FRACTIONAL INTEGRATION VERSUS TREND STATIONARY IN TIME SERIES ANALYSIS

Francesc Marmol

97-02

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 97-02
Statistics and Econometrics Series 02
January 1997

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

FRACTIONAL INTEGRATION VERSUS TREND STATIONARITY IN TIME SERIES ANALYSIS.

Francesc Marmol*

Abstract

The objective of this paper is to study the effects of spurious detrending of a nonstationary fractionally integrated process (NFI(d), $d \geq 5$) on the performance of the traditional least squares estimators and tests. We extend previous work on the subject undertaken by Durlauf and Phillips (1988) which considered only the leading difference stationary ($d=1$) case. Moreover, we also consider the possibility of a double misspecification both in the stochastic and in the nonstochastic trends. Standard t-Student tests are shown to diverge in distribution invalidating any inference concerning the presence of time trends. On the other hand, we prove that, even under this double misspecification, the Durbin-Watson statistic remains to be a useful misspecification test.

Keywords:
Spurious detrending, nonstationary fractionally integrated processes, specification tests, misspecification.

*Marmol, Department of Statistics and Econometrics, Universidad Carlos III de Madrid; I am indebted to Niels Haldrup for helpful comments and discussions and to Juan C. Reboredo for computer assistance. I also thanks participants in the various seminars and meetings at which previous versions of this paper were presented for their remarks. Financial support from the European Commission through the *Training and Mobility of Researches Program* is gratefully acknowledged. The usual disclaimers apply.

Fractional Integration Versus Trend Stationarity in Time Series Analysis

Francesc Marmol[1]

GREMAQ-IDEI, Université des Sciences Sociales, Place Anatole France, 31042 *Toulouse Cedex, France*

and

Departamento de Estadística y Econometría, Universidad Carlos III, 28903, *Getafe, Madrid, Spain.*

The objective of this paper is to study the effects of spurious detrending of a nonstationary fractionally integrated process ($NFI(d), d \geq .5$) *on the performance of the traditional least squares estimators and tests. We extend previous work on the subject undertaken by Durlauf and Phillips* (1988) *which considered only the leading difference stationary* ($d = 1$) *case. Moreover, we also consider the possibility of a double misspecification both in the stochastic and in the nonstochastic trends. Standard t-Student tests are shown to diverge in distribution invalidating any inference concerning the presence of time trends. On the other hand, we prove that, even under this double misspecification, the Durbin-Watson statistic remains to be a useful misspecification test.*

KEYWORDS: *Spurious detrending, nonstationary fractionally integrated processes, specification tests, misspecification.*

[1] I am indebted to Niels Haldrup for helpful comments and discussions and to Juan C. Reboredo for computer assistance. I also thanks participants in the various seminars and meetings at which previous versions of this paper were presented for their remarks. Financial support from the European Commission through the *Training and Mobility of Researches Program* is gratefully acknowledged. The usual disclaimers apply.

1 INTRODUCTION

When a given time series y_t becomes stationary after differencing d times, and the degree of differentiation or *memory parameter*, *d*, is a real number, then the series is said to be *fractionally integrated of order d*, denoted $y_t \sim FI(d)$. These processes have received an increasing attention because of their ability to provide a natural and flexible characterization of the nonstationary and persistent characteristics of economic time series.

A fractionally integrated process is both stationary and invertible if $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and (weakly) nonstationary if $d \geq \frac{1}{2}$. In spite of being nonstationary, the process is mean-reverting with *transitory memory*, i.e., with any random shock having only a temporary influence on the series, if $d < 1$, in contrast with the case when $d \geq 1$, where the process is both nonstationary and not mean-reverting with *permanent memory*, i.e., with any random shock having now a permanent effect on the present and future path of the series. On the other hand, a stationary fractionally integrated process has *short-memory*, with autocorrelations decaying at an exponential rate, if $d < 0$, whereas it has *long-memory*, with autocorrelations that die out at the slower hyperbolic rate, if $0 < d < 1/2$ and as such may be expected to be useful in modelling long-term persistence. When $d = 0$, the process is white noise, with zero correlations and constant spectral density.

The implications of the fractionally integrated processes in economic data are, at least potentially, profound. They allow for more parsimonious models. Moreover, this class of processes is naturally introduced when we consider the aggregation of heterogeneous time series (Granger, 1980, Gonçalves and Gourieroux, 1987). Lastly, by allowing a rich range of spectral behavior near the origin, they can provide superior approximations to the Wold representations of many economic time series (Granger, 1966).

Hence, it seems quite reasonable to assume that the macroeconomic time series achieve stationarity after applying a fractional filter. However, it is not by large an easy task to discriminate whether a real economic time series is trend stationary or fractionally integrated. Empirically, macroeconomic variables appear to be fractionally integrated in the post-Second World War quarterly data, whilst as for the historical data covering eighty or more years, it is difficult to offer unequivocal judgement as to whether many real economic variables are fractionally integrated or trend stationary. See Baillie (1996) for a recent survey and review of the major econometric work on fractionally integrated processes and their applications in economics and finance.

As an illustration of this difficulty, Chambers (1996) used fractional models to test trend and difference $(d = 1)$ stationarity in the logarithms of five U.K. macroeconomic time series, covering from 1955(2) to 1992(2), i.e., a total of 148 observations. The author concluded that three of these series (GNP, consumption and investment) were found to be consistent with the hypothesis of difference stationarity, while the remaining two series (exports and imports), however, rejected both trend and difference stationarity, suggesting, therefore, that these series were consistent with a fractional model.

Consequently, the empirical and theoretical debate of whether an economic series are best described as being trend stationary, difference stationary or fractionally integrated continues unresolved. The purpose of this paper is to analyze the effects of spuriously detrending a nonstationary fractionally integrated process (henceforth denoted *NFI*). In doing this, the asymptotic distributions of Durlauf and Phillips (1988), which regressed a difference stationary process onto a constant and a linear time trend, are embedded in our results.

Moreover, we proof that the Durbin-Watson statistic is a valid diagnostic against misspecification even in the case where we ignore not only the stochastic trends but also misspecify the non-deterministic component of the underlying series of interest. Therefore, the results of this paper continue strongly supporting the importance of combining hypothesis testing with specification analysis as a powerful method of exposing spurious relationships.

The paper is organized as follows. In Section 2 we derive the asymptotic distributions of the traditional least squares statistics in a time trend regression when in fact the time series is in reality a nonstationary fractionally integrated process. Section 3 extends these results to the case of nonstationary fractionally integrated process with drift, whereas in

Section 4 we provide some experimental evidence on the power of the Durbin-Watson statistic in finite samples. Finally, some concluding remarks are collected in Section 5. Proofs are given in the Appendix.

With respect to the notation employed, the symbols "$\Rightarrow$", "$\xrightarrow{p}$" and "$\equiv$" denote weak convergence, convergence in probability and equality in distribution, respectively, while $[\cdot]$ denotes "integer part". Stochastic processes such that $y_\infty(r)$ with $r \in [0, 1]$ are frequently written as y_∞. Similarly, we write integrals with respect to Lebesgue measure such as $\int_0^1 y_\infty(r)dr$ as $\int y_\infty$. The symbol $\sum_{t=1}^{T}$ is denoted simply as $\sum$ and $\Gamma(\bullet)$ denotes the gamma function. Finally, all limits given in the paper are as the sample size $T \to \infty$ unless otherwise stated.

2 TIME TRENDS VS. *NFI* PROCESSES

Let us initially consider the analysis of the following least squares regression:

(1) $$y_t = \hat{\alpha} + \hat{\beta} t + \hat{u}_t, \quad t = 1, 2, \ldots, T,$$

where y_t is assumed to be a stationary time trend process.

In 1988, Durlauf and Phillips considered the estimation of this model when in fact the true data generating process (*DGP*) for y_t is a difference stationary process, $\Delta y_t = \varepsilon_t$, with the $\{\varepsilon_t\}$ sequence assumed to satisfy a functional central limit theorem (*FCLT*) of the type discusses in Phillips (1987), allowing for weak dependence and some heterogeneity over time.

Under this *DGP*, Durlauf and Phillips (1988) found that the estimated time trend coefficient in (1) was consistent, $\hat{\beta} = O_p(T^{-1/2})$, converging to the true structural coefficient of zero and that the constant term $\hat{\alpha}$ diverged in distribution, $\hat{\alpha} = O_p(T^{1/2})$ as the significance $(H_0: \alpha = 0, \beta = 0)$ *t*-Student tests, $t_\alpha = O_p(T^{1/2})$, $t_\beta = O_p(T^{1/2})$. In the same manner, the coefficient of determination R^2 was found to have a nondegenerate limiting distribution, whilst the Durbin-Watson (*DW*) statistic converged in probability to zero at a super consistent $O_p(T^{-1})$ rate, indicating the presence of some kind of misspecification in the time regression (1) proposed.

