Further evidence on the statistical properties of Real GNP

Laura Mayoral*
Universidad Pompeu Fabra

First version: May 2005; This version: May 2006

Abstract

The well-known lack of power of unit root tests has often been attributed to the short length of macroeconomic variables and also to DGP's departing from the I(1)-I(0) models. This paper shows that by using long spans of annual real GNP and GNP per capita (133 years) high power can be achieved, leading to the rejection of both the unit root and the trend-stationary hypothesis. Then, more flexible representations are considered, namely, processes containing structural breaks (SB) and fractional orders of integration (FI). Economic justification for the presence of these features in GNP is provided. It is shown that both FI and SB formulations are in general preferred to the ARIMA (I(1) or I(0)) ones. As a novelty in this literature, new techniques are applied to discriminate between FI and SB. It turns out that the FI specification is preferred, implying that GNP and GNP per capita are non-stationary, highly persistent but mean-reverting series. Finally, it is shown that the results are robust when breaks in the deterministic component are allowed for in the FI model. Some macroeconomic implications of these findings are also discussed.

JEL Classification: C22, E23, C12.

Keywords: GNP, unit roots, fractional integration, structural change.

I. Introduction

Questions about the nature of the trend component and the persistence of shocks in macroeconomic series and, in particular, in GNP have occupied a very important place in economics and have given rise to a vast literature on the subject. In spite of this fact, important issues remain unclear. Until the early 80's, movements in output were traditionally viewed as representing temporary fluctuations about a stable linear time trend (the so-called trend-stationary models, T-ST henceforth). According

*The author wants to thank Anindya Banerjee and two anonimous referees for valuable comments and suggestions. She also acknowledges financial support from the Spanish Ministry of Education through grants SEC2003-04429 and SEC2003-04476 and also from the Barcelona Economics Program of CREA.

to this view, innovations have no impact on long-run forecasts of GNP. Nevertheless after the influential article of Nelson and Plosser (1982) many economists argued that GNP was better characterized as a differenced-stationary model (D-ST) given by the sum of a stochastic trend that shifts every period (a unit root component) and a transitory term. Under this characterization, there is not reversion to a deterministic trend path because any stochastic shock has a permanent effect on the process and therefore, they will affect output forecasts into the indefinite future.

The widespread acceptance of the unit root hypothesis that followed Nelson and Plosser's article was mainly due to the fact that unit root tests could not reject the latter hypothesis. However, various explanations, not related with the existence of a unit root, were soon put forward to account for such phenomenon. Firstly, some authors claimed that the main difference between these models is basically of an asymptotic character and, given the relative short length of the available macroeconomic data, unit root tests lack power to distinguish between T-ST and D-ST models. Moreover, this is not only true when the models are close (local alternatives) but also when the hypotheses are distant, as shown by Rudebusch (1993) . This author used quarterly real GNP per capita to show that the best-fitting T-ST and D-ST models implied very different medium and long term dynamics. Then, he computed the small sample distributions of the Dickey-Fuller (DF henceforth) test corresponding to both best-fitting models and showed that they were very similar despite the very different dynamics of these processes. This finding contributed to the "we don't know" literature initiated by Christiano and Eichenbaum (1990), since it implied that the DF test has very low power to distinguish between the relevant hypotheses.

Secondly, other group of authors considered that neither the T-ST nor the D-ST models were suitable in some circumstances. The former implies that all shocks are transitory and that the trend component is always the same (deterministic) while the latter has the opposite predictions: all shocks are permanent and the trend component shifts every period. Thus, different models were postulated as DGPs and it was shown that the classical unit root tests have very low power under these new formulations. Among the processes departing from the traditional models, those containing structural breaks and fractionally orders of integration (as opposed to integer ones) have attracted a great deal of attention (see Banerjee and Urga (2005) for a recent survey on both topics).

Perron (1989) initiated the literature on structural breaks by arguing that the movement in the trend component could be well-explained by a few permanent shocks related to very significant events (such as wars, deep economic crisis, etc.), the so-called structural breaks (SB hereafter), in an otherwise stable linear trend. All remaining shocks had a transitory character. He showed that when the DGP is of this type, standard unit root tests cannot reject the D-ST model. On the other hand, fractionally integrated (FI) models were introduced by Granger and Joyeux (1980) and Hosking (1981) who showed that, by allowing for fractional orders of integration (and not only integer ones as in the ARIMA methodology), a richer description of the persistence of shocks could be achieved. These models are able to fill the gap between the short-lasting and the permanent effect of shocks in the T-ST and D-ST models, respectively, by allowing for intermediate behaviors (such as long memory, non-stationary mean-reversion, etc.).

The goal of this paper is to shed further light on the controversy about the statistical properties of real GNP and real GNP per capita taking into account simultaneously the two criticisms described above. Firstly, in order to have good power properties a data set that covers a very long time span (133 years) has been analyzed. Secondly, we enlarge the set of DGP's considered by exploring if there exists evidence in favor of FI and SB models in these data set. It turns out that when tested against the D-ST and T-ST, FI and SB models are preferred. A well-known statistical problem arises here: despite the fact that FI and SB imply very different medium and long term dynamics, it is difficult to distinguish between them.[1] Surprisingly, although there exist many papers that look for FI behavior or for the existence of structural breaks in output, to the best of our knowledge this is the first contribution that directly test these hypotheses. When the FI and the SB models are directly tested, the former is in general preferred, and this conclusion is robust across the different econometric methods employed, the inclusion of trends and/or breaks in the FI model and the variables examined.

The outline of the paper is as follows. Section II introduces the data and considers some classical unit roots tests of I(1) vs. T-ST and vice versa. Section III analyzes the existence of structural breaks in the data for the case where the innovations are weakly dependent. Section IV describes the main characteristics and the economic mechanisms that are able to generate FI in economic data. Several procedures for estimating and testing the FI hypothesis against both the I(1) and the I(0) ones are also considered. Section V, in turn, tests the hypotheses of FI versus a breaking-trend model. Some robustness checks that consider the possibility of breaks/trends and FI are also included. Finally, Section VI put forward some macroeconomic implications of the findings of this article and concludes.

