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
Early version, also known as pre-print

Spurious multivariate regressions under stationary fractionally integrated processes

Daniel Ventosa-Santaulària[a,*], J. Eduardo Vera-Valdés[b], Katarzyna Łasak[c], Ricardo Ramírez-Vargas[d]

[a]CIDE. daniel.ventosa@cide.edu
[b]Aalborg University. eduardo@math.aau.dk
[c]Tinbergen Institute. k.a.lasak@vu.nl
[d]CIDE. ricardo.ramirez@alumnos.cide.edu

Abstract

Spurious regression under long memory was studied by Tsay and Chung (2000) [JoE 96,pp. 155-182] for a univariate model. We extend their findings for the multivariate linear regression and find that inference drawn from the latter is also spurious. Our results hold for any finite number of independent stationary fractionally integrated explanatory variables. It is shown that each estimated parameter and its t-ratio collapse or diverge, depending on the persistence of the corresponding explanatory variable. Moreover, inference drawn from standard test statistics, such as the joint $\mathcal{F}$ test and the Durbin-Watson, is spurious. Nonetheless, the R-squared remains a correct goodness of fit measure. Comprehensive finite sample evidence shows that our asymptotic results hold even for small sample sizes such as 100 observations.

Keywords: Fractional integration, long memory, spurious regression, fractional cointegration.

*Corresponding author: CIDE, Carr. México-Toluca 3655, Col. Lomas de Santa Fe, Del. Álvaro Obregón, México D.F, C.P. 01210.

1. Introduction

Spurious regression in empirical econometrics is widely understood as the failure of conventional testing procedures when the series exhibit strong temporal properties. In economics, the level of many macroeconomic time series are known to behave as nonstationary processes, which is fertile ground for spurious inference.

Spurious regression was uncovered by Granger and Newbold (1974), and later explained by Phillips (1986), assuming independent driftless unit root processes as the data-generating process (DGP) for the regressor and the regressand in a univariate regression. Many extensions of Phillips' original results have been considered. For example, the spurious regression under mean and trend stationarity, or under independent higher-integration-order processes was studied by Hassler (1996), Marmol (1996), and Hassler (2003), respectively. Note also that the spurious regression phenomenon can be understood as a misspecified model from which misleading inference is drawn. In a sense, this interpretation encompasses the view of the spurious regression as a result of strong temporal properties, as the unit roots mentioned above, merely by assuming that the misspecification problem lies in the fact that such strong temporal properties have been ignored in the specification. Several authors, such as McCallum (2010), Kolev (2011), Agiakloglou (2013) and Zhang (2018), consider that non-resolved autocorrelation problems in the regression analysis lie at the origin of spurious regression. They supported their arguments with Monte Carlo evidence. Nonetheless, other studies, also based in Monte Carlo evidence, provide evidence that spurious regression is more than just poorly controlled autocorrelation; see Sollis (2011), Martínez-Rivera and Ventosa-Santaulària (2012) and Ventosa-Santaulària et al. (2016).

The spurious regression phenomenon[1] under long memory was first examined by Cappuccio and Lubian (1997), and Marmol (1998) using fractionally integrated processes of order d. Fractionally integrated processes were introduced in time series econometrics by Granger and Joyeux (1980) and Hosking (1981). 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. Their results include the earlier analytical studies of Phillips (1986), Haldrup (1994), and Marmol (1995, 1996). More precisely, TC showed that the OLS estimates have orders of convergence which vary depending on the order

[1]The terms "nonsense" and "spurious" regression are not necessarily equivalent. Historically, the former is associated with Granger and Newbold (1974), while the latter with Yule (1926), but lately the term "nonsense regression" appears in the context of regressions that use variables integrated of order 2. In order to avoid any possible confusion, we refer to the phenomenon as "spurious regression" and we thus consider that the "inference is misleading."

of integration of the processes; they may collapse, diverge, or even achieve non-degenerate limiting distributions. To be more precise, TC (p. 155) found that, "as long as [the variables'] orders of integration sum up to a value greater than $\frac{1}{2}$, the t-ratios become divergent and spurious effects occur."

Although Phillips (1986) and Durlauf and Phillips (1988) suggested that it is nonstationarity that causes spurious effects, TC's findings suggest, however, that misleading inference can occur in a regression between two stationary $I(d)$ processes, as long as their orders of integration sum up to a value greater than $\frac{1}{2}$. They thus considered 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). Stochastic and deterministic trending and structural breaks in a variable are therefore relevant sources of misleading inference, but not the only ones.[2]

In Granger and Newbold's (1974) original simulation exercise, spurious regression under stationarity was also uncovered. This case, not considered in Phillips' (1986) paper, was later studied by Granger et al. (2001) for *positively autocorrelated autoregressive series* and *long* moving average processes; this is, $MA(q)$ for $q > 20$ or even $q > 50$. Mikosch and De Vries (2006) also uncovered spurious regression when the innovations' distribution is fat-tailed. It is therefore important to consider the spurious regression is not an exclusively nonstationary phenomenon. However, this avenue has been scarcely considered. We aim to extend it 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 stationary fractionally integrated processes to the case of multivariate regression. Such an extension is important 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 empirical finance and macroeconomics. In a review of empirical literature, Baillie (1996) notes that price series and Consumer Price Index inflation 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) and Varneskov and Perron (2018)).