In this section, we will extend Durlauf and Phillips' findings from the assumption that the true *DGP* for y_t is a difference stationary process to the following generating mechanism:

DGP-A:

$\Delta^d y_t = \varepsilon_t$, $d \geq 1/2$, where $\varepsilon_t \sim i.i.d(0,\sigma^2)$, $E|\varepsilon_t|^q < \infty$, $q \geq max\{d-1/2, 2\}$.

Throughout this paper, we shall assume, without any loss of generality, that $\varepsilon_t = 0 \; for \; t \leq 0$. In order to derive the behavior of the *OLS* statistics in model (1), we need the following result:

LEMMA 1: *Under DGP-A,*

(2) $$T^{1/2-d} y_{[Tr]} \Rightarrow \frac{\sigma}{\Gamma(d)} \int_0^r (r-s)^{d-1} dW(s) \qquad [\equiv y_\infty(r), say],$$

(3) $$T^{-1/2-d} \sum_{t=1}^{T} y_t \Rightarrow \int_0^1 y_\infty(r) dr,$$

(4) $$T^{-3/2-d} \sum_{t=1}^{T} t y_t \Rightarrow \int_0^1 r y_\infty(r) dr,$$

(5) $$T^{-2d} \sum_{t=1}^{T} y_t^2 \Rightarrow \int_0^1 y_\infty^2(r) dr,$$

where $W(r) \sim N(0,r)$ *is a standard Brownian motion associated with the* $\{\varepsilon_t\}$ *sequence.*

Result (2) is a *FCLT* for $NFI(d)$ processes due to Akonom and Gourieroux (1988) and results (3)-(5) follow in a direct manner from (2) by applying the continuous mapping theorem (*CMT*). See also Marmol (1995).

THEOREM 1: *Under DGP-A,*

(6) $$T^{1/2-d}\hat{\alpha} \Rightarrow -6\int_0^1 (r-2/3)y_\infty(r)dr,$$

(7) $$T^{3/2-d}\hat{\beta} \Rightarrow 12\int_0^1 (r-1/2)y_\infty(r)dr,$$

(8) $$R^2 \Rightarrow \frac{12\left\{\int (r-1/2)y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2},$$

(9) $$T^{-1/2}t_\beta \Rightarrow \left\{\frac{12\left\{\int r y_\infty - 0.5\int y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2)y_\infty\right\}^2}\right\}^{1/2},$$

(10) $$T^{-1/2}t_\alpha \Rightarrow \left\{\frac{9\left\{\int (r-2/3)y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2)y_\infty\right\}^2}\right\}^{1/2},$$

(11) $$T^{2d-1}DW \Rightarrow \frac{\sigma^2\Gamma(3-2d)}{\Gamma^2(2-d)\left\{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2)y_\infty\right\}^2\right\}},$$

if $d \in \left[\tfrac{1}{2},\tfrac{3}{2}\right)$, *and*

(12) $$T^2 DW \Rightarrow \frac{\int (\Delta y_\infty)^2 + 144\left\{\int (r-1/2)y_\infty\right\}^2 - 24\int (r-1/2)y_\infty \int \Delta y_\infty}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2)y_\infty\right\}^2},$$

if $d \geq 3/2$, *where* $\Delta y_\infty = \frac{\sigma}{\Gamma(d-1)}\int_0^r (r-s)^{d-2}dW(s) \sim FI(d-1)$.

Thus, when the true *DGP* is assumed to be a *NFI*(d) process, we have that the constant term $\hat{\alpha}$ in regression (1) diverges in distribution with a rate of convergence $O_p\left(T^{d-1/2}\right)$ that depends on the memory parameter *d*, *except* in the particular $d = \frac{1}{2}$ case, where $\hat{\alpha}$ has a nondegenerate limiting distribution. With respect to the estimated time trend coefficient $\hat{\beta}$, from expression (7) we can see that it has a well-defined limiting distribution upon suitable standardization given by $T^{3/2-d}$. Consequently, this estimator will be consistent only if $d < \frac{3}{2}$. When $d > \frac{3}{2}$, $\hat{\beta}$ actually *diverges* in distribution at a rate that depends on d. Equally, in the particular case where $d = \frac{3}{2}$, $\hat{\beta}$ has a nondegenerate limiting distribution.

As regards the hypothesis testing, from expressions (9) and (10) we have that the distributions of both traditional t-Student tests diverge at a rate, $O_p\left(T^{1/2}\right)$, that does not depend on d. Consequently, the possible consistency of $\hat{\beta}$ $\left(\text{if } d < \frac{3}{2}\right)$ to the true structural coefficient of zero does not translate into desirable properties for these conventional significance tests. This latter result mirrors the asymptotic behavior of the t-Student statistic in the general spurious framework recently studied by Marmol (1996a), which allows for the presence of deterministic trends of general form in a multivariate set-up of nonstationary fractionally integrated processes with different memory parameters and possibly cointegrated in the Engle and Granger (1987) sense.

With respect to the R^2 statistic, we observe that it converges weakly to the nondegenerate random variable given in expression (8), independently, hence, of d. This result is also obtained in the case where we regress a set of stocastically independent integrated processes with different memory parameters (Marmol, 1996b) with or without the presence of deterministic trends (Marmol, 1996a). When we allow for the presence of cointegrating relationships, then $R^2 \xrightarrow{p} 1$, as proved by Haldrup (1994) and Marmol (1996a).

Finally, as regards the DW statistic, it converges in probability to zero for all $d \geq \frac{1}{2}$, even that at a different rates, depending on d. For $d \geq \frac{3}{2}$, we have from expression (12) that $DW = O_p\left(T^{-2}\right)$, whereas in the case where $d < \frac{3}{2}$, $DW = O_p\left(T^{1-2d}\right)$ depending, consequently, on the memory parameter d. This implies, in particular, that in the degenerate case $d = \frac{1}{2}$, $DW = O_p(1)$, having, therefore, a well-defined limiting distribution. Indeed, even in this case the DW statistic car be a valid method of exposing the spurious regression (see Section 4 below).

In this sense, Marmol (1995, 1996b) proved that when we regress two integrated processes (with or without the same memory parameters) with no deterministic components except for a constant term in the spurious regression, then $DW = O_p(T^{-1})$ if $d < \frac{3}{2}$ whereas $DW = O_p(T^{-2})$ otherwise. By contrast, Marmol (1996a) proved that if we regress a multivariate set of fractionally integrated processes with different memory parameters, possibly cointegrated, with deterministic components, then the results obtained in Theorem 1 for the DW statistic reemerge and $DW = O_p\left(T^{1-2d^*}\right)$ for $d^* < \frac{3}{2}$ and $DW = O_p(T^{-2})$ for $d^* \geq \frac{3}{2}$, with d^* being the minimum order of differentiation included in the underlying spurious regression.

To close this section, it is worth noting that $\hat{\alpha}$ and $\hat{\beta}$, suitable standardize, have Gaussian limiting distributions, in spite of be functionals of the fractional Brownian motion y_∞. To prove this claim, we need the following lemma, which is of interest in its own:

LEMMA 2: *Under DGP-A,*

(13) $$y_\infty(r) \sim N\left(0, \sigma^2 r^{2d-1}\right),$$

(14) $$\int_0^1 y_\infty(r)dr \sim N\left(0, \frac{2\sigma^2}{(d+1)(2d+1)}\right),$$

(15) $$\int_0^1 r y_\infty(r)dr \sim N\left(0, \frac{2\sigma^2}{(d+2)(2d+3)}\right),$$

and

(16) $$\int_0^1 (r-a)y_\infty(r)dr \sim N(0, \Theta_a),$$

where $a \in \Re$ *and*

$$\Theta_a \equiv 2\sigma^2\left[\frac{1}{(d+2)(2d+3)}+\frac{a^2}{(d+1)(2d+1)}-\frac{a(2d+3)}{(d+1)(d+2)(2d+2)}\right].$$

Now, with the help of the above lemma, it is straightforward to prove the following result:

COROLLARY 1: *Under DGP-A, the OLS estimators of* α,β *in* (1) *have the following limiting distributions*:

$$T^{1/2-d}\hat{\alpha} \equiv N(0,\Theta^{\alpha}),$$

$$T^{3/2-d}\hat{\beta} \equiv N(0,\Theta^{\beta}),$$

with $\Theta^{\alpha} = 36\Theta_{2/3}$, $\Theta^{\beta} = 144\Theta_{1/2}$ *and with* Θ_a *as defined in expression* (16).

