CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.

CIDE

SPURIOUS MULTIVARIATE REGRESSIONS UNDER STATIONARY FRACTIONALLY INTEGRATED PROCESSES

TESINA

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ECONOMÍA

PRESENTA

RICARDO RAMÍREZ VARGAS

DIRECTOR DE LA TESINA: DR. DANIEL VENTOSA-SANTAULÁRIA

MÉXICO, D.F.   AGOSTO, 2017

Para mi familia

Agradecimientos

Abstract

Spurious regression under stationary processes exhibiting 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 the t-statistics associated to the estimated parameters diverge if the processes underlying the dependent variable and the particular explanatory variable are sufficiently persistent. It is shown also that inference drawn from test statistics and goodness of fit measures, such as the Wald $\mathcal{F}$ statistic and the R^2 can be contradictory in the sense that the test of joint significance may reject the null hypothesis if the underlying variables are strongly persistent, indicating incorrectly that at least one of the explanatory variables affects the dependent variable, whereas the latter always converges to zero, supporting the correct assertion that the variables used as regressors do not explain the variable used as regressand. Comprehensive finite sample evidence is consistent with our asymptotic results and shows that they hold even for small sample sizes such as 100 observations.

Keywords: fractional integration, long memory, spurious regression
JEL classification: C12, C13, C15, C22

Contents

List of tables

Chapter 1

Introduction

Spurious regression in empirical econometrics is widely understood as the failure of conventional testing procedures when the series exhibit strong persistence. In economics, the levels of many macroeconomic time series are known to behave as nonstationary processes, which is in turn known to lead to spurious inference. Granger and Newbold (1974) first studied the problem of spurious regression through simulations. They generate series as independent driftless unit root processes and subsequently carry out a simple regression using these as regressand and the single regressor. They then test the significance of the coefficient associated to the slope and find that, contrary to what should occur in the absence of spurious effects, the rejection rate of the null hypothesis is quite high, which suggests a linear relationship between the variables that evidently does not exist in the underlying series. Furthermore, they find moderately high values for the coefficient of determination (R^2), which suggests that a fair amount of the variation in the dependent variable is explained by the independent variable, and low values for the Durbin-Watson statistic ($\mathcal{DW}$), which, for its part, suggests a high positive first autocorrelation of the residuals.

Phillips (1986) explained Granger and Newbold's results analytically by studying the asymptotic properties of the statistics associated to the regression. In a univariate linear regression with independent unit root processes, he shows that the t-ratio test statistic does not tend toward

a limiting distribution, but rather diverges as the sample size (T) approaches infinity, which implies that the null hypothesis will be rejected and thus a statistic relationship between the two independent variables will be erroneously established. Additionally, he finds that the spurious effects extend to the estimated coefficient associated to the intercept, which diverges, and to the estimated coefficient associated to the slope and the R^2, both of which have a non-degenerate limiting distribution. Further, he finds that the $\mathcal{DW}$ statistic converges in probability to 0.

As for fractionally integrated processes, these were introduced by Granger and Joyeux (1980) and Hosking (1981) as a generalization of Box and Jenkins (1976) ARIMA models where the integration parameter d is allowed to adopt any real value rather than being restricted to integers. This allows the modeling of long-range dependence with $0 < d < \frac{1}{2}$. A series is said to exhibit long memory, another term for long-range dependence, if absolute summability of the autocorrelations does not hold. When $0 < d < \frac{1}{2}$, the autocovariance function declines at a hyperbolic rate, in contrast with the faster exponential rate of the conventional stationary ARMA processes. Moreover, Hosking and Granger and Joyeux showed that if $0 < d < \frac{1}{2}$ then the process is stationary.

The spurious regression phenomenon under fractional integration was first examined by Cappuccio and Lubian (1997) and Marmol (1998), both of whom consider nonstationary fractionally integrated processes. Cappuccio and Lubian studied univariate linear regressions between $I(d+1)$ processes with $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Marmol, on the other hand, analyzed linear regressions in the levels of nonstationary fractionally integrated processes spuriously related in a multivariate single-equation setting. He finds that the $\mathcal{DW}$ statistic converges to 0 and notes that the rule of thumb proposed by Granger and Newbold to detect spurious regression, $R^2 > \mathcal{DW}$, is still effective in the setting he described.