We consider a specification with an arbitrary finite number of explanatory variables, although inferior to the sample size, independent of the regressand. We consider both stationary and nonstationary cases. For the stationary case, our results show that, for non-correlated regressors, when the sum of the persis-

[2] Another interesting line of research in spurious regression focus on the effects of erroneously estimating short memory models, such as $ARMA(p,q)$ or even $ARIMA(p,1,q)$ when the the variables are fractionally integrated processes. Importantly, out-of-sample forecast errors increase; see, for instance Crato and Taylor (1996) and Arranz and Marmol (2001).

tence parameter of the regressand and that of a specific regressor is above or equal $\frac{1}{2}$, the t-ratio associated to the estimated coefficient of said regressor diverges. Conversely, when the value of this sum is below $\frac{1}{2}$, the t-ratio does not diverge. For correlated regressors, possibly the most relevant scenario, the convergence/divergence depends on the maximum degree of persistence among the regressors. The behaviour of the $\mathcal{F}$ statistic is similar to that of the t-statistics, albeit dependent on the sum of the persistence parameter of the regressand and the highest persistence parameter of all regressors, whether they are correlated or not. Hence, if the underlying processes are persistent enough, nonsense inference could be drawn for at least some individual t-ratios and the $\mathcal{F}$ statistic. Likewise, 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 signalling the poor fit. We contrast this behaviour with the one from a correctly specified fractional cointegration case where the R-squared does not collapse. As for the Durbin-Watson statistic, it converges to a value in the interval $(0,2)$ that depends on the persistence parameter of the process underlying the regressand. For the nonstationary case, our results extend the classical spurious regression literature with nonstationary variables to the multivariate fractionally cointegrated case. That is, all t-statistics and $\mathcal{F}$ diverge, at an even faster rate than for the stationary case, the Durbin-Watson statistic collapses to zero, while the R-squared does not.

The paper proceeds as follows: Section 2 presents the theoretical framework and the main asymptotic results. Section 3 shows the finite sample evidence that confirms our theoretical results, while Section 4 concludes. Proofs for the Theorems are provided in Appendix A and Appendix B.

2. Asymptotic results

We follow TC's notation and define a fractionally integrated process, denoted $FI(d_z)$, as 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 differencing parameter, and $(1-L)^d$ is the fractional differencing operator, defined as $(1-L)^d = \sum_{j=0}^{\infty} \Psi_j L^j$, where $\Psi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$ 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 $\sigma^2_{a_z}$. A fractionally integrated process is denoted here by $FI(d_z)$. When $d_z < \frac{1}{2}$, z_t is stationary; while for $d_z \in (-\frac{1}{2}, \frac{1}{2})$, it is invertible. Its autocovariance function 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[3] 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 $\sum_{j=-\infty}^{\infty} \gamma_z(j) = \infty$. We consider both stationary and nonstationary variables.

For the stationary extension, 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 possible correlation between the regressors. Let $m \in \{1,2,\cdots,k\}$ regressors be correlated. Without loss of generality, let these correlated regressors appear first in the specification. Our multivariate specification is given by

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

and

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

where $\varepsilon_{z,t}$ are i.i.d. white noise with zero mean and finite variance $\sigma^2_{\varepsilon,z}$, and $d_z \in (0, \frac{1}{2})$ 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 with variance matrix Σ, $\forall t$, given by

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

that is, $\sigma_{i,j} \neq 0$ for $1 \leq i, j \leq m$, $i \neq j$, and $\sigma_{i,j} = 0$ otherwise.[4]

Note that our specification allows for correlation between the regressors while simultaneously each regressor still 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

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

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

the matrix required in OLS computationally unstable and possibly not invertible. Note that near perfect correlation would translate in our specification 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 further 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 moving average representation of the form $w_t = \sum_{i=0}^{\infty} \phi_i \nu_{t-i}$ where $\sum_{i=0}^{\infty} i|\phi_i|$ with ν_t a white noise process. Note that this restriction ensures that the noise process is less persistent than the signal. Moreover, it is satisfied by all stationary ARMA processes.[5]

This theoretical framework suffices to study the asymptotic behaviour of a multivariate regression under stationary long memory processes. 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.[6] We summarise the data generating processes considered in the stationary multivariate extension in Assumption 1.[7]

Assumption 1. *Let y_t be an independent stationary fractionally integrated process of order d_y as in (1), and let $x_{i,t}$ for $i = 1, 2, \ldots, k$ and $k < T$, be signal plus noise processes given by (2). Suppose also that $E\left[\varepsilon_{z,t}\right]^{q_z} < \infty$ 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 indeed be misleading.

