Spurious Multivariate Regressions Under Fractionally Integrated Processes

Ventosa-Santaulària, Daniel; Vera-Valdés, J. Eduardo; Lasak, Katarzyna; Ramírez-Vargas, Ricardo

Published in:
Communications in Statistics: Theory and Methods

DOI (link to publication from Publisher):
10.1080/03610926.2020.1758945

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Spurious multivariate regressions under fractionally integrated processes

Daniel Ventosa-Santaulària* J. Eduardo Vera-Valdés†
Katarzyna Łasak‡ Ricardo Ramírez-Vargas§

April 6, 2020

Abstract

This paper studies spurious regression in the multivariate case for any finite number of fractionally integrated variables, stationary or not. We prove that the asymptotic behavior of the estimated coefficients and their t-statistics depend on the degrees of persistence of the regressors and the regressand. Nonsense inference could therefore be drawn when the sum of the degrees of persistence of the regressor and regressand is greater or equal than $1/2$. Moreover, the asymptotic behavior from the most persistent regressor spreads to correlated regressors. Thus, the risk of uncovering spurious results increases as more regressors are included. Inference drawn from other test statistics such as the joint $\mathcal{F}$ test, the R-squared, and the Durbin-Watson is also misleading. Finite sample evidence supports our findings.

Keywords: Fractional Integration, Long Memory, Spurious Regression, Multivariate Regression.

*Corresponding author. Banco de México and CIDE [Carr. México-Toluca 3655, Col. Lomas de Santa Fe, Del. Álvaro Obregón, México D.F, C.P. 01210. E-mail address: daniel.ventosa@cide.edu, Ph. +52 (55) 57279800 x-2723].

†Aalborg University and CREATES. eduardo@math.aau.dk

‡VU Amsterdam and Tinbergen Institute, the Netherlands. k.a.lasak@vu.nl

§CIDE. ricardo.ramirez@alumnos.cide.edu

1 Introduction

Spurious regression in econometrics is widely understood as the failure of conventional testing procedures. It was uncovered by Granger and Newbold (1974), and later explained by Phillips (1986), assuming that independent driftless unit root processes generate the regressor and regressand in a univariate regression. Many extensions have been considered. For example, spurious regression under mean and trend stationarity, under independent higher-integration-order processes, or in out-of-sample predictability criteria were studied by Hassler (1996), Marmol (1996), Hassler (2003), and Martínez-Rivera et al. (2012), respectively.

Although the results by Phillips (1986) and Durlauf and Phillips (1988) suggest that it is nonstationarity that causes spurious effects, Granger and Newbold's (1974) simulation exercise also uncovered spurious regression under stationarity. This case, not considered in Phillips (1986), was later studied by Granger et al. (2001) for positively autocorrelated autoregressive series and long moving average processes. In this regard, several authors consider that non-resolved autocorrelation problems lie at the origin of spurious regression; see McCallum (2010), Kolev (2011), Agiakloglou (2013), and Zhang (2018). The authors supported their arguments with Monte Carlo simulations. Nonetheless, other studies, also based in Monte Carlo experiments, provide evidence that spurious regression is more than just poorly controlled autocorrelation; see Sollis (2011), Ferson et al. (2003a), Martínez-Rivera and Ventosa-Santaulària (2012) and Ventosa-Santaulària et al. (2016). Furthermore, Mikosch and De Vries (2006) uncovered spurious regression when the distribution of the innovations is fat-tailed, while Ferson et al. (2003b) and Deng (2013) studied spurious regression in the presence of measurement errors. It is therefore relevant to consider that spurious regression is not an exclusively nonstationary phenomenon. However, this avenue has been scarcely considered.

The spurious regression phenomenon under long memory was first examined by Cappuccio and Lubian (1997) and Marmol (1998) using fractionally integrated processes of order d. Tsay and Chung (2000), TC hereafter, also studied the asymptotic properties of a regression with independent stationary and nonstationary fractionally integrated processes in a univariate setting. TC showed that the t-statistics in an OLS regression have orders of convergence that vary depending on the orders of integration of the processes. To be more precise, TC (p. 155) found that, "as long as [the variables'] orders of integration sum up to a value greater than $1/2$, the t-ratios become divergent and spurious effects occur".

TC's results show that misleading inference can occur in a regression between two stationary $I(d)$ processes. The authors thus argued that strong dependence originates the spurious effects. In other words, the underlying causes of spurious regression can be better understood as "strong temporal properties", as explained by Granger et al. (2001).

We aim to extend this line of inquiry by providing further theoretical and finite sample evidence that spurious regression may be due to the persistence of the series. We therefore extend TC's findings concerning fractionally integrated processes to the multivariate case. Such an extension is relevant because: (i) the assumption that there is only one explanatory variable in the model is restrictive, and; (ii) fractionally integrated processes are quite common in finance and macroeconomics. In a review of the empirical literature, Baillie (1996) notes that price and inflation series for several countries behave as long memory processes. Moreover, he mentions applications of fractionally integrated models to asset prices, stock returns, exchange rates, and interest rates; a few recent examples can be found in Leccadito et al. (2015), Varneskov and Perron (2018), and Osterrieder et al. (2019). We consider a specification with an arbitrary finite number of explanatory variables, independent of the regressand, and we analyze both stationary and nonstationary cases.

For the stationary case, our results show that, for non-correlated regressors, when the sum of the persistence parameter of the regressand and that of a regressor is above or equal to $1/2$, the t-statistic associated to the estimated coefficient diverges. For correlated regressors, a more relevant scenario, the divergence depends on the maximum degree of persistence among the regressors. The behavior of the $\mathcal{F}$-statistic is similar to that of the t-statistics, albeit dependent on the sum of the persistence parameters of the regressand and the most persistent of all regressors, regardless of whether the regressors are correlated or not. Hence, if the underlying processes are persistent enough, nonsense inference could be drawn from at least some t-statistics and the joint $\mathcal{F}$-statistic. Moreover, we find that the R-squared collapses to 0 at a rate that also depends on the sum of the order of integration of the regressand and the highest order of integration among all regressors, correctly signaling the poor fit. As for the Durbin-Watson statistic, it converges to a value in the $(0, 2)$ interval that only depends on the degree of persistence of the regressand.

For the nonstationary case, our results extend the classical spurious regression literature with nonstationary variables to the multivariate fractionally integrated case. That is, all t-statistics and $\mathcal{F}$-statistic diverge, and they do so at an even faster rate than for the stationary case. Moreover, the Durbin-Watson statistic collapses to zero, while the R-squared does not, thus, unreliably signaling the poor fit.

The paper proceeds as follows: Section 2 presents the theoretical framework and finite sample evidence for the multivariate fractionally integrated stationary case, while Section 3 focuses on the nonstationary case. Section 4 presents our concluding remarks. Proofs for the theorems are provided in Appendices A and B, while Appendix C provides additional finite sample evidence.

2 Stationary case

2.1 Asymptotic results

Fractionally integrated processes were introduced in time series econometrics by Granger and Joyeux (1980), and Hosking (1981). A fractionally integrated process, denoted $FI(d_z)$, is a discrete-time stochastic process z_t that satisfies $(1-L)^{d_z} z_t = a_{z,t}$, where L is the lag operator, d_z is the fractional difference parameter, and $(1-L)^d$ is the fractional difference operator, defined as $(1-L)^d = \sum_{j=0}^{\infty} \Psi_j L^j$, where $\Psi_j = \Gamma(j-d)[\Gamma(j+1)\Gamma(-d)]^{-1}$, and $\Gamma(\cdot)$ is the gamma function. The innovations sequence $a_{z,t}$ is *i.i.d.* white noise with zero mean and finite variance $\sigma^2_{a_z}$. The autocovariance function of the fractionally integrated process is

$$\gamma_z(j) = \frac{\Gamma(1-2d_z)\Gamma(d_z+j)}{\Gamma(d_z)\Gamma(1-d_z)\Gamma(1-d_z+j)}\sigma^2_{a_z},$$

and its first autocorrelation[1] is

$$\rho_z(1) = \frac{d_z}{1-d_z}.$$

When $d_z > 0$, the process is said to have long memory since it exhibits long-range dependence in the sense that $\sum_{j=-\infty}^{\infty} \gamma_z(j) = \infty$; see Haldrup and Vera-Valdés (2017) for a recent review on long memory definitions. Moreover, when $d_z < 1/2$, z_t is stationary; while nonstationary for $d_z > 1/2$.

For the stationary case, we let the regressand be a fractionally integrated process with parameter d_y, while each of the k regressors follows a signal plus noise specification where the signal follows a fractionally integrated process with parameter d_{x_i}, and the noise models the eventual correlation between the regressors. Let $m \in \{1, 2, \cdots, k\}$ regressors be correlated. Without loss of generality, let the correlated regressors appear first in the specification. Our multivariate specification is thus given by

$$y_t = (1-L)^{d_y} \epsilon_{y,t}, \tag{1}$$

and

$$x_i = (1-L)^{d_{x_i}} \epsilon_{i,t} + w_{i,t}, \tag{2}$$

where $\epsilon_{z,t}$ are *i.i.d.* white noise with zero mean and finite variance $\sigma^2_{\epsilon,z}$, and $d_z \in (0, 1/2)$ for $z = y, x_1, \cdots, x_k$. Furthermore, $w_t = (w_{1,t}, w_{2,t}, \cdots, w_{k,t})'$ is a zero mean random vector

[1] The sample autocorrelation function for stationary fractionally integrated processes was studied by Hosking (1984); the nonstationary sample autocorrelation function was obtained by Hassler (1997).

with variance matrix Σ, $\forall t$, given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,m} & 0 & \cdots & 0 \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{2,m} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,m} & \sigma_{2,m} & \cdots & \sigma_m^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{m+1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix}; \tag{3}$$

that is, $\sigma_{i,j} \neq 0$ for $1 \leq i, j \leq m$, $i \neq j$, and $\sigma_{i,j} = 0$ otherwise.[2]

Note that our specification allows for correlation between the regressors while simultaneously each regressor provides independent information by the fractionally integrated signal. These assumptions are in line with the ones made by the estimation procedure; recall that near-perfect correlation would make OLS computationally unstable, and possibly not defined. Note that near-perfect correlation translates to the case where regressors share the signal.

Furthermore, our specification is quite general. On the one hand, it encloses the uncorrelated case by making $m = 1$, and the fully correlated one by making $m = k$. On the other hand, we can allow for some short memory autocorrelation on w_t to capture the dynamics in the correlation among regressors. The only restriction imposed on w_t is that it has a finite vector moving average representation. Note that this restriction ensures that the noise process is less persistent than the signal.

