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 permision yet)

October 28, 2010

Abstract

The order of integration is valid to characterize linear processes; but it is not appropriate for non-linear worlds. We propose the concept of summability (a re-scaled partial sum of the process being Op(1)) to handle non-linearities. The paper shows that this new concept, S(δ): (i) generalizes I(δ); (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, δ, are provided. The finite sample performance of these methods is analyzed via a Monte Carlo experiment. The paper finishes with the estimation of the degree of summability for the Nelson-Plosser extended database.

Keywords: Co-integration, Co-summability, 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 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 Teräsvirta (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 problem of misspecification, which is 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 check the balancedness of a 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 $f(x_t, \theta_0)$, and hence that 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. 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 Section 5, the use of the proposed tools is shown with an empirical application. Finally, Section 6 is devoted to some concluding remarks. All proofs are collected in the Appendix.

2 Order of Integration and Non-linear Processes

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 dth 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 $I(0)$ process? Attempts to give a definition for $I(0)$ processes exists in the literature. Engle and Granger (1987) give the following characterization.

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, ρ_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 $I(0)$ process, the theoretical and the practical:

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

Characterization 3: *Practical Definition of I(0)*: y_t is $I(0)$ if it is generated by a stationary autoregressive model $a(B)y_t = e_t$, 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 $I(0)$ as follows.

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

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

Definition 2 *: A stochastic process y_t which satisfies*

$$y_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \qquad C(L) = \sum_{j=0}^{\infty} c_j L^j,$$

is called I(0) if

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

ε_t is i.i.d. with zero mean and $\sigma_\varepsilon^2 = E(\varepsilon_0^2) < \infty$.

As stated in Davidson (1999) "it is clear that $I(0)$, as commonly understood, is a property of linear models. Let's state this observation more forcefully: $I(0)$, 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 $I(0)$ process is the first difference of an $I(1)$ 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 $\sigma_n^2 = Var(\sum_{t=1}^{n} y_t)$, converges weakly to standard Brownian motion B as $n \to \infty$.*

In other words, the standardized partial sums of the series must satisfy the functional central limit theorem (FCLT).

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

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

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

(i) $E[y_t] = 0$

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

(iii) $\gamma(k) = Cov(y_t, y_{t-k}) = \sigma_z^2$ for all $k > 0$.

Since it is a strictly stationary process, one could think that it is $I(0)$. However, the autocovariance function is not absolutely summable and its spectrum does not satisfy the above characterizations of an $I(0)$ process[1]. Moreover, it cannot be $I(0)$ as described in Definitions 2 and 3[2].

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

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

Hence, the spectral density is

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

which diverges for all λ.

[2]Assume that y_t is $I(0)$ as described in Definition 2. Then, $y_t = c(L)\varepsilon_t$, where ε_t is $i.i.d.$ Moreover, the following alternative autoregressive representation exists, $a(L)y_t = \varepsilon_t$, with $a(L) = c(L)^{-1}$. Equivalently, $\varepsilon_t = a(L)z + a(L)e_t$, 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

$$Y_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{[n\xi]} (y_t - E(y_t)) = \frac{1}{\sqrt{n}\sqrt{(n\sigma_z^2 + \sigma_e^2)}} \sum_{t=1}^{[n\xi]} (z + e_t) \nRightarrow B.$$

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 an i.i.d. and a Random Walk*

Let

$$w_t = x_t \eta_t,$$

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

$$x_t = x_{t-1} + \varepsilon_t, \tag{2}$$

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

(i) $E[w_t] = 0$

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

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

It 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 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,

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

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

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

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

As pointed out by Granger (1995), 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*

Let

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

where x_t and η_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,

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

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

and

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

where $\mu_4 = E[\eta_t^4]$. This means that not only the variances but 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 $I(1)$ too. However,

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

Again any of the above characterizations or definitions of $I(0)$ can be applied to Δx_t^2 or $\Delta^d x_t^2$.

Example 6 : *Product of Indicator Function and Random Walk*

Let

$$h_t = 1(v_t \leq \gamma)x_t,$$

where v_t is stationary and ergodic, $1(\cdot)$ is the indicator function, and x_t is the random walk defined in (2). This is another example where the concept of integrability is difficult to apply. Its variance and autocovariances depend on time, hence, one would think that h_t is $I(1)$. However, again, the variance of the first difference

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

depends on t. And, in fact, it can be considered, once again, that h_t has an infinite order of integration.

Example 7 : *Park and Phillips (1999, 2001)*

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

Example 8 : *Stochastic Unit Root and Explosive Processes*

Consider, on one hand, a stochastic unit root process

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \tag{3}$$

where $\rho_t \sim i.i.d.(\rho, \omega^2) \perp \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. On the other hand, think about an explosive process

$$z_t = \alpha z_{t-1} + \xi_t, \tag{4}$$

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

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.

As mentioned in the Introduction, this lack of a proper definition for non-linear univariate time series translates into 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. 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; and 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 straightforward generalized. 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.

2.3 Attempts to Solve the Problem

Problems highlighted in the previous section make necessary to extend the concept of integratedness to allow for more general types of processes and long run relationships. 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*

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

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 $c_h = O(h^{-\gamma})$, $\gamma > 0$, then the process under study can be said to be SMM of order γ. 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

3.1 Order of Summability Definition

The idea of order of summability of a stochastic process was initially introduced in a heuristic way in Gonzalo and Pitarakis (2006) when dealing with threshold effects in co-integrating regressions.

In this section, the concept of summability is formalized and its generality, usefulness and simplicity are asserted.

Definition 6 *: Summability of order δ: A stochastic process y_t with positive variance is said to be summable of order δ, symbolically represented as $S(\delta)$, if there exist a nonrandom sequence m_t such that*

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

where δ is the minimum number that makes S_n bounded in probability[3] *and $L(n)$ is a slowly-varying function*[4]*.*

Note that, when possible, the order of summability will be determined by some Central Limit result. In the standard Central Limit Theorem –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, by the Functional Central Limit Theorem –FCLT– and the Continuous Mapping Theorem –CMT–, $\delta = 1$ 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[5].

As we emphasized above, the concept of integrability does not apply to some leading models. And it is here where the concept of summability becomes a useful device.

3.2 Examples

From an univariate perspective, in all processes considered in Examples 1-8 the order of integration was difficult to establish. Next, it is shown that the order of summability can be directly obtained

[3] S_n is said to be bounded in probability if, for every $\epsilon > 0$, there exists a positive real number M_ϵ such that $P(|S_n| \geq M_\epsilon) \leq \epsilon$, for all n. It will be assumed in the following that boundedness in probability refers to $O_p(1)$ but not $o_p(1)$.

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

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

is said to be slowly varying (in Karatama's sense).

[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

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

This is a case where $L(n) = (1/\log n)^{1/2}$, and not just a constant.

for almost all of the above examples.

Summability in Example 1 (*α-stable distributed process*): Let y_t be an α-stable Levy distributed process. It can be shown that when the Levy distribution is symmetric with $0 < \alpha \leq 2$ the normalized sum

$$S_n = \frac{1}{n^{\frac{1}{\alpha}}} \sum_{t=1}^{n} y_t,$$

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 $S(0.5)$.

Summability in Example 2 (*An i.i.d. plus a random variable*): It is easy to see that

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

Therefore, y_t is $S(0.5)$ and $L(n) = 1$.

In Examples 3-6, we considered several non-linear models. Next, we show that the order of summability of almost all 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

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

This means that w_t is $S(0.5)$ with $L(n) = 1/\sigma_\varepsilon$.

Summability in Example 4 (*Product of i.i.d. squared and random walk*): For q_t note that,

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

Then, by the Chebyshev's inequality,

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

which implies that q_t is $S(1)$.

Comparing Examples (3) and (4), we can see that summability is taken into account not only the covariance structure but 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

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

Hence, we can conclude that x_t^2 is $S(1.5)$.

Summability in Example 6 (*Product of indicator function and random walk*): In this case,

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

meaning that h_t is $S(1)$ with $L(n) = 1/p\sigma_\varepsilon$.