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

$$ii) \;\; \hat{\beta}_i = \begin{cases} O_p(T^{-\frac{1}{2}}) & \text{for } \bar{d}_{x1:m} + d_y < \frac{1}{2}, \\ O_p\left[\left(T^{-1}\log T\right)^{\frac{1}{2}}\right] & \text{for } \bar{d}_{x1:m} + d_y = \frac{1}{2}, \\ O_p(T^{\bar{d}_{x1:m} + d_y - 1}) & \text{for } \frac{1}{2} < \bar{d}_{x1:m} + d_y, \end{cases}$$

[5] See Hamilton (1994) pp. 504.

[6] Note that both type-I and type-II fractionally integrated processes could be encompassed in Assumption 1 by making $\varepsilon_{z,t} = 0$, for $z = y, x_i$, $\forall t \leq 0$, for type-II. See Marinucci and Robinson (2000) for further details. We would like to thank an anonymous referee for pointing out this to us.

[7] Note that both type-I and type-II fractionally integrated processes could be encompassed in Assumption 1 by making $\varepsilon_{z,t} = 0$, for $z = y, x_i$, $\forall t \leq 0$, for type-II. See Marinucci and Robinson (2000) for further details. We would like to thank an anonymous referee for pointing out this to us.

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

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

v)
$$t_{\beta_i} = \begin{cases} O_p(1) & \text{for } \bar{d}_{x1:m} + d_y < \frac{1}{2}, \\ O_p\left[(\log T)^{\frac{1}{2}}\right] & \text{for } \bar{d}_{x1:m} + d_y = \frac{1}{2}, \\ O_p(T^{\bar{d}_{x1:m}+d_y-\frac{1}{2}}) & \text{for } \frac{1}{2} < \bar{d}_{x1:m} + d_y, \end{cases}$$
for $i = 1, \ldots, m$*;*

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

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

Furthermore,

vii)
$$R^2 = \begin{cases} O_p\left(T^{-1}\right) & \text{for } \bar{d}_{x1:k} + d_y < \frac{1}{2}, \\ O_p\left(T^{-1}\log T\right) & \text{for } \bar{d}_{x1:k} + d_y = \frac{1}{2}, \\ O_p\left[T^{2(\bar{d}_{x1:k}+d_y-1)}\right] & \text{for } \frac{1}{2} < \bar{d}_{x1:k} + d_y; \end{cases}$$

viii)
$$\mathcal{F} = \begin{cases} O_p(1) & \text{for } \bar{d}_{x1:k} + d_y < \frac{1}{2}, \\ O_p(\log T) & \text{for } \bar{d}_{x1:k} + d_y = \frac{1}{2}, \\ O_p\left[T^{2(\bar{d}_{x1:k}+d_y)-1}\right] & \text{for } \frac{1}{2} < \bar{d}_{x1:k} + d_y; \end{cases}$$

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

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

Proof: See Appendix A.

Theorem 1 shows that, independently of the persistence of the series, all of the OLS-estimated coefficients for the stationary case collapse to zero asymptotically. This is not surprising given that there is no linear relationship between the variables. Nonetheless, the rate of convergence of each estimator varies depending on the values of d_{x_i}, d_y, and whether they are correlated with other regressors.

For the non-correlated regressors with $d_{x_i} + d_y < \frac{1}{2}$, the convergence rate is the usual $T^{-\frac{1}{2}}$, while for $d_{x_i} + d_y = \frac{1}{2}$, the convergence case is slightly slower. If $d_{x_i} + d_y > \frac{1}{2}$, 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.[8]

For correlated regressors, these orders of convergence are maintained by considering the rate of the more persistent series. Nonetheless, note that adding regressors to the specification does not make the estimates diverge in both cases.

[8]The estimate of the constant term $\hat{\beta}_0$ also collapses at a rate dependent solely on d_y; the more persistent the regressand, the slower the estimate collapses.

Moreover, Theorem 1 shows that nonsense inference could be drawn via the t-statistics. The slower rate of convergence of the estimators determines the rate of convergence of the t-statistics.

We focus first on the non-correlated regressors. Note that for regressors with $d_y + d_{x_i} < \frac{1}{2}$, the t-ratio associated to $\hat{\beta}_i$ does not diverge. This does not necessarily mean there are no distortions; the limiting distribution may well depart from the standard normal, as we illustrate through finite sample evidence. For $d_y + d_{x_i} = \frac{1}{2}$, the t-ratios slowly diverge at rate $(\log T)^{\frac{1}{2}}$. Such a rate ensures, asymptotically, that the null hypothesis is eventually rejected, although the required sample size could be relatively large due to the rather slow convergence rate. Finally, for $d_y + d_{x_i} > \frac{1}{2}$, the t-ratios diverge, and the rate of divergence is directly dependent on $d_y + d_{x_i}$.

For correlated regressors, the distortions get propagated. Given that the rate of convergence for correlated regressors depends on the maximum degree of persistence, it takes only one regressor to have memory such that $d_y + d_{x_i} \geq \frac{1}{2}$, to make all t-statistics diverge.