As for the notation we employ, let $\hat{\beta}_j$, for $j = 0, 1, \ldots, k$, denote the OLS estimators of the parameters, where $\hat{\beta}_0$ is the estimator of the constant, and t_{β_j} their associated t-statistics. Furthermore, let $\mathcal{F}$, R^2, and $\mathcal{DW}$ denote the joint significance test statistic, the R-squared measure of fit, and the Durbin Watson statistic, respectively.

We summarise the data generating processes considered in the stationary multivariate extension in Assumption 1.[3]

Assumption 1. *Let y_t be an independent stationary fractionally integrated process of order $d_y \in (0, 1/2)$ as in (1), and let $x_{i,t}$ for $i = 1, 2, \ldots, k$, be signal plus noise processes given by (2). Suppose also that $E\left[\epsilon_{z,t}\right]^{q_z} < \infty$ with $q_z \geq \max\left\{4, -\frac{8d_z}{1+2d_z}\right\}$ for all $z = y, x_1, \cdots, x_k$.*

Theorem 1 shows that inference drawn from a regression involving such processes can be misleading.

[2]We write σ_i^2 instead of $\sigma_{i,i}$ $\forall i$ to ease notation.

[3]Note that both type-I and type-II fractionally integrated processes could be encompassed in Assumption 1 by making $\epsilon_{z,t} = 0$, for $z = y, x_i$, $\forall t \leq 0$, for type-II. See Marinucci and Robinson (2000) for further details.

Theorem 1. *Let Assumption 1 hold. Suppose that the linear specification* $y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{i,t} + u_t$ *is estimated by OLS. Then, as* $T \to \infty$,

i) $\hat{\beta}_0 = O_p(T^{d_y - \frac{1}{2}})$;

$$ii)\ \hat{\beta}_i = \begin{cases} O_p(T^{-\frac{1}{2}}) & \text{for } \bar{d}_{x_{1:m}} + d_y < \frac{1}{2}, \\ O_p\left[(T^{-1}\log T)^{\frac{1}{2}}\right] & \text{for } \bar{d}_{x_{1:m}} + d_y = \frac{1}{2}, \\ O_p(T^{\bar{d}_{x_{1:m}} + d_y - 1}) & \text{for } \frac{1}{2} < \bar{d}_{x_{1:m}} + d_y, \end{cases}$$

for $i = 1, \ldots, m$;

$$iii)\ \hat{\beta}_i = \begin{cases} O_p(T^{-\frac{1}{2}}) & \text{for } d_{x_i} + d_y < \frac{1}{2}, \\ O_p\left[(T^{-1}\log T)^{\frac{1}{2}}\right] & \text{for } d_{x_i} + d_y = \frac{1}{2}, \\ O_p(T^{d_{x_i} + d_y - 1}) & \text{for } \frac{1}{2} < d_{x_i} + d_y, \end{cases}$$

for $i = m+1, \ldots, k$;

vi) $t_{\beta_0} = O_p(T^{d_y})$;

$$v)\ t_{\beta_i} = \begin{cases} O_p(1) & \text{for } \bar{d}_{x_{1:m}} + d_y < \frac{1}{2}, \\ O_p\left[(\log T)^{\frac{1}{2}}\right] & \text{for } \bar{d}_{x_{1:m}} + d_y = \frac{1}{2}, \\ O_p(T^{\bar{d}_{x_{1:m}} + d_y - \frac{1}{2}}) & \text{for } \frac{1}{2} < \bar{d}_{x_{1:m}} + d_y, \end{cases}$$

for $i = 1, \ldots, m$;

$$vi)\ t_{\beta_i} = \begin{cases} O_p(1) & \text{for } d_{x_i} + d_y < \frac{1}{2}, \\ O_p\left[(\log T)^{\frac{1}{2}}\right] & \text{for } d_{x_i} + d_y = \frac{1}{2}, \\ O_p(T^{d_{x_i} + d_y - \frac{1}{2}}) & \text{for } \frac{1}{2} < d_{x_i} + d_y, \end{cases}$$

for $i = m+1, \ldots, k$.

Furthermore,

$$vii)\ R^2 = \begin{cases} O_p\left(T^{-1}\right) & \text{for } \bar{d}_{x_{1:k}} + d_y < \frac{1}{2}, \\ O_p\left(T^{-1}\log T\right) & \text{for } \bar{d}_{x_{1:k}} + d_y = \frac{1}{2}, \\ O_p\left[T^{2(\bar{d}_{x_{1:k}} + d_y - 1)}\right] & \text{for } \frac{1}{2} < \bar{d}_{x_{1:k}} + d_y; \end{cases}$$

$$viii)\ \mathcal{F} = \begin{cases} O_p\left(1\right) & \text{for } \bar{d}_{x_{1:k}} + d_y < \frac{1}{2}, \\ O_p\left(\log T\right) & \text{for } \bar{d}_{x_{1:k}} + d_y = \frac{1}{2}, \\ O_p\left[T^{2(\bar{d}_{x_{1:k}} + d_y) - 1}\right] & \text{for } \frac{1}{2} < \bar{d}_{x_{1:k}} + d_y; \end{cases}$$

ix) $\mathcal{DW} \xrightarrow{P} 2 - 2\rho_y(1) = \frac{2(1-2d_y)}{1-d_y}$;

where $\bar{d}_{x_{r:s}} := \max\{d_{x_i} \mid r \leq i \leq s\}$; *and* $\xrightarrow{P}$, *and* $O_p(\cdot)$ *are short for convergence in probability, and order in probability, respectively.*

Proof: See Appendix A.

In the following subsections, we study separately the asymptotic properties of coefficients, t-statistics, and other statistics, to provide a better understanding of the results in Theorem 1.

Coefficients: Theorem 1 shows that the OLS-estimated coefficients for the stationary case collapse to zero asymptotically regardless of the degrees of persistence of the associated regressors. Nonetheless, the rate of convergence of each estimator, $\hat{\beta}_i$, varies depending on the values of d_{x_i} and d_y, and whether they are correlated with other regressors.

For non-correlated regressors, if $d_{x_i} + d_y < 1/2$, the convergence rate is the usual $T^{-\frac{1}{2}}$, while for $d_{x_i} + d_y = 1/2$, the convergence rate is slightly slower. If $d_{x_i} + d_y > 1/2$, the convergence rate explicitly depends on the value of $d_{x_i} + d_y$; as the value of this sum approaches 1, the order in probability of the estimator approaches $O_p(1)$. In other words, the more persistent the processes are, the slower the rate of convergence of the estimators, in line with TC's results.

For correlated regressors, the most relevant scenario, the orders of convergence for all coefficients depend solely on the degree of memory of the most persistent regressor. This is, in contrast to TC's results, the orders of convergence for the estimators of less persistent regressors, $\hat{\beta}_i$, do not longer depend on the sum $d_{x_i} + d_y$. Finally, the estimate of the constant term, $\hat{\beta}_0$, collapses at a rate that only depends on d_y in both cases. The more persistent the regressand, the slower the estimate for the constant collapses.

t-statistics: Theorem 1 also shows that misleading inference could be drawn via the t-statistics. We focus first on the non-correlated regressors. Note that for any regressor with $d_y + d_{x_i} < 1/2$, the t-statistic associated to $\hat{\beta}_i$ does not diverge. This does not necessarily mean that there are no distortions since the limiting distribution may differ from the standard normal, as we illustrate through finite sample evidence. For $d_y + d_{x_i} = 1/2$, the t-statistic slowly diverges at rate $(\log T)^{\frac{1}{2}}$. Finally, for $d_y + d_{x_i} > 1/2$, the t-statistic diverges, and the rate of divergence is directly dependent on $d_y + d_{x_i}$. The greater the degree of memory of the variable, the faster its t-statistic diverges.

For correlated regressors, the distortions are propagated among variables. Specifically, note that the rate of convergence for the t-statistics of correlated regressors depends on the maximum degree of persistence among them. Thus, it takes only one regressor with memory such that $d_y + d_{x_i} \geq 1/2$ to make all remaining t-statistics diverge. Importantly, this implies that the asymptotic behavior for the t-statistics in the multivariate case cannot be derived from the univariate scenario of TC. Under correlated regressors, the rate of divergence of a

t-statistic associated to a regressor with $d_y + d_{x_i} < 1/2$ in a univariate regression will differ from the one obtained in a multivariate regression if at least one of the regressors is such that $d_y + d_{x_j} \geq 1/2$. This shows that the risk of uncovering spurious results increases as we add more regressors to the specification. In particular, coefficients that would be found insignificant in a univariate setting may be spuriously found to be significant in a multivariate setting, as we will illustrate with finite sample evidence. Finally, the t-statistic for the constant term diverges at a rate that solely depends on the degree of persistence of the regressand.

Other statistics: As for the complementary statistics, Theorem 1 shows that the standard statistical tools to draw inference provide conflicting information. On the one hand, the joint $\mathcal{F}$-statistic diverges if $\max_k\{d_{x_k}\}+d_y \geq 1/2$, which would spuriously indicate that at least one of the explanatory variables is related to the regressand. On the other hand, note that the coefficient of determination, R^2, converges in probability to zero regardless of the correlation structure between the regressors. Consequently, as the sample size increases, the declining R^2 correctly reflects the fact that the regressors do not explain the variation in the regressand. Furthermore, the asymptotic value of $\mathcal{DW}$ depends solely on the memory parameter d_y, such that $\mathcal{DW}$ is in the $(0, 2)$ interval. Hence, Granger and Newbold's (1974) rule of thumb for detecting a spurious regression, $R^2 > \mathcal{DW}$, no longer applies in view of Theorem 1.

2.2 Finite sample evidence under stationarity

Our theoretical findings point the risk of drawing misleading inference from an OLS-estimated regression model using fractionally integrated series. Furthermore, this risk increases as the sample size and number of regressors grow; we confirm this in finite samples. In the simulations,[4] all error terms are sampled from Normal distributions, with a multivariate Normal for the correlated case. We generate the fractionally integrated processes using the algorithm based on the fast Fourier transform, see Jensen and Nielsen (2014), with initial values set to zero. The simulated sample sizes are $T = 50$, $T = 100$, and $T = 1000$.