3 *NFI* PROCESSES WITH DRIFT

Consider now the following *DGP* for the y_t sequence, where we allow for the presence of a non-zero drift, acting as a nuisance parameter:

DGP-B:

(*i*) $\Delta^d y_t = \mu + \varepsilon_t$, $d \geq 1/2, \varepsilon_t \sim i.i.d(0,\sigma^2)$, $E|\varepsilon_t|^q < \infty$, $q \geq max\{d-1/2, 2\}$.

Moreover, throughout this section we shall assume the following truncated property for the Δ^d operator:

ASSUMPTION 1:

$$\Delta^{-d} \sim \sum_{j=1}^{t}\frac{\Gamma(d+j)}{\Gamma(d)\Gamma(1+j)},\ t = 1,2,\ldots.$$

Durlauf and Phillips (1988) were also concerned with the effect of estimating (1) when, in fact, the true DGP was a difference stationary process with drift, $\Delta y_t = \mu + \varepsilon_t$, showing that the regression theory for this *DGP* was identical to the driftless case.

The intuition behind this result is that a difference stationary process with drift can be converted into a time trend plus a driftless difference stationary process

$$\Delta y_t = \mu + \varepsilon_t \Leftrightarrow y_t = \Delta^{-1}\mu + \Delta^{-1}\varepsilon_t = \mu t + \tilde{y}_t,$$

assuming zero initial conditions. Consequently, in the difference stationary case, regressing this sum against a time trend will generate identical results to those obtained in Section 2 where $\mu = 0$.

The same results will be obtained in the fractional case if we assume that y_t has been generated according to

$$y_t = \alpha + \beta t + y_t^0 = \gamma' \varsigma_t + y_t^0$$

where $\gamma = (\alpha, \beta)'$, $\varsigma_t = (1, t)'$ and y_t^0 is a $NFI(d)$ process generated under *DGP-A* In this case, least squares estimation of γ in (1) yields

$$\hat{\gamma} - \gamma = \left(\sum \varsigma_t \varsigma_t'\right)^{-1}\left(\sum \varsigma_t y_t^0\right)$$

and by defining the diagonal matrix $\Im_T = diag\{1, T\}$ jointly with the *CMT* and Lemma 1, we have that

$$T^{1/2-d}\Im_T^{-1}(\hat{\gamma} - \gamma) \Rightarrow \left(\int \tau\tau'\right)^{-1}\left(\int \tau y_\infty\right)$$

where $\tau = (1, r)'$ so that

$$T^{1/2-d}(\hat{\alpha} - \alpha) \Rightarrow -6\int_0^1 (r - 2/3) y_\infty(r) dr$$

and

$$T^{3/2-d}\left(\hat{\beta} - \beta\right) \Rightarrow 12\int_0^1 (r - 1/2) y_\infty(r) dr$$

as in Theorem 1. Consequently, if we only erroneously ignore stochastic regressors in model (1) but we correctly specified the deterministic components, then the regression theory is identical to the case where y_t is assume to be generated without these nonstochastic trends. Moreover, this result can be extended in a straightforward manner to general polynomial trends as noted by Haldrup (1991a, b) in the $d = 2$ case.

However, if we assume true *DGP-B*, then things change drastically. In effect, in this case, a reparametrization of y_t allows us to write it as the sum of a driftless $NFI(d)$ process plus a deterministic function of time

$$\Delta^d y_t = \mu + \varepsilon_t$$
$$\Leftrightarrow y_t = \Delta^{-d}\mu + \Delta^{-d}\varepsilon_t$$
$$\Leftrightarrow y_t = \xi_t + \bar{y}_t,$$

where $\xi_t = \Delta^{-d}\mu$ and $\bar{y}_t = \Delta^{-d}\varepsilon_t$. As regards the ξ_t term, under Assumption 1, it follows that

$$\xi_t = \Delta^{-d}\mu = \frac{\mu}{\Gamma(d)}\sum_{j=1}^{t}\frac{\Gamma(d+j)}{\Gamma(1+j)} \cong \frac{\mu}{\Gamma(d)}\sum_{j=1}^{t} j^{d-1} \equiv \frac{\mu}{\Gamma(d)}\tilde{\xi}_t ,$$

where the approximation follows from Sheppard's formula. See Granger (1988) for a detailed justification of this truncated deterministic filter.

Hence, by assuming true *DGP-B*, if we implement a least squares testing procedure in model (1) we are mistakenly ignore stochastic regressors and misspecifying the non-stochastic component. In this section we will explore the consequences of this double misspecification on the asymptotic behavior of the traditional least squares statistics. For this, let us first announce the following results.

LEMMA 3:

(17) $$T^{-d}\tilde{\xi}_t \rightarrow \frac{r^d}{d},$$

(18) $$T^{-d-1}\sum\tilde{\xi}_t \rightarrow \frac{1}{d(d+1)},$$

(19) $$T^{-d-2}\sum t\tilde{\xi}_t \rightarrow \frac{1}{d(d+2)},$$

(20) $$T^{-2d-1}\sum\left(\tilde{\xi}_t\right)^2 \rightarrow \frac{1}{d^2(2d+1)},$$

(21) $$T^{-2d-1/2}\sum \tilde{\xi}_t \tilde{y}_t \Rightarrow \frac{1}{d}\int_0^1 r^d y_\infty(r)dr.$$

LEMMA 4: *Under DGP-B,*

(22) $$T^{-d-1}\sum y_t \xrightarrow{p} \frac{\mu}{\Gamma(d+2)},$$

(23) $$T^{-1-2d}\sum y_t^2 \xrightarrow{p} \frac{\mu^2}{\Gamma^2(d+1)(2d+1)},$$

(24) $$T^{-1-2d}\sum (y_t-\bar{y})^2 \xrightarrow{p} \frac{\mu^2}{\Gamma^2(d)(d+1)^2(2d+1)},$$

(25) $$T^{-2-d}\sum t y_t \xrightarrow{p} \frac{\mu}{\Gamma(d+2)},$$

(26) $$T^{-2-d}\sum (y_t-\bar{y})(t-\bar{t}) \xrightarrow{p} \frac{\mu}{2\Gamma(d+3)}.$$

Using the above results and the methods employed in Theorem 1, it is now easy to find the relevant asymptotic theory for model (1).

THEOREM 2:

(*i*) *Under DGP-B, if* $d \neq 1$,

(27) $$T^{-d}\hat{\alpha} \xrightarrow{p} \mu\frac{2(1-d)}{\Gamma(d+3)},$$

(28) $$T^{1-d}\hat{\beta} \xrightarrow{p} \mu\frac{6d}{\Gamma(d+3)},$$

(29) $$R^2 \xrightarrow{p} \frac{6d-3}{(d+2)^2} \quad (\geq 0),$$

(30) $$T^{-1/2}t_\beta \xrightarrow{p} \frac{d(6d+3)^{1/2}}{d-1},$$

(31) $$T^{-1/2}t_{\alpha}\xrightarrow{p}\frac{(2d+1)^{1/2}}{d},$$

(32) *if* $1/2 \le d < 1$,

$$T^{2d}DW\xrightarrow{p}\frac{\sigma^2\Gamma(3-2d)\Gamma^2(d+3)(2d+1)}{\mu^2\Gamma(2-d)(d-1)^2},$$

(33) *if* $d > 1$,

$$T^2 DW\xrightarrow{p}\frac{(2d+1)}{(d-1)^2}\left[\frac{\Gamma^2(d+3)+36\Gamma^2(d+1)(2d-1)}{\Gamma^2(d)(2d-1)}-\right.$$

$$\left.\frac{-12\Gamma(d)\Gamma(d+3)(2d-1)}{\Gamma^2(d)(2d-1)}\right].$$

(*ii*) *Under DGP-B, in the case where* $d = 1$,

(34) $$T^{-1/2}\hat{\alpha}\equiv N\left(0,\frac{2\sigma^2}{15}\right),$$

(35) $$T^{1/2}\left(\hat{\beta}-\mu\right)\equiv N\left(0,\frac{6\sigma^2}{15}\right),$$

whilst R^2, t_{α}, t_{β} *and* DW *have the same limiting distributions as in Theorem* 1.

First at all, notice that, when $d = 1$, in model (1) we only misspecify the stochastic trends. Consequently, and according to aforementioned comments, the asymptotic behavior of the *OLS* statistics obtained in the second part of the theorem is the same than in the driftless case.