Note that the t-ratios diverge in TC because they assume that the orders of integration of y_t and x_t,the sole regressor, are either always superior to $\frac{1}{4}$, or the sum of d_x and d_y is superior to $\frac{1}{2}$, which allows one of the orders in integration to be below $\frac{1}{4}$. In our case, it can be seen clearly that whether the t-ratios diverge or not depends on the persistence of the regressand, and that of the specific regressor to which the t-ratio considered is associated to if it is a non-correlated regressor, or to the persistence of the maximum regressor if they are correlated. In other words, when all FI variables have a relatively low persistence parameter, the risk of drawing misleading inference from the regression analysis is rather low; this is, the estimators of the parameters collapse faster towards zero, and the t-ratios do not diverge. Nonetheless, if the regressors are correlated, and the persistence of at least one of them is marginally stronger such that the sum of its order of integration and that of the regressand is greater or equal to $\frac{1}{2}$, all t-ratios diverge.

Furthermore, Theorem 1 shows that the standard statistical tools to draw inference from the regression provide contradictory information. On the one hand, note that the coefficient of determination R^2 converges in probability to zero for all cases. Consequently, as the sample size increases, the declining R^2 correctly reflects the fact that the regressors do not explain the variation in the variable used as regressand. On the other hand, observe that the joint $\mathcal{F}$ test diverges if $\max_k\{d_{x_k}\} + d_y \geq \frac{1}{2}$, which would falsely indicate that at least one of the explanatory variables is linearly related to the regressand. Finally, 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 given that the asymptotic value of $\mathcal{DW}$ depends solely on the memory parameter d_y such that $\mathcal{DW}$ is in the interval $(0,2)$.

It is useful to contrast the results of Theorem 1 with those of a fractionally cointegrated regression, arguably, the antipode of a spurious regression; see Johansen and Nielsen (2012) for a technical explanation of fractional cointegration. The processes are generated as stated in Assumption 2.

Assumption 2. *Let $x_{i,t}$, for $i = 1,2,\ldots,k$ with $k < T$, and ε_t be independent stationary fractionally integrated processes that satisfy $(1-L)^{d_z} z_t = a_{z,t}$, for $z = x_i, \varepsilon$, where $a_{z,t}$ are i.i.d. white noises with finite variance $\sigma^2_{a,z}$, and d_x is the order of integration of processes $x_{i,t}$ and d_y is the order of integration of process ε_t, such that $0 \leq d_y < d_x < \frac{1}{2}$. Assume also that $E\left[a_{z,t}\right]^{q_z} < \infty$ with $q_z \geq \max\left\{4, -\frac{8d_z}{1+2d_z}\right\}$ for all $z = x_i, \varepsilon$. Finally, let y_t be generated as a linear combination of the previous processes: $y_t = \alpha + \sum_{i=1}^{k} \beta_i x_{i,t} + \varepsilon_t$.*

Given these data generating processes, Theorem 2 shows that the R^2 does not collapse but instead converges to a numerical value within zero and one, correctly reflecting the proportion of variance mimicked by the linear combination of the 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) $\hat{\beta}_0 \xrightarrow{P} \alpha$;

ii) $\hat{\beta}_i \xrightarrow{P} \beta_i$, *for* $i = 1,\ldots,k$;

iii) $t_{\beta_0} = T \frac{1}{\gamma_\varepsilon(0)} \alpha$;

iv) $t_{\beta_i} = T \frac{\gamma_{x_i}(0)}{\gamma_\varepsilon(0)} \beta_i$, *for* $i = 1,\ldots,k$;

v) $R^2 \xrightarrow{P} \left[1 - \frac{\gamma_\varepsilon(0)}{\sum_{i=1}^{k} \beta_i^2 \gamma_{x_i}(0) + \gamma_\varepsilon(0)}\right]$.

Proof: See Appendix A.

Theorem 2 show that OLS estimation of parameters α and β_i, $i = 1,\ldots,k$, converge in probability to their true values. As for the R^2, it asymptotically takes a value greater than 0 as long as at least one coefficient is different from 0. The R^2 is increasing in β_i and $\gamma_{x_i}(0)$, and decreasing in $\gamma_\varepsilon(0)$. The intuition behind this result is straightforward: the variance of the estimated residuals, ε_t, is truly the regressand's variance not explained by the linear combination of the regressors. In other words, the R^2 works as traditionally expected; this is, as a goodness of fit measure. The above result, coupled with the one obtained in Theorem 1, allows us to suggest to the practitioner to employ the R^2 as a vehicle to assess the goodness of fit, and even the validity of the regression. It could be argued that the R^2 could further be used to test whether there is a genuine linear relationship between the variables or not. Unfortunately, the limit expressions of the R^2 in both cases, spurious and fractionally integrated, make it clear that there are many unknown nuisance parameters that would make it rather hard to design a reliable test.

In keeping with our intent of generalising TC's results, we now turn our attention to the case of spurious multivariate regression under fractionally integrated nonstationary processes. The assumptions for the nonstationary case are detailed in Assumption 3.