Univariate versus multivariate regressions: Table 1 shows the simulation results for a scenario that allows us to contrast our multivariate extension against the univariate case considered by TC. The regressand is generated as an $FI(d_y)$ process, Equation (1), while the regressors as $FI(d_{x_i})$ plus correlated noise, Equation (2). Two regressors are considered, x_1, and x_2, such that, $d_{x_1} + d_y < 1/2$, and $d_{x_2} + d_y > 1/2$. The table considers three regressions

[4] Codes for the Monte Carlos analysis are available at: *https://github.com/everval/Spurious_Multivariate*

using these two regressors: a multivariate regression using both regressors, and two univariate regressions using just one of the regressors.

Table 1: Spurious regression, univariate against multivariate case.

$y_t \sim FI(d_y)$		$x_{1,t} \sim FI(d_{x_1}) + w_1$			$x_{2,t} \sim FI(d_{x_2}) + w_2$			
d_y	$\sigma^2_{\epsilon,y}$	d_{x_1}	σ^2_{ϵ,x_1}	σ^2_1	d_{x_2}	σ^2_{ϵ,x_2}	σ^2_2	$\sigma_{1,2}$
0.35	0.5	0.10	0.25	1	0.45	0.25	1	0.8

	$y_t = \beta_0 + \beta_1 x_{1,t}$			$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t}$			$y_t = \beta_0 + \beta_2 x_{2,t}$		
T	50	100	1000	50	100	1000	50	100	1000
$RR_{t\beta_0}$	0.5289	0.6053	0.8029	0.4979	0.5867	0.7969	0.5107	0.5974	0.7987
$RR_{t\beta_1}$	0.0508	0.0533	0.0576	0.0685	0.0823	0.1621	-	-	-
$RR_{t\beta_2}$	-	-	-	0.0851	0.1176	0.2877	0.0685	0.0888	0.2232
$RR_{\mathcal{F}}$	0.0508	0.0533	0.0576	0.0739	0.1014	0.2507	0.0685	0.0888	0.2232
R^2	0.0207	0.0104	0.0011	0.0472	0.0262	0.0046	0.0236	0.0135	0.0026
$\mathcal{DW}$	1.3340	1.2322	1.0557	1.3667	1.2565	1.0627	1.3446	1.2415	1.0597

RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests at a 5% nominal size, respectively. The number of replications is 10,000.

Table 1 shows that in both univariate cases, the t-statistics follow a rate of divergence that depends on the degree of memory of the regressor. Particularly, t_{β_2} diverges as the sample size increases given that $d_{x_2} + d_y > 1/2$, while t_{β_1} is $O_p(1)$ given that $d_{x_1} + d_y < 1/2$. Importantly, note that t_{β_1} diverges as the sample size increases in the multivariate regression, in line with Theorem 1. This is, the multivariate regression spuriously finds a linear relation between x_1 and y that was not uncovered in the univariate case. In this sense, Table 1 shows that TC's univariate results do not trivially generalize to the multivariate setting. In a multivariate regression, a practitioner would determine that all regressors are significant asymptotically, whilst she may find a particular regressor to be insignificant using a univariate regression framework. The practitioner could also find a subset of regressors (in a smaller, but still multivariate regression) to be insignificant if the persistence of the variables in such subset is smaller than the persistence of the excluded variables. In short, these results show that the risk of drawing spurious inference increases as we add more regressors to the specification because the persistence of one of such added regressors may be high enough to drag the divergence rate of the remaining t-stats.

Regarding the other statistics, note that the rates of divergence for the $\mathcal{F}$-statistics also differ between the univariate and multivariate cases. The $\mathcal{F}$-statistic is bigger for the multivariate scenario than for both univariate scenarios for all sample sizes. In this regard, the table shows that the risk of uncovering spurious results increases as we increase the number of regressors. Furthermore, in Appendix C it is shown that the risk of spurious regression increases with the degree of correlation among regressors in finite samples. Finally, the R^2 and the $\mathcal{DW}$ statistic converge to the expressions shown in Theorem 1 for all specifications.

Correlated versus uncorrelated regressors: Table 2 shows the simulation results for the multivariate stationary scenario under different correlation structures between the regressors. The regressand is generated as an $FI(d_y)$ process, Equation (1), while the regressors as $FI(d_{x_i})$ plus possibly correlated noise, Equation (2). Three regressors are considered, x_1, x_2 and x_3, such that, $d_{x_1} + d_y > 1/2$, $d_{x_2} + d_y = 1/2$, and $d_{x_3} + d_y < 1/2$. We explore three correlation scenarios among regressors: *i*) they are all independent, *ii*) the first two are correlated while the last one is not, and *iii*) they are all correlated.

Table 2: Spurious regression, stationary variables, $\max_k\{d_{x_k}\} + d_y \geq \frac{1}{2}$.

$y_t \sim FI(d_y)$		$x_{1,t} \sim FI(d_{x_1}) + w_1$			$x_{2,t} \sim FI(d_{x_2}) + w_2$			$x_{3,t} \sim FI(d_{x_3}) + w_3$		
d_y	$\sigma^2_{\epsilon,y}$	d_{x_1}	σ^2_{ϵ,x_1}	σ^2_1	d_{x_2}	σ^2_{ϵ,x_2}	σ^2_2	d_{x_3}	σ^2_{ϵ,x_3}	σ^2_3
0.35	0.5	0.45	0.25	1	0.15	0.25	1	0.10	0.25	1

	$\sigma_{1,2} = 0$; $\sigma_{1,3} = 0$; $\sigma_{2,3} = 0$			$\sigma_{1,2} = 0.8$; $\sigma_{1,3} = 0$; $\sigma_{2,3} = 0$			$\sigma_{1,2} = 0.8$; $\sigma_{1,3} = 0.8$; $\sigma_{2,3} = 0.4$		
T	50	100	1000	50	100	1000	50	100	1000
$RR_{t\beta_0}$	0.5076	0.5947	0.8033	0.4898	0.5883	0.8015	0.4885	0.5915	0.8003
$RR_{t\beta_1}$	0.0657	0.0920	0.2294	0.0859	0.1217	0.2870	0.1093	0.1400	0.3456
$RR_{t\beta_2}$	0.0520	0.0607	0.0630	0.0699	0.0830	0.1665	0.0722	0.0869	0.1784
$RR_{t\beta_3}$	0.0565	0.0539	0.0592	0.0524	0.0541	0.0600	0.0684	0.0833	0.1703
$RR_{\mathcal{F}}$	0.0620	0.0794	0.1845	0.0752	0.0998	0.2339	0.0859	0.1094	0.2857
R^2	0.0657	0.0348	0.0049	0.0689	0.0369	0.0058	0.0713	0.0388	0.0067
$\mathcal{DW}$	1.3786	1.2636	1.0612	1.3841	1.2668	1.0649	1.3888	1.2729	1.0641

RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests at a 5% nominal size, respectively. The number of replications is 10,000.

As for the uncorrelated case, Table 2 shows the dependence of the rate of divergence of the t-statistic on the persistence level of its associated regressor. The results show that a regression estimated with as few as 50 observations may obtain spurious results for highly persistent regressors, akin to TC's results. For the least persistent regressor, the rejection rates of the t-tests remain relatively stable. Nonetheless, they show that the distribution has heavier tails, since the actual rejection rates are systematically above the nominal 5%.
Turning to the correlated case, Table 2 shows how the spurious results get propagated to correlated regressors. In particular, note the increase in rejection rates for each variable

once it is allowed to be correlated to the more persistent regressor. In the most severe case, the table shows that even if the sum of the degree of memory of the regressor and the regressand is less than $1/2$, as it is the case for $x_{3,t}$, its t-statistic diverges asymptotically due to the correlation with the more persistent regressor. This contrasts with TC's results where spurious regression does not occur in regressions where the sums of the degrees of memory of regressand and regressor is less than $1/2$. The table shows that, under correlated regressors, the most realistic scenario, it takes only one regressor with a degree of memory such that its sum with that of the regressand is greater or equal than $1/2$, to make all t-statistics diverge. Thus, Table 2 further shows the increase of spurious results as we expand the specification to include more regressors.

Finally, Table 2 exhibits how the behavior of the $\mathcal{F}$-statistic depends on the sum $\max_k\{d_{x_k}\}+d_y$. Given that the value of this sum is greater than $1/2$, $\mathcal{F}$ starts to grow as the sample size increases, regardless of the correlation structure among regressors. Nonetheless, note that in finite samples, the $\mathcal{F}$ grows faster, the more correlation there is among regressors. Additional tables with $\max_k\{d_{x_k}\}+d_y < 1/2$ showing that $\mathcal{F}$ is $O_p(1)$ in that case, and thus does not diverge as the sample size increases, are shown in Appendix C. Finally, the R^2 statistic collapses to zero, and the Durbin-Watson statistics converge to the value obtained in Theorem 1 which only depends on the degree of memory of the regressand, regardless of the correlation structure between the regressors.

Signal-to-noise ratio: The multivariate stationary specification assumes a signal plus noise form for the regressors. In Table 3, we explore how the signal-to-noise ratio within the regressors affect the finite sample results. The table shows the simulation results for a multivariate stationary scenario where $\max_k\{d_{x_k}\}+d_y > 1/2$. The regressand is generated as an $FI(d_y)$ process, while the regressors as $FI(d_{x_i})$ plus correlated noise. Let $\sigma^2_{x_j}$ be the variance of the fractionally integrated signal in x_j; that is,

$$\sigma^2_{x_j} = \frac{\Gamma(1-2d_{x_j})}{\Gamma(d_{x_j})[\Gamma(1-d_{x_j})]^2}\sigma^2_{\epsilon,x_j},$$

and define the signal-to-noise ratio by σ^2_{xj}/σ^2_j. We consider three regressors under three signal-to-noise ratio scenarios: *i)* $\sigma^2_{xj}/\sigma^2_j = 4$, $j = 1,2,3$; *ii)* $\sigma^2_{xj}/\sigma^2_j = 1$, $j = 1,2,3$; and *iii)* $\sigma^2_{xj}/\sigma^2_j = 1/4$, $j = 1,2,3$.

Table 3 allows us to assess the effect that the signal-to-noise ratio has in uncovering spurious results in finite samples. The table shows that the smaller the signal-to-noise ratio; that is, the greater the variance of the noise is in relation to the variance of the signal, the slower the t-statistics and $\mathcal{F}$-statistic diverge. Nonetheless, note that we uncover spurious results for sample sizes as smalls as 50 observations for cases where the variance of the noise is four times

Table 3: Spurious regression, stationary variables, $\max_k\{d_{x_k}\} + d_y > \frac{1}{2}$.