Tsay and Chung (2000), TC henceforth, studied the asymptotic properties of a regression when independent stationary and nonstationary fractionally integrated processes are spuriously related in a univariate single-equation setting. More precisely, TC showed that the ordinary least squares (OLS) estimates have orders of convergence which vary depending on the orders of

integration of the processes; they may converge, diverge or even achieve non-degenerate limiting distributions. Specifically, in regards to stationary processes, TC found that if the processes are "strongly persistent," such that the sum of their orders of integration is a value greater than $\frac{1}{2}$, then spurious effects are present in the divergent t-ratio, but absent in the R^2, which converges to 0, albeit at a slower rate with respect to that under short memory processes.

Although Phillips (1986) and Durlauf and Phillips (1988) suggested that it is nonstationarity that causes spurious effects, TC's findings indicate that misleading inference can occur in a regression between two stationary $I(d)$ processes (as long as the sum of their orders of integration is a value greater than $\frac{1}{2}$). They thus considered that strong persistence originates the spurious effects. In other words, the causes of spurious regression can be better understood as "strong temporal properties", as explained by Granger et al. (2001). It is therefore important to consider spurious regression not as a phenomenon exclusively associated with nonstationarity or lack of ergodicity. It is our aim to provide further theoretical and finite sample evidence that spurious effects are due to the persistence of the series. We therefore extend TC's findings concerning stationary fractionally integrated processes to the case of a multivariate regression (TC's corresponding results deal only with a simple regression with one explanatory variable and a constant term).

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 present behavior that appears to exhibit long memory. Moreover, he mentions applications of fractionally integrated models to asset prices, stock returns, exchange rates and interest rates.

We consider a specification with an arbitrary finite number of explanatory variables (although inferior to the sample size), independent of each other and the dependent variable. Our results show that when the sum of the persistence parameter of the dependent variable and that of a particular regressor is above $\frac{1}{2}$, the t-ratio associated to the estimated coefficient of said regressor diverges. Conversely, when the value of this sum is lower than $\frac{1}{2}$, the t-ratio does not diverge.

3

The behavior 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 the regressor processes. Likewise, we find that the R^2 does converge 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 of the regressor processes. Hence, if the underlying processes are persistent enough, spurious effects are present in the divergent t and $\mathcal{F}$ statistics, but absent in the collapsing R^2. As for the $\mathcal{DW}$ statistic, it converges to a value in the interval $(0,2)$ that depends negatively on the persistence parameter of the process underlying the regressand.

The dissertation proceeds as follows: section 2 presents the theoretical framework, the main asymptotic results, and the finite sample evidence that confirms our theoretical results. Section 3 concludes. Appendix A contains the proof of the theorem. Appendix B contains Tables 1-6 which display the Monte Carlo simulation results.

Chapter 2

Asymptotic results and finite sample evidence

We follow TC's notation and define a fractionally integrated process, denoted $FI(d_z)$, as a discrete-time stochastic process z_t (for $z = y, x_1, \ldots, x_k$) 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 Gaussian white noise with zero mean and finite variance $\sigma_{a_z}^2$. A stationary fractionally integrated process is denoted here by $SFI(d_z)$; when $d_z < 1/2$, z_t is stationary. Moreover, for $d_z \in (-1/2, 1/2)$, z_t is invertible. Its autocovariance function is

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

and its first autocorrelation,

$$\rho_z(1) = \frac{d_z}{1 - d_z}. \tag{2.1}$$

When $d_z > 0$ the process is said to possess long memory since it exhibits long-range dependence in the sense that $\sum_{j=-\infty}^{\infty} \gamma_z(j) = \infty$.