Assumption 3. *Let y_t and $x_{i,t}$, for $i = 1, 2, \ldots, k$ and $k < T$, be independent nonstationary fractionally integrated processes of orders d_y and d_{x_i}, respectively, that satisfy $(1-L)^{d_z} z_t = \varepsilon_{z,t}$, where $\varepsilon_{z,t}$ are i.i.d. white noises with zero mean and finite variance $\sigma^2_{\varepsilon,z}$, and $d_z \in \left(\frac{1}{2}, 1\right)$ for $z = y, x_i$. Suppose also that $E\left[\varepsilon_{z,t}\right]^{q_z} < \infty$ with $q_z \geq \max\left\{4, -\frac{8d_z}{1+2d_z}\right\}$ for all $z = y, x_i$.*

Theorem 3 shows that, for the nonstationary case, the behaviour 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 3. *Let Assumption 3 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* $i = 1, \ldots, k$;

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

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

Furthermore,

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

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

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

Proof: See Appendix B.

Theorem 3 is in line with the results from classical studies 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 $O_p(1)$. In this sense, Theorem 3 extends the results from spurious regressions with nonstationary variables to the multivariate fractionally integrated case.

3. Finite sample results

Our theoretical findings point the risk of drawing misleading inference from an OLS-estimated regression model using long-range dependent series. Furthermore, this risk increases as the sample size grows; we confirm this in finite samples. For the simulations,[9] all error terms are sampled from Normal distributions, with a multivariate Normal for the correlated case. We generate the fractionally integrated processes using

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

the algorithm based on the fast Fourier transform, see Jensen and Nielsen (2014), with initial values set to zero. The simulated sample sizes are $T = 100$; $T = 1,000$; and $T = 10,000$.

Table 1 shows the simulation results for the stationary scenario where $\max_k\{d_{x_k}\} + d_y < \frac{1}{2}$. The regressand is generated as an $FI(d_y)$ processes, equation (1), while the regressors as $FI(d_{x_i})$ plus possibly correlated noise, equation (2). In all simulations, 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 1: 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	$\sigma^2_{\varepsilon,y}$	d_{x_1}	$\sigma^2_{\varepsilon,x_1}$	σ^2_1	d_{x_2}	$\sigma^2_{\varepsilon,x_2}$	σ^2_2	d_{x_3}	$\sigma^2_{\varepsilon,x_3}$	σ^2_3
0.25	2	0.20	1	1	0.15	0.75	0.75	0.10	0.75	0.40

	$\sigma_{1,2} = 0$; $\sigma_{1,3} = 0$; $\sigma_{2,3} = 0$			$\sigma_{1,2} = 0.4$; $\sigma_{1,3} = 0$; $\sigma_{2,3} = 0$			$\sigma_{1,2} = 0.4$; $\sigma_{1,3} = 0.6$; $\sigma_{2,3} = 0.3$		
T	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
$RR_{t\beta_0}$	0.4554	0.6736	0.8188	0.4514	0.6737	0.8176	0.4534	0.6855	0.8175
$RR_{t\beta_1}$	0.0618	0.0792	0.0914	0.0622	0.0861	0.0977	0.0707	0.0905	0.1036
$RR_{t\beta_2}$	0.0568	0.0656	0.0735	0.0569	0.0656	0.0791	0.0597	0.0688	0.0744
$RR_{t\beta_3}$	0.0583	0.0642	0.0637	0.0563	0.0567	0.0672	0.0600	0.0705	0.0759
R^2	0.0326	0.0035	0.0004	0.0326	0.0035	0.0004	0.0333	0.0036	0.0004
$RR_{\mathcal{F}}$	0.0647	0.0789	0.0927	0.0630	0.0811	0.1010	0.0686	0.0895	0.1047
$\mathcal{DW}$	1.4973	1.3797	1.3467	1.4977	1.3808	1.3465	1.5010	1.3789	1.3464

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-tests 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 behaviour is analogous to that of the t-ratios. This provided that all degrees of memory are such that their sum with the degree of memory of the regressand is less than $\frac{1}{2}$. 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 1 shows that under lightly persistent series, the spurious regression problem may be less acute.

The spurious phenomena is nonetheless more severe when at least one the regressors shows high persistence, and it spreads throughout to correlated series. This can be seen in Table 2 which shows the results for the scenario where $\max_k\{d_{x_k}\}+d_y \geq \frac{1}{2}$. The Table presents the results for a specification with three regressors, x_1, x_2 and x_3, such that, $d_{x_1}+d_y > \frac{1}{2}$, $d_{x_2}+d_y = \frac{1}{2}$, and $d_{x_3}+d_y < \frac{1}{2}$. We consider the same three correlation scenarios as for Table 1.