$y_t \sim FI(d_y)$		$x_{1,t} \sim FI(d_{x_1}) + w_1$		$x_{2,t} \sim FI(d_{x_2}) + w_2$		$x_{3,t} \sim FI(d_{x_3}) + w_3$	
d_y	$\sigma^2_{\epsilon,y}$	d_{x_1}	σ^2_{ϵ,x_1}	d_{x_2}	σ^2_{ϵ,x_2}	d_{x_3}	σ^2_{ϵ,x_3}
0.35	0.5	0.45	0.25	0.15	0.25	0.10	0.25

	$\sigma^2_i = \frac{1}{4}\sigma^2_{x_i}; \sigma_{1,2} = 0.6;$ $\sigma_{1,3} = 0.6;\ \sigma_{2,3} = 0.6$			$\sigma^2_i = \sigma^2_{x_i}; \sigma_{1,2} = 0.6;$ $\sigma_{1,3} = 0.6;\ \sigma_{2,3} = 0.6$			$\sigma^2_i = 4\sigma^2_{x_i}; \sigma_{1,2} = 0.6;$ $\sigma_{1,3} = 0.6;\ \sigma_{2,3} = 0.6$		
T	50	100	1000	50	100	1000	50	100	1000
$RR_{t\beta_0}$	0.4856	0.5742	0.8002	0.4887	0.5906	0.8030	0.5082	0.6091	0.8090
$RR_{t\beta_1}$	0.0984	0.1475	0.3525	0.0818	0.1046	0.2629	0.0633	0.0705	0.1605
$RR_{t\beta_2}$	0.0660	0.0758	0.1160	0.0633	0.0721	0.1102	0.0550	0.0632	0.0801
$RR_{t\beta_3}$	0.0615	0.0663	0.0894	0.0623	0.0651	0.0919	0.0581	0.0547	0.0757
$RR_{\mathcal{F}}$	0.0950	0.1321	0.3306	0.0768	0.0993	0.2324	0.0645	0.0671	0.1342
R^2	0.0745	0.0418	0.0074	0.0694	0.0373	0.0057	0.0649	0.0329	0.0042
$\mathcal{DW}$	1.4025	1.2822	1.0658	1.3875	1.2685	1.0650	1.3695	1.2584	1.0594

RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests at a 5% nominal size, respectively. The number of replications is 10,000.

as big as that of the signal. Thus, the simulations show that we can expect spurious results even when the fractionally integrated signal is small compared to the random component. In this sense, it takes only a relatively small long-range dependent component in the regressors for spurious results to occur.

To summarise, the simulations show that extra consideration should be taken when dealing with multivariate specifications in relation to univariate ones. For fractionally integrated stationary regressors, it takes only one highly persistent regressor for spurious results to appear. In particular, the $\mathcal{F}$-statistic and t-statistics of highly persistent regressors diverge asymptotically. Moreover, the asymptotic divergence is propagated to the t-statistics of correlated regressors, regardless of their degree of persistence. The rate of divergence, and thus the uncovering of spurious results, is directly related to the level of dependence of the regressors, the correlation between them, and the signal-to-noise ratio.

Notwithstanding, the R-squared remains a relatively reliable tool for identifying misleading inference. It is quite small under any circumstance where there is not a real linear relationship between the regressors and the regressand. Should the practitioner encounter a small R-squared in a regression with stationary variables, she should be wary that the regression might be of little use.

3 Nonstationary case

3.1 Asymptotic results

We now turn our attention to the case of spurious multivariate regression under fractionally integrated nonstationary processes. Cappuccio and Lubian (1997) and TC are among the first works to study the phenomenon of spurious regression under fractional nonstationarity. Capuccio and Lubian (1997) study the asymptotics of the OLS estimates of a univariate regression where both y_t and x_t are assumed to be $I(1)$ processes with fractionally integrated errors, this is $I(1+d)$ with $-0.5 < d < 0.5$; while TC present the asymptotics of a more varied set of data-generating processes: y_t and x_t are (i) both stationary fractionally integrated processes; (ii) both nonstationary fractionally integrated processes; (iii) both variables are fractionally integrated but only one is stationary, and; (iv) a fractionally integrated variable, stationary or not, is regressed against a deterministic trend. The main finding across these two works is that the phenomenon of spurious regressions seems to relate to the strong long memory of the series rather than stationarity, as noted by Granger et al. (2001). We extend these results to the multivariate nonstationary case and show that the asymptotics of the statistics associated to individual regressors do not depend on the order of integration of other regressors, nor do they depend on the number of regressors. Below we list the assumptions for the multivariate nonstationary case.

Assumption 2. *Let y_t and $x_{i,t}$, for $i = 1, 2, \dots, k$, be independent nonstationary fractionally integrated processes of orders d_y and d_{x_i}, respectively, that satisfy $(1-L)^{d_z} z_t = \epsilon_{z,t}$, where $\epsilon_{z,t}$ are i.i.d. white noises with zero mean and finite variance $\sigma^2_{\epsilon,z}$, and $d_z \in (1/2, 1)$ for $z = y, x_i$. Suppose also that $E\left[\epsilon_{z,t}\right]^{q_z} < \infty$ with $q_z \geq \max\left\{4, -\frac{8d_z}{1+2d_z}\right\}$ for all $z = y, x_i$.*

Theorem 2 shows that, for the nonstationary case, the behavior of statistics associated to individual regressors does not depend on the order of integration of other regressors, nor does it depend on the number of regressors.

Theorem 2. *Let Assumption 2 hold. Suppose that the linear specification $y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{i,t} + u_t$ is estimated by OLS. Then, as $T \to \infty$,*

i) $\hat{\beta}_0 = O_p(T^{d_y - \frac{1}{2}})$;

ii) $\hat{\beta}_i = O_p\left(T^{d_y - d_{x_i}}\right)$, *for* $i = 1, \dots, k$;

iii) $t_{\beta_0} = O_p(T^{\frac{1}{2}})$;

iv) $t_{\beta_i} = O_p\left(T^{\frac{1}{2}}\right)$, *for* $i = 1, \dots, k$;

Furthermore,

v) $R^2 = O_p(1)$;

vi) $\mathcal{F} = O_p(T)$;

vii) $\mathcal{DW} \xrightarrow{P} 0$;

where $\xrightarrow{P}$*, and* $O_p(\cdot)$ *are short for convergence in probability, and order in probability, respectively.*

Proof: See Appendix B.

Theorem 2 is in line with classical results on spurious regressions with nonstationary variables. All t-statistics diverge at rate $O_p\left(T^{\frac{1}{2}}\right)$, the $\mathcal{F}$ diverges at rate $O_p(T)$, and the R^2 is $O_p(1)$, which shows that for nonstationary variables we cannot reliably use the R^2 to filter spurious results. In this sense, Theorem 2 extends the results from spurious regressions with nonstationary variables to the multivariate fractionally integrated case.

3.2 Finite sample evidence under nonstationarity

Turning our attention to the finite sample evidence, Table 4 shows the results of simulations with nonstationary fractionally integrated variables. On the one hand, the table is in line with the results from Theorem 2, this is: the rejection rates of the t-ratios shown in Table 4 grow as the sample size increases, but at a slower pace (less than linear); the R^2 is nondegenerate and the $\mathcal{F}$ joint test diverges, as Theorem 2 states.

On the other hand, Theorem 2 and the finite-sample evidence are in line with classical results on spurious regressions with nonstationary variables. The divergence rate of the t-ratios, $T^{1/2}$, the non-degenerate R-square, and the divergent $\mathcal{F}$ statistic are common findings in the literature of spurious regression since Phillips' (1986) seminal work. See Ventosa-Santaulària (2009) for a review.

More precisely, Table 4 shows that the rejection rates of the t-statistics are higher than those obtained with stationary variables. This can be explained given the faster rate of divergence than for the stationary case. Similarly, the rejection rates for the $\mathcal{F}$ test are higher than for the stationary case. Furthermore, Table 4 shows the asymptotic collapse of the Durbin-Watson statistic. In finite samples, the rate of convergence of the Durbin-Watson statistic depends on the degree of persistence of the regressand, the closer the process is to a unit root process, the faster the statistic collapses to zero. This, coupled with the asymptotic behavior of the R-squared statistic, $R^2 = O_p(1)$, allows us to extend Granger and Newbold's (1974) rule of thumb for detecting a spurious regression, $R^2 > \mathcal{DW}$, to nonstationary long

Table 4: Spurious regression, nonstationary variables.

	$y_t \sim FI(d_y)$			$x_{1,t} \sim FI(d_{x_1})$			$x_{2,t} \sim FI(d_{x_2})$			$x_{3,t} \sim FI(d_{x_3})$		
	$\sigma^2_{\epsilon,y}$=0.5			σ^2_{ϵ,x_1}=0.25			$\sigma^2_{\epsilon,x_2} = 0.25$			$\sigma^2_{\epsilon,x_3} = 0.25$		

	$d_y = 0.60$; $d_{x_1} = 0.60$; $d_{x_2} = 0.60$; $d_{x_3} = 0.60$			$d_y = 0.80$; $d_{x_1} = 0.80$; $d_{x_2} = 0.80$; $d_{x_3} = 0.80$			$d_y = 0.75$; $d_{x_1} = 0.70$; $d_{x_2} = 0.65$; $d_{x_3} = 0.60$		
T	50	100	1000	50	100	1000	50	100	1000
$RR_{t\beta_0}$	0.5882	0.6913	0.8935	0.6112	0.7217	0.9088	0.6112	0.7559	0.9219
$RR_{t\beta_1}$	0.2596	0.3911	0.7396	0.4235	0.5655	0.8510	0.4235	0.5232	0.8201
$RR_{t\beta_2}$	0.2648	0.3827	0.7314	0.4323	0.5734	0.8524	0.4323	0.4936	0.8110
$RR_{t\beta_3}$	0.2683	0.3890	0.7378	0.4306	0.5574	0.8522	0.4306	0.4485	0.7857
$RR_{\mathcal{F}}$	0.5137	0.7104	0.9757	0.7973	0.9130	0.9968	0.7973	0.8345	0.9914
R^2	0.1830	0.1534	0.1064	0.3314	0.3134	0.2977	0.3314	0.2257	0.1889
$\mathcal{DW}$	1.0506	0.8151	0.3892	0.7903	0.5181	0.1241	0.7903	0.5740	0.1754

RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests at a 5% nominal size, respectively. The number of replications is 10,000.

memory variables. Nonetheless, note that for the rule of thumb to perform satisfactorily in small samples, the degree of memory of the regressand has to be significantly above $1/2$.

