains log(m2), and $\hat{\delta}^* = 0.229$, in the reverse specification; and zero lies in the estimated confidence intervals in both cases. When an intercept is introduced in these regressions, opposite conclusions are drawn. In particular, we see in columns 3 and 5 of Table 17 that $\hat{\delta}^* = 1.445$, when regressing log(gnp) on log(m2), and $\hat{\delta}^* = 1.553$, when log(m2) is dependent. However, these regressions, at least those in columns 2 and 4, are not balanced as it can be seen in Table 14.

To confirm or refuse this first set of results with an alternative treatment of the deterministic elements, we run two more sets of regressions. In the second and third sets, we deal with the demeaned and detrended time series, respectively. The associated evidence to these regression can be obtained from Tables 18 and 19.

From both tables, we found feeble evidence supporting the quantitative theory of money. The
most favorable results are found when demeaned variables are related and a constant term is introduced. The point estimates in those cases, reported in columns 3 and 5 of Table 18, are $\hat{\delta}_u^* = 0.637$ and $\hat{\delta}_u^* = 0.569$ when $dem(\log(gnp))$ and $det(\log(m2))$ are treated as endogenous, respectively. Besides, the corresponding confidence intervals hardly contain the estimated order of summability of $dem(\log(gnp))$ and $det(\log(m2))$. In all the other cases, these estimated orders of summability are definitely contained in the confidence intervals. Hence, at most, we found a weak co-summable relationship between money and GNP; which agrees with the estimated order of summability of the velocity of money in previous section and disagree the monetarist position.

### 7 Conclusions

The order of integration of non-linear stochastic processes is not always well defined. Hence, stochastic properties of non-linear time series cannot be summarized using the concept of order of integration. Additionally, in a multivariate environment, this lack of a proper definition has, at least, two important worrying consequences. First, it is not possible to characterize the balancedness of a non-linear regression, which is a necessary condition for an appropriate model specification. And, second, co-integration cannot be directly extended to deal with non-linear long run relationships.
Shortly, non-stationarities in non-linear environments cannot be directly studied using the standard ideas of integration and co-integration.

In this paper, we have proposed to use the concept of order of summability. It has been proved that it is a generalization of the order of integration, measures the persistence as well as the evolution of the variance of stochastic processes, controls the balancedness of non-linear regressions, and can be used to generalize co-integration for non-linear processes by defining co-summability.

On the practical side of our proposal, econometric tools have been proposed to estimate and carry out inferences on the unknown order of summability of observed time series. The performance of this machinery has been investigated through Monte Carlo experiments, which show a successfully effectiveness in practice. An empirical application has shown how to use these new techniques to test for economic hypothesis that involve non-linear transformations of non-stationary time series.
References