Summability in Example 7 (*Park and Phillips (1999, 2001)*): The order of summability of the processes considered in this example can be obtained by using the asymptotic theory developed in Park and Phillips (1999). Specifically, it can be shown that $e^{-x_t^2} \sim S(0)$, $1/(1+x_t^2) \sim S(0)$, $\log(|x_t|) \sim S(0.5)$ and $(1+e^{-x_t})^{-1} \sim S(0.5)$.

Summability in Example 8 (*STUR and Explosive processes*): Consider the STUR process defined in (3). To simplify matters, let $\rho_t \sim i.i.d.(1,1)$, i.e. set $\rho = \omega^2 = 1$. From Leybourne, McCabe and Tremayne (1996), it can be shown that

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

With respect the explosive process (4), from White (1958) it can be shown that

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

Strictly speaking, the order of summability of y_t and z_t will depend on n; hence, a unique δ does not exist for these two processes.

As Example 8 states, the order of summability of some processes will not be properly defined. Nevertheless, as Examples 1-7 show, summability allows for studying the stochastic properties of many interesting non-stationary time series without imposing linear structures.

3.3 Integrability implies Summability

In this subsection, the relationship between integrability and summability is discussed. For the former and the latter concepts Definitions 2 and 6 will be used, respectively. Definition 2 can be considered as the most general practical definition of an $I(0)$ process. Furthermore, assuming that,

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

it 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 $S(d)$.*

Remark: Proposition 1 states that $I(d)$ processes are $S(d)$ when d is a non-negative real number. The same is not true when the order of integration is negative as it will be shown in the following proposition.

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

Remark: Since negative integer orders of integration are not the most important and/or interesting ones, only cases where $d \geq 0$ will be considered. In this sense, integrated processes can be considered as particular cases of summable processes.

3.4 Algebra of Summability

By definition, the d cumulation of an $I(d_1)$ process is $I(d_1 + d)$ and the d difference $I(d_1 - d)$. In this section we investigate whether this properties are true for summable processes.

Proposition 3 *: A stochastic process y_t such that $\inf(var(y_t)) > 0$ cannot have an order of summability $\delta < -1/2$.*

Remark: Proposition 3 is the reason for which $I(-d)$ processes, for $d = 1, 2, ... < \infty$, are all $S(-1/2)$. This shows that in general we cannot say that cumulating (differencing) d times a stochastic process its order of summability will increase (decrease) in d units.

Integrated processes with negative orders of integration are not the unique processes for which cumulating or taking differences the order of summability will not change accordingly. This feature is also present in nonlinear processes that do not reach the lower bound $-1/2$. Consider, for instance, the process $y_t = t\varepsilon_t$, with $\varepsilon_t \sim iid(0, 1)$, which is $S(1)$ given that

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} y_t = \frac{1}{n^{3/2}} \sum_{t=1}^{n} t\varepsilon_t \Longrightarrow \int_0^1 r dW(r).$$

The sum of its first difference is

$$\sum_{t=1}^{n} \Delta y_t = \sum_{t=1}^{n} t\Delta\varepsilon_t + \sum_{t=1}^{n} \varepsilon_{t-1} = -\sum_{t=1}^{n} \varepsilon_{t-1} + n\varepsilon_n + \sum_{t=1}^{n} \varepsilon_{t-1} = n\varepsilon_n,$$

meaning that Δy_t is $S(1/2)$. Hence, the first difference of a $S(1)$ process has not to be $S(0)$. The same example shows that if we cumulate once a time series, Δy_t in this case, its order of summability could not increase in one unit; in fact, in the present example, it only increases in 0.5 units given that $y_t \sim S(1)$.

Remark: Although in general the first difference of a time series will not reduce its order of summability in one unit, if we knew the order of boundedness of the time series, the order of summability of its first difference can be directly determined through the fact that

$$\sum_{t=1}^{n} y_t = y_n - y_0.$$

Hence, if $y_n - y_0 = O_p(n^{\gamma})$ then $\Delta y_t \sim S(\gamma - 1/2)$. With respect cumulations, the following proposition can be stated.

Proposition 4 *: Let $y_t \sim S(\delta)$; then, $\Delta^{-1} y_t \sim S(\delta + \alpha)$ where $\alpha \in [0, 1]$.*

Remark: Proposition 4 ensures that cumulating once a time series its order of summability will increase, at most, in one unit.

3.5 Balancedness and Co-summability

Next, it is shown how the concept of summability can be used to define balancedness of non-linear regressions and directly generalize the standard concept of co-integration to non-linear long run relationships.

Definition 7 *: A regression model of the form*

$$y_t = f(x_t, \theta) + u_t, \tag{5}$$

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

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 –a direct extension of the idea of co-integration.

Definition 8 *: Two summable stochastic processes, $y_t \sim S(\delta_y)$ and $x_t \sim S(\delta_x)$, will be said to be co-summable if there exists a transformation $f(x_t, \theta)$ such that $f(x_t, \theta) \sim S(\delta_y)$ and $u_t = y_t - f(x_t, \theta)$ is $S(\delta_u)$, with $\delta_u = \delta_y - \delta$ and $\delta > 0$. In short, $(y_t, z_t) \sim CS(\delta_y, \delta_y - \delta)$.*

Co-summable processes will share an equilibrium relationship in the long run, i.e. an attractor $y_t = f(x_t, \theta)$ that can be linear or not. This type of equilibrium relationships will be usually established by the economic theory and have 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: Estimation and Inference

4.1 Order of Summability Estimation

If a stochastic process, y_t, satisfies

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1), \tag{6}$$

we say that y_t is summable of order δ. For a δ-summable stochastic process, it should be true that

$$U_n = \log S_n^2 = \log \left(n^{-(1+2\delta)} L^2(n) \left(\sum_{t=1}^{n} (y_t - m_t) \right)^2 \right) = O_p(1). \tag{7}$$

Equation (7) can be written as

$$\begin{aligned} U_n &= -(1+2\delta)\log n + 2\log L(n) + \log \left(\sum_{t=1}^{n} (y_t - m_t) \right)^2 \\ &= -\beta \log n - \alpha + Y_n, \end{aligned}$$

where $\beta = 1 + 2\delta$, $\alpha = -2\log L(n)$, and $Y_n = \log \left(\sum_{t=1}^{n} (y_t - m_t) \right)^2$. In regression model form,

$$Y_k = \alpha + \beta \log k + U_k, \tag{8}$$

with $U_k = O_p(1)$.

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

$$\hat{\beta}_1 = \frac{\sum_{k=1}^{n} Y_k \log k}{\sum_{k=1}^{n} \log^2 k} = \beta + \frac{\sum_{k=1}^{n} (\alpha + U_k) \log k}{\sum_{k=1}^{n} \log^2 k}, \tag{9}$$

or

$$\hat{\beta}_2 = \frac{\sum_{k=1}^{n} (Y_k - \bar{Y})(\log k - \overline{\log n})}{\sum_{k=1}^{n} (\log k - \overline{\log n})^2} = \beta + \frac{\sum_{k=1}^{n} (U_k - \bar{U})(\log k - \overline{\log n})}{\sum_{k=1}^{n} (\log k - \overline{\log n})^2}, \tag{10}$$

where $\bar{Y} = (1/n) \left(\sum_{k=1}^{n} Y_k \right)$, $\overline{\log n} = (1/n) \left(\sum_{k=1}^{n} \log k \right)$. Given that $\beta = 1 + 2\delta$, the estimator of δ is given by

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

Since $\log L(k) = o(\log k)$, $\alpha = -2\log L(k)$ 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 α is constant; i.e. $\alpha = -2\log \omega$, ω being some finite real number.

4.2 Asymptotic Properties

In this section, the asymptotic properties of $\hat{\beta}_1$ and $\hat{\beta}_2$ will be studied. Let $x_t = y_t - m_t$. We start focusing on

$$\hat{\beta}_1 - \beta = \frac{\sum_{k=1}^{n} V_k \log k}{\sum_{k=1}^{n} \log^2 k},$$

with $V_k = \alpha + U_k$.

