Working Paper 99-65
Statistics and Econometrics Series 22
September 1999

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax 34 - 91- 624.9849

HOW SPURIOUS FEATURES ARISE IN CASE OF FRACTIONAL COINTEGRATION

Francesc Marmol *

Abstract

It is well-known that a linear regression among the levels of independent highly persistent processes yields high values of the corresponding coefficient of determination along with divergent t-ratios and low values of the Durbin-Watson statistic. In fact, such a behaviour of the customary OLS statistics has become a sort of definition of the so-called spurious regressions in econometrics. In this paper, however, we show how these *spurious stylized facts* also arise among nonstationary (fractionally) cointegrated processes.

Key Words
Spurious regressions; cointegration; fractionally integrated processes.

*Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain, e-mail: fmarmol@est-econ.uc3m.es; Web page: www.halweb..uc3m.es/fmarmol. J.E.L. Classifications: C15, C22. I thank Miguel A. Arranz, Juan J. Dolado and Uwe Hassler, whose helpful comments and suggestions have improved this paper. All remaining errors are mine.

"Econometricians have found their Philosophers' Stone; it is called regression analysis and is used for transforming data into "significant" results! Deception is easily practised from false recipes intended to simulate useful findings, and these are derogatively referred by the profession as "nonsense regressions" (...)".
(David F. Hendry, "Econometrics - Alchemy or Science?")

1. INTRODUCTION

If we run a regression among the levels of independent nonstationary time series, then we shall obtain divergent t-ratios along with high values of the coefficient of multiple correlation R^2 or the corrected coefficient $\overline{R}^2$ and an extremely low value for the Durbin-Watson (DW) statistic. This is the so-called *spurious regression* problem. Granger and Newbold (1974, 1986) through simulations, and Phillips (1986, 1998) analytically, showed how these findings apply in the case where such a spurious or nonsense regression is composed by independent $I(1)$ or near $I(1)$ processes.

In recent years, it has been known that these *spurious stylized facts* do not hold only for independent $I(1)$ processes but also for other persistent processes, such as higher order integrated processes (Haldrup, 1994, Marmol, 1995), nonstationary fractionally integrated processes (Tsay and Chung, 1995, Cappuccio and Lubian, 1997, Marmol, 1998) or stochastic unit root processes (Granger and Swanson, 1997). Furthermore, Haldrup (1994), Tsay and Chung (1995), Hassler (1996) and Marmol (1996) show that the spurious problem also appears in situations where the underlying processes are allowed to have different orders of integration or memory parameters.

On the other hand, Tsay and Chung (1995) show that if we regress a stationary long memory processes on a constant term and another independent stationary long memory process, then as long as their memory parameters sum up to a value greater than $1/2$, the t-ratios become divergent and spurious effects occur. Moreover, recently Granger et al. (1998) find evidence that similar results are found with positively autocorrelated series on long moving averages. Therefore, it seems that it is the long memory or strong temporal dependence, instead of only nonstationarity or lack of ergodicity, that causes such spurious effects. Nonstationarity in one or all of the variables only helps to accelerate the divergence rates.

By contrast, it is well-known (see, e.g., Banerjee et al., 1986; Phillips and Durlauf, 1986) that when the processes are $I(1)$ but *cointegrated* with $I(0)$ errors, the regression coefficients and $\left(1-R^2\right)$ are T-consistent while the t-ratios and the DW are bounded in probability. Similar qualitative results have been provided by Phillips (1988) and Dolado and Marmol (1998) for near-integrated and nonstationary fractionally integrated processes, respectively. Of course, for some values of the sample size, T, and for some values of the nuisance parameters of the underlying Data Generating Process (*DGP*), one can obtain low values of the DW statistic jointly with high values of R^2 and the t-ratios, but *in general*, the spurious stylized facts are no longer present in cointegrated systems with weakly stationary $I(0)$ errors.