This theoretical framework suffices to study the asymptotic behavior 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. Further, let s^2 and $s^2_{\hat{\beta}_j}$ denote the estimated variance of the residuals and the estimated variance of $\hat{\beta}_j$, respectively. We make the following assumption as to the data generating processes:

Assumption 1: *Let y_t and $x_{i,t}$, for $i = 1, 2, \ldots, k$ and $k < T$, be independent stationary fractionally integrated processes of orders d_y and d_{x_i}, respectively, that satisfy $(1 - L)^{d_z} z_t = \varepsilon_{z,t}$, for $z = y, x_i$, where $\varepsilon_{z,t}$ are gaussian white noises with zero mean and finite variance $\sigma^2_{\varepsilon,z}$, and $d_z \in \left(0, \frac{1}{2}\right)$. 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. Finally, we define $\overline{d_x} \equiv \max\{d_{x_1}, d_{x_2}, \ldots, d_{x_k}\}$.*

The theorem 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 $T \to \infty$:*

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

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

3. $s^2 \xrightarrow{P} \gamma_y(0)$.

4. $T s^2_{\hat{\beta}_0} \xrightarrow{P} \gamma_y(0)$,
   $T s^2_{\hat{\beta}_i} \xrightarrow{P} \frac{\gamma_y(0)}{\gamma_{x_i}(0)}$, for $i = 1, \ldots, k$.

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

6. $T^{-d_y} t_{\beta_0} = O_p(1)$.

7. $R^2 = \begin{cases} O_p\left(T^{-1}\right) & \text{for } \overline{d_x} + d_y < \frac{1}{2}, \\ O_p\left(T^{-1}\ln T\right) & \text{for } \overline{d_x} + d_y = \frac{1}{2}, \\ O_p\left[T^{2(\overline{d_x}+d_y-1)}\right] & \text{for } \frac{1}{2} < \overline{d_x} + d_y. \end{cases}$

8. $\mathcal{F} = \begin{cases} O_p(1) & \text{for } \overline{d_x} + d_y < \frac{1}{2}, \\ O_p(\ln T) & \text{for } \overline{d_x} + d_y = \frac{1}{2}, \\ O_p\left[T^{2(\overline{d_x}+d_y)-1}\right] & \text{for } \frac{1}{2} < \overline{d_x} + d_y. \end{cases}$

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

We have $\xrightarrow{P}$ and $O_p(\cdot)$ denote convergence in probability and order in probability, respectively.

Proof: See Appendix A.

Item 1 of the theorem shows that, independently of the persistence of the series, as long as it remains stationary, all of the OLS-estimated coefficients collapse to zero as $T \to \infty$, as could be expected given that there is no linear relationship between the variables (i.e., the population parameters are equal to zero). Nonetheless, the rate of convergence of each estimator $\hat{\beta}_i$, for $i = 1, \ldots, k$, considerably varies depending on the value of $d_{x_i} + d_y$. For $0 < d_{x_i} + d_y < \frac{1}{2}$, the convergence rate is the usual $T^{-\frac{1}{2}}$, but for $d_{x_i} + d_y = \frac{1}{2}$, the convergence case is slightly slower. If $\frac{1}{2} < d_{x_i} + d_y < 1$, the convergence rate depends explicitly 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 $O_p(T^0)$, and, conversely, as it approaches $\frac{1}{2}$ its order in probability approaches instead $O_p\left(T^{-1/2}\right)$. In other words, the more persistent the process is, the slower the rate of convergence of the estimators. From item 2 we can see that the estimate of the constant, $\hat{\beta}_0$, also collapses, at a rate dependent

solely on d_y; the more persistent the process used as regressand, the slower the constant estimate converges to 0.