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	$\sigma^2_{\varepsilon,y}$	d_{x_1}	$\sigma^2_{\varepsilon,x_1}$	σ^2_1	d_{x_2}	$\sigma^2_{\varepsilon,x_2}$	σ^2_2	d_{x_3}	$\sigma^2_{\varepsilon,x_3}$	σ^2_3
0.35	2	0.25	1	1	0.15	0.75	0.75	0.10	0.75	0.40

	$\sigma_{1,2}=0$; $\sigma_{1,3}=0$; $\sigma_{2,3}=0$			$\sigma_{1,2}=0.4$; $\sigma_{1,3}=0$; $\sigma_{2,3}=0$			$\sigma_{1,2}=0.4$; $\sigma_{1,3}=0.6$; $\sigma_{2,3}=0.3$		
T	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
$RR_{t\beta_0}$	0.6005	0.8076	0.9155	0.5929	0.8058	0.9151	0.5982	0.8135	0.9098
$RR_{t\beta_1}$	0.0800	0.1360	0.2182	0.0816	0.1458	0.2195	0.0922	0.1614	0.2499
$RR_{t\beta_2}$	0.0612	0.0768	0.0972	0.0620	0.0806	0.1155	0.0646	0.0867	0.1077
$RR_{t\beta_3}$	0.0602	0.0754	0.0817	0.0597	0.0665	0.0855	0.0687	0.0935	0.1192
R^2	0.0347	0.0042	0.0005	0.0349	0.0042	0.0005	0.0361	0.0044	0.0006
$RR_{\mathcal{F}}$	0.0781	0.1320	0.2109	0.0789	0.1365	0.2220	0.0888	0.1541	0.2391
$\mathcal{DW}$	1.2627	1.0599	0.9841	1.2628	1.0615	0.9839	1.2666	1.0595	0.9837

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.

Focusing on the uncorrelated case, the Table shows the different rate of divergence of each t-statistic depending on the persistence level of its associated regressor. The results show that a practitioner with as few as 100 observations may obtain spurious results for highly persistent regressors. Moreover, the Table shows how the spurious problem gets propagated to correlated regressors. In particular, note the increase in rejection rates for each extra variable once we allow it to be correlated to the highly 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 $\frac{1}{2}$, as it is the case for the third regressor, but it is correlated with another whose memory does add to more than $\frac{1}{2}$, its t-statistic starts to diverge. Thus, under correlated regressors, it takes only one regressor to have a degree of memory such that its sum with that of the regressand is more than $\frac{1}{2}$, to obtain divergent t-statistics for all of them.

Moreover, comparing Tables 1 and 2 exhibits how the behaviour of the $\mathcal{F}$ statistic depends on the sum $\max_k\{d_{x_k}\}+d_y$. When the value of this sum is lower than $\frac{1}{2}$, Table 1, the statistic is relatively stable. When the value of this sum is instead greater or equal to $\frac{1}{2}$, Table 2, $\mathcal{F}$ starts to grow as the sample size increases. Nevertheless, the R^2 statistic always collapses to zero, although at different speeds depending on the persistence of the regressors. Finally, the Durbin-Watson statistics converge to the value obtained in Theorem 1, which only depends on the degree of memory of the regressand.

The results for the R^2 statistic are of particular interest when compared with results from a correctly specified regression in Table 3. The Tables shows four variables, y_t, $x_{1,t}, x_{2,t}$, and $x_{3,t}$, where there is a fractional cointegration relationship between them: the three regressors are $FI(d_{x_i})$, with $d_{x_i} > 0.2$, but the linear combination $(y_t - 0.70 - 0.2x_{1,t} - 0.3x_{2,t} - 0.4x_{3,t}) \sim FI(0.2)$ is less persistent. This data-generating process reflects a genuine relationship between the variables, at odds with the spurious ones simulated previously. Table 3 shows that in this case the rejection rates of the t-statistics are far greater than those obtained with independent variables, and, notably, the R^2 are away from zero, correctly indicating that the relationship is not spurious.

Table 3: Correctly specified fractional cointegration.

$$y_t = 0.70 + 0.2x_{1,t} + 0.3x_{2,t} + 0.4x_{3,t} + u_t;$$
$$z_t \sim FI(d_z);\ \sigma^2_{\varepsilon,x_1} = 1;\ \sigma^2_{\varepsilon,x_2} = 0.75;\ \sigma^2_{\varepsilon,x_3} = 0.75;\ \sigma^2_u = 0.5;\ d_u = 0.2$$

	$d_z = 0.25$ $z = x_1, x_2, x_3$			$d_z = 0.35$ $z = x_1, x_2, x_3$			$d_z = 0.45$ $z = x_1, x_2, x_3$		
T	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
$RR_{t\beta_0}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$RR_{t\beta_1}$	0.958	1.000	1.000	0.966	1.000	1.000	0.974	1.000	1.000
$RR_{t\beta_2}$	0.984	1.000	1.000	0.987	1.000	1.000	0.990	1.000	1.000
$RR_{t\beta_3}$	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
R^2	0.443	0.434	0.436	0.468	0.480	0.492	0.506	0.554	0.595

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.

To summarise, for fractionally integrated stationary regressors, the less persistent the series, the smaller the spurious regression phenomenon. Nonetheless, it takes only one highly persistent regressor for spurious phenomena to appear. 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, she should be wary that her regression might be of little use.

Turning our attention to the nonstationary case, Table 4 shows the results from simulations with nonstationary fractionally integrated variables. The Table is in line with the results from Theorem 3, which is in turn in line with classical results on spurious regressions with nonstationary variables, see Ventosa-Santaulària (2009) for a review.