On the other hand, when $d \neq 1$ all the statistics converge in probability to some constant after suitable standardization. This constitutes a kind of return to more classical asymptotic results with stationary variables, and, of course, this fact is due to the asymptotically dominant ξ_t term. The influence of this deterministic component on the relevant asymptotic theory for the *OLS* statistics that we are examining *only cancels* exactly when $d = 1$.

Otherwise, in the case where $d \neq 1$, we can see in the first part of the theorem that the constant term is $O_p(T^d)$, i.e., it diverges and a rate that depends on d. By contrast, the estimated time trend coefficient $\hat{\beta}$ diverges when $d > 1$ but converges to the true structural coefficient of zero when $d < 1$. Further, the coefficient of determination R^2 converges in probability to a non-negative constant and the t-Student tests diverge at the rate $O_p(T^{1/2})$, as always. Finally, as in the driftless case, $DW \xrightarrow{p} 0$ for all $d \geq \frac{1}{2}$, and, hence, it remains to be a valid tool against misspecification.

4 SOME MONTE CARLO EVIDENCE

In this section we perform some Monte Carlo experiments to examine the power of the DW statistic as a misspecification test against spurious detrending in small to moderate samples. The parameter space consider in this study is the following:

$$\{T = 50;100;200\} \otimes \{d = 0.5;0.6;0.8;1;1.2;1.5;1.8;2\} \otimes \{\mu = 0;1\} \otimes \{\vartheta = 0;0.4;0.8\}.$$

Observations on the $NFI(d)$ process were generated in the following manner: First, we simulate a stationary $\eta_t \sim SFI(\delta)$ process for $\delta \in [-\frac{1}{2}, \frac{1}{2})$. In this case we have that $\eta_t = \Delta^{-\delta} u_t$, where the perturbation term u_t is generated as $u_t = \varepsilon_t - \vartheta \varepsilon_{t-1}$, where ε_t is a sequence of identically and independently distributed $N(0,1)$ variables. The fractional difference operator $\Delta^{-\delta}$ is defined as an infinite lag polynomial, so that we must truncate it in some point m. Herein we choose $m = 10,000$. Now to mimic the sample path of the $NFI(d)$ process y_t, for $d \in [\frac{1}{2}, 2]$, we simply take partial sums of η_t with initial condition $y_0 = 0$. Thus, as a matter of definition, y_t is $NFI(1.4)$ if $\Delta y_t = \eta_t \sim SFI(0.4)$. Using this procedure, for each sample size T, we generate $T + 100$ observations. Then, the first 100 observations are discarded in order to eliminate the influence of the initial conditions.

The results of our experiments are given in Tables 1-3. These results were generated by a simulation using 20,000 replications. As is well-known, a serious shortcoming of the DW test is that its exact distribution depends on the data matrix. Notwithstanding, the true distribution of the DW statistic lies between that of two other statistics, d_l (the lower bound) and d_u (the upper bound), which only depend on T and the number of regressors. The null hypothesis of no autocorrelation is rejected against the alternative of positive autocorrelation if $DW < d_l^*$ (*Region* 1) against the alternative of negative autocorrelation if $DW > 4 - d_l^*$ (*Region* 5) and not rejected if $d_u^* < DW < 4 - d_u^*$ (*Region* 3), where asterisks indicate tabulated values at appropriate significance levels (e.g., Savin and White, 1977). If $d_l^* < DW < d_u^*$ (*Region* 2) or if $4 - d_u^* < DW < 4 - d_l^*$ (*Region* 4) the test is inconclusive. For each point of the parameter space, tables give percent of times each of these five region contains the *DW* statistic.

Table 1 gives the power of the *DW* statistic in the most favorable situation, namely, where the true model is a pure *NFI* with white noise innovations. In Tables 2 and 3 we compare power in the presence of a *NFI* with moderate to large $MA(1)$ innovations, respectively. Some results are clear and in accord with the asymptotics of the previous sections. With other things held constant, (*i*) power increases as T increases This is a reflection of the consistency of the test. (*ii*) Power is higher when d is larger, i.e., as the alternative hypothesis becomes further from the null.

Other important practical conclusions that can be drawn from our simulations are the following. First, the presence or absence of the drift term μ does not change the power properties of the *DW* statistic. Durlauf and Phillips (1988) and our Theorem 2 prove that this must be the case for $d = 1$. However, Tables 1-3 show that this result remains true for all $d \geq \frac{1}{2}$. Consequently, it seems that the power properties of the *DW* statistic are independent of misspecifications in the deterministic part. Second, the percent of times the *DW* statistic lies in Regions 2 and 4 (the inconclusive regions) are almost zero in all cases (perhaps to a surprising degree, in fact).

On the other hand, the *DW* statistic has good power properties for all $d \geq \frac{1}{2}$ in the presence of moderate moving average innovations (Table 2). Indeed, this power can be very low for $d < 1$ against *NFI* processes with large moving average innovations (Table 3), pointing out the presence of *severe identification problems*. For instance, when $d = 0.8$ the null hypothesis of correct specification is accepted in about 70% for $T = 200$.

...:'y, note that, for this sample size, for $d = 0.5$ or $d = 0.6$, the *DW* statistic lies in on 5 in almost all the occasions.

5 CONCLUSIONS

.. aim of this paper has been to provide a generalization of the available results on e behavior of difference stationary processes which are misspecified as trend stationary ne series to the more general framework of misspecified nonstationary fractionally egrated processes.

.eral conclusions can be drawn from our study. First, depending on the memory rameter d, the estimated time trend coefficient can either diverge or converge to the .e structural parameter of zero. Second, when we assume that the true generating echanism includes a non-zero drift, then the asymptotic behavior of the standard *OLS* ..istics changes with respect to the driftless case, with the only exception of the case nere $d = 1$, i.e., in the difference stationary case. This is due to the fact that, except in e unit root case, for st of values of d, a least squares regression in model (1) .plies to commit *two different misspecifications,* namely, to ignore the stochastic trends d to misspecify the deterministic component of the underlying series of interest. .ird, in any case, the conventional t-Student statistics diverge at the rate $O_p(T^{1/2})$. onsequently, they will, with probability one, reject the null hypothesis of no gnificance. Hence, standard *OLS* inference remains invalid, as in the difference .tionary case. Finally, we showed that, in any case, $DW \xrightarrow{p} 0$ for all $d \geq \frac{1}{2}$, and, herefore, this statistic will, with probability one, reject the hypothesis of correct model specification, as $T \to \infty$. This, in turn, constitutes an useful diagnostic against isspecification. However, from our Monte Carlo experiments we conclude that, in oderate samples and for $d < 1$, the *DW* statistic has important identification problems cept in the simplest case of a pure *NFI* process with white noise innovations.

TABLE 1

Power of the *DW* statistic against spurious detrending

			Value of d							
T	μ	*Region*	0.5	0.6	0.8	1	1.2	1.5	1.8	2
50	0	1	0.8770	0.9560	0.9970	1.0000	0.9990	1.0000	1.0000	1.0000
50	0	2	0.0370	0.0160	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
50	0	3	0.0860	0.0280	0.0010	0.0000	0.0010	0.0000	0.0000	0.0000
50	0	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
50	0	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
50	1	1	0.8670	0.9550	0.9970	0.9990	1.0000	1.0000	0.9990	1.0000
50	1	2	0.0350	0.0170	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000
50	1	3	0.0980	0.0280	0.0030	0.0000	0.0000	0.0000	0.0010	0.0000
50	1	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
50	1	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	1	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	0	2	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	3	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	1	0.9950	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	3	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
200	0	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
200	1	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

True model: $\Delta^d y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim NIID(0,1)$.

Estimated model: $y_t = \hat{\alpha} + \hat{\beta} t + res.$

Regions 1 and 5: regions of rejection of the null hypothesis of correct specification. *Region* 3: region of no rejection of the null hypothesis of correct specification. *Regions* 2 and 4: inconclusive regions. 5% significance level.