4 Concluding remarks

We studied the asymptotic and finite sample behavior of the OLS-estimated multivariate regression with an arbitrary finite number of fractionally integrated regressors. Both the stationary and nonstationary cases are considered. We show that nonsense inference may be drawn in this setting: t-statistics and the $\mathcal{F}$ joint test diverge, the R-square does not collapse to zero (only in the nonstationary case), thus providing the practitioner with evidence of an otherwise nonexistent linear relationship between the variables. Our findings extend the literature of spurious regression in several directions.

For the stationary case, our multivariate approach shows that the asymptotic behaviors of the parameter estimates and the associated t-statistics depend on the persistence of the regressors, and on whether there is correlation among them. Under correlated regressors, the most relevant scenario, increasing or decreasing the number of regressors in a specification may alter the asymptotic properties of the t-statistics: when a more persistent regressor is included in the specification, the remaining t-statistics diverge. In particular, our multivariate extension shows that spurious results can be found in a multivariate setting for regressors for which no spurious results would occur in a univariate setting. Hence, the risk of drawing spurious inference increases as the number of regressors grows.

We also show that irrespective of the correlation scenario, the standard joint $\mathcal{F}$ test provides misleading inference when at least one of the regressors is highly persistent, while the R-squared works properly. Moreover, the simulation exercise confirms our asymptotic results and shows that the phenomenon of spurious regression is more acute when the number and persistence of the variables grows.

The behavior of the R-squared under stationary regressors is particularly revealing. The R-squared collapses to zero when the regressors are independent of the regressand. In this sense, our results suggest that a practitioner may be able to detect spurious results in the stationary case by looking at the R-squared.

For the nonstationary case, our results extend the classical spurious regression results. All t-statistics and the $\mathcal{F}$-statistic diverge as the sample size grows, and they do so at a faster rate than for the stationary case. The $\mathcal{DW}$ collapses to zero, while R-squared statistic does not; which allows us to extend Granger's rule of thumb for spurious regression to the multivariate nonstationary fractionally integrated case; this is, the regression may be spurious when the R^2 is greater than the $\mathcal{DW}$ statistic.

Overall, we show that when the variables behave as long memory processes, whether stationary or not, inference drawn from the t-statistics or the $\mathcal{F}$ joint test is unreliable. The conjecture that spurious regression is a persistence problem, rather than a stationarity problem, is therefore supported.

Disclosure statement: The authors state that no financial interest or benefit has arisen from the direct applications of our research.

Acknowledgements: The authors would like to thank the anonymous referee, as well as Niels Haldrup, Ignacio Lobato, Michael Massmann, and the participants of the Long Memory Conference 2018 for their valuables comments. All remaining errors are ours.

A Proof of Theorem 1

To obtain the OLS estimators, along with the associated t-statistics, it is necessary to obtain the limit expression of the sums that define them. These are summarized in Table 5, along with their respective convergence rates. All of the convergence rates, the under-braced expressions, can be found in Tsay and Chung (2000) or Hayashi (2000) except for the normalization ratios of products of fractionally integrated processes and short memory processes which follow from Osterrieder et al. (2019).

$\sum y_t$	$=$	$\sum(1-L)^{d_y}\epsilon_{y,t}$	$=$	$O_p\left(T^{\frac{1}{2}+d_y}\right);$
$\sum y_t^2$	$=$	$\sum\left((1-L)^{d_y}\epsilon_{y,t}\right)^2$	$=$	$O_p\left(T\right);$
$\sum x_{i,t}$	$=$	$\sum(1-L)^{d_{x_i}}\epsilon_{i,t}+\underbrace{\sum w_{i,t}}_{O_p(T^{\frac{1}{2}})}$	$=$	$O_p\left(T^{\frac{1}{2}+d_{x_i}}\right);$
$\sum x_{i,t}^2$	$=$	$\sum\left((1-L)^{d_{x_i}}\epsilon_{i,t}\right)^2 + \underbrace{\sum w_{i,t}^2}_{O_p(T)}+2\underbrace{\sum(1-L)^{d_{x_i}}\epsilon_{i,t}w_{i,t}}_{O_p(T^{\frac{1}{2}})}$	$=$	$O_p\left(T\right);$
$\sum x_{i,t}y_t$	$=$	$\sum(1-L)^{d_y}\epsilon_{y,t}(1-L)^{d_{x_i}}\epsilon_{i,t}+\sum(1-L)^{d_y}\epsilon_{y,t}w_{i,t}$	$=$	$\begin{cases} O_p\left(T^{\frac{1}{2}}\right) & \text{if } 0<d_{x_i}+d_y<1/2; \\ O_p\left(\sqrt{T\ln T}\right) & \text{if } d_{x_i}+d_y=1/2; \\ O_p\left(T^{d_{x_i}+d_y}\right) & \text{if } \frac{1}{2}<d_{x_i}+d_y<1; \end{cases}$
$\sum x_{i,t}x_{j,t}$ $\sigma_{i,j}=0$	$=$	$\sum(1-L)^{d_{x_i}}\epsilon_{i,t}(1-L)^{d_{x_j}}\epsilon_{j,t}+\sum(1-L)^{d_{x_i}}\epsilon_{i,t}w_{j,t}+\sum(1-L)^{d_{x_j}}\epsilon_{j,t}w_{i,t}+\underbrace{\sum w_{i,t}w_{j,t}}_{O_p\left(T^{\frac{1}{2}}\right)}$	$=$	$\begin{cases} O_p\left(T^{\frac{1}{2}}\right) & \text{if } 0<d_{x_i}+d_{x_j}<1/2; \\ O_p\left(\sqrt{T\ln T}\right) & \text{if } d_{x_i}+d_{x_j}=1/2; \\ O_p\left(T^{d_{x_i}+d_{x_j}}\right) & \text{if } \frac{1}{2}<d_{x_i}+d_{x_j}<1; \end{cases}$
$\sum x_{i,t}x_{j,t}$ $\sigma_{i,j}\neq 0$	$=$	$\sum(1-L)^{d_{x_i}}\epsilon_{i,t}(1-L)^{d_{x_j}}\epsilon_{j,t}+\sum(1-L)^{d_{x_i}}\epsilon_{i,t}w_{j,t}+\sum(1-L)^{d_{x_j}}\epsilon_{j,t}w_{i,t}+\underbrace{\sum w_{i,t}w_{j,t}}_{O_p(T)}$	$=$	$O_p\left(T\right).$

Table 5: Expressions for sums in Theorem 1 with $i \neq j$; $i,j=1,\cdots,k$. All sums range from $t=1$ to $t=T$. $O_p(\cdot)$ is short for order in probability.

Items $i)$ to $iii)$

Recall the OLS formula:

$$\hat{\beta}=\left(X'X\right)^{-1}X'Y,$$

where $\dim\left(X\right)=T\times(k+1)$, $\dim\left(Y\right)=T\times 1$, and $\dim\left(\hat{\beta}\right)=(k+1)\times 1$.
Let

$$\overrightarrow{x}:=\begin{bmatrix}\sum x_{1,t}\\ \sum x_{2,t}\\ \vdots\\ \sum x_{k,t}\end{bmatrix},\qquad \text{and}\qquad \Omega:=\begin{bmatrix}\sum x_{1,t}^2 & \sum x_{1,t}x_{2,t} & \dots & \sum x_{1,t}x_{k,t}\\ \sum x_{1,t}x_{2,t} & \sum x_{2,t}^2 & \dots & \sum x_{2,t}x_{k,t}\\ \vdots & \vdots & \ddots & \vdots\\ \sum x_{1,t}x_{k,t} & \sum x_{2,t}x_{k,t} & \dots & \sum x_{k,t}^2\end{bmatrix},$$

we can rewrite $X'X$ as

$$X'X = \begin{bmatrix} T & \sum x_{1,t} & \dots & \sum x_{k,t} \\ \sum x_{1,t} & \sum x_{1,t}^2 & \dots & \sum x_{1,t}x_{k,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{k,t} & \sum x_{1,t}x_{k,t} & \dots & \sum x_{k,t}^2 \end{bmatrix} = \begin{bmatrix} T & \overrightarrow{x}' \\ \overrightarrow{x} & \Omega \end{bmatrix}.$$

Hence,

$$\begin{aligned} plim\left(\hat{\beta}\right) &= plim\left[(X'X)^{-1}X'Y\right] \\ &= plim\left[\left(\frac{1}{T}X'X\right)^{-1}\frac{1}{T}X'Y\right] \\ &= plim\left\{\begin{bmatrix} 1 & \frac{1}{T}\overrightarrow{x}' \\ \frac{1}{T}\overrightarrow{x} & \frac{1}{T}\Omega \end{bmatrix}^{-1}\begin{bmatrix} \frac{1}{T}\sum y_t \\ \frac{1}{T}\sum x_{1,t}y_t \\ \vdots \\ \frac{1}{T}\sum x_{k,t}y_t \end{bmatrix}\right\}. \end{aligned}$$

where *plim* is short for probability limit. Define $\Pi := (\frac{1}{T}\Omega - \frac{1}{T}\overrightarrow{x}\frac{1}{T}\overrightarrow{x}')^{-1}$ and using blockwise inversion,

$$\begin{aligned} plim\left(\hat{\beta}\right) &= plim\left\{\begin{bmatrix} 1+\frac{1}{T}\overrightarrow{x}'\Pi\frac{1}{T}\overrightarrow{x} & -\frac{1}{T}\overrightarrow{x}\Pi \\ \Pi\frac{1}{T}\overrightarrow{x} & \Pi \end{bmatrix}\begin{bmatrix} \frac{1}{T}\sum y_t \\ \frac{1}{T}\sum x_{1,t}y_t \\ \vdots \\ \frac{1}{T}\sum x_{k,t}y_t \end{bmatrix}\right\} \\ &= plim\begin{bmatrix} 1+\frac{1}{T}\overrightarrow{x}'\Pi\frac{1}{T}\overrightarrow{x} & -\frac{1}{T}\overrightarrow{x}\Pi \\ \Pi\frac{1}{T}\overrightarrow{x} & \Pi \end{bmatrix} plim\begin{bmatrix} \frac{1}{T}\sum y_t \\ \frac{1}{T}\sum x_{1,t}y_t \\ \vdots \\ \frac{1}{T}\sum x_{k,t}y_t \end{bmatrix}. \end{aligned}$$

From Table 5, note that $plim\frac{1}{T}\overrightarrow{x} = 0$. This in turn implies that $plim(\Pi) = (plim(\frac{1}{T}\Omega))^{-1} = \Sigma_X^{-1}$, where Σ_X is the variance matrix of $(x_1, \cdots, x_k)$. Given Assumption 1, it has the same structure as Σ, the variance matrix of w_t, Equation (3).