II. The data and preliminary tests

We consider the same data set employed in Diebold and Senhadji (1996) (DS henceforth): the annual real GNP series reported in table 1.10 of the National Income and Product Accounts of the United States, measured in billions of 1987 dollars, ranging from 1929 to 2001 (8 new observations have been added with respect to DS analysis). As in DS, these series has been spliced to the 1869-1929 real GNP series of Balke and Gordon (1989) or Romer (1989) given rise to two different series of Real GNP, each containing 133 annual observations. Per capita GNP has also been considered and in order to construct

[1] This is so because short memory models containing some type of trends and/or breaks display persistence properties that look very similar to those of FI(d) models. The opposite is also true, that is, conventional procedures for detecting and dating structural changes tend to find spurious breaks, usually in the middle of the sample, when in fact there is only fractional integration in the data.

the series, total population residing in the United States (in thousands of people) has been taken from table A-7 of *Historical Statistics of the United States* for years ranging from 1869 to 1970 and for the years 1971-2001, data has been taken from the *Census Bureau's Current Population Reports*, Series P-25. All series are in natural logs.

We keep the same notation as in DS and define:

GNP-BG ("GNP-Balke-Gordon"): Gross national product, pre-1929 values from Balke-Gordon.

GNP-R ("GNP-Romer"): Gross national product, pre-1929 values from Romer.

GNP-BGPC ("GNP-Balke-Gordon, per capita"): Gross national product per capita, pre-1929 values from Balke-Gordon.

GNP-RPC ("GNP-Romer, per capita"): Gross national product per capita, pre-1929 values from Romer.

Some preliminary unit root tests

Before beginning the analysis, it is illustrative to look at some basic properties of D-ST and T-ST models. A trended process may be represented by

$$y_t = \mu + \beta t + u_t. \tag{1}$$

If u_t is a linear weakly-dependent process, that is,

$$u_t = C(L)\,\varepsilon_t, \tag{2}$$

with $\varepsilon_t \sim i.i.d\,(0, \sigma^2)$ and $C(L) = \sum_{j=0}^{\infty} c_j L^j$ such that $\sum_{j=0}^{\infty} j\,|c_j| < \infty$, then y_t is a T-ST process. In this case, the trend component $\tau_t^{\text{T-ST}} = \mu + \beta t$ is completely deterministic and stable over time, while the cycle $c_t = u_t$ has a short effect on the trend because the correlation function of u_t decays to zero at an exponential rate. On the other hand, a unit root in y_t with drift β and initial condition μ can be modelled as,

$$(1 - L)\,u_t = C(L)\,\varepsilon_t. \tag{3}$$

The well-known trend-cycle decomposition allows to write,

$$y_t = \mu + \beta t + C(1)\sum_{j=0}^{t-1} \varepsilon_{t-j} + a_t, \tag{4}$$

where $a_t = C^*(L)\,\varepsilon_t$ is a weakly-dependent stationary processes.[2] The trend component is given by $\tau_t^{\text{D-ST}} = \mu + \beta t + C(1)\sum_{j=0}^{t-1} \varepsilon_{t-j}$ and, as opposed to $\tau_t^{\text{T-ST}}$, has a stochastic character since it shifts every period in an unpredictable way.

[2] This is so because the polynomial $C(z)$ can always be written as $C(z) = C(1) + (1-z)\,C^*(z)$, where $C^*(z) = \sum_{i=0}^{\infty} c_i^* z^i$ and $c_i^* = -\sum_{j=i+1}^{\infty} c_j$, see Johansen (1995), Lemma 4.1.

Since the major difference between T-ST and D-ST models is the duration of shocks, it is clear that data covering long data spans are needed in order to have a reasonable power. This is the approach followed by DS (1996) who contested Rudebusch (1993) conclusions by using the above-described long data set. They computed the best fitting T-ST and D-ST models for each of the four series. Then, from each of the estimated models they calculated the exact finite sample distribution of the corresponding t-statistics from an augmented Dickey-Fuller (ADF) regression. They showed that the p-value associated to the former statistic was very small under the D-ST best-fitting model but quite large under the T-ST one and therefore it was possible to reject the null hypothesis of a unit root with reasonable power.

Nevertheless, as noticed by the authors, rejecting the null hypothesis does not mean that the alternative is a good characterization of this data set. The first column of Table I below presents the results of the KPSS test for the null hypothesis of trend stationarity versus a unit root. It shows that, when the hypotheses are reversed, the T-ST hypothesis is rejected for this data set. This table also presents the output from other classical unit root tests. Columns two to four report the results of testing the unit root as null hypothesis using some popular approaches: the Augmented Dickey-Fuller test, the Phillips-Perron (P-P) test and the efficient DF-GLS method proposed by Elliott et al. (1996). Not surprisingly, the I(1) hypothesis is rejected for the four extended series employed in this article, confirming DS findings.

[**Table I about here**]

Figure 1 is similar to DS' Figure 2 but contains in this case the exact finite-sample distribution of the KPSS statistic under the best-fitting T-ST and D-ST models for real GNP per capita (Pre-1929 values are from Romer, (1989)). This plot confirms DS conclusions in the sense that the KPSS value is very unlikely under the best-fitting D-ST formulation: only 4% of the values coming from the D-ST distribution are smaller than that one. But remarkably, it shows that the sample KPSS is also unlikely under the T-ST model since only 19% of the values are bigger than that quantity.

[**Figure I about here**]

It is interesting to notice that the chronic problem of lack of power does not seem to apply here since both hypotheses can be rejected. In the following sections we explore the plausibility of other formulations that can be considered to be mid-way between the T-ST and the D-ST formulations: models containing structural breaks and fractional orders of integration.

III. Models containing Structural Breaks

It follows from the discussion below equation (1) that the unit root versus T-ST problem can be viewed as addressing the question "do the data support the view that the trend never changes or is it changing every period?". Clearly, these alternatives are not mutually exclusive and can be seen as too extreme in some contexts. A more interesting question to ask would be "how frequent permanent shocks are?". Perron (1989) initiated this line of research postulating that only few shocks, those related with very significant events, might have a permanent character while the remaining ones would only last for a short time. The specific number of permanent shocks becomes then case-specific. Under these assumptions, the process y_t can be represented by the sum of two components: a deterministic trend whose parameters are allowed to change at each break date and a cycle, similar to that described in (2). More specifically, using quarterly data for the postwar U.S. GNP, Perron (1989) favored the model $y_t = \tau_t + u_t$ where the trending component is given by,

$$\tau_t = \alpha_1 + \beta_1 t + (\beta_2 - \beta_1)(t - T^*) 1_{(t>T^*)},$$

where T^* coincides with the 1973 oil crisis, $1_{(t>T^*)}$ is an indicator function and β_2 captures the slow down in the growth rate after the crisis.

We now explore if there exists evidence in favor of a breaking trend in the data set considered in this article. Given the long time span considered here, it seems reasonable to allow for several permanent shocks in the data occurring at unknown break dates. We employ the method proposed by Bai and Perron (1998, 2003), for multiple structural breaks. For the sake of brevity, details about the testing procedure have been skipped but are available in the working paper version of this paper (see Mayoral, 2006a).