TABLE 2

Power of the *DW* statistic against spurious detrending

			Value of d							
T	μ	*Region*	0.5	0.6	0.8	1	1.2	1.5	1.8	2
50	0	1	0.0960	0.2570	0.7540	0.9600	0.9680	0.9520	0.9600	0.9490
50	0	2	0.0340	0.0880	0.0640	0.0140	0.0130	0.0180	0.0090	0.0200
50	0	3	0.8050	0.6410	0.1790	0.0260	0.0190	0.0300	0.0310	0.0310
50	0	4	0.0340	0.0060	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000
50	0	5	0.0310	0.0080	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
50	1	1	0.1010	0.2490	0.7380	0.9460	0.9610	0.9590	0.9680	0.9560
50	1	2	0.0390	0.0810	0.0580	0.0200	0.0160	0.0140	0.0110	0.0120
50	1	3	0.7870	0.6500	0.2040	0.0340	0.0230	0.0260	0.0210	0.0320
50	1	4	0.0290	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
50	1	5	0.0440	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	1	0.2790	0.6530	0.9890	1.0000	1.0000	1.0000	1.0000	1.0000
100	0	2	0.0550	0.0470	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	3	0.6360	0.2970	0.0080	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	4	0.0110	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	0	5	0.0190	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	1	0.2700	0.6680	0.9910	1.0000	1.0000	1.0000	1.0000	1.0000
100	1	2	0.0500	0.0490	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	3	0.6630	0.2800	0.0070	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	4	0.0070	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	1	5	0.0100	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	1	0.6410	0.9590	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
200	0	2	0.0250	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	3	0.3330	0.0370	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	4	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	1	0.6510	0.9670	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
200	1	2	0.0380	0.0080	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	3	0.3080	0.0250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	5	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

True model: $\Delta^d y_t = \mu + u_t$, $u_t = \varepsilon_t - 0.4\varepsilon_{t-1}$, $\varepsilon_t \sim NIID(0,1)$.

Estimated model: $y_t = \hat{\alpha} + \hat{\beta} t + res.$

Regions 1 and 5: regions of rejection of the null hypothesis of correct specification. *Region* 3: region of no rejection of the null hypothesis of correct specification. *Regions* 2 and 4: inconclusive regions. 5% significance level.

TABLE 3

Power of the *DW* statistic against spurious detrending

			Value of d							
T	μ	*Region*	0.5	0.6	0.8	1	1.2	1.5	1.8	2
50	0	1	0.0000	0.0000	0.0130	0.1580	0.2040	0.1850	0.1860	0.1810
50	0	2	0.0000	0.0020	0.0150	0.0730	0.0640	0.0580	0.0660	0.0730
50	0	3	0.2880	0.4650	0.8110	0.7200	0.7060	0.7280	0.7220	0.7160
50	0	4	0.1300	0.1350	0.0580	0.0220	0.0140	0.0150	0.0090	0.0160
50	0	5	0.5820	0.3980	0.1030	0.0270	0.0120	0.0140	0.0170	0.0140
50	1	1	0.0000	0.0000	0.0170	0.1860	0.1920	0.1760	0.1630	0.1940
50	1	2	0.0000	0.0000	0.0120	0.0680	0.0650	0.0600	0.0890	0.0550
50	1	3	0.3010	0.4680	0.7960	0.7160	0.7210	0.7440	0.7250	0.7280
50	1	4	0.1150	0.1320	0.0790	0.0080	0.0150	0.0100	0.0130	0.0110
50	1	5	0.5840	0.4000	0.0960	0.0220	0.0070	0.0100	0.0100	0.0120
100	0	1	0.0000	0.0000	0.0530	0.5730	0.6130	0.5790	0.5990	0.5760
100	0	2	0.0000	0.0010	0.0140	0.0440	0.0490	0.0530	0.0550	0.0460
100	0	3	0.0860	0.2670	0.8030	0.3820	0.3380	0.3640	0.3450	0.3770
100	0	4	0.0380	0.0780	0.0300	0.0000	0.0000	0.0020	0.0010	0.0000
100	0	5	0.8760	0.6540	0.1000	0.0010	0.0000	0.0020	0.0000	0.0010
100	1	1	0.0000	0.0000	0.0520	0.5860	0.6430	0.5890	0.6050	0.5600
100	1	2	0.0000	0.0000	0.0300	0.0540	0.0450	0.0550	0.0490	0.0500
100	1	3	0.0760	0.2640	0.7840	0.3560	0.3090	0.3520	0.3440	0.3890
100	1	4	0.0430	0.0530	0.0320	0.0020	0.0010	0.0020	0.0010	0.0000
100	1	5	0.8810	0.6830	0.1020	0.0020	0.0020	0.0020	0.0010	0.0010
200	0	1	0.0000	0.0000	0.2000	0.9610	0.9540	0.9550	0.9480	0.9530
200	0	2	0.0000	0.0000	0.0310	0.0090	0.0100	0.0060	0.0130	0.0110
200	0	3	0.0090	0.1190	0.7190	0.0300	0.0360	0.0390	0.0390	0.0360
200	0	4	0.0010	0.0300	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
200	0	5	0.9900	0.8510	0.0440	0.0000	0.0000	0.0000	0.0000	0.0000
200	1	1	0.0000	0.0010	0.2270	0.9580	0.9670	0.9530	0.9410	0.9480
200	1	2	0.0000	0.0000	0.0240	0.0070	0.0070	0.0080	0.0150	0.0080
200	1	3	0.0060	0.1060	0.6940	0.0340	0.0260	0.0390	0.0430	0.0440
200	1	4	0.0010	0.0320	0.0070	0.0010	0.0000	0.0000	0.0010	0.0000
200	1	5	0.9930	0.8610	0.0480	0.0000	0.0000	0.0000	0.0000	0.0000

True model: $\Delta^d y_t = \mu + u_t$, $u_t = \varepsilon_t - 0.8\varepsilon_{t-1}$, $\varepsilon_t \sim NIID(0,1)$

Estimated model: $y_t = \hat{\alpha} + \hat{\beta} t + res.$

Regions 1 and 5: regions of rejection of the null hypothesis of correct specification. *Region* 3: region of no rejection of the null hypothesis of correct specification. *Regions* 2 and 4: inconclusive regions. 5% significance level.

APPENDIX

PROOF OF THEOREM 1: First, consider the *OLS* estimators of α and β in model (1)

$$\hat{\alpha} = \theta^{-1}\left[\sum t^2 \sum y_t - \sum t y_t \sum t\right],$$

and

$$\hat{\beta} = \theta^{-1}\left[T\sum t y_t - \sum t \sum y_t\right],$$

where $\theta = T\sum t^2 - \left(\sum t\right)^2$. Given the general result

(A1) $$T^{-m}\sum t^{m-1} = T^{-1}\sum (t/T)^{m-1} \to \int_0^1 r^{m-1} dr = \frac{r^m}{m}\Big|_0^1 = \frac{1}{m},$$

it follows that

(A2) $$T^{-4}\theta = T^{-3}\sum t^2 - \left(T^{-2}\sum t\right)^2 \to \frac{1}{12}.$$

Therefore, using Lemma 1, (A2) and the *CMT*, we have that

$$T^{1/2-d}\hat{\alpha} = \left(T^{-4}\theta\right)^{-1}\left[T^{-3}\sum t^2 T^{-1/2-d}\sum y_t - T^{-3/2-d}\sum t y_t T^{-2}\sum t\right]$$

$$\Rightarrow 12\left[\frac{1}{3}\int y_\infty - \frac{1}{2}\int r y_\infty\right] = -6\int (r - 2/3) y_\infty .$$

and

$$T^{3/2-d}\hat{\beta} = \left(T^{-4}\theta\right)^{-1}\left[T^{-3/2-d}\sum t y_t - T^{-2}\sum t T^{-1/2-d}\sum y_t\right]$$

$$\Rightarrow 12\left[\int r y_\infty - \frac{1}{2}\int y_\infty\right] = 12\int (r - 1/2) y_\infty .$$

With respect to the R^2 statistic,

$$R^2 = \frac{\sum (\hat{y}_t - \bar{y})^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{\sum \hat{u}_t^2}{\sum (y_t - \bar{y})^2},$$

using Lemma 1, (A1) and the *CMT*, we get the following results

(A3) $$T^{-2d}\sum (y_t - \bar{y})^2 = T^{-2d}\sum y_t^2 - \left(T^{-1/2-d}\sum y_t\right)^2 \Rightarrow \int y_\infty^2 - \left\{\int y_\infty\right\}^2,$$

(A4) $$T^{-3/2-d}\sum (y_t - \bar{y})(t - \bar{t}) \Rightarrow \int (r - 1/2) y_\infty ,$$

and

(A5) $$T^{-2d}\sum \hat{u}_t^2 = T^{-2d}\sum (y_t - \bar{y})^2 - T^{3/2-d}\hat{\beta} T^{-3/2-d}\sum (y_t - \bar{y})(t - \bar{t})$$

$$\Rightarrow \int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2) y_\infty\right\}^2,$$

where $\bar{y} = T^{-1}\sum y_t$ and $\bar{t} = T^{-1}\sum t$. Now, (A3)- (A5) and the *CMT* together imply

$$R^2 \Rightarrow 1 - \frac{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2) y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2} = \frac{12\left\{\int (r-1/2) y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2}.$$