Proposition 5 *: McElroy and Politis:* $\hat{\beta}_1 - \beta = o_p(1)$.

Remark: 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, to the best of our knowledge, an asymptotic distribution for $\hat{\beta}_1$ has not been derived. We address this issue in the following proposition.

Proposition 6 *: If*

$$S_n(r,\delta) = \frac{1}{n^{1/2+\delta}\omega} \sum_{t=1}^{[nr]} x_t \Longrightarrow D_x(r,\delta),$$

where $D_x(r,\delta)$ *is some random variable with positive variance, then*

$$\log n(\hat{\beta}_1 - \beta) \Longrightarrow \alpha + \int_0^1 U_x(r,\delta)\, dr,$$

where $\alpha = -2\log\omega$ *and* $U_x(r,\delta) = \log\left[\left(r^{-1/2-\delta} D_x(r,\delta)\right)^2\right]$.

Remark: When the series under study is i.i.d.(0,1), for instance, by the FCLT

$$S_n = \frac{1}{n^{1/2}} \sum_{t=1}^{[nr]} x_t \Longrightarrow W(r).$$

Therefore,

$$\log n(\hat{\beta}_1 - \beta) \Longrightarrow \int_0^1 \log\left[\left(r^{-1/2} W(r)\right)^2\right] dr.$$

Similarly if the time series under consideration was a standard random walk, then

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

and

$$\log n(\hat{\beta}_1 - \beta) \Longrightarrow \int_0^1 \log\left[\left(r^{-3/2} \int_0^r W(r) dr\right)^2\right] dr.$$

As it is shown in Proposition 6, the asymptotic distribution of $\hat{\beta}_1$ will depend on the nuisance parameter α, unless $\alpha = 0$, i.e. $L(n) = 1$. Next, we study the asymptotic properties of

$$\hat{\beta}_2 = \frac{\sum_{k=1}^{n} (Y_k - \bar{Y})(\log k - \overline{\log n})}{\sum_{k=1}^{n} (\log k - \overline{\log n})^2},$$

Proposition 7 *: If*

$$S_n(r,\delta) = \frac{1}{n^{1/2+\delta}}\omega \sum_{t=1}^{[nr]} x_t \Longrightarrow D_x(r,\delta),$$

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

$$(\hat{\beta}_2 - \beta) \Longrightarrow \int_0^1 U_x(r,\delta)(1+\log r)\,dr,$$

where $U_x(r,\delta) = \log\left[\left(r^{-1/2-\delta} D_x(r,\delta)\right)^2\right]$.

Remark: When the series under study is i.i.d.(0,1), for instance, by the FCLT

$$S_n = \frac{1}{n^{1/2}} \sum_{t=1}^{[nr]} x_t \Longrightarrow W(r).$$

Therefore,

$$(\hat{\beta}_2 - \beta) \Longrightarrow \int_0^1 \log\left[\left(r^{-1/2} W(r)\right)^2\right](1+\log r)\,dr.$$

Similarly if the time series under consideration was a standard random walk, then

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

and

$$(\hat{\beta}_2 - \beta) \Longrightarrow \int_0^1 \log\left[\left(r^{-3/2}\int_0^r W(r)dr\right)^2\right](1+\log r)\,dr.$$

Note that there is still room for consistency of $\hat{\beta}_2$ if the limiting distribution is degenerated. We have not been able to show that. In fact, Monte Carlo simulations points out to inconsistency of $\hat{\beta}_2$; a similar result to the one encountered in spurious regressions with I(1) processes. It is worthmentioning at this point that the constant term, α, will not be identified, since boundedness in probability holds for any value of α. Since $\hat{\alpha} = \bar{Y} - \hat{\beta}_2\overline{\log n}$, the non-identification of α could be behind the findings concerning $\hat{\beta}_2$. These inconsistency results in the i.i.d. case seems to contradict results in McElory and Politis (2007).

Attending to this evidence, $\hat{\beta}_1$ has been chosen to be applied in the following. Remember, however, that when a constant must be introduced in the regression model, although $\hat{\beta}_1$ is still consistent its asymptotic distribution depends on the unknown nuisance parameter α. In order to get rid of it, we propose to estimate, instead of

$$Y_k = \alpha + \beta \log k + U_k,$$

the modified regression model

$$Y_k^* = \beta \log k + U_k^*,$$

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

$$\hat{\beta}_1^* = \frac{\sum_{k=1}^n Y_k^* \log k}{\sum_{k=1}^n \log^2 k},$$

satisfies the same asymptotic properties than those of $\hat{\beta}_1$ when $L(n) = 1$; i.e. it will be $\log n$-consistent with an asymptotic distribution which does not depend on the nuisance parameter α.

4.3 Subsampling Confidence Intervals

Although the asymptotic distribution of $\hat{\beta}_1^*$ does not depend on α, it depends on the true order of summability. Therefore, a unique limiting distribution is not available. Nevertheless, subsampling methods could be an alternative way to undertake 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 α-mixing coefficients sequence of $Z_{n,b,t} = \log b(\hat{\beta}_{n,b,t} - \beta)$.

Since the α-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

Let $x_t = \left(\sum_{j=1}^t \varepsilon_j\right)$ with $\varepsilon_t \sim iidN(0,1)$. The following DGPs will be studied.

Table 1: Data Generating Processes. Without Deterministic Components

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

Performance of subsampling is mainly measured by coverage probability, denoted CP, of two-sided nominal 95% symmetric[6] intervals for the parameter $\delta = (\beta - 1)/2$. We also look at the mean of the estimated order of summability and the median lower and upper bounds of the estimated confidence intervals. These measures are denoted by $\bar{\delta}$, I_{low}, and I_{up}, respectively. The experiment is based on 1000 replicas and we use three different sample sizes $n = \{100, 200, 500\}$. A subsample size $b = \sqrt{n}$ has been chosen[7]. Results are collected in Table 2.

Table 2: Performance of the subsampling methodology. Without Deterministic Components

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

CP denotes the coverage probability of two-sided nominal 95% symetric intervals. $\bar{\delta}$ denotes the mean of the $\hat{\delta}$'s. I_{low} and I_{up} denotes the median of the lower and upper bounds of the intervals, respectively.

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 increases the

[6] In the simulations we ran, two-sided symmetric confidence intervals performed much better than equally tailed confidence intervals.

[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.

coverage probability decreases for a given sample size. However, as expected, the higher the sample size the better the coverage probability. It is remarkable the precision of the point estimation $\bar{\delta}$ in all the cases, even with small sample sizes. The confidence intervals are somehow wide, at least in small samples. Nevertheless, the higher the sample size the narrower the confidence intervals. As a whole, the proposed machinery seems to work resonably well.

4.3.2 With Deterministic Components

Let

$$y_t = m_t + x_t,$$

where

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

Up to now, we have assumed that $m_t = 0$; however, it is not the case in practice. Actually, to deal with the effect of the deterministic components will be crucial to correctly estimate and infer the true order of summability δ. To be more precise, consider the following two situations:

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

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

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

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

Therefore, when $\delta < \gamma$ the order of the deterministic component dominates and it will be confused with the order of summability. Admittedly, even when $\delta > \gamma$, the deterministic components, when not properly considered, will affect the order of summability estimation, at least in finite samples. Although not reported here to save space, Monte Carlo experiments reveal the existence of an important bias effect when deterministic components are present and not suitably taken into consideration. Hence, a proper technique to deal with them is needed; mainly in economic application were deterministic components are known to be an important issue.

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

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - \hat{m}_t) \Longrightarrow D_x^*(\delta). \tag{11}$$

In this way, the true order of summability δ can be recovered.

Three usual parametric forms for m_t will be considered here. The constant term case $m_t = m$; a linear deterministic trend, $m_t = m_0 + m_1 t$, with m_0 and m_1 unknown; and a quadratic trend in which $m_t = m_0 + m_1 t + m_2 t^2$, with m_0, m_1 and m_2 unknown. For these three cases, proper treatment of the deterministic components, in the sense of equation (11), will be derived.