Table II gathers the main results.[3] They can be summarized as follows. Real GNP shows little evidence of structural breaks. The null of no-break is only rejected at the 10% level versus the alternative of two breaks, occurring, for both GNP-R and GNP-BG around the 1929 crisis and the end of World War II. On the other hand, one break is found in per-capita series located around 1939, coinciding with the beginning of World War II. Since there is no evidence of a change in real GNP around that date, the latter finding suggests that it may be due to a demographic shock. Tests of no-break versus an unknown number of breaks and sequential tests were performed and their conclusions were similar as the above-described ones.

[Table II about here]

IV. Fractionally Integrated models

As described above, memory is infinite in D-ST models (past shocks are perfectly remembered) while it is very short (with an exponential correlation decay) in T-ST ones. Fractionally integrated (FI)

[3] Only the results of the tests with three breaks or less are reported. The hypotheses of 4 and 5 breaks were also considered and rejected in favor of the no-break hypothesis in all cases.

processes constitute an interesting alternative to this dichotomy since they are able to bridge the gap between these two possibilities. To achieve this, the model in (1) is considered, where in this case u_t is defined as,

$$(1-L)^d u_t = C(L)\,\varepsilon_t. \tag{5}$$

where d is a real number. This parameter determines the integration order of u_t and governs the memory of the process. The higher the value of d, the more persistent shocks are. Notice that if $d = \{0,1\}$, the T-ST and D-ST models are found, respectively. However, if fractional values of d are considered, richer memory properties can be obtained. This provides for parsimonious, yet flexible, modeling of the low frequency variation. The process u_t is stationary and invertible if -1/2< $d < 1/2$ (see Hosking (1981)). *Long memory* occurs whenever d belongs to the (0,0.5) interval, since in this case the autocorrelation function, $\rho(k)$, verifies that,

$$\rho(k) \sim ck^{2d-1} \text{ for large } k, d \in (0, 0.5), \tag{6}$$

where c is a constant. In other words, long memory is implied by a hyperbolic decay of correlations, as opposed to the case $d = 0$ where correlations decay exponentially fast. For values of $d \in [0.5, 1)$, the process is non-stationary since the variance is unbounded. Shocks are very persistent but do not have a permanent effect implying that the process is mean reverting despite its non-stationary character. Values of d greater or equal than 1 imply a permanent character of shocks.

There exists appealing theoretical underpinnings for the existence of FI in output. Michelacci and Zaffaroni (2000) showed that FI in GNP per capita arises in a Solow-Swan growth model just by allowing for cross sectional heterogeneity in the speed with which different units in the same countries adjust. Abadir and Talmain (2002) consider a monopolistically-competitive Real Business Cycle model and, by allowing for heterogeneity at the firm level, show that GDP turns out to be very persistent although mean-reverting. Haubrich and Lo (2001) discuss a multiple-sector real business cycle model and show analytically that GNP behaves as a FI process.

From an empirical point of view, several papers have also tested the existence of FI in output with somehow mixed conclusions. Both Diebold and Rudebusch (1989) and Sowell (1992*a*) analyzed quarterly post-war data and obtain estimates of d below unity. Nevertheless, their results are in line with Rudebusch's (1993) and Christiano and Eichembaum (1990) conclusions in the sense that the confidence interval of the estimated value of d includes the unit root and, in the Sowell's case, also the T-ST model.

In the next subsection, we contribute to this literature by analyzing the empirical plausibility of the FI hypothesis in real GNP and real GNP per capita. The use of data that covers longer time spans and a wider set of econometric techniques allow us to obtain more robust conclusions.

Estimation of $FI\,(d)$ models

There is a broad literature on the parametric and semiparametric estimation of FI models. In the following we consider some of the most representative techniques in both fields. ARFIMA processes have been chosen to specify the parametric model. This amounts to consider that the polynomial C(L) in expression (5) admits an ARMA representation. Exact Maximum Likelihood (ML) (Sowell, 1992b), Minimum Distance (MD) (Mayoral, 2004), and the Whittle estimator with tapered data (WT) (Velasco and Robinson, 2000) are employed. The semiparametric techniques proposed Shimotsu (2006) and by Teverovsky and Taqqu (TT) (1997) have also been applied. Table III gathers the main results.

Two main conclusions can be drawn from the inspection of the table below. Firstly, the memory parameter is a fractional number below unity for all series across all techniques employed. Secondly, although the values differ slightly across techniques, values of d in the interval (0.5,1) are found in general. The finding of fractional integration seems to be very robust in all the four series across the different methods employed, with an integration order around 0.7. This implies that the series are non-stationary with very persistent although mean-reverting shocks.

[**Table III about here**]

Figure 2 reproduces Figure 1 above but the exact distribution of the best-fitting FI(d) model (according to the exact ML procedure) has been included. It can be seen that the sample KPSS occupies a central position in the distribution under the FI hypothesis, implying that this is a very likely value under this distribution. More specifically, the probability mass on the left side of this value is equal to 0.44.

[**Figure 2 about here**]

Although Table III provides some evidence in favor of the hypothesis of non-stationary mean-reversion ($0.5 \leq d < 1$), tests based on confidence intervals around the estimated values are known to have very low power against integer alternatives. We next report the output of different and more powerful approaches for testing FI versus the T-ST and the D-ST alternatives.

Testing Fractional versus Integer integration

We consider first tests of I(1) versus fractional orders of integration. Two popular tests have been employed: Lagrange Multiplier (LM) and Wald type tests. The former was introduced by Robinson (1994) and Tanaka (1999) in the frequency and time domain respectively.[4] A different approach was

[4] Tanaka's (1999) time domain version has been computed instead of Robinson's (1994) original frequency domain test since Monte Carlo simulations show that the former sligthly outperforms its frequency domain counterpart in finite samples

introduced by Dolado et al. (2002, 2004) who generalized the well-known Wald-type Dickey-Fuller test of $I(1)$ against $I(0)$ to the more general framework of $I(1)$ versus FI(d), with $d < 1$.

Table IV presents the outcome of the tests. With respect to the FDF test, the invariant regression described in Dolado et al. (2004) has been performed against several non-stationary fractional hypotheses (see Mayoral (2006a) for futher details). For all the values of d considered under the alternative (from 0.6 to 0.9), the null hypothesis of a unit root was rejected. Similar results were obtained by applying the time domain LM test. Then, both tests support the hypothesis that these series do not contain a unit root.