All the above results have led many authors to claim that a high value of R^2 or $\overline{R}^2$ combined with a low value of DW and a highly significant value of the corresponding t-statistic is no indication of a true relationship. In this paper, however, we demonstrate, both theoretically and by means of a small Monte Carlo experiment that these *spurious stylized facts* do also apply *in general* in the cointegrated case as long as the equilibrium errors are *fractionally integrated*. More precisely, we show that in this case the *t-ratios always diverge and* $R^2 \xrightarrow{p} 1$, whereas the *DW converges in probability to zero* whenever the error term has memory parameter, say δ, in the nonstationary range, i.e., $\delta > 1/2$. Otherwise, i.e. if $\delta < 1/2$, the DW statistic has a well-defined limiting distribution though even in this latter case DW takes values close to zero except for very small values of δ. This paper, hence, proves the observational equivalence on the basis of the customary OLS statistics between spurious and a large and important class of genuine regressions.

2. THE MODEL AND ASUMPTIONS

To keep things simple (and without loss of generality), our main concern will be with a cointegrated system of fractionally integrated processes that are generated by the triangular representation

(1) $$y_t = \beta x_t + z_t, \qquad (t = 1, 2, \ldots),$$

where

(2) $$\Delta^d x_t = u_{1t}, \quad d > \tfrac{1}{2},$$

and

(3) $$\Delta^\delta z_t = u_{2t}, \quad d > \delta > 0.$$

Let $\{u_t\}_{-\infty}^{\infty} = \{(u_{1t}, u_{2t})'\}_{-\infty}^{\infty}$ be a zero mean weakly stationary $I(0)$ sequence of random 2×1 vectors satisfying that each element of $\{u_t\}_{-\infty}^{\infty}$ is either L_r − bounded for $r > 2$, and $L_2 - NED$ of size $-1/2$ on a α −mixing sequence of size $-r/(r-2)$, or uniformly square-integrable, and $L_2 - NED$ of size $-1/2$ on a $\phi-$mixing sequence of size $-r/(2(r-1))$, $r \geq 2$.[1]

Define now

(4) $\quad W_t = (x_t, z_t)' = \Delta(L)^{-1} u_t,$

where $\Delta(L)$ is a diagonal 2×2 lag polynomial matrix having diagonal elements $(1-L)^d$ and $(1-L)^\delta$. Further, let $D_T = diag(T^{1/2+d}, T^{1/2+\delta})$ and $W_T(\xi) = \sum_{t=1}^{[T\xi]} D_T^{-1} W_t$, $\xi \in [0,1]$. Then, following Davidson and de Jong (1999), we have that as the sample size $T \uparrow \infty$,

(5) $\quad W_T(\xi) \Rightarrow W(\xi),$

where $W(\xi)$ is a functional of a 2×1 vector of fractional Brownian motions as introduced by Mandelbrot and Van Ness (1968), and with the symbol "$\Rightarrow$" denoting weak convergence of the associated probability measures.

Therefore, from expression (5) we obtain that

(6) $\quad x_t = O_p(T^{d-1/2})$

and

(7) $\quad z_t = O_p(T^{\delta-1/2}).$

3. OLS ASYMPTOTICS UNDER FRACTIONAL COINTEGRATION

Assume now that we are interested in the limiting distribution of the *OLS* estimator of β in the *DGP* given by expressions (1)-(3),

[1] While this assumption rules out nonstationarity in the form of trending sequences of L_2- and L_r-norms, for example, it allows for many weakly dependent time series including a broad class of data generating mechanisms such a finite order *ARMA* models under very general conditions.

(8) $$\hat{\beta}-\beta=\frac{\sum x_t z_t}{\sum x_t^2},$$

and associated statistics, namely,

(9) $$t_\beta=\frac{\hat{\beta}-\beta}{s_{\hat{\beta}}},\quad s_{\hat{\beta}}^2=\frac{T^{-1}\sum \hat{z}_t^2}{\sum x_t^2},\qquad \hat{z}_t=y_t-\hat{\beta}x_t,$$

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

and

(11) $$DW=\frac{\sum_{t=2}^{T}(\Delta\hat{z}_t)^2}{\sum \hat{z}_t^2},$$

where all summations Σ run from $t=1$ to T.