To study the performance of the subsampling confidence intervals for δ when using $\hat{x}_t = y_t - \hat{m}_t$ will be studied through simulations. The following DGPs will be considered.

Table 3: Data Generating Processes. With Deterministic Components

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

Case 1 : ***Constant Term:*** *Let*

$$y_t = m + x_{yt},$$

where m is a constant and $x_{yt} \sim S(\delta)$ in the sense that

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} x_{yt} \Longrightarrow D_x(\delta), \tag{12}$$

where $D_x(\delta)$ is a random variable with positive variance. The time series of interest is y_t. By definition of order of summability, we know that $y_t \sim S(\delta)$. Let us assume that δ is unknown and we only observe y_t. To demean y_t with its arithmetic mean, say $\bar{y}$, is problematic in this context because

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

Next proposition shows that the partial mean

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

is an alternative operational choice.

Proposition 8 : *Let*

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

If

$$y_t = m + x_{yt},$$

where m is a constant and

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{[nr]} x_{yt} \Longrightarrow D_x(r, \delta),$$

then

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - \hat{m}_t) \Longrightarrow D_x(1, \delta) - \int_0^1 r^{-1} D_x(r, \delta) dr.$$

Remark: As Proposition 8 shows,

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

effectively deals with the presence of a constant term in the DGP. Table 4 reports the performance of the subsampling confidence intervals after partially demeaning the processes described in Table 3 when $m_t = m = 10$. Results are independet to the particular choice of m.

Table 4: Performance of the subsampling methodology. With Constant Term

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

CP denotes the coverage probability of two-sided nominal 95% symetric intervals. $\bar{\delta}$ denotes the mean of the $\hat{\delta}$'s. I_{low} and I_{up} denotes the median of the lower and upper bounds of the intervals, respectively.

As it can be seen in Table 4, the results are similar or even better than those obtained without deterministic components. For this reason, we recommend to demean always the processes by using partial means.

Case 2 *: **Linear Trend:** Let*

$$y_t = m_0 + m_1 t + x_{yt},$$

where m_0 *and* m_1 *are unknown parameters and* $x_{yt} \sim S(\delta)$ *in the sense that*

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} x_{yt} \Longrightarrow D_x(\delta),$$

as before. Next Proposition shows how to deal with the deterministic components in this case.

Proposition 9 *: Let*

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

If

$$y_t = m_0 + m_1 t + x_{yt},$$

where m_0 *and* m_1 *are unknown parameters and*

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{[nr]} x_{yt} \Longrightarrow D_x(r, \delta),$$

then

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - \hat{m}_t) \Longrightarrow D_x\left(1, \delta\right) - 3 \int_0^1 r^{-1} D_x\left(r, \delta\right) dr + 2 \int_0^1 r^{-3/2+\delta} \left(\int_0^r s^{-1/2+\delta} D_x\left(s, \delta\right) ds \right) dr.$$

Remark: Note that the linear trend case, the appropriate $\hat{m}_t$ consists, basically, in a doble partial demeaning procedure. Table 5 summarizes the performance of the subsampling confidence intervals after properly detrending the DGPs in Table 3 when $m_t = m_0 + m_1 t = 10 + 2t$. As the in the previous case, results are independent of the particular choices for m_0 and m_1. All measures are as in Table 4. Again, results resemble those encoutered in the non-deterministic components case. This means, as stated in Proposition 9, that the proposed $\hat{m}_t$ in this case works correctly.

Table 5: Performance of the subsampling methodology. With Linear Trend

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

CP denotes the coverage probability of two-sided nominal 95% symetric intervals. $\bar{\delta}$ denotes the mean of the $\hat{\delta}$'s. I_{low} and I_{up} denotes the median of the lower and upper bounds of the intervals, respectively.

Case 3 : ***Quadratic Trend:*** *Let*

$$y_t = m_0 + m_1 t + m_2 t^2 + x_{yt},$$

where m_0 and m_1 are unknown parameters and $x_{yt} \sim S(\delta)$ in the sense that

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} x_{yt} \Longrightarrow D_x(\delta),$$

as before. It will be shown by means of simulations that

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

is a proper estimator to deal with the deterministic components in this case. Essentially, this transformation implies a triple partial demeaning procedure.

Results collected in Table 6 show similar conclusions to those described with the previous example. In all cases,

$$m_t = 10 + 2t + 3t^2,$$

although as before, results are invariant to the particular values of m_0, m_1 and m_2.

Table 6: Performance of the subsampling methodology. With Quadratic Trend

DGP	CP	$\bar{\delta}$	I_{low}	I_{up}	CP	$\bar{\delta}$	I_{low}	I_{up}	CP	$\bar{\delta}$	I_{low}	I_{up}
$S(\delta)$	$n=100$				$n=200$				$n=500$			
$1-S(0)$	0.749	0.530	-0.204	1.200	0.721	0.500	-0.146	1.081	0.695	0.466	-0.099	1.001
$2-S(1)$	0.954	0.892	0.216	1.488	0.958	0.904	0.319	1.431	0.953	0.933	0.422	1.373
$3-S(2)$	0.744	1.513	0.709	2.220	0.825	1.565	0.786	2.272	0.913	1.639	0.892	2.328
$4-S(0.5)$	0.920	0.674	-0.072	1.346	0.926	0.660	0.015	1.230	0.913	0.617	0.054	1.086
$5-S(1.5)$	0.761	1.111	0.328	1.817	0.790	1.151	0.434	1.814	0.818	1.220	0.557	1.852
$6-S(1)$	0.974	0.731	-0.134	1.483	0.982	0.765	-0.050	1.478	0.999	0.797	0.039	1.485
$7-S(0.7)$	0.982	0.743	0.053	1.351	0.983	0.735	0.146	1.244	0.990	0.735	0.227	1.174
$8-S(0.5)$	0.976	0.526	-0.198	1.197	0.971	0.542	-0.115	1.124	0.965	0.478	-0.089	1.005
$9-S(0.5)$	0.993	0.591	-0.191	1.290	0.992	0.586	-0.128	1.242	0.994	0.566	-0.072	1.146
$10-S(1)$	0.886	0.801	0.003	1.544	0.882	0.812	0.084	1.483	0.905	0.835	0.180	1.441
$11-S(1)$	0.898	0.755	-0.029	1.461	0.888	0.755	0.028	1.400	0.910	0.779	0.125	1.380
$12-S(0.5)$	0.822	0.728	0.169	1.211	0.816	0.714	0.245	1.138	0.769	0.712	0.333	1.036

CP denotes the coverage probability of two-sided nominal 95% symetric intervals. $\bar{\delta}$ denotes the mean of the $\hat{\delta}$'s. I_{low} and I_{up} denotes the median of the lower and upper bounds of the intervals, respectively.

Overall, it can be said that the methodology we have proposed in this section works reasonably well to determine the order of summability of univariate time series.

5 Empirical Application

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[8]. As a novelty, we do not impose any linearity assumption.

More precisely, we estimate the order of summability of the fourteen macroeconomic aggregates with $\hat{\delta}_1^*$ and derive the subsampling confidence intervals, denoted by (I_L^*, I_U^*). The same quantities

[8]The data have been taken from P.C.B. Phillips' webpage.

have been computed for the demeaned, linearly and quadratically detrended time series, denoted by $\hat{\delta}_m^*$, $\hat{\delta}_t^*$ and $\hat{\delta}_{t^2}^*$, respectively. The associated bounds of the confidence intervals are denoted by (I_L^*, I_U^*) as well. Results corresponding to its levels and logarithms are shown in Tables 7 and 8, respectively.