[Table IV about here]

Finally, it has also been checked whether the T-ST hypothesis can be rejected once FI has been allowed for. Again, two procedures have been employed: Firstly, LM tests have again been applied, setting in this case the T-ST as null hypothesis versus the alternative of a higher integration order.[5] Secondly, a likelihood ratio test that allows to reverse the previous hypotheses, was also employed (see Mayoral, (2006*b*)). This test considers the composite null hypothesis that the process is $FI(d)$, with $d > 0.5$, (allowing for deterministic components) versus a T-ST specification.[6] Table V reports the output of this analysis. Both tests support the hypothesis of FI in this data set. The LM test rejects the null hypothesis of T-ST for all the series while the LR test cannot reject the null of non-stationary fractionally integration.

[Table V about here]

V. Fractional Integration or Structural Breaks

So far, two important conclusions can be drawn. First, the rejection of the D-ST and the T-ST hypotheses is very robust across the various alternatives considered, the different econometric techniques employed and the type of pre-1929 data (Balke-Gordon or Romer). This suggests that the problem of lack of power is not very relevant here. Second, empirical evidence supporting the hypothesis of non-stationary fractional integration has been found and similar evidence in favor of a breaking trend in an otherwise weakly dependent data has also been reported.

(cf. Tanaka, (1999)). The test statistic is given by $\sqrt{T}\sum_{k=1}^{T-1}\frac{1}{k}\hat{\rho}_k$, where $\hat{\rho}_k$ is the autocorrelation function of the residuals of a FI(d) parametric model, and it is asymptotically normally distributed.

[5] Following Bhattacharya et al. (1983) different specifications for the trend, different than the line, were also tried and similar results were obtained. The interested reader is referred to Mayoral (2006*a*) for further details.

[6] According to the likelihood principle, an estimated value of d under H_0 is needed to evaluate the likelihood under that hypothesis. These estimated values have been taken from Table III (second row) and the corresponding critical values are provided in Mayoral (2006b).

In spite of the very different medium and long term dynamics that these two formulations imply, the latter finding is not surprising since it is well-known that these models share some statistical properties that make their identification a difficult task. Some authors have shown analytically that the existence of trends and/or changes in some parameter values that have not been explicitly accounted for can produce spurious persistence properties similar to those of FI processes (see Bhattacharya et al. (1983), Künsch (1986), Teverovsky and Taqqu (1997), Giraitis et al. (2001), Davidson and Sibbertsen (2005), etc.).[7] The opposite effect is also well-documented, that is, standard methods for detecting and locating structural breaks tend to find spurious breaks when the DGP is a FI process (see Hsu (2001) and Nunes et al. (1995)).

An intuitive way to see why this happens is to notice that both FI(d), with $d < 1$, and SB models share the property of being able to accommodate a few number of "permanent" shocks while the remaining ones are transitory. In this context, "permanent" should be understood in a broad sense as in Perron (2005), that is, a shock is permanent if, given a sample of data, lasts for a long time and, in particular, is still in effect at the end of that sample. To see this, consider the decomposition of a FI process obtained by using the property $C(L) = C(1) + (1-L)C^*(L)$. The process y_t, defined as in (1) and (5) can be rewritten as,

$$y_t = \mu + \beta t + C(1)\sum_{j=0}^{t-1} \pi_j(-d)\,\varepsilon_{t-j} + a_t^*, \tag{7}$$

where $a_t^* = (1-L)^{1-d} C^*(L)\,\varepsilon_t$ is a stationary process. As a particular case, notice that if $d = 1$, then $\pi_j(-1) = 1$ for all j implying that all shocks are permanent and expression (7) is identical to the trend-cycle decomposition in (4). Otherwise, $\pi_j(-d) \approx \Gamma(d)^{-1} j^{d-1}$ for large j. This means that for values of $d < 1$, shocks tend to vanish but at a very slow rate, specially for high values of d.

The process $x_t = C(1)\sum_{j=0}^{t-1}\pi_j(-d)\,\varepsilon_{t-j}$ is very persistent while a_t^* is not. This implies that a shock happening at time T^* has an effect on y_{T^*+h} given approximately by $C(1)\,\pi_h(-d)\,\varepsilon_{T^*} \approx C(1)\Gamma(d)^{-1} h^{d-1}\varepsilon_{T^*}$ for large h. This quantity can be large even at very distant h if ε_{T^*} is large. For instance, if $d = 0.7$, the effect of ε_{T^*} on y_{T^*+h} is approximately given by $0.32\mathrm{C}(1)\varepsilon_{T^*}$, $0.24\mathrm{C}(1)\varepsilon_{T^*}$ and $0.20\mathrm{C}(1)\varepsilon_{T^*}$ for $h = 20$, 50 and 100, respectively. This property allows FI processes (with values of $d < 1$) to mimic the behavior of SB models. After a sufficiently large number of periods, the effect of most shocks will be small and not significantly different from zero. However, large shocks can retain a considerable impact on the process even at very distant horizons. In particular, given the relative short length of most macroeconomic variables, large shocks can still be in effect at end of the sample.

[7] The latter authors also point out that cross-sectional aggregation of a fairly general class of nonlinear processes produces a model that not only has the same correlation patterns as FI processes but is also observationally equivalent to FI, in the sense that the aggregated model is linear and converges to fractional Brownian motion.

Parke (1999) reached similar conclusions using a different approach. He introduced the so-called Error Duration (ED) model where the process y_t is given by the cumulation of shocks that switch to 0 after a random delay following a power law distribution. He showed that under certain conditions y_t presents a behavior similar to a FI(d) process with $d \in (0,1]$.[8] In this framework, it is easy to compute the number of shocks that, on average, will remain alive at the end of the sample. For instance, he showed that in a sample of 100 observations and a value of $d = 0.6$, on average, 4.5 shocks would still be in effect at the end of the sample while the rest would have already switched to zero.

The problem of distinguishing between FI and SB models has attracted considerable attention in the last few years and several techniques have been recently proposed to deal with this issue, specially for the case of stationary FI processes (see Heyde and Dai (1996), Sibbertsen and Venetis (2004), Berkes et al. (2005) , among others). Even more recently, this framework has been extended to cover also non-stationary FI processes (cf. Shimotsu (2005), Dolado et al. (2006) and Mayoral (2006b)). See Barnerjee and Urga (2005) for a comprehensive survey. To the best of our knowledge, this paper is the first contribution that directly tests the hypothesis of FI versus SB in GNP data without any prior information of the date of the possible breaks.