THEOREM 1: *Given the DGP* (1)-(3), *then as* $T\uparrow\infty$,

(12) $$\hat{\beta}\xrightarrow{p}\beta,\ with\ \hat{\beta}=\begin{cases}O_p\left(T^{\delta-d}\right) & if\ \delta>\tfrac{1}{2}\ or\ 0<\delta<\tfrac{1}{2}\ and\ d+\delta>1,\forall d,\\ O_p\left(T^{1-2d}\right) & if\ 0<\delta<\tfrac{1}{2}\ and\ d+\delta<1,\forall d\end{cases},$$

(13) $$t_\beta\to\infty,\ with\ t_\beta=\begin{cases}O_p\left(T^{1/2}\right) & if\ \delta>\tfrac{1}{2},\forall d\\ O_p\left(T^{\delta}\right) & if\ 0<\delta<\tfrac{1}{2}\ and\ d+\delta>1,\forall d,\\ O_p\left(T^{1-d}\right) & if\ 0<\delta<\tfrac{1}{2}\ and\ d+\delta<1,\forall d\end{cases}$$

(14) $$R^2\xrightarrow{p}1,\ with\ \left(1-R^2\right)=\begin{cases}O_p\left(T^{-2(d-\delta)}\right) & if\ \delta>\tfrac{1}{2},\forall d\\ O_p\left(T^{1-2d}\right) & if\ \delta<\tfrac{1}{2},\forall d\end{cases},$$

(15) $$DW\xrightarrow{p}0\ if\ \delta>1/2,\ with\ DW=\begin{cases}O_p\left(T^{-2}\right) & if\ d>\tfrac{3}{2}\ and\ \delta>\tfrac{3}{2}\\ O_p\left(T^{1-2\delta}\right) & if\ \tfrac{1}{2}<\delta<\tfrac{3}{2},\forall d\end{cases},$$

and

(16) $DW \xrightarrow{p} 2(1-\rho_1(\delta))$ *if* $0<\delta<1/2$, $\forall d$,

where $\rho_1(\delta)$ *stands for the first lag correlation coefficient of* z_t *and where the symbol* "$\xrightarrow{p}$" *denotes convergence in probability.*

PROOF: See Appendix.

From this theorem the following comments are in order. First, the least squares regression coefficient $\hat{\beta}$ is a consistent estimator of the corresponding theoretical counterpart for $\forall\ d,\delta$. Second, if we focus the attention on the inferential results, we can observe from (13) that the *t-ratio always diverges*, at a rate that depends on whether the error term is a nonstationary or stationary fractionally integrated process. *Only when* $\delta=0$, i.e., only when the error term is a weakly stationary *I*(0) process, the *t*-ratio has a well-defined limiting distribution. Third, as regards the coefficient of multiple correlation, we have from (14) that it tends to one in probability for $\forall\ d,\delta$, as in the standard $d=1$, $\delta=0$ case; see, i.e., Banerjee et al. (1986). Therefore, in a fractionally cointegrated system, with an increasing probability, we are likely to obtain high values of R^2. Please, notice that the rate of convergence to unity is smaller in the nonstationary $\delta>1/2$ case.

Lastly, consider the performance of the *DW* statistic. From (15) we deduce that it converges to zero (at different rates) whenever $\delta>1/2$. Low values of *DW* in this case, thus, must be expected in fractionally cointegrated systems with nonstationary equilibrium errors. On the other hand, from (16) we learn that *DW* converges in probability to $2(1-\rho_1(\delta))$ as in the conventional *I*(0) case if $\delta<1/2$. Notice, however, the dependency of the DW statistic on the memory parameter δ of the error term through ρ_1. In fact, it can be proved that DW and δ are inversely related, and hence, the smaller the value of δ, the closer is the DW statistic of zero, other things held constants.