Let $\Sigma_{X,1:m}$ be the square matrix composed of rows 1 to m, columns 1 to m of Σ_X, and let $\Sigma_{X,m+1:k}$ be the square matrix composed of rows $m+1$ to k, columns $m+1$ to k of Σ_X. By the argument above, note that $\Sigma_{X,1:m}$ is a full matrix while $\Sigma_{X,m+1:k}$ is diagonal.

By blockwise inversion,

$$\Sigma_X^{-1} = \begin{bmatrix} \Sigma_{X,1:m} & \mathbb{O}_{m,k-m} \\ \mathbb{O}_{m,k-m} & \Sigma_{X,m+1:k} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{X,1:m}^{-1} & \mathbb{O}_{m,k-m} \\ \mathbb{O}_{m,k-m} & \Sigma_{X,m+1:k}^{-1} \end{bmatrix},$$

where $\mathbb{O}_{r,s}$ is a matrix of zeros with r rows and s columns. Thus,

$$plim\left(\hat{\beta}\right) = \begin{bmatrix} 1 & \mathbb{O}_{1,m} & \mathbb{O}_{1,k-m} \\ \mathbb{O}_{m,1} & \Sigma^{-1}_{X,1:m} & \mathbb{O}_{m,k-m} \\ \mathbb{O}_{k-m,1} & \mathbb{O}_{m,k-m} & \Sigma^{-1}_{X,m+1:k} \end{bmatrix} plim \begin{bmatrix} \frac{1}{T}\sum y_t \\ \frac{1}{T}\sum x_{1,t}y_t \\ \vdots \\ \frac{1}{T}\sum x_{m,t}y_t \\ \frac{1}{T}\sum x_{m+1,t}y_t \\ \vdots \\ \frac{1}{T}\sum x_{k,t}y_t \end{bmatrix}.$$

The result follows from plugging the appropriate rate of convergence from Table 5. Item i) follows directly from $\frac{1}{T}\sum y_t = O_p\left(T^{\frac{1}{2}+d_y-1}\right) = O_p\left(T^{d_y-\frac{1}{2}}\right)$, while the results for items ii) and iii) depend on the correlation scenario. Given that $\Sigma^{1}_{X,1:m}$ is a full matrix, the rate of convergence of $\hat{\beta}_1$ to $\hat{\beta}_m$ depends on the memory of all m regressors, inheriting that of the maximum. On the contrary, being $\Sigma^{-1}_{X,m+1:k}$ a diagonal matrix, the rates of convergence of $\hat{\beta}_{m+1}$ to $\hat{\beta}_k$ only depend on the memory of the associated regressor.

Items iv) to vi)

First, note that

$$\begin{aligned} s^2 &= \frac{1}{T}\sum \hat{u}_t^2 = \frac{1}{T}\sum\left(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t} - \ldots - \hat{\beta}_k x_{k,t}\right)^2 \\ &= \frac{1}{T}\Bigg(\sum y_t^2 - 2\hat{\beta}_0\sum y_t + T\hat{\beta}_0^2 - 2\sum_{i=1}^{k}\hat{\beta}_i\sum x_{i,t}y_t + \sum_{i=1}^{k}\hat{\beta}_i^2\sum x_{i,t}^2 + \\ &\quad 2\hat{\beta}_0\sum_{i=1}^{k}\hat{\beta}_i\sum x_{i,t} + 2\sum_{i=1}^{k}\sum_{j>i}\hat{\beta}_i\hat{\beta}_j\sum x_{i,t}x_{j,t}\Bigg). \end{aligned} \tag{4}$$

From items i) to iii), note that the term with highest order of probability is $\sum y_t^2$, which is $O_p(T)$, all other terms are an order in probability strictly lower. Thus, $s^2 = O_p(1)$.
Now, from the variance matrix of the estimators

$$t_{\beta_i} = \frac{\hat{\beta}_i}{\left(s^2(X'X)^{-1}_{(i,i)}\right)^{\frac{1}{2}}},$$

where the sub-index denotes the i-th element in the diagonal. From the computations above, we know that they all are $O_p(T)$. Substituting all components shows that each t-statistic has the same order of convergence as its estimator minus one half.

Items $vii)$ to $ix)$

From the R^2 formula

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

we will show the order of convergence for both the numerator and denominator.
On the one hand, using (4), the numerator can be written as

$$\begin{aligned}\sum (y_t - \bar{y})^2 - \sum \hat{u}_t^2 &= \sum y_t^2 - \frac{1}{T}\left(\sum y_t\right)^2 - \sum \hat{u}_t^2 \\ &= -\frac{1}{T}\left(\sum y_t\right)^2 + 2\hat{\beta}_0 \sum y_t - T\hat{\beta}_0^2 + 2\sum_{i=1}^{k} \hat{\beta}_i \sum x_{i,t} y_t - \\ &\quad \sum_{i=1}^{k} \hat{\beta}_i^2 \sum x_{i,t}^2 - 2\hat{\beta}_0 \sum_{i=1}^{k} \hat{\beta}_i \sum x_{i,t} - 2\sum_{i=1}^{k}\sum_{j>i} \hat{\beta}_i \hat{\beta}_j \sum x_{i,t} x_{j,t},\end{aligned}$$

which shows that $\hat{\beta}_i^2 \sum x_{i,t}^2$ for $i = \max\{d_i \mid d_i \geq d_j \forall j\}$ has the highest order of probability.
On the other hand, the denominator is given by

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

Replacing both, proves item $vii)$.
To prove $viii)$, recall that

$$\mathcal{F} = \frac{\left[\sum (y_t - \bar{y})^2 - \sum \hat{u}_t^2\right]/k}{\sum \hat{u}_t^2 / [T-(k+1)]} = \frac{[T-(k+1)]\, R^2}{k \sum \hat{u}_t^2 / \sum (y_t - \bar{y})^2},$$

which shows the desired result once we replace the orders of probability obtained above.
Finally, to prove $ix)$, recall the definition of the Durbin-Watson statistic:

$$\mathcal{DW} = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2} = \frac{\sum_{t=2}^{T} \hat{u}_t^2 + \sum_{t=2}^{T} \hat{u}_{t-1}^2 - 2\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \approx 2 - 2\frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2},$$

where in the last expression we use the fact that $\frac{\hat{u}_1^2 + \hat{u}_T^2}{\sum \hat{u}_t^2}$ is negligible as $T \to \infty$.
Now,

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} = \frac{1}{T}\Bigg[&\sum_{t=2}^{T} y_t y_{t-1} - \hat{\beta}_0 \sum_{t=2}^{T} y_t - \hat{\beta}_0 \sum_{t=2}^{T} y_{t-1} + (T-1)\hat{\beta}_0^2 - \sum_{i=1}^{k} \hat{\beta}_i \sum_{t=2}^{T} x_{i,t} y_{t-1} - \\ &\sum_{i=1}^{k} \hat{\beta}_i \sum_{t=2}^{T} x_{i,t-1} y_t + \sum_{i=1}^{k} \hat{\beta}_i^2 \sum_{t=2}^{T} x_{i,t} x_{i,t-1} + \hat{\beta}_0 \sum_{i=1}^{k} \hat{\beta}_i \sum_{t=2}^{T} x_{i,t} + \hat{\beta}_0 \sum_{i=1}^{k} \hat{\beta}_i \sum_{t=2}^{T} x_{i,t-1} + \\ &\sum_{i=1}^{k}\sum_{j>i} \hat{\beta}_i \hat{\beta}_j \sum_{t=2}^{T} x_{i,t} x_{j,t-1} + \sum_{i=1}^{k}\sum_{j>i} \hat{\beta}_i \hat{\beta}_j \sum_{t=2}^{T} x_{i,t-1} x_{j,t}\Bigg].\end{aligned}$$

Using items i) to iii), which show that all estimators have negative orders of probability, the highest order of probability is the one of $\sum_{t=2}^{T} y_t y_{t-1}$. Noting that $\frac{1}{T}\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} \xrightarrow{P} \gamma_y(1)$, and $\frac{1}{T}\sum \hat{u}_t^2 \xrightarrow{P} \gamma_y(0)$, we find that:

$$\mathcal{DW} \to 2 - 2\rho_y(1).$$

B Proof of Theorem 2

Analogous to the stationary case, to obtain the OLS estimators, along with the associated t-statistics, it is necessary to obtain the limit expression of the sums that define them. These are summarized in Table 6, along with their respective convergence rates. Table 6 draws upon the results of TC that correspond to nonstationary processes.

$\sum z_t$	$=$	$O_p\left(T^{\frac{1}{2}+d_z}\right)$;
$\sum z_t^2$	$=$	$O_p\left(T^{2d_z}\right)$;
$\sum x_{i,t} y_t$	$=$	$O_p\left(T^{d_{x_i}+d_y}\right)$, for $i = 1, \ldots, k$;
$\sum x_{i,t} x_{j,t}$	$=$	$O_p\left(T^{d_{x_i}+d_{x_j}}\right)$, for $i, j = 1, \ldots, k$ and $i \neq j$.

Table 6: Expressions for sums in Theorem 2 with $i \neq j$; $i, j = 1, \cdots, k$. Here, $z = y, x_1, \ldots, x_k$. All sums range from $t = 1$ to $t = T$.

Items i) and ii)

From the OLS estimator formula, it follows that

$$\hat{\beta}_0 = T^{-1}\left(\sum y_t - \sum_{i=1}^{k} \hat{\beta}_i \sum x_{i,t}\right); \tag{5}$$

while for the rest of the estimators, by Cramer's rule, we have that

$$\hat{\beta}_k = \frac{\Delta_k}{\Delta}, \tag{6}$$

where

$$\Delta = \det\left(X'X\right), \tag{7}$$

and

$$\Delta_k = \begin{vmatrix} T & \sum x_{1,t} & \sum x_{2,t} & \ldots & \sum y_t \\ \sum x_{1,t} & \sum x_{1,t}^2 & \sum x_{1,t}x_{2,t} & \ldots & \sum x_{1,t}y_t \\ \sum x_{2,t} & \sum x_{1,t}x_{2,t} & \sum x_{2,t}^2 & \ldots & \sum x_{2,t}y_t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{k,t} & \sum x_{1,t}x_{k,t} & \sum x_{2,t}x_{k,t} & \ldots & \sum x_{k,t}y_t \end{vmatrix}. \tag{8}$$

To find the order in probability of $\hat{\beta}_k$ we triangulate the matrices whose determinant is equal to Δ and Δ_k, so as to exploit the fact that the determinant of any triangular matrix is merely the product of the elements along the main diagonal. We find that the order in probability of said elements after triangulation to be the same as that prior to triangulation for both Δ and Δ_k.