Testing for FI versus SB

Two procedures have been implemented to test for FI vs. SB. The first one is a likelihood ratio (LR) test that extends the LR technique used in Section IV (see Mayoral (2006b)). It can be used to test the composite null hypotheses $H_0 : d_0 > 0.5$ versus the alternative of $H_1 : d = 0$, allowing in both cases for deterministic components an for a single break in (some or all of the components of) the deterministic trend under the alternative.[9] The break date is unknown and is estimated as the value that maximizes the likelihood under H_1. A non-parametric correction has been introduced to account for the short-term correlation (see Mayoral (2006a, b) for further details on the computation of this test).

Table VI present the results. The main conclusion is that, for all the series considered, the null hypothesis of non-stationary fractional integration cannot be rejected.

[Table VI about here]

Next, we apply a different technique that permits to reverse the hypotheses. We use an asymptotically equivalent version of the LM test of Robinson (1994) and Tanaka (1999) recently introduced by

[8] More specifically, if the probability that a shock survives for k periods, p_k, decreases with k at the rate $p_k = k^{2d-2}$ for $d \in (0,1]$, Parke shows that the ED model has the same correlation patterns as a FI process.

[9] As mentioned in Section IV, an estimated value of d under H_0 is needed to evaluate the likelihood under this hypothesis. These values have taken from Table III (second row) and the corresponding critical values are reported in Mayoral (2006b).

Demetrescu et al. (2005). This technique has the advantage of being very simple to implement since is completely regression-based and the autocorrelation of the short memory component can be captured by introducing an increasing number of lagged terms of the dependent variable, in the spirit of Said and Dickey (1984). As a drawback, breaks in the trend component can only be introduced at dates that should be pre-specified and this could be an important limitation of this procedure (see Banerjee et al. (1992) and Zivot and Andrews (1992) for a discussion). We consider the three dates that have usually been postulated as candidates for break dates, namely, the 1929 crash, World War II and the oil crisis of 1973. Breaks in the level, in the constant or both have been considered. For the sake of brevity, only values corresponding to GNP-R and GNP-RPC are reported (those related to GNP-BG and GNP-BGPC were very similar). For comparison, values of the test obtained without allowing the possibility of breaks in the trend are also reported (bottom row). Critical values are taken from a standard normal distribution and large values of the statistic favor an order of integration greater than 0.

The results are presented in Table VII. In agreement with the results in Section II, the null hypothesis of T-ST is rejected for both GNP and GNP per-capita (bottom row). This is also the case for real GNP across the different types of breaks and dates considered. Values of the statistic are large and positive which implies that the test favors higher values of d. The result is more ambiguous in the case of real GNP per capita. In this case, the test only rejects a T-ST model with a break in the slope occurring in the first half of the sample at a 10% signification level. In spite of this result, it can be concluded that in general FI models are preferred to SB ones.

[Table VII about here]

Some robustness checks

In line with most of the $I(1)$ vs. SB testing literature, in the section above FI and SB models have been treated as mutually exclusive alternatives. Considering a pure FI model is economically meaningful since there exists economic underpinnings for the existence of such a behavior in output (Section IV). It is also justified from a statistic point of view since, as explained above, FI models are able to represent processes that posses a few number of "permanent" shocks while the rest are transitory. Since this is the main characteristic of the type of SB models considered here, it makes sense to consider which model is preferred for these data.

Nevertheless, it could be the case that both FI and SB coexist in the data. If this was the case and breaks in the deterministic components were not taken into account, estimates of the order of integration, d, would probably be biased upwards. This would lead to incorrect forecasts and biased measures of persistence.

We have checked whether there is evidence of overestimation of d stemming from unaccounted shifts in the deterministic components using different techniques. Firstly, the data have been estimated according to the exact ML and MD procedures described in Section IV, but this time dummy variables have been introduced in the trend component to capture possible breaks in the trend parameters. Changes in the level and/or the rate of growth have been allowed at three pre-specified dates: the 1929 crash, the beginning of World War II and the 1973 oil crisis. In all cases the introduction of dummy variables do no change significative the outcome of the estimation procedures.[10]

The results above depend on the break date that has been selected in a rather arbitrary way. More robust methods are also available. Iacone (2005) has recently shown that the local Whittle estimator can still yield consistent estimates of d in the presence of some trending components and shifts in the mean.[11] Since stationary of the stochastic component is required, first differences have been taken before applying this technique.[12] Table VIII presents the estimated values of d (unity has been added to the original estimation). It can be seen that the estimated values do not change significantly from those that have been presented along the article.[13]

[Table VIII about here]

The previous results suggest that the estimated orders of integration provided in Table III, where no breaks were allowed for, are not overestimated. Clearly, this does not rule out the possibility of breaks in the trend component. There is an incipient literature dealing with the problem of testing for structural breaks in the presence of (stationary) long memory innovations (see Hidalgo and Robinson, (1996) and Lazarova (2005)) . Unfortunately, these techniques are not well-suited here since non-stationary FI processes are suspected and adapting them for this problem is well beyond the scope of the paper. Nevertheless, this will an interesting avenue for future research.

[10] See Mayoral (2006a) for the tables containing the new estimates.

[11] More specifically, the estimator is consistent if trends of the form $\kappa t^{\rho_0 - 1/2}$ for some $\rho_0 < d_0 + 1/2$ are included and also in the presence of mean shifts.

[12] Estimation of d_0 is carried out by maximixing the Whittle log-likelihood in a neighborhood of the zero frequency. Following Iacone (2005) , the number of frecuencies included in the criterion function was $m = 16$ whereas the parameter in charge of the trimming of the lowest frequencies, l, was set equal to 3.

[13] The modified version of the variance-type estimator introduced in Teverovsky and Taqqu (1997) was also tried. Nevertheless, given the short length of the data, the resulting plots were too scattered to obtain any conclusion.

VI. Concluding remarks

This paper has tried to shed further light on the controversy about the statistical properties of real GNP. Taking as starting point the conclusions in Diebold and Senhadji (1996), we have complemented their analysis by first, considering a wider range of models, both under the null and under the alternative hypothesis. In agreement to their results, the unit root hypothesis was robustly rejected for all the alternatives and across the different techniques. But, interestingly, when the hypotheses were reversed also trend-stationarity was rejected. This led us to analyze in depth some generalizations of the considered models: FI and processes containing breaks. Applying a wide set of recent techniques, various tests of FI vs. structural break have been implemented. The final conclusion is that the finding of fractional integration is robust.

From an economic point of view, the implications of these findings are important. As we have seen, long memory can appear in macroeconomic series after aggregating heterogeneous individual entities. This suggests that moving from the representative agent assumption to a multiple-sector real business cycle model introduces not unmanageable complexity, but qualitatively new behavior that should be taken into account. On the other hand, calibrations aimed at matching only a few first and second order moments can similarly hide major differences between models and the data, missing long range dependence properties (which is basically characterized by the slow rate of decay of covariances). Finally, the lack of structural breaks in the data together with the finding of integration orders of around 0.7 for per capita series imply that the growth rate of GDP per capita (first differences of the logarithm of real GNP per capita) is well characterized as a process with little persistence and a constant mean. As Jones (1995) first suggested, this evidence is inconsistent with endogenous growth theories for which permanent changes in certain policy variables have permanent effects on the rate of economic growth.