Table 7: Order of Summability. Estimation and Inference I

Variable	levels			mean			trend			quadratic trend		
	$\hat{\delta}_1^*$	I_L^*	I_U^*	$\hat{\delta}_m^*$	I_L^*	I_U^*	$\hat{\delta}_t^*$	I_L^*	I_U^*	$\hat{\delta}_{t^2}^*$	I_L^*	I_U^*
cpi	0.588	0.514	0.663	0.644	0.236	1.051	0.950	0.497	1.403	1.576	0.713	2.439
employ	0.638	0.554	0.722	1.601	0.962	2.240	0.759	0.212	1.305	0.649	0.242	1.056
gnpdefl	0.623	0.541	0.705	1.399	0.792	2.007	1.352	0.640	2.064	0.889	0.271	1.506
nomgnp	0.915	0.681	1.150	1.658	0.861	2.455	1.671	0.762	2.579	1.524	0.688	2.360
interest	0.546	0.486	0.605	1.035	0.453	1.617	0.994	0.330	1.657	1.092	0.485	1.698
indprod	1.011	0.736	1.286	1.458	0.933	1.982	1.384	0.694	2.074	1.345	0.550	2.139
gnppc	0.580	0.512	0.649	1.509	0.528	2.490	1.784	0.661	2.907	0.875	0.022	1.728
realgnp	0.681	0.574	0.788	1.452	0.812	2.091	2.064	0.814	3.315	1.273	0.390	2.156
wages	0.803	0.628	0.978	1.512	0.913	2.110	2.003	0.977	3.030	1.576	0.753	2.399
rwages	0.614	0.545	0.683	1.109	0.746	1.472	1.040	0.254	1.826	1.312	0.475	2.149
S&P	0.675	0.534	0.816	1.138	0.544	1.732	1.055	0.266	1.845	1.037	0.273	1.802
unemploy	0.660	0.367	0.953	0.401	-0.126	0.929	0.448	-0.190	1.088	0.670	0.124	1.216
velocity	0.345	0.238	0.453	0.837	0.602	1.071	1.142	0.484	1.799	0.377	-0.152	0.908
money	1.070	0.760	1.380	1.713	1.066	2.360	1.884	0.798	2.969	1.613	0.765	2.462

$\hat{\delta}_1^*$, $\hat{\delta}_m^*$, $\hat{\delta}_t^*$ and $\hat{\delta}_{t^2}^*$ are the estimated orders of summability of the time series in levels, demeaned, linearly and quadratically detrended, respectively. I_L^* and I_U^* denotes the lower and upper bounds of the corresponding subsampling intervals.

Table 8: Order of Summability. Estimation and Inference II

Variable	log			log mean			log trend			log quadratic trend		
	$\hat{\delta}^*_{l1}$	I^*_L	I^*_U	$\hat{\delta}^*_{lm}$	I^*_L	I^*_U	$\hat{\delta}^*_{lt}$	I^*_L	I^*_U	$\hat{\delta}^*_{lt^2}$	I^*_L	I^*_U
cpi	0.521	0.502	0.540	0.541	0.092	0.991	0.846	0.479	1.213	2.369	1.112	3.625
employ	0.512	0.504	0.519	1.486	0.916	2.056	0.803	0.129	1.477	0.579	0.185	0.973
gnpdefl	0.527	0.507	0.547	1.280	0.759	1.802	1.119	0.532	1.706	0.900	0.168	1.631
nomgnp	0.528	0.511	0.545	1.297	0.777	1.817	0.952	0.390	1.514	1.031	0.557	1.505
interest	0.546	0.486	0.605	0.976	0.485	1.467	0.934	0.359	1.509	1.031	0.353	1.709
indprod	1.097	0.783	1.411	1.038	0.737	1.339	0.418	-0.114	0.950	0.738	0.082	1.393
gnppc	0.509	0.501	0.517	1.423	0.446	2.400	1.669	0.602	2.737	0.938	0.278	1.599
realgnp	0.530	0.511	0.548	1.287	0.739	1.836	1.634	0.608	2.660	0.898	0.287	1.510
wages	0.536	0.514	0.557	1.249	0.799	1.698	1.707	0.718	2.695	0.961	0.341	1.580
rwages	0.531	0.513	0.550	1.009	0.699	1.318	0.805	0.107	1.504	1.070	0.320	1.821
S&P	0.561	0.504	0.618	0.972	0.485	1.459	0.817	0.100	1.534	0.702	0.121	1.283
unemploy	0.563	0.323	0.802	0.264	-0.439	0.967	0.162	-0.603	0.928	0.566	-0.040	1.173
velocity	0.366	0.227	0.505	0.958	0.613	1.303	1.044	0.404	1.685	0.576	-0.010	1.163
money	0.705	0.594	0.816	1.237	0.838	1.636	0.818	0.237	1.398	0.913	0.279	1.548

$\hat{\delta}^*_{l1}$, $\hat{\delta}^*_{lm}$, $\hat{\delta}^*_{lt}$ and $\hat{\delta}^*_{lt^2}$ are the estimated orders of summability of the time series in logs, in logs and demeaned, in logs and linearly and quadratically detrended, respectively. I^*_L and I^*_U denotes the lower and upper bounds of the corresponding subsampling intervals.

Focusing on the second column of Table 7, it is immediately seen that the estimated order of summability is similar, around 0.5-0.6, for almost all the fourteen macroeconomic variables. 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 untreated deterministic components introduce biases in the estimation process of the true order of summability, at least in finite samples. Comparing this experimental evidence with the empirical results in columns two to four of Table 7, we really believe that some attention must be paid to the deterministic components of the time series. With this objective, columns five to thirteen of Table 7 show the point and interval estimates of δ for the demeaned, linearly and quadratically detrended cases. It is remarkable that the variable with a lower order of summability is the unemployment rate in almost all cases; while several variables, like nominal and real GNP, industrial production, wages or stock of money, share the highest orders of summability.

It is noteworthy that for some of the aggregates the estimated orders of summability increase as deterministic components of higher orders are removed; nevertheless, they stabilizes at some degree of the deterministic components. This fact is in line with some simulation experiments which show that substracting trends of higher degree than the true one is not a drawback of the procedure. In effect, after the true degree has been considered, substracting trends of higher degrees implies stable estimated orders of summability around its true value.

Consider, for instance, real GNP. Its estimated order of summability increases as deterministic components are substracted up to the quadratic trend in which a reasonable order of summability is estimated. This circumstance agrees with a significant coefficient of a regression of the growth rate of real GNP on a linear trend that is found with these data. As another example, examine the unemployment rate. In this case, the estimated order of summability seems to be stable for all the studied deterministic specifications. As emphasized above, to partially demean the time series is always an advisable practice. Therefore, results concerning to the mean case seem to be a reliable choice. Similar conclusions can be drawn when analyzing the logarithms of the time series as it can be seen in Table 8.

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 exercise, 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.

6 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, this lack of a proper definition has two important multivariate repercussions. 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 designed 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 reasonable effectiveness in practice. An empirical application has shown how to use these new techniques to infer the order of summability of the main macroeconomic aggregates of the U.S. economy.

7 Appendix

Proof of Proposition 1: The proof will be divided in four parts.

(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

$$y_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t, \tag{13}$$

where

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

Now, the sum of x_t can be computed as

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

and the following CLT holds

$$\frac{1}{n^{\frac{1}{2}}} \sum_{t=1}^{n} y_t \xrightarrow{d} N(0, \sigma_\varepsilon^2 C(1)^2),$$

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

For parts (ii)-(iv) let

$$\Delta^d y_t = u_t,$$

where u_t have zero mean, are *i.i.d.*, and $E\left|u_t\right|^4 < \infty$ for $r \geq \max\left[4, -8d_0/\left(1+2d_0\right)\right]$ with $d_0 \in (-1/2, 1/2]$.

(ii) $d \in (0, 1/2)$. It was shown in Hosking (1996) that

$$n^{\alpha/2}\bar{y} \Longrightarrow N\left(0, \frac{2\lambda}{(1-\alpha)(2-\alpha)}\right),$$

where $\alpha = 1 - 2d$ and λ is determined by the power law decay of the autocovariance function of y_t. Hence, $y_t \sim S(d)$.