Table 4: Spurious regression, nonstationary variables.

$$y_t \sim FI(d_y);\ x_{1,t} \sim FI(d_{x_1});\ x_{2,t} \sim FI(d_{x_2});\ x_{3,t} \sim FI(d_{x_3})$$

	$d_y = 0.60;\ d_{x_1} = 0.60;$ $d_{x_2} = 0.60;\ d_{x_3} = 0.60$			$d_y = 0.80;\ d_{x_1} = 0.80;$ $d_{x_2} = 0.80;\ d_{x_3} = 0.80$			$d_y = 0.75;\ d_{x_1} = 0.70;$ $d_{x_2} = 0.65;\ d_{x_3} = 0.60$		
T	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
$RR_{t\beta_0}$	0.6931	0.8931	0.9643	0.7216	0.9081	0.9732	0.7617	0.9204	0.9760
$RR_{t\beta_1}$	0.3906	0.7395	0.9041	0.5701	0.8538	0.9530	0.5250	0.8354	0.9434
$RR_{t\beta_2}$	0.3841	0.7331	0.9024	0.5665	0.8482	0.9515	0.4854	0.8164	0.9374
$RR_{t\beta_3}$	0.3904	0.7395	0.9039	0.5644	0.8517	0.9522	0.4524	0.7860	0.9248
R^2	0.1537	0.1064	0.0867	0.3140	0.2955	0.2931	0.2257	0.1914	0.1780
$RR_{\mathcal{F}}$	0.7082	0.9762	0.9984	0.9096	0.9969	1.0000	0.8342	0.9917	1.0000
$\mathcal{DW}$	0.8174	0.3891	0.2112	0.5143	0.1244	0.0306	0.5726	0.1767	0.0597

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 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 far higher than for the stationary case. Furthermore, the Table 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, the closer the process is to a unit root process, the closer the statistic is to zero. This, coupled with the asymptotic behaviour of the R-squared statistic, $R^2 \sim O_p(1)$, allows us to partly recover Granger and Newbold's (1974) rule of thumb for detecting a spurious regression, $R^2 > \mathcal{DW}$, for nonstationary variables with high degree of persistence. Nonetheless, note that if the degree of memory is only slightly above $\frac{1}{2}$, and thus close to the stationary region, the rule of thumb still fails to suggest that the regression may be spurious. Thus, care should be taken with using the rule of thumb when the practitioner suspects that the nonstationary variables may be fractionally integrated.

4. Concluding remarks

We studied the asymptotic and finite sample behaviour of the OLS-estimated multivariate regression with an arbitrary finite number of fractionally integrated regressors. We consider both stationary and nonstationary cases. Our findings extend the literature in several directions.

For the stationary case, our multivariate approach shows that the asymptotic behaviour of each parameter estimate and its associated t-ratio depends on the specific persistence of the regressors, and on whether they are correlated. Under correlated regressors, probably the most relevant scenario, increasing or decreasing the number of regressors in a specification may alter the asymptotic properties of the estimates. This in the sense that if a more persistent regressor, correlated with the rest, is introduced, all t-statistics start to diverge. 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 persistence of the variables grows, and it gets propagated when they are correlated.

The behaviour of the R-squared under stationary regressors is particularly interesting when compared with the results from a correctly specified regression. The R-squared collapses to zero when the regressors are independent of the regressand, but not when there is a genuine fractionally cointegrated relationship. In this sense, these results might prove useful to distinguish spurious regressions from genuine ones when the variables behave as stationary fractionally integrated processes.

For the nonstationary case, our results extend the classical spurious regression results. All t-statistics and the $\mathcal{F}$ 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.

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

Acknowledgements

We would like to thank Niels Haldrup, Ignacio Lobato, Michael Massmann, two anonymous referees, and the participants of the Long Memory Conference 2018 for their valuables comments. All remaining

errors are ours.

References

Agiakloglou, C., 2013. Resolving spurious regressions and serially correlated errors. Empirical Economics 45, 1361–1366. doi:10.1007/s00181-012-0647-4.

Arranz, M.A., Marmol, F., 2001. Out-of-sample forecast errors in misspecific perturbed long memory processes. Statistical papers 42, 423–436.

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

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

Crato, N., Taylor, H.M., 1996. Stationary persistent time series misspecified as nonstationary arima. Statistical Papers 37, 215–223.

Durlauf, S.N., Phillips, P.C.B., 1988. Trends versus random walks in time series analysis. Econometrica 56, 1333–1354.

Granger, C.W.J., Hyung, N., Jeon, Y., 2001. Spurious regressions with stationary series. Applied Economics 33, 899–904.

Granger, C.W.J., 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., Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics 2, 11–20.

Haldrup, N., 1994. The asymptotics of the single-equation cointegration regressions with I(1) ans I(2) variables. Journal of Econometrics 63, 153–181.

Hamilton, J.D., 1994. Time Series Analysis. 1 ed., Princeton University Press.

Hassler, U., 1996. Spurious regressions when stationary regressors are included. Economics Letters 50, 25–31.

Hassler, U., 1997. Sample autocorrelations of nonsationary fractionally integrated series. Statistical Papers 38, 43–62.