References

Abadir, K. and Talmain, G. (2002). ‘Aggregation, Persistence and Volatility in a Macro model’, *Review of Economic Studies,* 69, pp. 749-779.

Andrews, D. W. K. (1991). ‘Heteroskedasticity and autocorrelation consistent covariance matrix estimation”, *Econometrica,* 59, pp. 817-854.

Bai J. and Perron, P. (1998). ‘Testing for and estimation of multiple structural changes’, *Econometrica*, 66, pp. 47-79.

Bai J. and Perron, P. (2003) . ‘Computation and Analysis of Multiple Structural Change Models,’ *Journal of Applied Econometrics,* 18, pp. 1-22.

Balke, N. and Gordon, R. J. (1989). ‘The estimation of prewar gross national product: methodology and new evidence’, *Journal of Political Economy,* 97 (1), pp. 38-92.

Banerjee, A., Lumsdaine, R.L., and Stock, J.H. (1992). ‘Recursive and sequential tests of the unit root and trendbreak hypothesis,’ *Journal of Business and Economic Statistics,* 10, pp. 271-288.

Banerjee A. and Urga, G. (2005). ‘Modelling structural breaks, long memory and stock market volatility: an overview,’ *Journal of Econometrics,* 129, pp. 1-34.

Bhattacharya, R.N., Gupta, V.K., Waymire, E. (1983). ‘The Hurst effect under trends,’ *Journal of Applied Probability,* 20, pp. 649-662.

Berkes, I., Horvath, L., and Kokoszka, P.S. (2006). “On discriminating between long-range dependence and changes in mean”, *Annals of Statistics* Vol. 34, No. 3.

Christiano, L. J. and Eichenbaum, M. (1990). ‘Unit root in real GNP: do we know and do we care?’, *Carnegie-Rochester Conference Series on Public Policy*, 32, pp. 7-61.

Davidson, J. and Sibbertsen, P. (2005). ‘Generating schemes for long memory processes: regimes, aggregation and linearity’, *Journal of econometrics,* 128, pp. 253-282.

Demetrescu, M., Kuzin, V. and Hassler, U. (2005). ‘Long Memory Testing in the Time Domain,’ Preprint.

Diebold, F.X, and Rudebusch, G. (1989). ‘Long memory and persistence in aggregate output’, *Journal of Monetary Economics*, 24, pp. 189-209.

Diebold, F. X. and Senhadji, A.S. (1996) . ‘The uncertain unit root in Real GNP: Comment’, *The American Economic Review*, Vol.86, N ^{o}5, pp. 1291-1298.

Dolado, J.J., Gonzalo, J. and Mayoral, L. (2002) . ‘A Fractional Dickey-Fuller test for unit roots’, *Econometrica,* Vol 70, pp. 5.

Dolado J.J., Gonzalo J. and Mayoral L. (2004) . ‘Testing I(1) against I(d) alternatives in the presence of deterministic components’, Mimeo.

Dolado J.J., Gonzalo J. and Mayoral L. (2006). ‘What is what?: A simple time-domain test for long-memory vs. structural breaks', Mimeo.

Elliott, G., Rothenberg, T.J. and Stock, J.H.(1996). 'Efficient tests for an autoregressive unit root,' *Econometrica,* 64, pp. 813-836.

Giraitis, L., P. Kokoszka, R. Leipus (2001), 'Testing for long memory in the presence of a general trend', *Journal of Applied Probability,* 38, pp.1033 - 1054.

Granger, C.W.J and Joyeux, K. (1980). 'An introduction to long memory series', *Journal of Time series Analysis*, 1, 15-30.

Haubrich, J. G and Lo, A. W. (2001). 'The sources and nature of long-term memory in aggregate output', *Economic Review (Q II),* pp. 15-30, Federal Reserve Bank of Cleveland.

Heyde, C. C. and Dai, W. (1996). "On the robustness to small trends of estimation based on the smoothed periodogram", *Journal of Time Series Analysis,* 17, pp. 141-150.

Henry, M. and Zaffaroni P. (2002). 'The Long Range Dependence paradigm for Macroeconomics and Finance', in *Long range dependence: Theory and applications*, P. Doukhan, G. Oppenheim and M. Taqqu (ed). Birkhäuser, Boston.

Hidalgo, J. and Robinson, P. (1996).'Testing for Structural change in a long-memory environment', *Journal of Econometrics*, 70, (1), pp. 159-174.

Hosking, J.R.M. (1981). 'Fractional differencing', *Biometrika,* 68, pp. 165-176.

Hsu, C. C. (2001). 'Change point estimation in regressions with I(d) variables,' *Economics Letters* 70, pp. 147-155.

Iacone, F. (2005). 'Local Whittle estimation of the memory parameter in presence of deterministic components', Preprint.

Jones, C. I, (1995). 'Time series of endogenous growth models', *The Quarterly Journal of Economics*, 110, 2, pp. 495-525.

Künsch, H. (1986). 'Discrimination between deterministic trends and long range dependence', *Journal of Applied Probability,* 23, pp. 1025-1030.

Kydland, F. E. and Prescott, E.C (1982). 'Time to build an aggregate fluctuation', *Econometrica,* vol. 50, N^o 6, pp. 1279-313.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin,Y., (1992). 'Testing the null hypothesis of stationarity against the alternative of a unit root', *Journal of Econometrics,* 54, pp. 159-178.

Lazarova, S. (2005). 'Testing for structural change in regression with long memory processes', *Journal of Econometrics*, 129, pp. 329-372.

Mayoral, L. (2004). 'A new minimum distance estimator for ARFIMA processes', Working paper, Universidad Pompeu Fabra.

Mayoral, L.(2006*a*). 'Further evidence of the statistical properties of Real GNP', Working paper, Universidad Pompeu Fabra, avalaible at http://www.econ.upf.edu/docs/papers/downloads/955.pdf

Mayoral, L.(2006*b*). 'Is the observed persistence spurious or real? A test for fractional integration versus structural breaks', Mimeo.

Michelacci, C. and Zaffaroni P. (2000). 'Fractional Beta-convergence', *Journal of Monetary Economics,* 45, pp. 129-153.