(iii) $d = 1/2$. Let $S_n(s)$ be the D-space analog of the partial sum process of y_t. Theorem 2.2. in Liu (1998) shows that

$$\frac{\sqrt{2}}{\left(2\sigma_u^2/\pi\right)^{1/2} n \log^{1/2} n} S_n(s) \xrightarrow{d} sB(1),$$

which implies that $y_t \sim S(0.5)$ with $L(n) = \sqrt{2}/\left(2\sigma_u^2/\pi\right)^{1/2} \log^{1/2} n$.

(iv) $d > 1/2$. Let $Y_n(s)$ denote the D-space analog of y_t. Theorem 2.3 in Liu (1998) shows that

$$\kappa(n,d)^{-1/2} n^{-(d-1/2)} Y_n(s) \xrightarrow{d} W_{d_0}^{d-d_0}(s),$$

where

$$W_{d_0}^{d-d_0}(s) \equiv \int_0^s \int_0^{s_{d-d_0-1}} \cdots \int_0^{s_2} W_{d_0}^1(s_1) ds_1 ds_2 \cdots ds_{d-d_0-1},$$

denotes the $(d - d_0 - 1)\, th$ integrated fractional Brownian motion,

$$\kappa(n,d) = \begin{cases} \frac{\sigma_u^2 \Gamma(1-2d_0)}{(1+2d_0)\Gamma(1+d_0)\Gamma(1-d_0)} & if\ d > 1/2\ and\ d \neq \frac{2k+1}{2}\ \forall k \in \mathbb{N} \\ \frac{\left(2\sigma_u^2/\pi\right)}{2} & if\ d = \frac{2k+1}{2}\ \forall k \in \mathbb{N} \end{cases},$$

and $\Gamma(\cdot)$ denotes the gamma function.

Applying the Continuous Mapping Theorem

$$\begin{aligned} \int_0^1 \kappa(n,d)^{-1/2} n^{-(d-1/2)} Y(s) ds &= \frac{1}{n} \sum_{t=1}^n \frac{1}{\kappa(n,d)^{1/2} n^{d-1/2}} y_t \\ &= \frac{1}{\kappa(n,d)^{1/2}} \frac{1}{n^{1/2+d}} \sum_{t=1}^n y_t \xrightarrow{d} \int_0^1 W_{d_0}^{d-d_0}(s) ds, \end{aligned}$$

and the desired result holds. **Q.E.D.**

Proof of Proposition 2: Let y_t be $I(0)$ as described by Definition 2. Let $d = 1, 2, ... < \infty$. By definition, $\Delta^d y_t \sim I(-d)$. The sum of $\Delta^d y_t$ is

$$\sum_{t=1}^n \Delta^d y_t = C(1) \sum_{t=1}^n \Delta^d \varepsilon_t + \Delta^d \tilde{\varepsilon}_0 - \Delta^d \tilde{\varepsilon}_n = A_n + B_n,$$

where $A_n = C(1) \sum_{t=1}^n \Delta^d \varepsilon_t$ and $B_n = \Delta^d \tilde{\varepsilon}_0 - \Delta^d \tilde{\varepsilon}_n$. By definition of $\tilde{\varepsilon}_t$,

$$B_n = \Delta^d \tilde{\varepsilon}_0 - \Delta^d \tilde{\varepsilon}_n = O_p(1).$$

With respect A_n note that,

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

and by Definition 2,

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

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

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

as well. And, all together implies that

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

or equivalently that $y_t \sim S(-0.5)$, as stated **Q.E.D.**

Proof of Proposition 3: By contradiction. Assume $y_t \sim S(\delta)$ with $\delta < -1/2$ and $\inf(var(y_t)) > 0$. By definition of summability,

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

Given that $1/2 + \delta < 0$,

$$\sum_{t=1}^{n} y_t = o_p(1).$$

But, then

$$y_t = o_p(1),$$

and $\inf(var(y_t)) = 0$, which is a contradiction. **Q.E.D.**

Proof of Proposition 4: By definition of summability

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

The first cumulation of x_t is

$$z_{1t} = \sum_{j=1}^{t} x_j,$$

and its sum is

$$\sum_{t=1}^{n} z_{1t} = \sum_{t=1}^{n} \left(\sum_{j=1}^{t} x_j \right) = \sum_{t=1}^{n} (n-t+1) x_t = \sum_{t=1}^{n} (n+1) x_t - \sum_{t=1}^{n} t x_t.$$

Note that

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

and hence $tx_t \sim S(\delta + 1)$. Consider

$$\sum_{t=1}^{n} z_{1t} + \sum_{t=1}^{n} t x_t = \sum_{t=1}^{n} (n+1) x_t.$$

Again

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

which means that $(n+1)x_t = (z_{1t} + tx_t) \sim S(\delta + 1)$. Therefore, z_{1t} cannot have an order of summability strictly greater than $\delta + 1$. And because

$$z_{1t} = \sum_{j=1}^{t} x_j,$$

it cannot have an order of summability strictly lower than δ. **Q.E.D.**

Proof of Proposition 5: U_k is $O_p(1)$ by definition of summable processes. Hence, Theorem 3.1. in McElroy and Politis (2007) applies. **Q.E.D.**

Proof of Proposition 6: The OLS estimator that does not consider the presence of a constant term can be written as

$$\hat{\beta}_1 = \beta + \frac{\sum_{k=1}^{n} (\alpha + U_k) \log k}{\sum_{k=1}^{n} \log^2 k}.$$

The denominator satisfies

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

hence, we write

$$\begin{aligned}
\hat{\beta}_1 - \beta &= \frac{\frac{1}{n\log^2 n} \sum_{k=1}^{n} V_k \log k}{\frac{1}{n\log^2 n} \sum_{k=1}^{n} \log^2 k} = \frac{\frac{1}{n\log^2 n} \sum_{k=1}^{n} (\alpha + U_k) \log k}{\frac{1}{n\log^2 n} \sum_{k=1}^{n} \log^2 k} \\
&= \alpha \frac{\frac{1}{n\log^2 n} \sum_{k=1}^{n} \log k}{\frac{1}{n\log^2 n} \sum_{k=1}^{n} \log^2 k} + \frac{\frac{1}{n\log^2 n} \sum_{k=1}^{n} U_k \log k}{\frac{1}{n\log^2 n} \sum_{k=1}^{n} \log^2 k}.
\end{aligned}$$

Now, we concentrate on

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

Let

$$U_{nk} = \log \left[\left(\left(\frac{n}{k}\right)^{1/2+\delta} \frac{1}{n^{1/2+\delta}} \omega \sum_{t=1}^{k} x_t \right)^2 \right]$$

and its D-space analog

$$U_n(r,\delta) = \log\left[\left(r^{-1/2-\delta}\frac{1}{n^{1/2+\delta}}\omega\sum_{t=1}^{[nr]}x_t\right)^2\right].$$

By assumption,

$$U_n(r,\delta) \Longrightarrow \log\left[\left(r^{-1/2-\delta}D_x(r,\delta)\right)^2\right].$$

Now consider,

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n}U_{nk}\log\left(\frac{k}{n}\right) &= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}U_n(r,\delta)\left[\log\left(\frac{k}{n}\right)+\log r-\log r\right]dr \\
&= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}U_n(r,\delta)\log r dr+\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}U_n(r,\delta)\left[\log\left(\frac{k}{n}\right)-\log r\right]dr \\
&= \int_0^1 U_n(r,\delta)\log r dr+\sum_{k=1}^{n}U_{nk}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left[\log\left(\frac{k}{n}\right)-\log r\right]dr.
\end{aligned}$$

Let

$$a_k = \int_{\frac{k-1}{n}}^{\frac{k}{n}}\left[\log\left(\frac{k}{n}\right)-\log r\right],$$

hence,

$$\frac{1}{n}\sum_{k=1}^{n}U_{nk}\log\left(\frac{k}{n}\right) = \int_0^1 U_n(r,\delta)\log r dr+\sum_{k=1}^{n}U_{nk}a_k.$$