For instance, assume the empirically relevant case where z_t is a stationary $ARFIMA(0, \delta, 1)$ process,

(17) $\Delta^{\delta} z_t = (1 - \theta B)\varepsilon_t, \quad |\theta| < 1, \; \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \; 0 < \delta < 1/2.$

Then, it follows from Hosking (1981, Lemma 2) that

(18) $$\rho_1(\delta,\theta) = \frac{(1+\theta^2)\delta(2-\delta) - 2\theta(1-\delta+\delta^2)}{(1-\delta)(2-\delta)\{1+\theta^2 - 2\theta\delta/(1-\delta)\}},$$

and after some tedious algebra, we get

$$\frac{\partial DW(\delta,\theta)}{\partial\theta} > 0 \qquad \text{and} \qquad \frac{\partial DW(\delta,\theta)}{\partial\delta} < 0,$$

uniformly in $|\theta| < 1$ and $0 < \delta < 1/2$. Table 1 below shows the theoretical values of the DW statistic for selected values of θ and δ.

TABLE 1 ABOUT HERE

4. SOME MONTE CARLO EVIDENCE

In order to gain further insights into the finite sample performance of the OLS statistics in a fractionally cointegrated system, we have conducted the following small experiment. Let x_t, y_t be two $I(1)$ processes in the conditional model

(19) $y_t = \beta x_t + z_t \quad \beta = 1,$

(20) $\Delta x_t = u_{1t},$

(21) $\Delta^{\delta} z_t = u_{2t}, \qquad 0 < \delta < 1,$

where u_{1t}, u_{2t} are generated as i.i.d. standard Gaussian variables, such that $E(u_{1t} u_{2s}) = 0$ for all *t*, s, in order to avoid second-order bias problems. Further, the sample size was set equal to 100. The results obtained over 5,000 simulations are collected in Table 2. It is clear from this table that diverging t-ratios associated with high values of R^2 and low values of *DW* are not (only), *in general*, spurious stylized facts.

TABLE 2 ABOUT HERE

---

5. CONCLUSIONS

Herein we have shown, both, theoretically and by means of some experimental evidence, how the so-called spurious stylized facts do also arise in case of fractional cointegration. In this sense, perhaps the most important finding of the paper is that conventional *inference remains invalid* not only in the spurious but in the cointegrated case whenever the equilibrium errors are fractionally integrated.

This, in turn, leads naturally to the conclusion that tests of statistical hypothesis of interest crucially depend on a pre-classification of the variables of interest by their degrees of integration as well as on the memory parameters of the corresponding equilibrium errors. Incorrect estimation of the persistent characteristics of the underlying processes will lead to an inappropriate choice of critical values and thus lead to incorrect inference.

APPENDIX: PROOF OF THEOREM 1

Consider first the case where $\delta > 1/2$. Then, from (6), (7) and the Continuous Mapping Theorem (*CMT*) we have that $\sum x_t^2 = O_p\left(T^{2d}\right)$, $\sum z_t^2 = O_p\left(T^{2\delta}\right)$ and $\sum z_t = O_p\left(T^{1/2+\delta}\right)$, so that by the Cauchy-Schwarz inequality, it follows that $\left(\sum x_t z_t\right)^2 \le \left(\sum x_t^2\right)\left(\sum z_t^2\right) = O_p\left(T^{2(d+\delta)}\right)$ and thus, $\sum x_t z_t = O_p\left(T^{d+\delta}\right)$, implying that $\hat{\beta} = O_p\left(T^{\delta-d}\right)$. By contrast, if $0 < \delta < 1/2$ then $\sum z_t^2 = O_p(T)$ by the Ergodic Theorem, whereas from Robinson and Marinucci (1998) we obtain that $\sum x_t z_t = O_p\left(T^{d+\delta}\right)$ (and hence $\hat{\beta} = O_p\left(T^{\delta-d}\right)$) if $d + \delta > 1$ and $\sum x_t z_t = O_p(T)$ (and thus $\hat{\beta} = O_p\left(T^{1-2d}\right)$) if $d + \delta < 1$.