Mikosch, T. and Starica C. (2004). 'Non-stationarities in financial time series, the long range dependence and the IGARCH effects, *Review of Economics and Statistics*, 86 (1), pp. 378-390.

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

Nunes , L. C., Kuan, C. M. and Newbold, P.(1995). 'Spurious Breaks', *Econometric Theory,* 11, pp. 736-749.

Parke, W.R. (1999). 'What is fractional Integration', *The Review of Economics and Statistics*, 81 (4), pp. 632-638.

Perron, P. (1989). The Great Crash, the Oil Price shock and the unit root hypothesis', *Econometrica,* 58, pp. 1361-1401.

Perron, P. (2005). "Dealing with Structural Breaks", Forthcoming in the P*algrave Handbook of Econometrics, Volume 1: Econometric Theory.*

Robinson, P.M. (1994). 'Efficient tests of nonstationary hypotheses,' *Journal of the American Statistical Association, 89,* pp. 1420-1437.

Romer, C. D. (1989). 'The prewar business cycle reconsidered: New estimates of Gross National Product, 1869-1908', *Journal of Political Economy* 97(1), pp.1-37.

Rudebusch, G. D., (1993). 'The uncertain unit root in real GNP', *American Economic Review,* 83, 1, pp. 264-272.

Said, S.E., and Dickey D. A. (1984). 'Testing for unit roots in autoregressive moving average models of unknown order,' *Biometrika* , 71, pp. 599-608.

Shimotsu, K (2006). 'Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend', Queens Economics Department Working Paper No. 1061.

Shimotsu, K. (2005). 'Simple (but effective) tests of long memory versus structural breaks', Mimeo.

Sibbertsen, P. and Venetis, I.(2004). "Distinguishing between long-range dependence and deterministic trends", Mimeo.

Sowell, F. (1992*a*). 'Modeling long-run behavior with the fractional ARIMA model', *Journal of Monetary Economics*, 29, pp. 227-302.

Sowell, F. (1992*b*). 'Maximum likelihood estimation of stationary univariate fractionally integrated time series', *Journal of Econometrics*, 53.

Tanaka, K. (1999). 'The nonstationary fractional unit root,' *Econometric Theory*, 15, pp. 549-582.

Teverovsky V. and Taqqu M. (1997) . 'Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator', *Journal of Time Series Analysis*, Vol. 18, N^o 3, pp. 279-304

Velasco, C., and Robinson, P. M. (2000). 'Whittle pseudo-Maximum likelihood estimation for non-stationary time series', *Journal of the American Association,* 95**,** pp. 1229-1243.

Zivot, E. and Andrews, D.W.K. (1992). 'Further evidence on the Great Crash, the oil-price shock and the unit root hypothesis', *Journal of Business and Economic Statistics*, Vol. 10, 3, pp. 251-270.

TABLE I

Unit root tests

	H_0: T-ST	H_0: Unit Root		
Data | Test	KPSS	ADF	P-P	DF-GLS
GNP-R	0.190^{**}	-4.70^{***}	-3.53^{**}	-3.34^{**}
GNP-RPC	0.168^{**}	-4.63^{***}	-4.27^{**}	-4.61^{***}
GNP-BG	0.181^{**}	-4.16^{***}	-3.65^{**}	-3.34^{**}
GNP-BGPC	0.182^{**}	-4.79^{***}	-3.65^{**}	-4.76^{***}

*Reject. 10% level;**Reject. 5% level; ***Reject. 1% level.

Note: The number of lags for the ADF and the DF-GLS tests was selected using the AIC. The bandwidth for the Newey-West estimator needed to compute the P-P and KPSS test was chosen according to Andrew's (1991) procedure.

TABLE II

Sup F tests for a fixed number of breaks. H_0 : *no break*

H_1 | Data	GNP-R	GNP-RPC	GNP-BG	GNP-BGPC
H_1 : 1 break	6.53	15.84^{***}	5.53	16.26^{***}
H_1 : 2 breaks	9.44^{*}	8.06	9.43^{*}	7.39
H_1 : 3 breaks	7.35	6.13	7.07	5.49
Break date (BD)	-	1941	-	1939
Conf. Int. of Break Dates	(1927-1929) (1945-1956)	(1938-1956)	(1928-1930) (1945-1956)	(1936-1958)

*Reject. at the 10% level;**Reject. at the 5% level; ***Reject. at the 1% level.

TABLE III

Estimation of $FI\,(d)$ *models*

Method | Data	GNP-R	GNP-BG	GNP-RPC	GNP-BGPC
MD	$d = 0.59$ (0.21)	$d = 0.711$ (0.15)	d=0.570 (0.20)	d=0.581 (0.21)
ML	$d = 0.68$ (0.16)	d=0.65 (0.16)	d=0.65 (0.15)	d=0.63 (0.15)
WT	d=0.731 (0.22)	d=0.627 (0.24)	d=0.731 (0.24)	d=0.731 (0.26)
FELW	d=0.642 (0.21)	d=0.618 (0.21)	d=0.43 (0.21)	d=0.421 (0.21)
TT	$d = 0.66$	$d = 0.66$	$d = 0.83$	$d = 0.83$

Note: All parametric models have chosen according to the AIC. The exact ML and the TT estimator have been computed in first differences and unity has been added to the estimated value of d. Tapering has been employed to compute the WT estimator since non-stationary was suspected.

TABLE IV

Tests of $I(1)$ versus $FI(d)$. $H_0 : d = 1$.

	FD-F Test				LM Test (Tanaka,1999)
Data | H_1 :	d=0.6	d=0.7	d=0.8	d=0.9	d<1
GNP-R	-4.06**	-3.92**	-3.75**	-3.58**	-1.71**
GNP-RPC	-3.85**	-3.67**	-3.49**	-3.32**	-1.91**
GNP-BG	-2.89**	-2.78**	-2.66**	-3.51**	-1.91**
GNP-BGPC	-3.99**	-3.81**	-3.63**	-3.46**	-1.89**

*Reject. at the 10% level;**Reject. at the 5% level; ***Reject. at the 1% level.

TABLE V

Tests of T-ST vs. $FI(d)$ and viceversa

	LM test, (Tanaka, 1999)	LR test (Mayoral, 2006b)
	$H_0 : d_0 = 0$ $vs.$ $H_1 : d_1 > 0$	$H_0 : d_0 > 0.5$ $vs.$ $H_1 : d_1 = 0$
GNP-R	1.95**	0.361
GNP-RPC	1.82**	0.395
GNP-BG	1.99**	0.402
GNP-BGPC	1.88**	0.462

*Reject. at the 10% level;**Reject. at the 5% level; ***Reject. at the 1% level.