Now,

$$\begin{aligned}
a_k &= \int_{\frac{k-1}{n}}^{\frac{k}{n}}\left[\log\left(\frac{k}{n}\right)-\log r\right]dr \\
&= \int_{\frac{k-1}{n}}^{\frac{k}{n}}\log\left(\frac{k}{n}\right)dr-\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr \\
&= \frac{1}{n}\log\left(\frac{k}{n}\right)-\frac{k}{n}\log\left(\frac{k}{n}\right)+\left(\frac{k-1}{n}\right)\log\left(\frac{k-1}{n}\right)+\frac{1}{n} \\
&= -\left(\frac{k-1}{n}\right)\log\left(\frac{k}{k-1}\right)+\frac{1}{n}.
\end{aligned}$$

Thus, $a_1 = 1/n$. For $k > 1$, the series expansion

$$\log x = \frac{x-1}{x}+\frac{1}{2}\left(\frac{x-1}{x}\right)^2+\frac{1}{3}\left(\frac{x-1}{x}\right)^3+...$$

will be used to show that

$$\log\left(\frac{k}{k-1}\right) = \frac{1}{k}+\frac{1}{2}\left(\frac{1}{k}\right)^2+\frac{1}{3}\left(\frac{1}{k}\right)^3+...$$

and hence

$$\begin{aligned}
a_k &= -\left(\frac{k-1}{n}\right)\left[\frac{1}{k}+O\left(\left(\frac{1}{k}\right)^2\right)\right]+\frac{1}{n} \\
&= O\left(\frac{1}{(k-1)n}\right),
\end{aligned}$$

that is,

$$\begin{aligned}
(k-1)na_k &= -(k-1)^2\left[\frac{1}{k}+O\left(\left(\frac{1}{k}\right)^2\right)\right]+(k-1) \\
&= (k-1)-\frac{(k-1)^2}{k}+O(1) \\
&= \frac{k(k-1)-(k-1)^2}{k}+O(1) \\
&= \frac{(k-1)}{k}+O(1) \\
&= O(1).
\end{aligned}$$

Given that $U_{nk}=O_p(1)$ and

$$\begin{aligned}
n\sum_{k=1}^{n}a_k \sim \sum_{k=1}^{n}\frac{1}{k-1} \sim \log n, \\
\sum_{k=1}^{n}U_{nk}a_k = O_p\left(\frac{\log n}{n}\right),
\end{aligned}$$

which implies that

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n}U_{nk}\log\left(\frac{k}{n}\right) &= \int_0^1 U_n(r,\delta)\log r dr+\sum_{k=1}^{n}U_{nk}a_k \\
&= \int_0^1 U_n(r,\delta)\log r dr+o_p(1) \\
&\Longrightarrow \int_0^1 \log r U_x(r,\delta)dr
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n\log n}\sum_{k=1}^{n}U_k\log k &= \frac{1}{\log n}\left(\frac{1}{n}\sum_{k=1}^{n}U_{nk}\log\left(\frac{k}{n}\right)\right)+\frac{1}{n}\sum_{k=1}^{n}U_{nk}+o(1) \\
&= \frac{1}{n}\sum_{k=1}^{n}U_{nk}+o_p(1) \\
&= \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}U_n(r,\delta)dr \\
&= \int_0^1 U_n(r,\delta)dr \\
&\Longrightarrow \int_0^1 U_x(r,\delta)dr.
\end{aligned}$$

All together gives the stated result

$$\begin{aligned}
\log n(\hat{\beta}_1-\beta) &= \frac{\frac{1}{n\log n}\sum_{k=1}^{n}(\alpha+U_k)\log k}{\frac{1}{n\log^2 n}\sum_{k=1}^{n}\log^2 k} \\
&= \alpha\frac{\frac{1}{n\log n}\sum_{k=1}^{n}\log k}{\frac{1}{n\log^2 n}\sum_{k=1}^{n}\log^2 k}+\frac{\frac{1}{n\log n}\sum_{k=1}^{n}U_k\log k}{\frac{1}{n\log^2 n}\sum_{k=1}^{n}\log^2 k} \\
&\Longrightarrow \alpha+\int_0^1 U_x(r,\delta)dr.
\end{aligned}$$

Q.E.D.

Proof of Proposition 7: The OLS estimator that takes into account the presence of a constant term can be written as

$$\hat{\beta}_2 = \beta + \frac{\sum_{k=1}^{n} U_k(\log k - \overline{\log n})}{\sum_{k=1}^{n} (\log k - \overline{\log n})^2}.$$

The denominator satisfies

$$\frac{1}{n}\sum_{k=1}^{n}(\log k - \overline{\log n})^2 \rightarrow 1.$$

With respect the numerator, note that

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

Now,

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n} U_k \left(\log\left(\frac{k}{n}\right) - \left(\frac{1}{n}\sum_{k=1}^{n}\log\left(\frac{k}{n}\right)\right)\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} \log\left[\left(\frac{1}{k^{1/2+\delta}}\omega\sum_{t=1}^{k} x_t\right)^2\right]\left(\log\left(\frac{k}{n}\right) - \left(\frac{1}{n}\sum_{k=1}^{n}\log\left(\frac{k}{n}\right)\right)\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} \log\left[\left(\frac{n^{1/2+\delta}}{k^{1/2+\delta}}\frac{1}{n^{1/2+\delta}}\omega\sum_{t=1}^{k} x_t\right)^2\right]\left(\log\left(\frac{k}{n}\right) - \left(\frac{1}{n}\sum_{k=1}^{n}\log\left(\frac{k}{n}\right)\right)\right), \\
&= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log\left(\frac{k}{n}\right) + \log r - \log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right) + \log r - \log r\right)dr\right)\right)dr \\
&= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right) + \log r - \log r\right)dr\right)\right)dr \\
&\quad + \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log\left(\frac{k}{n}\right) - \log r\right)dr \\
&= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right) + \log r - \log r\right)dr\right)\right)dr + \sum_{k=1}^{n} U_k a_k,
\end{aligned}$$

where

$$a_k = \int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right) - \log r\right)dr.$$

As before,

$$\sum_{k=1}^{n} U_k a_k = o_p(1),$$

and hence, the expression of interest becomes

$$\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right)+\log r - \log r\right)dr\right)\right)dr.$$

Rearranging terms

$$\begin{aligned}
&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right)+\log r - \log r\right)dr\right)\right)dr\\
=\;&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr\\
&+\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log r-\log\left(\frac{k}{n}\right)\right)dr\right)dr\\
=\;&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr\\
&-\sum_{k=1}^{n}U_t\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\log\left(\frac{k}{n}\right)-\log r\right)dr\right)dr\\
=\;&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr-\sum_{k=1}^{n}U_k\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left(\sum_{k=1}^{n}a_k\right)dr\\
=\;&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr-\sum_{k=1}^{n}U_k\left(\frac{1}{n}\sum_{k=1}^{n}a_k\right).
\end{aligned}$$

As before,

$$\sum_{k=1}^{n}a_k = O\left(\frac{\log n}{n}\right)=o\left(1\right),$$

and

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

Therefore,

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n}U_k(\log k-\overline{\log n}) \;=\;&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr\\
&+\sum_{k=1}^{n}U_k a_k-\left(\frac{1}{n}\sum_{k=1}^{n}U_k\right)\left(\sum_{k=1}^{n}a_k\right)\\
=\;&\sum_{k=1}^{n}\int_{\frac{kt-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr\\
&+\sum_{k=1}^{n}a_k\left(U_k-\left(\frac{1}{n}\sum_{k=1}^{n}U_k\right)\right)\\
=\;&\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left(\log r - \left(\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr\right)\right)dr+o_p(1)\\
=\;&\int_0^1 U_n(r,\delta)\left(\log r-\int_0^1\log r dr\right)dr+o_p(1).
\end{aligned}$$

Next, note that

$$\left(\int_0^1 \log r dr\right) = -1,$$

hence,

$$U_n(r,\delta)\left(\log r - \int_0^1 \log r dr\right) \Longrightarrow U_x(r,\delta)\left(\log r - \left(\int_0^1 \log r dr\right)\right) = U_x(r,\delta)\left(1+\log r\right).$$

Finally, by the Continuous Mapping Theorem

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} U_k(\log k - \overline{\log n}) &= \int_0^1 U_n(r,\delta)\left(\log r - \int_0^1 \log r dr\right) dr + o_p(1) \\
&= \int_0^1 U_n(r,\delta)\left(1+\log r\right) dr + o_p(1) \\
&\Longrightarrow \int_0^1 U_x(r,\delta)\left(\log r - \left(\int_0^1 \log r dr\right)\right) dr \\
&= \int_0^1 U_x(r,\delta)\left(1+\log r\right) dr.
\end{aligned}$$

Q.E.D.