Let us first look at (7), the denominator in expression (6). Note from Table 6 that, prior to triangulation,

$$\Delta = \begin{vmatrix} T & O_p\left(T^{\frac{1}{2}+d_{x_1}}\right) & O_p\left(T^{\frac{1}{2}+d_{x_2}}\right) & \dots & O_p\left(T^{\frac{1}{2}+d_{x_k}}\right) \\ O_p\left(T^{\frac{1}{2}+d_{x_1}}\right) & O_p\left(T^{2d_{x_1}}\right) & O_p\left(T^{d_{x_1}+d_{x_2}}\right) & \dots & O_p\left(T^{d_{x_1}+d_{x_k}}\right) \\ O_p\left(T^{\frac{1}{2}+d_{x_2}}\right) & O_p\left(T^{d_{x_1}+d_{x_2}}\right) & O_p\left(T^{2d_{x_2}}\right) & \dots & O_p\left(T^{d_{x_2}+d_{x_k}}\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O_p\left(T^{\frac{1}{2}+d_{x_k}}\right) & O_p\left(T^{d_{x_1}+d_{x_k}}\right) & O_p\left(T^{d_{x_2}+d_{x_k}}\right) & \dots & O_p\left(T^{2d_{x_k}}\right) \end{vmatrix}.$$

If we add the first row multiplied by scalar $-\frac{\sum x_{i-1,t}}{T}$ to the i-th row, for $i = 2, \dots, k+1$, we arrive at the following

$$\Delta = \begin{vmatrix} T & \sum x_{1,t} & \sum x_{2,t} & \dots & \sum x_{k,t} \\ 0 & \sum x_{1,t}^2 - \frac{\left(\sum x_{1,t}\right)^2}{T} & \sum x_{1,t}x_{2,t} - \frac{\sum x_{1,t}\sum x_{2,t}}{T} & \dots & \sum x_{1,t}x_{k,t} - \frac{\sum x_{1,t}\sum x_{k,t}}{T} \\ 0 & \sum x_{1,t}x_{2,t} - \frac{\sum x_{1,t}\sum x_{2,t}}{T} & \sum x_{2,t}^2 - \frac{\left(\sum x_{2,t}\right)^2}{T} & \dots & \sum x_{2,t}x_{k,t} - \frac{\sum x_{2,t}\sum x_{k,t}}{T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sum x_{1,t}x_{k,t} - \frac{\sum x_{1,t}\sum x_{k,t}}{T} & \sum x_{2,t}x_{k,t} - \frac{\sum x_{2,t}\sum x_{k,t}}{T} & \dots & \sum x_{k,t}^2 - \frac{\left(\sum x_{k,t}\right)^2}{T} \end{vmatrix}.$$

Table 6 shows that all non-zero elements retain their original order in probability.

Should we continue this process for a total of k steps akin to the one described above,[5] we would obtain an expression for Δ as the determinant of an upper triangular matrix.

Let $\Delta_{i,j}^{(m)}$ denote the element of the i-th row and j-th column at step m in the triangulation process. Then, for $m = 1, \dots, k$,

$$\Delta_{i,j}^{(m)} = \begin{cases} \Delta_{i,j}^{(m-1)} - \frac{\Delta_{i,m}^{(m-1)}\Delta_{m,j}^{(m-1)}}{\Delta_{m,m}^{(m-1)}} & \text{if } m \leq i-1, \\ \Delta_{i,j}^{(m-1)} & \text{if } m > i-1, \end{cases}$$

with $\Delta_{i,j}^{(0)} = \sum x_{i-1,t}x_{j-1,t}$ (the element of the i-th row and j-th column before the triangulation process), and $x_{0,t} = 1$ for all t.

All non-zero elements retain their order in probability at each step in the triangulation process and, consequently, once triangulation is completed. We prove this statement through

[5] At each step m, the elements of the m-th column from the $(m+1)$-th row onward become 0.

induction. We have already shown the first step in triangulation conforms to the previous statement. Then, at an arbitrary m-th step, the term added to elements (i,j) for which $i > m-1$ is

$$-\frac{\Delta_{i,m}^{(m-1)}\Delta_{m,j}^{(m-1)}}{\Delta_{m,m}^{(m-1)}},$$

which is non-zero for $j > m-1$ and zero otherwise.

The induction hypothesis allows us to determine the order in probability of the term being added when it differs from zero: $O_p\left(T^{d_{x_{i-1}}+d_{x_{j-1}}}\right)$. Therefore, at this step, the element under consideration has itself retained its original order in probability as well if it is not rendered zero.

Hence, we have that

$$\Delta = O_p\left(T^{1+2d_{x_1}+2d_{x_2}+\ldots+2d_{x_k}}\right). \tag{9}$$

A similar argument may be applied analogously to (8) to find that

$$\Delta_k = O_p\left(T^{1+2d_{x_1}+2d_{x_2}+\ldots+d_{x_k}+d_y}\right). \tag{10}$$

Dividing (10) by (9), in accordance to (6), concludes the proof of item *ii*).

As for $\hat{\beta}_0$, note that all the terms in Equation (5) now have known orders in probability; plugging them in concludes the proof of item *i*).

Items *iii*) and *iv*)

The proof of items *iii*) and *iv*) makes use of the following equation

$$s^2 = \frac{1}{T}\sum \hat{u}_t^2 = \frac{1}{T}\Bigg(\sum y_t^2 - 2\hat{\beta}_0\sum y_t + T\hat{\beta}_0^2 - 2\sum_{i=1}^{k}\hat{\beta}_i\sum x_{i,t}y_t + \sum_{i=1}^{k}\hat{\beta}_i^2\sum x_{i,t}^2 + 2\hat{\beta}_0\sum_{i=1}^{k}\hat{\beta}_i\sum x_{i,t} + 2\sum_{i=1}^{k}\sum_{j>i}\hat{\beta}_i\hat{\beta}_j\sum x_{i,t}x_{j,t}\Bigg).$$

By applying the orders of convergence obtained above, we observe that all terms inside the parentheses are $O_p\left(T^{2d_y}\right)$.

To show item *iv*), we make use of the formula for the estimator of the variance-covariance matrix of the estimators, which may be written as

$$\widehat{\text{Var}}\left(\hat{\beta}\right) = s^2\frac{1}{\det\left(X'X\right)}\text{adj}\left(X'X\right),$$

where $\text{adj}\left(X'X\right)$ denotes the adjunct of $X'X$.

The order in probability of $\det(X'X)$ was previously determined at (9), whereas the order in probability of s^2 is provided above. As for the elements along the main diagonal of the adjunct, which are composed of the determinants of minors of the matrix $X'X$, we draw upon the previous triangulation argument to determine their order in probability: if the minors were to be triangulated, the elements of the main diagonal would retain their order in probability, and the determinant of any triangular matrix is the product of the elements along its main diagonal.

As regards the first element along the main diagonal of $\operatorname{adj}(X'X)$, we have that

$$\begin{vmatrix} \sum x_{1,t}^2 & \sum x_{1,t}x_{2,t} & \dots & \sum x_{1,t}x_{k,t} \\ \sum x_{1,t}x_{2,t} & \sum x_{2,t}^2 & \dots & \sum x_{2,t}x_{k,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{1,t}x_{k,t} & \sum x_{2,t}x_{k,t} & \dots & \sum x_{k,t}^2 \end{vmatrix} = O_p\left(T^{\sum_{i=1}^k 2d_{x_i}}\right).$$

Moreover, the i-th element along the main diagonal of $\operatorname{adj}(X'X)$, for $i = 2, \dots, k+1$, is

$$\begin{vmatrix} T & \sum x_{1,t} & \dots & \sum x_{i-2,t} & \sum x_{i,t} & \dots & \sum x_{k,t} \\ \sum x_{1,t} & \sum x_{1,t}^2 & \dots & \sum x_{1,t}x_{i-2,t} & \sum x_{1,t}x_{i,t} & \dots & \sum x_{1,t}x_{k,t} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{i-2,t} & \sum x_{1,t}x_{i-2,t} & \dots & \sum x_{i-2,t}^2 & \sum x_{i-2,t}x_{i,t} & \dots & \sum x_{i-2,t}x_{k,t} \\ \sum x_{i,t} & \sum x_{1,t}x_{i,t} & \dots & \sum x_{i-2,t}x_{i,t} & \sum x_{i,t}^2 & \dots & \sum x_{i,t}x_{k,t} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{k,t} & \sum x_{1,t}x_{k,t} & \dots & \sum x_{i-2,t}x_{k,t} & \sum x_{i,t}x_{k,t} & \dots & \sum x_{k,t}^2 \end{vmatrix} = O_p\left(T^{1+\sum_{j\neq i} 2d_{x_j}}\right).$$

Proofs of items $iii)$ and $iv)$ come from the orders of convergence computed above and the formula $t_{\beta_i} = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$ for $i = 0, 1, \dots, k$.

Items $v)$ to $vii)$

The proofs of items $v)$ and $vi)$ are analogous to those of these same items in Theorem 1 and are therefore omitted for reason of space.

Proof of item $vii)$ comes from the fact that

$$\mathcal{DW} = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}.$$

For which, we have that

$$\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2 = \sum_{t=2}^T \left[(y_t - y_{t-1}) - \hat{\beta}_1 (x_{1,t} - x_{1,t-1}) - \dots - \hat{\beta}_k (x_{k,t} - x_{k,t-1})\right]^2.$$

Note that $z_t - z_{t-1}$ are fractionally integrated processes of order $(d_z - 1) \in (0, 1/2)$. Consequently, $\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2 = O_p(T)$, as was shown for the stationary case, coupled with $\sum \hat{u}_t^2 = O_p(T^{1+2d_y})$ shown above, concludes the proof of item $vii)$.

C Additional finite sample evidence

Correlation between regressors

Table 7 shows simulation results for a scenario that allows us to examine the effect that the correlation between regressors have on the results. The regressand is generated as an $FI(d_y)$ process, Equation (1), while the regressors as $FI(d_{x_i})$ plus correlated noise, Equation (2). Three regressors are considered, x_1, x_2 and x_3, such that, $d_{x_1} + d_y > 1/2$, $d_{x_2} + d_y = 1/2$, and $d_{x_3} + d_y < 1/2$. We explore three correlation scenarios among regressors.