On the other hand, since $\sum \hat{z}_t^2 = \sum z_t^2 + \left(\hat{\beta} - \beta\right)^2 \sum x_t^2 - 2\left(\hat{\beta} - \beta\right)\sum x_t z_t$, then from the previous analysis it follows that for $\delta > \frac{1}{2}$, $\sum \hat{z}_t^2 = O_p\left(T^{2\delta}\right)$ and therefore $s^2 = T^{-1}\sum \hat{z}_t^2 = O_p\left(T^{2\delta-1}\right)$, $s_{\hat{\beta}}^2 = O_p\left(T^{-2(d-\delta)-1}\right)$ and $t_\beta = O_p\left(T^{1/2}\right)$, whereas that if $0 < \delta < 1/2$, then $\sum \hat{z}_t^2 = O_p(T)$, $s^2 = O_p(1)$ and $s_{\hat{\beta}}^2 = O_p\left(T^{-2d}\right)$ so that $t_\beta = O_p\left(T^{\delta}\right)$ $d + \delta > 1$ and $t_\beta = O_p\left(T^{1-d}\right)$ if $d + \delta < 1$, proving (13).

With regard to the R^2 statistic, since $\sum \left(y_t - \bar{y}\right)^2 = O_p\left(T^{2d}\right)$ for all δ, it is immediate to deduce that $R^2 = O_p\left(T^{-2(d-\delta)}\right)$ if $\delta > 1/2$ and $R^2 = O_p\left(T^{1-2d}\right)$ otherwise.

Lastly, let us be concerned with the behavior of the *DW* statistic. After some manipulations, the numerator becomes

$$\sum(\Delta\hat{z}_t)^2 = \sum(\Delta z_t)^2 + (\hat{\beta}-\beta)^2\sum(\Delta x_t)^2 - 2(\hat{\beta}-\beta)\sum(\Delta x_t)(\Delta z_t).$$

Consequently, if $d > 3/2$ and $\delta > 3/2$, it follows from the previous analysis that $\sum(\Delta\hat{z}_t)^2 = O_p(T^{2\delta-2})$ so that $DW = O_p(T^{-2})$. On the other hand, if $d > 3/2$ but $1/2 < \delta < 3/2$, we obtain that $\sum(\Delta\hat{z}_t)^2 = O_p(T)$ implying that $DW = O_p(T^{1-2\delta})$. In the same manner, in the case where $1/2 < d < 3/2$ and $1/2 < \delta < 3/2$, $\sum(\Delta\hat{z}_t)^2 = O_p(T)$ given that $\sum(\Delta x_t)(\Delta z_t) = O_p(T)$ by the Ergodic Theorem and then $DW = O_p(T^{1-2\delta})$.

Likewise, by compiling all the above results, we have that in the case where $0 < \delta < 1/2$, $\sum(\Delta\hat{z}_t)^2 = O_p(T)$ and thereby $DW = O_p(1)$, whether $d > 3/2$ or $1/2 < d < 3/2$. In fact, since $\sum(\Delta\hat{z}_t)^2 = \sum(\Delta z_t)^2 + o_p(T)$ and $\sum\hat{z}_t^2 = \sum z_t^2 + o_p(T)$, it turns out that

$$DW = \frac{T^{-1}\sum(\Delta\hat{z}_t)^2}{T^{-1}\sum\hat{z}_t^2} \xrightarrow{p} \frac{\mathrm{var}(\Delta z_t)}{\mathrm{var}(z_t)} = 2(1-\rho_1(\delta)),$$

which is the assertion of the theorem. ■

REFERENCES

BANERJEE, A., DOLADO, J.J., HENDRY, D.F. & SMITH, G.W. (1986) Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics*, 48, pp. 253-277.

CAPPUCCIO, N. & LUBIAN, D. (1997) Spurious regressions between $I(1)$ processes with long memory errors, *Journal of Time Series Analysis*, 18, pp. 341-354.