Consider now the *t*-Student statistics,

$$t_\beta^2 = \frac{\hat{\beta}^2}{\frac{T^{-1}\sum \hat{u}^2}{\sum (t-\bar{t})^2}},$$

and

$$t_\alpha^2 = \frac{\hat{\alpha}^2}{\frac{T^{-1}\sum \hat{u}^2 \sum t^2}{T\sum (t-\bar{t})^2}}.$$

In this case, from (6), (7), (A1), (A5) and the *CMT*, it is straightforward to show that

$$T^{-1}t_\beta^2 = \frac{T^{3-2d}\hat{\beta}^2}{\frac{T^{-2d}\sum \hat{u}^2}{T^{-3}\sum (t-\bar{t})^2}} \Rightarrow \frac{12\left\{\int r y_\infty - 0.5\int y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2) y_\infty\right\}^2},$$

and

$$T^{-1}t_\alpha^2 = \frac{T^{1-2d}\hat{\alpha}^2}{\frac{T^{-2d}\sum \hat{u}^2 T^{-3}\sum t^2}{T^{-3}\sum (t-\bar{t})^2}} \Rightarrow \frac{9\left\{\int (r-2/3) y_\infty\right\}^2}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int (r-1/2) y_\infty\right\}^2}.$$

Finally, consider the asymptotic behavior of the *DW* statistic, defined as

$$DW = \frac{\sum (\Delta \hat{u}_t)^2}{\sum \hat{u}_t^2}.$$

Manipulating the numerator, we get

$$\sum(\Delta\hat{u}_t)^2 = \sum(\Delta y_t)^2 + T\hat{\beta}^2 - 2\hat{\beta}\sum\Delta y_t.$$

Let us first assume that $d \geq 3/2$. In this case, Δy_t is a $NFI(d-1)$ process and hence, using Lemma 1, it follows that

(A6) $$T^{2-2d}\sum(\Delta y_t)^2 \Rightarrow \int(\Delta y_\infty)^2,$$

and

(A7) $$T^{1/2-d}\sum\Delta y_t \Rightarrow \int\Delta y_\infty.$$

Consequently, using (7), (A5)- (A7) and the *CMT*, it can be deduced that

$$T^2 DW \Rightarrow \frac{\int(\Delta y_\infty)^2 + 144\left\{\int(r-1/2)y_\infty\right\}^2 - 24\int(r-1/2)y_\infty\int\Delta y_\infty}{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int(r-1/2)y_\infty\right\}^2}.$$

In the same manner, consider the case where $d < 3/2$. Now, observe that Δy_t is a $SFI(d-1)$ process, and standard asymptotic results can be applied to its sample moments, yielding

(A8) $$T^{-1}\sum\Delta y_t \xrightarrow{p} 0$$

and

(A9) $$T^{-1}\sum(\Delta y_t)^2 \xrightarrow{p} \operatorname{var}(\Delta y_t) = \frac{\sigma^2\Gamma(3-2d)}{\Gamma(2-d)}.$$

Hence, by virtue of expressions (7), (A5), (A8), (A9) and the *CMT*, it is direct to show that $T^{-1}\sum(\Delta\hat{u}_t)^2 \approx T^{-1}\sum(\Delta y_t)^2$ which, in turn, implies that

$$T^{2d-1}DW \Rightarrow \frac{\sigma^2\Gamma(3-2d)}{\Gamma^2(2-d)\left\{\int y_\infty^2 - \left\{\int y_\infty\right\}^2 - 12\left\{\int(r-1/2)y_\infty\right\}^2\right\}},$$

as required.

Q.E.D.

PROOF OF LEMMA 2: Given that $W(r)$ is a Brownian motion, we know, by linearity with respect to $W(r)$, that $y_\infty, \int y_\infty, \int r y_\infty$ and $\int(r-a)y_\infty$ are also Gaussian processes.

Indeed, expression (13) follows from Jonas (1983). To prove expression (14), notice from (13) that

(A10) $$\int_0^1 y_\infty(r)dr = \sigma\int_0^1 r^{d-1}W(r)dr.$$

Now, given that

$$E\int_0^1 r^{d-1}W(r)dr = \int_0^1 r^{d-1}E(W(r))dr = 0$$

and

$$\operatorname{var}\int_0^1 r^{d-1}W(r)dr = E\int_0^1\int_0^1 r^{d-1}s^{d-1}W(r)W(s)drds$$

$$=\int_0^1\int_0^1 r^{d-1}s^{d-1}E(W(r)W(s))drds$$

$$=\int_0^1\int_0^1 r^{d-1}s^{d-1}min\{r,s\}drds = \int_0^1 r^{d-1}\left[\int_0^r s^d ds + r\int_r^1 s^{d-1}ds\right]dr$$

$$=\int_0^1 r^{d-1}\left[\frac{s^{d+1}}{d+1}\bigg|_0^r + r\frac{s^d}{d}\bigg|_r^1\right]dr = \int_0^1 r^{d-1}\left[\frac{r^{d+1}}{d+1}+\frac{r-r^d}{d}\right]dr = \frac{2}{(d+1)(2d+1)},$$

we get $\int_0^1 r^{d-1}W(r)dr \sim N(0,\ 2/(d+1)(2d+1))$, meaning that

$$\int_0^1 y_\infty(r)dr \sim N\left(0,\ \frac{2\sigma^2}{(d+1)(2d+1)}\right).$$

Similarly, it can be proved that $\int_0^1 r^d W(r)dr \sim N(0,\ 2/(d+2)(2d+3))$, implying

$$\int_0^1 ry_\infty(r)dr \sim N\left(0,\ \frac{2\sigma^2}{(d+2)(2d+3)}\right).$$

Finally, to prove expression (16), notice that

$$\int_0^1 (r-a)y_\infty(r)dr = \int_0^1 ry_\infty(r)dr - a\int_0^1 y_\infty(r)dr.$$

Hence,

$$E\int_0^1 (r-a)y_\infty(r)dr = \int_0^1 rE(y_\infty(r))dr - a\int_0^1 E(y_\infty(r))dr = 0$$

and

$$\operatorname{var}\int_0^1 (r-a)y_\infty(r)dr \equiv \Theta_a = \operatorname{var}\int_0^1 ry_\infty(r)dr + a^2 \operatorname{var}\int_0^1 y_\infty(r)dr$$

$$-2a\operatorname{cov}\left(\int_0^1 ry_\infty(r)dr, \int_0^1 y_\infty(r)dr\right).$$

Thus, given that

$$\operatorname{cov}\left(\int_0^1 ry_\infty(r)dr, \int_0^1 y_\infty(r)dr\right) = E\int_0^1 ry_\infty(r)dr\int_0^1 y_\infty(r)dr$$

$$= \int_0^1\int_0^1 rE(y_\infty(r)y_\infty(s))drds,$$

with $E(y_\infty(r)y_\infty(s)) = \sigma^2 r^{d-1}s^{d-1}E(W_\infty(r)W_\infty(s))$, we obtain after some manipulations

$$\operatorname{cov}\left(\int_0^1 ry_\infty(r)dr, \int_0^1 y_\infty(r)dr\right) = \sigma^2\int_0^1\int_0^1 r^{d-1}s^{d-1}min\{r,s\}drds$$

$$= \frac{\sigma^2(2d+3)}{(2d+2)(d+1)(d+2)}.$$

Therefore,

$$\Theta_a = \frac{2\sigma^2}{(d+2)(2d+3)} + \frac{2a^2\sigma^2}{(d+1)(2d+1)} - \frac{2a\sigma^2(2d+3)}{(2d+2)(d+1)(d+2)}$$

$$= 2\sigma^2\left[\frac{1}{(d+2)(2d+3)} + \frac{a^2}{(d+1)(2d+1)} - \frac{a(2d+3)}{(d+1)(d+2)(2d+2)}\right].$$

Q.E.D.