TABLE VI

LR Tests of $FI(d)$ vs SB. (Mayoral 2006b)

$H_0 : d_0 > 0.5; H_1$: SB (1 break)				
Data	GNP-R	GNP-RPC	GNP-BG	GNP-BGPC
LR test	0.623	0.651	0.734	0.763

*Reject. at the 10% level;**Reject. at the 5% level; ***Reject. at the 1% level.

TABLE VII

LM tests of SB vs. FI; Demestrescu et. al. (2005).

$H_0 : d_0 = 0$+Breaking trend; H_1: $d_1 > 0$.						
	GNP-R			GNP-RPC		
Break date\Type of break	*Level*	*Slope*	*Both*	*Level*	*Trend*	Both
1929	2.63***	2.30**	2.25**	1.76**	1.75**	1.73**
1939	2.97***	2.20**	2.38***	1.68**	1.61*	1.58*
1973	2.25**	2.30**	2.28**	1.79**	1.78**	1.76**
T-ST (no break)		2.53***			2.01**	

*Rejection at the 10% level;**Rejection at the 5% level;***Rejection at the 1% level

TABLE VIII

Local Whittle estimation of d, Iacone (2005)

Data	GNP-R	GNP-BG	GNP-RPC	GNP-BGPC
$\hat{d}$	0.78	0.79	0.69	0.67

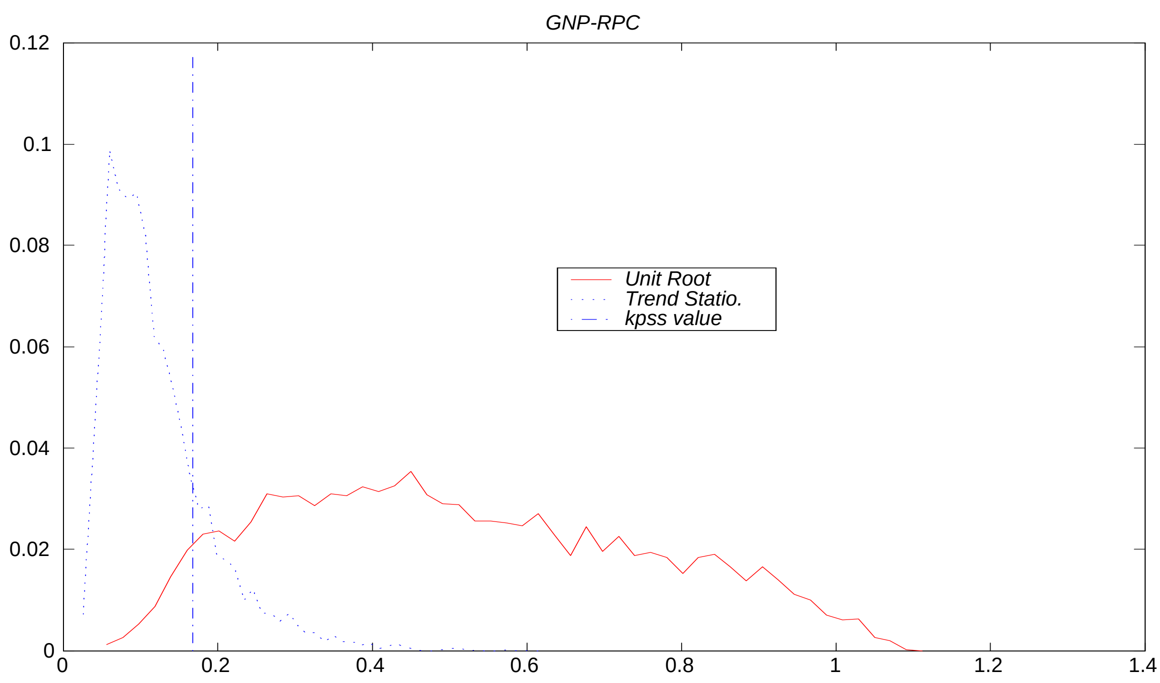

FIGURE I

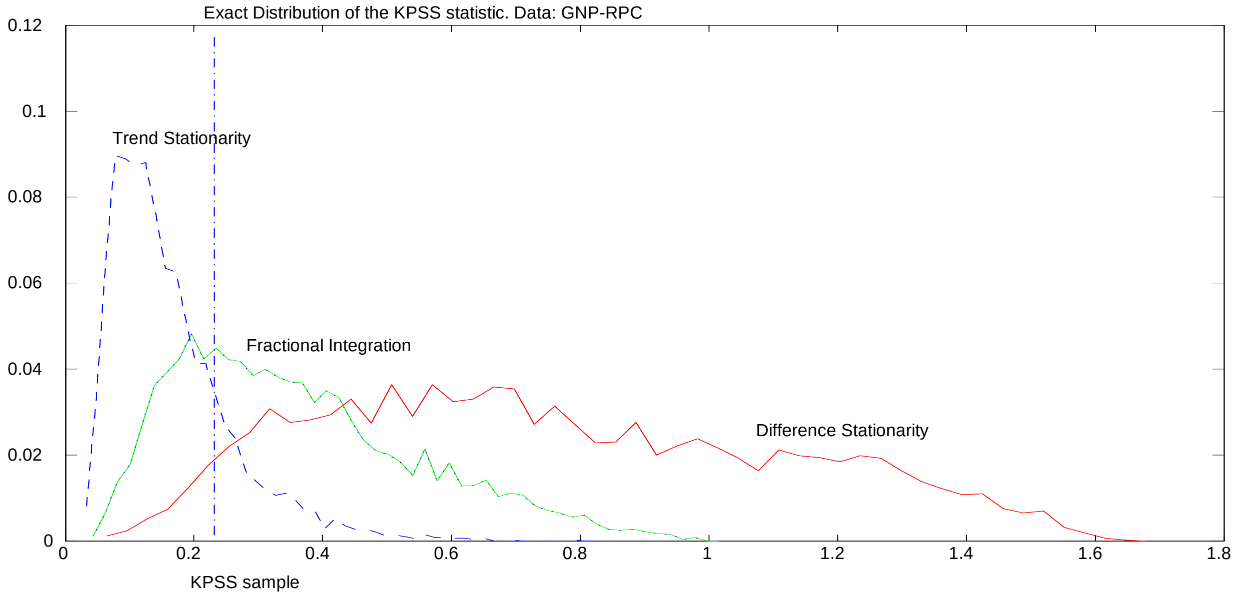
Exact Distribution of the KPSS statistic. Data: GNP-RPC
0.12
0.1
0.08
0.06
0.04
0.02
0
Trend Stationarity
Fractional Integration
Difference Stationarity
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
KPSS sample

FIGURE II