DAVIDSON, J. & DE JONG, R.M. (1999) The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes, preprint, Michigan State University.

DOLADO, J.J. & MARMOL, F. (1998) FM-OLS estimation of cointegrating relationships among nonstationary fractionally integrated processes, Working Paper 98-41, Department of Statistics and Econometrics, Universidad Carlos III de Madrid.

GRANGER, C.W.J., HYUNG, N. & JEON, Y. (1998) Spurious regressions with stationary series, Discussion Paper 98-25, Department of Economics, University of California at San Diego.

GRANGER, C.W.J. & NEWBOLD, P. (1974) Spurious regressions in econometrics, *Journal of Econometrics*, 2, pp. 111-120.

GRANGER, C.W.J. & NEWBOLD, P. (1986) *Forecasting economic time series*, 2nd ed., London: Academic Press.

GRANGER, C.W.J. & SWANSON, N. (1997) An introduction to stochastic unit root processes, *Journal of Econometrics*, 80, pp. 35-62.

HALDRUP, N. (1994) The asymptotics of single-equation cointegration regressions with $I(1)$ and $I(2)$ variables, *Journal of Econometrics,* 63, pp. 153-181.

HASSLER, U. (1996) Spurious regressions when stationary regressors are included, *Economics Letters*, 50, pp. 25-31.

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

MANDELBROT, B.B & VAN NESS, J.W. (1968) Fractional brownian motions, fractional brownian noises and applications, *SIAM Review,* 10, pp. 422-437.

MARMOL, F. (1995) Spurious regressions between $I(d)$ processes, *Journal of Time Series Analysis,* 16, pp. 313-321.

MARMOL, F. (1996) Nonsense regressions between integrated processes of different orders, *Oxford Bulletin of Economics and Statistics*, 58, pp. 525-536.

MARMOL, F. (1998) Spurious regression theory with nonstationary fractionally integrated processes, *Journal of Econometrics*, 84, pp. 232-250.

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

PHILLIPS, P.C.B. (1988) Regression theory for near-integrated processes, *Econometrica,* 56, pp.1021-1043.

PHILLIPS, P.C.B. (1998) New tools for understanding spurious regressions, *Econometrica,* 66, pp. 1299-1325.

PHILLIPS, P.C.B. & DURLAUF, S.N. (1986) Multiple time series regression with integrated processes, *Review of Economic Studies,* 53, pp. 473-495.

ROBINSON, P.M. & MARINUCCI, D. (1998) Semiparametric frequency domain analysis of fractional cointegration, STICERD Discussion Paper # 348 LSE.

TSAY, W-J. & CHUNG, C.-F. (1995) The spurious regression of fractionally integrated processes", Econometrics and Economic Theory Paper No. 9503, Michigan State University.

TABLE 1. Asymptotic values of the *DW* statistic when z_t is generated by expression (17)

δ θ	-0.9	-0.5	0	0.5	0.9
0	1.01	1.20	2.00	2.80	2.99
0.1	0.85	1.02	1.78	2.69	2.94
0.2	0.67	0.81	1.50	2.54	2.88
0.3	0.48	0.57	1.14	2.31	2.81
0.4	0.25	0.30	0.67	1.86	2.72
0.49	0.03	0.04	0.08	0.43	2.35

TABLE 2. Features of regressions among $I(1)$ cointegrated processes with fractionally integrated of order δ errors.

δ	Average $\hat{\beta}$	$\Pr\left[\lvert t_\beta \rvert \geq 2\right]$	Average R^2	Average *DW*
0.01	0.99961	0.87	0.85	2.021
0.10	0.99873	0.88	0.80	1.842
0.40	0.99904	0.89	0.81	0.721
0.49	0.99863	0.90	0.77	0.101
0.51	1.00116	0.90	0.78	0.091
0.80	0.98482	0.94	0.70	0.042
0.90	0.98220	0.97	0.65	0.024
0.99	1.01816	0.97	0.62	0.017