PROOF OF LEMMA 3: Straightforward using Lemma 1 and the *CMT*

$$T^{-d}\tilde{\xi}_t = T^{-d}\tilde{\xi}_{[Tr]} = T^{-1}\sum_{t=1}^{[Tr]}(t/T)^{d-1} \to \int_0^r s^{d-1}ds = \frac{r^d}{d},$$

$$T^{-d-1}\sum \tilde{\xi}_t = T^{-1}\sum\left(T^{-d}\tilde{\xi}_t\right) \rightarrow \frac{1}{d}\int_0^1 r^d dr = \frac{1}{d(d+1)},$$

$$T^{-d-2}\sum t\tilde{\xi}_t = T^{-1}\sum\left(T^{-1}t\right)\left(T^{-d}\tilde{\xi}_t\right) \rightarrow \frac{1}{d}\int_0^1 r^{d+1} dr = \frac{1}{d(d+2)},$$

$$T^{-2d-1}\sum\left(\tilde{\xi}_t\right)^2 = T^{-1}\sum\left(T^{-d}\tilde{\xi}_t\right)^2 \rightarrow \frac{1}{d^2}\int_0^1 r^{2d} dr = \frac{1}{d^2(2d+1)},$$

$$T^{-2d-1/2}\sum \tilde{\xi}_t\tilde{y}_t = T^{-1}\sum\left(T^{-d}\tilde{\xi}_t\right)\left(T^{1/2-d}\tilde{y}_t\right) \Rightarrow \frac{1}{d}\int_0^1 r^d y_\infty(r) dr.$$

Q.E.D.

PROOF OF LEMMA 4. Using Lemma 1, Lemma 3 and the *CMT*, expressions (22)-(26) can be proved in the following manner:

$$T^{-d-1}\sum y_t = \frac{\mu}{\Gamma(d)}T^{-d-1}\sum \tilde{\xi}_t + T^{-1/2}T^{-1/2-d}\sum \tilde{y}_t \xrightarrow{p} \frac{\mu}{\Gamma(d)d(d+1)},$$

$$T^{-1-2d}\sum y_t^2 = \frac{\mu^2}{\Gamma^2(d)}T^{-1-2d}\sum \tilde{\xi}_t^2 + T^{-1}T^{-2d}\sum \tilde{y}_t^2$$

$$+2\frac{\mu}{\Gamma(d)}T^{-1/2}T^{-1/2-2d}\sum \tilde{\xi}_t\tilde{y}_t \xrightarrow{p} \frac{\mu^2}{\Gamma^2(d+1)(2d+1)},$$

$$T^{-1-2d}\sum(y_t-\bar{y})^2 = T^{-1-2d}\sum y_t^2 - \left(T^{-d}\bar{y}\right)^2$$

$$\xrightarrow{p} \frac{\mu^2}{\Gamma^2(d+1)(2d+1)} - \frac{\mu^2}{\Gamma^2(d+2)} = \frac{\mu^2}{\Gamma^2(d)(d+1)^2(2d+1)},$$

$$T^{-2-d}\sum ty_t = \frac{\mu}{\Gamma(d)}T^{-2-d}\sum t\tilde{\xi}_t + T^{-1/2}T^{-3/2-d}\sum t\tilde{y}_t \xrightarrow{p} \frac{\mu}{\Gamma(d)d(d+2)},$$

and

$$T^{-2-d}\sum(y_t-\bar{y})(t-\bar{t}) = T^{-2-d}\sum ty_t - T^{-2}\sum t T^{-1-d}\sum y_t$$

$$\xrightarrow{p} \frac{\mu}{\Gamma(d)d(d+2)} - \frac{\mu}{2\Gamma(d)d(d+1)} = \frac{\mu}{2\Gamma(d+3)}.$$

Q.E.D.

PROOF OF THEOREM 2. Under *DGP-B*,

$$\hat{\alpha} = \theta^{-1}\left[\sum t^2 \sum y_t - \sum ty_t \sum t\right]$$

$$= \theta^{-1}\left[\sum t^2 \sum \tilde{y}_t - \sum t\tilde{y}_t \sum t\right] + \theta^{-1}\frac{\mu}{\Gamma(d)}\left[\sum t^2 \sum \tilde{\xi}_t - \sum t\tilde{\xi}_t \sum t\right].$$

If $d \neq 1$, then $\sum t^2 \sum \tilde{\xi}_t \neq \sum t\tilde{\xi}_t \sum t$. Hence, using (A1), Lemma 1, Lemma 2 and the *CMT*

$$T^{-d}\hat{\alpha} = \left(T^{-4}\theta\right)^{-1} T^{-1/2}\left[T^{-3}\sum t^2 T^{-1/2-d}\sum \tilde{y}_t - T^{-3/2-d}\sum t\tilde{y}_t T^{-2}\sum t\right]$$

$$+\left(T^{-4}\theta\right)^{-1}\frac{\mu}{\Gamma(d)}\left[T^{-3}\sum t^2 T^{-1-d}\sum \tilde{\xi}_t - T^{-2-d}\sum t\tilde{\xi}_t T^{-2}\sum t\right]$$

$$\xrightarrow{p} \frac{\mu}{\Gamma(d)}\left[\frac{4}{d(d+1)} - \frac{6}{d(d+2)}\right] = \frac{\mu}{\Gamma(d)}\frac{2(1-d)}{d(d+1)(d+2)} = \mu\frac{2(1-d)}{\Gamma(d+3)},$$

where in the last equality we have employed the well-known recursion formula $\Gamma(z+1) = z\Gamma(z), z \in \Re^+$.

With respect to $\hat{\beta}$, we have that, under *DGP-B*,

$$\hat{\beta} = \theta^{-1}\left[T\sum ty_t - \sum t \sum y_t\right]$$

$$= \theta^{-1}\left[T\sum t\tilde{y}_t - \sum t \sum \tilde{y}_t\right] + \theta^{-1}\frac{\mu}{\Gamma(d)}\left[T\sum t\tilde{\xi}_t - \sum t \sum \tilde{\xi}_t\right].$$

When $d \neq 1$, $T\sum t\tilde{\xi}_t - \sum t \sum \tilde{\xi}_t \neq \theta\Gamma(d)$. Therefore, using again Lemma 1, Lemma 2 and the *CMT* we get

$$T^{1-d}\hat{\beta} = \left(T^{-4}\theta\right)^{-1} T^{-1/2}\left[T^{-3/2-d}\sum t\tilde{y}_t - T^{-2}\sum t T^{-1/2-d}\sum \tilde{y}_t\right]$$

$$+\left(T^{-4}\theta\right)^{-1}\frac{\mu}{\Gamma(d)}\left[T^{-2-d}\sum t\tilde{\xi}_t - T^{-2}\sum t T^{-1-d}\sum \tilde{\xi}_t\right]$$

$$\xrightarrow{p} \frac{\mu}{\Gamma(d)}\frac{6}{(d+1)(d+2)} = \mu\frac{6d}{\Gamma(d)d(d+1)(d+2)} = \mu\frac{6d}{\Gamma(d+3)}.$$

As regards the coefficient of determination given that

$$\sum \hat{u}_t^2 = \sum (y_t - \bar{y})^2 - \hat{\beta}\sum (y_t - \bar{y})(t - \bar{t}),$$

then, under *DGP-B*, if $d \neq 1$, from (24), (26) and the *CMT* it follows that

$$T^{-1-2d}\sum \hat{u}_t^2 = T^{-1-2d}\sum (y_t - \bar{y})^2 - T^{1-d}\hat{\beta}T^{-2-d}\sum (y_t - \bar{y})(t - \bar{t})$$

$$\xrightarrow{p} \frac{\mu}{\Gamma^2(d)(d+1)^2(2d+1)} - \frac{3\mu^2}{\Gamma^2(d)(d+1)^2(d+2)^2}$$

(A11) $$= \frac{\mu^2(d-1)^2}{\Gamma^2(d+3)(2d+1)}.$$

Hence, using (24), (A11) and the *CMT* yields

$$R^2 = 1 - \frac{T^{-1-2d}\sum \hat{u}_t^2}{T^{-1-2d}\sum (y_t - \bar{y})^2} \xrightarrow{p} \frac{6d-3}{(d+2)^2}.$$

With respect to the t-ratios, using (A1), (A2), (28), (A11) and the *CMT*, it is direct to show that

$$T^{-1}t_{\beta}^2 = \frac{T^{2-2d}\hat{\beta}^2}{\frac{T^{-2d-1}\sum \hat{u}^2}{T^{-3}\sum (t-\bar{t})^2}} \xrightarrow{p} \frac{d(6d+3)^{1/2}}{d-1}$$

and

$$T^{-1}t_{\alpha}^2 = \frac{T^{-2d}\hat{\alpha}^2}{\frac{T^{-2d-1}\sum \hat{u}^2 T^{-3}\sum t^2}{T^{-3}\sum (t-\bar{t})^2}} \xrightarrow{p} \frac{(2d+1)^{1/2}}{d}.$$

Finally, with respect to the *DW* statistic, under *DGP-B*, when $d \geq 3/2$,

$$\Delta y_t = \Delta \xi_t + \Delta \tilde{y}_t = \frac{\mu}{\Gamma(d)} t^{d-1} + \Delta \tilde{y}_t.$$