Proof of Proposition 8: The DGP of interest is

$$y_t = m + x_{yt},$$

hence,

$$\begin{aligned}
y_t - \hat{m}_t &= y_t - \frac{1}{t}\sum_{j=1}^{t} y_j \\
&= x_{yt} - \frac{1}{t}\sum_{j=1}^{t} x_{yj}.
\end{aligned}$$

By assumption it holds that

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{[nr]} x_{yt} \Longrightarrow D_x(r,\delta).$$

By the CMT,

$$\int_0^1 \left(\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{[nr]} x_{yj}\right) dr \Longrightarrow \int_0^1 D_x(r,\delta) dr.$$

Now,

$$\begin{aligned}
\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}(y_t-\hat{m}_t) &= \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(x_{yt}-\frac{1}{t}\sum_{j=1}^{t}x_{yj}\right)\\
&= \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_{yt}-\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{n^{1/2+\delta}}{t}\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{t}x_{yj}\\
&= \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_{yt}-\frac{n^{\delta-1/2}}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{n}{t}\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{t}x_{yj}\\
&= \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_{yt}-\frac{1}{n}\sum_{t=1}^{n}\frac{n}{t}\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{t}x_{yj}\\
&= \\
&\Longrightarrow D_x(1,\delta)-\int_0^1 r^{-1}D_x(r,\delta)dr.
\end{aligned}$$

Q.E.D.

Proof of Proposition 9: The proof will be divided in five steps.

(i) First, the partial mean is computed

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

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

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

(iii) Third, compute

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

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

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

(v) Finally, the asymptotic behavior of the following re-scaled sum is analyzed

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(y_t-\frac{1}{t}\sum_{j=1}^{t}y_j-\frac{2}{t}\sum_{j=1}^{t}\left(y_j-\frac{1}{j}\sum_{i=1}^{j}y_i\right)\right)=\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(x_{yt}-\frac{3}{t}\sum_{j=1}^{t}x_{yj}+\frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j}x_{yi}\right).$$

With respect the first summand, by assumption

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}x_{yt}\Longrightarrow D_x\left(1,\delta\right).$$

For the second and third summand, the CMT will be used as well. With respect the former

$$\frac{3}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{1}{t}\sum_{j=1}^{t}x_{yj}=\frac{3}{n}\sum_{t=1}^{n}\frac{n}{t}\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{t}x_{yj}\Longrightarrow 3\int_0^1 r^{-1}D_x\left(r,\delta\right)dr,$$

and the later

$$\begin{aligned}\frac{2}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{1}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j}x_{yi} &= \frac{2}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{1}{t}\sum_{j=1}^{t}\frac{t^{1/2+\delta}}{j}\frac{1}{t^{1/2+\delta}}\sum_{i=1}^{j}x_{yi}\\ &= \frac{2}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{t^{-1/2+\delta}}{t}\sum_{j=1}^{t}\frac{t}{j}\frac{1}{t^{1/2+\delta}}\sum_{i=1}^{j}x_{yi}\\ &= \frac{2}{n^{1/2+\delta}}\sum_{t=1}^{n}n^{-3/2+\delta}\frac{t^{-3/2+\delta}}{n^{-3/2+\delta}}\sum_{j=1}^{t}\frac{t}{j}\frac{1}{t^{1/2+\delta}}\sum_{i=1}^{j}x_{yi}\\ &= \frac{2}{n}\sum_{t=1}^{n}\frac{t^{-3/2+\delta}}{n^{-3/2+\delta}}\sum_{j=1}^{t}\frac{t}{j}\frac{1}{t^{1/2+\delta}}\sum_{i=1}^{j}x_{yi}\\ &\Longrightarrow 2\int_0^1 r^{-3/2+\delta}\left(\int_0^r s^{-1/2+\delta}D_x\left(s,\delta\right)ds\right)dr.\end{aligned}$$

Therefore,

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}(y_t-\hat{m}_t)\Longrightarrow D_x\left(1,\delta\right)-3\int_0^1 r^{-1}D_x\left(r,\delta\right)dr+2\int_0^1 r^{-3/2+\delta}\left(\int_0^r s^{-1/2+\delta}D_x\left(s,\delta\right)ds\right)dr.$$

Q.E.D.

References

Box, G. E .P., and G. M. Jenkins (1970): “Time Series Analysis Forecasting and Control,”. San Francisco: Holden-Day.

Davidson, J. (1999): “When is a time series I(0)? Evaluating the memory properties of non-linear dynamic models,” Discussion Paper, Cardiff University.

de Jong, R. M. (2004):"Addendum to ‘Asymptotics for non-linear transformations of integrated time series," *Econometric Theory*, 20, 627-635.

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

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

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

Escanciano, J. C., and A. Escribano (2008): "Econometrics: non-linear co-integration," Encyclopedia of complexity and systems science. Springer.

Franses, P. H., and D. van Dijk (2000): "non-linear time series models in empirical finance," Cambridge: Cambridge University Press.

Granger, C. W. J. (1983): "Co-Integrated variables and error-correcting models," unpublished UCSD Discussion Paper, 83-13.

Granger, C. W. J. (1986): "Developments in the study of cointegrated economic variables," *Oxford Bulletin of Economics & Statistics*, 48, 231-228.

Granger, C. W. J. (1995): "Modelling non-linear Relationships between Extended-Memory Variables," *Econometrica*, 63, 265-279.

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

Granger, C. W. J., and J. Hallman (1991): "non-linear transformations of integrated time series," *Journal of Time Series Analysis*, 12, 207-224.

Granger, C. W. J., and T. Teräsvirta (1993): "Modelling non-linear economic relationships," Oxford: Oxford University Press.

Gonzalo, J., and J. Y. Pitarakis (2006): "Threshold Effects in co-integrating Regressions," *Oxford Bulletin of Economics & Statistics*, 68, 813-833.

Gouriéroux, C., and J. Jasiak, (1999): "non-linear Persistence and Copersistence," Working Paper, York University, Department of Economics.

Gouriéroux, C., F. Maurel and A. Monfort (1987): "Regression and non stationarity", INSEE, Working Paper 8708.

Hosking, J. R. M. (1996): "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series," *Journal of Econometrics*, 68, 165-176.

Johansen, S. (1991): "Estimation and hypothesis testing of co-integration vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 551-1580.

Johansen, S. (1995): "Likelihood-Based Inference in Cointegrated Vector Autoregressive Models," Oxford: Oxford University Press.

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

Liu, M. (1998): "Asymptotic of non-stationary Fractional Integrated Series," *Econometric Theory*, 14, 641-662.

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

Nelson, Ch. R., and Ch. I. Plosser (1982): "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.

Park, J. Y., and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–497.

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

Phillips, P. C. B. (1986): "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311-340.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

Politis, D. N., J. P. Romano, and M. Wolf (1999): "Subsampling," New York: Springer.

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

Romano, J. P., and A. F. Siegel (1986): "Counterexamples in probability and statistics," Monterey, California: Wadsworth and Brooks/Cole.

White, J. S. (1958): "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Mathematical Statistics*, 29, 1188-1197.

Yoon, G. (2006): "A note on some properties of STUR processes," *Oxford Bulletin of Economics & Statistics*, 68, 253-260.

Yoon, G. (2003): "Stochastic unit roots, long memory and I(1.5)," Discussion Paper, Department of Economics and Related Studies, University of New York.