Table 7: Spurious regression, stationary variables, $\max_k\{d_{x_k}\} + d_y \geq \frac{1}{2}$.

$y_t \sim FI(d_y)$		$x_{1,t} \sim FI(d_{x_1}) + w_1$			$x_{2,t} \sim FI(d_{x_2}) + w_2$			$x_{3,t} \sim FI(d_{x_3}) + w_3$		
d_y	$\sigma^2_{\epsilon,y}$	d_{x_1}	σ^2_{ϵ,x_1}	σ^2_1	d_{x_2}	σ^2_{ϵ,x_2}	σ^2_2	d_{x_3}	σ^2_{ϵ,x_3}	σ^2_3
0.35	0.5	0.45	0.25	1	0.15	0.25	1	0.10	0.25	1

	$\sigma_{1,2} = 0.2;\ \sigma_{1,3} = 0.2;$ $\sigma_{2,3} = 0.2$			$\sigma_{1,2} = 0.5;\ \sigma_{1,3} = 0.5;$ $\sigma_{2,3} = 0.5$			$\sigma_{1,2} = 0.8;\ \sigma_{1,3} = 0.8;$ $\sigma_{2,3} = 0.8$		
T	50	100	1000	50	100	1000	50	100	1000
$RR_{t\beta_0}$	0.4881	0.5802	0.8008	0.4790	0.5788	0.7983	0.4786	0.5826	0.7984
$RR_{t\beta_1}$	0.0940	0.1371	0.3438	0.1023	0.1437	0.3455	0.1179	0.1574	0.3754
$RR_{t\beta_2}$	0.0626	0.0649	0.0883	0.0632	0.0724	0.1006	0.0666	0.0756	0.1258
$RR_{t\beta_3}$	0.0577	0.0608	0.0684	0.0603	0.0675	0.0879	0.0610	0.0699	0.1127
$RR_{\mathcal{F}}$	0.0821	0.1178	0.3025	0.0879	0.1231	0.3092	0.0988	0.1311	0.3417
R^2	0.0719	0.0398	0.0069	0.0729	0.0403	0.0071	0.0747	0.0418	0.0077
$\mathcal{DW}$	1.3956	1.2769	1.0649	1.3984	1.2758	1.0671	1.4002	1.2807	1.0658

RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests at a 5% nominal size, respectively. The number of replications is 10,000.

Table 7 allows us to quantify the finite sample behavior of the t-statistics and $\mathcal{F}$-statistics for different levels of correlations between the regressors. Note that for low levels of correlation, the statistics diverge at a slower rate. Thus, the table show that the risk of uncovering spurious regressions in finite samples increases with the correlation between regressors.

Slightly persistent regressors

Table 8 shows the simulation results for the multivariate stationary scenario where $\max_k\{d_{x_k}\} + d_y < 1/2$; this is, all of the regressors are only slightly persistent. The regressand is generated as an $FI(d_y)$ process, while the regressors as $FI(d_{x_i})$ plus possibly correlated noise. We consider three regressors under three scenarios: i) they are all independent, ii) the first two are correlated while the last one is not, and *iii*) they are all correlated.

Table 8: Spurious regression, stationary variables, $\max_k\{d_{x_k}\} + d_y < \frac{1}{2}$.

$y_t \sim FI(d_y)$		$x_{1,t} \sim FI(d_{x_1}) + w_1$			$x_{2,t} \sim FI(d_{x_2}) + w_2$			$x_{3,t} \sim FI(d_{x_3}) + w_3$		
d_y	$\sigma^2_{\epsilon,y}$	d_{x_1}	σ^2_{ϵ,x_1}	σ^2_1	d_{x_2}	σ^2_{ϵ,x_2}	σ^2_2	d_{x_3}	σ^2_{ϵ,x_3}	σ^2_3
0.20	0.5	0.20	0.25	1	0.15	0.25	1	0.10	0.25	1

	$\sigma_{1,2} = 0$; $\sigma_{1,3} = 0$; $\sigma_{2,3} = 0$			$\sigma_{1,2} = 0.8$; $\sigma_{1,3} = 0$; $\sigma_{2,3} = 0$			$\sigma_{1,2} = 0.8$; $\sigma_{1,3} = 0.8$; $\sigma_{2,3} = 0.4$		
T	50	100	1000	50	100	1000	50	100	1000
$RR_{t\beta_0}$	0.3039	0.3730	0.5789	0.2937	0.3749	0.5739	0.2974	0.3730	0.5751
$RR_{t\beta_1}$	0.0458	0.0548	0.0585	0.0568	0.0621	0.0665	0.0599	0.0637	0.0735
$RR_{t\beta_2}$	0.0500	0.0549	0.0526	0.0573	0.0579	0.0636	0.0554	0.0607	0.0670
$RR_{t\beta_3}$	0.0541	0.0539	0.0530	0.0505	0.0524	0.0548	0.0525	0.0567	0.0628
$RR_{\mathcal{F}}$	0.0508	0.0545	0.0545	0.0564	0.0608	0.0614	0.0556	0.0547	0.0699
R^2	0.0621	0.0314	0.0031	0.0636	0.0315	0.0032	0.0633	0.0316	0.0033
$\mathcal{DW}$	1.6655	1.6075	1.5250	1.6643	1.6064	1.5263	1.6643	1.6094	1.5254

RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests at a 5% nominal size, respectively. The number of replications is 10,000.

As the table shows, the rejection rates of the t-statistics remain relatively stable for all sample sizes and all correlation cases. Nonetheless, they also show that the distribution has heavier tails, since the actual rejection rates are systematically above the nominal 5% for relatively high values of d_{x_i}. This is in line with Theorem 1 given that under the three scenarios, the t-statistics are always $O_p(1)$. As for the $\mathcal{F}$ joint significance test statistic, its behavior is analogous to that of the t-statistics. Moreover, our simulations show that, as the sample size increases, the R^2 collapses to zero, whilst the $\mathcal{DW}$ approaches the value shown in Theorem 1. Thus, Table 8 shows that if all series are only slightly persistent, the spurious regression problem may be less acute, regardless of the correlation structure.

References

Agiakloglou, C. (2013). Resolving spurious regressions and serially correlated errors. *Empirical Economics*, 45(3):1361–1366.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59.

Cappuccio, N. and Lubian, D. (1997). Spurious regressions between I(1) processes with long memory errors. *Journal of Time Series Analysis*, 18:341–354.

Deng, A. (2013). Understanding Spurious Regression in Financial Economics. *Journal of Financial Econometrics*, 12(1):122–150.

Durlauf, S. N. and Phillips, P. C. B. (1988). Trends versus random walks in time series analysis. *Econometrica*, 56(6):1333–1354.

Ferson, W., Sarkissian, S., and Simin, T. (2003a). Is stock return predictability spurious? *Journal of Investment Management*, 1(3):1–10.

Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003b). Spurious regressions in financial economics? *The Journal of Finance*, 58(4):1393–1413.

Granger, C. W. J., Hyung, N., and Jeon, Y. (2001). Spurious regressions with stationary series. *Applied Economics*, 33(7):899–904.

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

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120.

Haldrup, N. and Vera-Valdés, J. E. (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics*, 199(1):1–11.

Hassler, U. (1996). Spurious regressions when stationary regressors are included. *Economics Letters*, 50(1):25–31.

Hassler, U. (1997). Sample autocorrelations of nonsationary fractionally integrated series. *Statistical Papers*, 38(1):43–62.

Hassler, U. (2003). Nonsense regressions due to neglected time-varying means. *Statistical Papers*, 44(2):169.

Hayashi, F. (2000). Econometrics. Princeton. *New Jersey, USA: Princeton University.*

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–76.

Hosking, J. R. M. (1984). Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series. *University of Winsconsin-Madison*, 2752.

Jensen, A. N. and Nielsen, M. Ø. (2014). A Fast Fractional Difference Algorithm. *Journal of Time Series Analysis*, 35(5):428–436.

Kolev, G. (2011). The "spurious regression problem" in the classical regression model framework. *Economics Bulletin*, 31(1):925–937.

Leccadito, A., Rachedi, O., and Urga, G. (2015). True versus spurious long memory: Some theoretical results and a monte carlo comparison. *Econometric Reviews*, 34(4):452–479.

Marinucci, D. and Robinson, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their applications*, 86(1):103–120.

Marmol, F. (1996). Nonsense regressions between integrated processes of different orders. *Oxford Bulletin of Economics & Statistics*, 58(3):525–36.

Marmol, F. (1998). Spurious regression theory with nonstationary fractionally integrated processes. *Journal of Econometrics*, 84:233–250.

Martínez-Rivera, B. and Ventosa-Santaulària, D. (2012). A comment on 'Is the spurious regression problem spurious?'. *Economics Letters*, 115(2):229–231.

Martínez-Rivera, B., Ventosa-Santaulària, D., and Vera-Valdés, J. E. (2012). Spurious forecasts? *Journal of Forecasting*, 31(3):245–259.

McCallum, B. T. (2010). Is the spurious regression problem spurious? *Economics Letters*, 107(3):321–323.

Mikosch, T. and De Vries, C. G. (2006). Tail probabilities for regression estimators. Technical report, Tinbergen Institute Discussion Paper. http://hdl.handle.net/10419/86396.

Osterrieder, D., Ventosa-Santaulària, D., and Vera-Valdés, J. E. (2019). The VIX, the Variance Premium, and Expected Returns*. *Journal of Financial Econometrics*, 17(4):1–42.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3):311–340.

Sollis, R. (2011). Spurious regression: A higher-order problem. *Economics Letters*, 111(2):141–143.

Tsay, W. J. and Chung, C. F. (2000). The spurious regression of fractionally integrated processes. *Journal of Econometrics*, 96(1):155–182.

Varneskov, R. T. and Perron, P. (2018). Combining long memory and level shifts in modelling and forecasting the volatility of asset returns. *Quantitative Finance*, 18(3):371–393.

Ventosa-Santaulària, D. (2009). Spurious regression. *Journal of Probability and Statistics*, 2009(802975):1–27.

Ventosa-Santaulària, D., Vera-Valdés, J. E., and Martínez-Olmos, A. I. (2016). A comment on 'resolving spurious regressions and serially correlated errors'. *Empirical Economics*, 51(3):1289–1298.

Zhang, L. (2018). Spurious regressions with high-order models: A reconsideration. *Economics Letters*, 168(1):70–72.