Hassler, U., 2003. Nonsense regressions due to neglected time-varying means. Statistical Papers 44, 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., Nielsen, M.Ø., 2014. A Fast Fractional Difference Algorithm. Journal of Time Series Analysis 35, 428–436. doi:10.1111/jtsa.12074.

Johansen, S., Nielsen, M.Ø., 2012. Likelihood inference for a fractionally cointegrated vector autoregressive model. Econometrica 80, 2667–2732.

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

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

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

Marmol, F., 1995. Spurious regressions between I(d) processes. Journal of Time Series Analysis 16, 313–321.

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

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

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

McCallum, B.T., 2010. Is the spurious regression problem spurious? Economics Letters 107, 321–323.

Mikosch, T., 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., Vera-Valdés, J.E., 2018. The VIX, the variance premium, and expected returns. Journal of Financial Econometrics doi:10.1093/jjfinec/nby008.

Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. Journal of Econometrics 33, 311–340.

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

Tsay, W.J., Chung, C.F., 2000. The spurious regression of fractionally integrated processes. Journal of Econometrics 96, 155–182.

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

Ventosa-Santaulària, D., 2009. Spurious regression. Journal of Probability and Statistics 2009. doi:10.1155/2009/802975.

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

Yule, G.U., 1926. Why do we sometimes get nonsense-correlations between Time-Series?–A study in sampling and the nature of time-series. Journal of the royal statistical society 89, 1–63.
Zhang, L., 2018. Spurious regressions with high-order models: A reconsideration. Economics Letters 168, 70–72.

Appendix A. Proof of Theorems 1 and 2

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 A.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. (2018).

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

Table A.5: Expressions for sums in Theorem 1 with $i\neq j$; $i,j=1,\cdots,k$. All sums range from $t=1$ to $t=T$.

Items $i)$ to $iii)$

Recall the OLS formula:

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

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

Let

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

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,

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

Define $\Pi := (\frac{1}{T}\Omega - \frac{1}{T}\overrightarrow{x}\frac{1}{T}\overrightarrow{x}')^{-1}$ and using blockwise inversion,

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

From Table A.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,

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

The result follows from plugging the appropriate rate of convergence from Table A.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 case. Given that $\Sigma_{X,1:m}^1$ 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 $\Sigma_{X,m+1:k}^{-1}$ 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

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

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 (A.1), the numerator can be written as

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

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,

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

Using items $i)$ to $iii)$, which show that all estimators have negative orders of probability, the highest order of probability is the one of $\sum_{t=2}^{T}y_ty_{t-1}$. 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:

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

Appendix B. Proof of Theorem 3

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 B.6, along with their respective convergence rates. Table B.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 B.6: Expressions for sums in Theorem 3 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

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

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

$$\hat{\beta}_k=\frac{\Delta_k}{\Delta}, \tag{B.2}$$

where

$$\Delta=\det\left(X'X\right), \tag{B.3}$$

and

$$\Delta_k=\begin{vmatrix} T & \sum x_{1,t} & \sum x_{2,t} & \cdots & \sum y_t \\ \sum x_{1,t} & \sum x_{1,t}^2 & \sum x_{1,t}x_{2,t} & \cdots & \sum x_{1,t}y_t \\ \sum x_{2,t} & \sum x_{1,t}x_{2,t} & \sum x_{2,t}^2 & \cdots & \sum\sum x_{2,t}y_t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{k,t} & \sum x_{1,t}x_{k,t} & \sum x_{2,t}x_{k,t} & \cdots & \sum x_{k,t}y_t \end{vmatrix}. \tag{B.4}$$

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 (B.3), the denominator in expression (B.2). Note from Table B.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 B.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,[10] 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 $m = 1,\dots,k$,

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

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 (i.e., once we are left with an upper triangular matrix). We prove this statement through 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.

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

Hence, we have that

$$\Delta = O_p\left(T^{1+2d_{x_1}+2d_{x_2}+\ldots+2d_{x_k}}\right). \tag{B.5}$$

A similar argument may be to applied analogously to (B.4) to find that

$$\Delta_k = O_p\left(T^{1+2d_{x_1}+2d_{x_2}+\ldots+d_{x_k}+d_y}\right). \tag{B.6}$$

Dividing (B.6) by (B.5), in accordance to (B.2), concludes the proof of item *ii*).

As for $\hat{\beta}_0$, note that all the terms in equation (B.1) 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

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

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 once again make use of the formula for the estimator of the variance-covariance matrix of the estimators, which may be written as

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

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

The order in probability of $\det\left(X'X\right)$ was previously determined at (B.5), 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 $\mathrm{adj}\left(X'X\right)$, we have that

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

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

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

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 $i = 0,1,\ldots,k$.

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

$$\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2 = \sum_{t=2}^{T} \left[(y_t - y_{t-1}) - \hat{\beta}_1 (x_{1,t} - x_{1,t-1}) - \hat{\beta}_2 (x_{2,t} - x_{2,t-1}) - \ldots - \hat{\beta}_k (x_{k,t} - x_{k,t-1}) \right]^2.$$

Note that $z_t - z_{t-1}$ are fractionally integrated processes of order $(d_z - 1) \in \left(0, \frac{1}{2}\right)$. 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*).