Consequently, using (A1), Lemma 1 and the *CMT*, it follows that

$$T^{-d}\sum \Delta y_t = \frac{\mu}{\Gamma(d)} T^{-d}\sum t^{d-1} + T^{-1/2}T^{1/2-d}\sum \Delta \tilde{y}_t \xrightarrow{p} \frac{\mu}{\Gamma(d+1)}$$

and

$$T^{1-2d}\sum (\Delta y_t)^2 = \frac{\mu^2}{\Gamma^2(d)} T^{1-2d}\sum t^{2d-2} + T^{-1}T^{-2d+2}\sum (\Delta \tilde{y}_t)^2$$

$$+2\frac{\mu}{\Gamma(d)}T^{-1/2}T^{3/2-2d}\sum t^{d-1}\Delta\tilde{y}_t \xrightarrow{p} \frac{\mu^2}{\Gamma^2(d)(2d-1)},$$

which implies, using (A11) and the *CMT*, that

$$T^2 DW \xrightarrow{p} \frac{(2d+1)}{\Gamma^2(d)(2d-1)(d-1)^2}\left[\Gamma^2(d+3)+\right.$$

$$\left.36\Gamma^2(d+1)(2d-1)-12\Gamma(d)\Gamma(d+3)(2d-1)\right].$$

When $d<1$,

$$T^{-1}\sum \Delta y_t = \frac{\mu}{\Gamma(d)}T^{d-1}T^{-d}\sum t^{d-1} + T^{-1}\sum \Delta\tilde{y}_t \xrightarrow{p} 0$$

and

$$T^{-1}\sum(\Delta y_t)^2 = \frac{\mu^2}{\Gamma^2(d)}T^{2d-2}T^{1-2d}\sum t^{2d-2} + T^{-1}\sum(\Delta\tilde{y}_t)^2$$

$$+2\frac{\mu}{\Gamma(d)}T^{-1}\sum t^{d-1}\Delta\tilde{y}_t \xrightarrow{p} \operatorname{var}(\Delta\tilde{y}_t) = \frac{\sigma^2\Gamma(3-2d)}{\Gamma(2-d)}.$$

Hence,

$$T^{2d} DW = \frac{T^{-1}\sum(\Delta\hat{u}_t)^2}{T^{-2d-1}\sum \hat{u}_t^2} = \frac{T^{-1}\sum(\Delta\tilde{y}_t)^2}{T^{-2d-1}\sum \hat{u}_t^2} + o_p(1)$$

$$\xrightarrow{p} \frac{\sigma^2\Gamma(3-2d)\Gamma^2(d+3)(2d+1)}{\mu^2\Gamma(2-d)(d-1)^2}.$$

Similarly, when $1<d<3/2$,

$$T^{-d}\sum \Delta y_t = \frac{\mu}{\Gamma(d)}T^{-d}\sum t^{d-1} + T^{1-d}T^{-1}\sum \Delta\tilde{y}_t \xrightarrow{p} \frac{\mu}{\Gamma(d+1)}$$

and

$$T^{1-2d}\sum(\Delta y_t)^2 = \frac{\mu^2}{\Gamma^2(d)}T^{1-2d}\sum t^{2d-2} + T^{2-2d}T^{-1}\sum(\Delta\tilde{y}_t)^2$$

$$+2\frac{\mu}{\Gamma(d)}T^{2-2d}T^{-1}\sum t^{d-1}\Delta\tilde{y}_t \xrightarrow{p} \frac{\mu^2}{\Gamma^2(d)(2d-1)},$$

so that

$$T^{1-2d}\sum(\Delta\hat{u}_t)^2 = T^{1-2d}\sum(\Delta\tilde{y}_t)^2 + T^{2-2d}\hat{\beta}^2 - 2T^{1-d}\hat{\beta}T^{-d}\sum\Delta\tilde{y}_t$$

$$\xrightarrow{p} \mu^2\left[\frac{\Gamma^2(d+3)+36\Gamma^2(d+1)(2d-1)-12\Gamma(d+3)\Gamma(d)(2d-1)}{\Gamma^2(d+3)\Gamma^2(d)(2d-1)}\right].$$

Consequently, collecting the above results we obtain

$$T^2DW \xrightarrow{p} \frac{(2d+1)}{(d-1)^2}\left[\frac{\Gamma^2(d+3)+36\Gamma^2(d+1)(2d-1)}{\Gamma^2(d)(2d-1)}\right.$$

$$\left.\frac{-12\Gamma(d+3)\Gamma(d)(2d-1)}{\Gamma^2(d)(2d-1)}\right].$$

Finally, it is straightforward to prove that, if $d=1$, the *OLS* statistics have the same limiting distributions as in Theorem 3 under *DGP-A*. For instance, when $d=1$, $\tilde{\xi}_t = t$ and $\Gamma(d)=\Gamma(1)=1$. Hence,

$$\sum t^2\sum\tilde{\xi}_t - \sum t\tilde{\xi}_t\sum t = \sum t^2\sum t - \sum t^2\sum t = 0$$

and

$$\frac{T\sum t\tilde{\xi}_t - \sum t\sum\tilde{\xi}_t}{\theta\Gamma(d)} = \frac{T\sum t^2 - \left(\sum t\right)^2}{T\sum t^2 - \left(\sum t\right)^2} = 1,$$

so that

$$\hat{\alpha} = \theta^{-1}\left[\sum t^2\sum\tilde{y}_t - \sum t\tilde{y}_t\sum t\right]$$

and

$$\hat{\beta} - \mu = \theta^{-1}\left[T\sum t\tilde{y}_t - \sum t\sum\tilde{y}_t\right],$$

as in the driftless case, i.e., under *DGP-A*. Consequently, when $d=1$, Corollary 1 applies and, then, using expression (16) in Lemma 2, we obtain $T^{-1/2}\hat{\alpha} \equiv N(0,\Theta^{\alpha})$ and $T^{1/2}(\hat{\beta}-\mu) \equiv N(0,\Theta^{\beta})$, where $\Theta^{\alpha} = 2\sigma^2/15$ and $\Theta^{\beta} = 6\sigma^2/15$.

Q.E.D.

REFERENCES

AKONOM, J. AND C. GOURIEROUX (1988): "A Functional Limit Theorem for Fractional Processes", CEPREMAP, Discussion Paper Series No. 88-01.

BAILLIE, R. (1996): "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5-59.

CHAMBERS, M.J. (1996): "Fractional Integration, Trend Stationarity and Difference Stationarity. Evidence from Some U.K. Macroeconomic Time Series," *Economics Letters*, 50, 19-24.

DURLAUF, S.N. AND P.C.B. PHILLIPS (1988): "Trends versus Random Walks in Time Series Analysis," *Econometrica*, 56, 1333-1354.

ENGLE, R. AND C.W.J. GRANGER (1987): "Cointegration and Error-Correction: Representation, Estimation and Testing," *Econometrica*, 35, 251-276.

GONÇALVES, E. AND C. GOURIEROUX (1987): "Agrégation de Processus Autorégressifs d'Ordre 1," *Annales d'Economie et de Statistique*, 12, 127-147.

GRANGER, C.W.J. (1966): "The Typical Spectral Shape of an Economic Variable",. *Econometrica*, 34, 150-161.

_______ (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227-238.

_______ (1988): "Models that Generate Trends", *Journal of Time Series Analysis*, 9, 329-343.

HALDRUP, N. (1991a): "Testing for Double Unit Roots," mimeo, Aarhus University.

_______ (1991b): "Multivariate Regression Models with $I(2)$ Variables," mimeo, Aarhus University.

_______ (1994): "The Asymptotics of Single-Equation Cointegration Regressions with $I(1)$ and $I(2)$ Variables," *Journal of Econometrics*, 63, 153-181.

JONAS, A. (1983): "Persistent Memory Random Processes", Unpublished Ph.D. Dissertation, Department of Statistics, Harvard University.

MARMOL, F. (1995): "Spurious Regressions between $I(d)$ Processes," *Journal of Time Series Analysis*, 16, 313-321.

_______ (1996a): "Spurious Regression Theory with Nonstationary Fractionally Integrated Processes," *Journal of Econometrics*, in press.

_______ (1996b): "Nonsense Regressions between Integrated Processes of Different Orders," *Oxford Bulletin of Economics and Statistics*, 58, 525-536.

PHILLIPS, P.C.B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-301.

SAVIN, N.E. and K.J. WHITE (1977): "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors", *Econometrica*, 45, 1989-1996.