Furthermore, observe from item 3 that the estimator of the variance of the residuals is consistent. Item 4 shows that the squares of the standard errors $s^2_{\hat{\beta}_j}$ converge at the usual rate T^{-1}. As in TC, it is the slower rate of convergence of the estimators which causes the t-ratios to diverge. Spurious effects are observable in the t-ratio test statistics, as can be seen in item 5 of the theorem. For $0 < d_y + d_{x_i} < \frac{1}{2}$, the t-ratio associated to $\hat{\beta}_i$ does not diverge. Note, however, that this does not necessarily mean there are no distortions, since 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 $(\ln T)^{\frac{1}{2}}$. Such a rate *ensures*, asymptotically, that the null hypothesis is eventually rejected, although the required sample size should be relatively large (greater than 500 according to our simulations) due to the rather slow convergence rate. Finally, for $\frac{1}{2} < d_y + d_{x_i} < 1$, the t-ratios diverge at rate T^c, where $c \in \left(0, \frac{1}{2}\right)$, and the rate of divergence is directly dependent on $d_y + d_{x_i}$.

The t-ratios diverge in TC because the authors assumed that the orders of integration of y_t and x_t, the single 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 process used as the regressand and that of the particular regressor to which the t-ratio considered is associated. These results are quite intuitive and consistent with TC. On the one hand, when the independent SFI variables have a relatively low persistence parameters, the inference drawn from the regression analysis tends to be more accurate; this is, the estimators of the parameters collapse faster towards zero and the t-ratios do not diverge. When the persistence of the series is marginally stronger, such that the sum of the orders of integration of the regressand and the regressor is equal to $\frac{1}{2}$, the t-ratio diverges, albeit slowly, and the estimated coefficient collapses at a slightly slower rate. On the other hand, when the persistence of the series is stronger, such that the value of this sum is strictly greater than $\frac{1}{2}$, the risk of making spurious inference is higher, since the estimates collapse

to zero at a slower rate whilst the t-ratio diverges and the rates of convergence and divergence, respectively, are directly dependent on the sum of the persistence parameters.

As for the standard statistical tools to draw inference from the regression, these provide contradictory information. On the one hand, note from item 7 that the coefficient of determination R^2 converges in probability to zero for any $d_z \in \left(0, \frac{1}{2}\right)$. Consequently, as the sample size increases, the declining R^2 correctly reflects the fact that the regressors do not explain the variations of the variable used as regressand. On the other hand, observe from item 8 that the Wald statistic diverges if $\overline{d_x} + d_y \geq \frac{1}{2}$, indicating that at least one of the explanatory variables has a considerable influence on the behavior of the regressand variable, which is incorrect (though the divergence rate is slow if the condition holds with equality). In this sense, the divergent $\mathcal{F}$ statistic and the collapsing R^2 are contradictory. Finally, Granger and Newbold's (1974) rule of thumb for detecting a spurious regression, $R^2 > \mathcal{DW}$, no longer applies in view of item 9, because the asymptotic value of the $\mathcal{DW}$ depends solely on the memory parameter d_y such that $\mathcal{DW}$ is in the interval $(0, 2)$.

The theoretical results show that, under specific persistence properties, the risk of drawing spurious inference from an OLS-estimated regression model using (long-range dependence) stationary independent series increases as the sample size grows. We confirm this in finite samples. All the series are generated as independent SFI(d_z) processes, $z_t = (1 - L)^{-d_z} a_{z,t} \sim I(d_z)$ for $z = y, x_1, x_2, x_3$, and x_4 and $a_{z,t} \sim$ i.i.d.$\mathcal{N}(0, \sigma_{a,z}^2)$. We then use these series to estimate by OLS the following three specifications: **(1)** $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$; **(2)** $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$; **(3)** $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$. In Tables B.1 and B.2 all the variables have the same persistence parameter, this is, $d_y = d_{x_1} = \ldots = d_{x_4} < 1/2$. Specifically, Table B.1 shows results for $d = 1/5, 1/4$ and $3/10$, whilst Table B.2 does it for $d = 7/20, 2/5$ and $9/20$. In Tables B.4, B.5, and B.6 the persistence parameter is different for each variable, though always inferior to $1/2$. In these Tables we isolate the three cases found in our theoretical analysis: $\overline{d_x} + d_y < \frac{1}{2}, \overline{d_x} + d_y = \frac{1}{2}$, and $\overline{d_x} + d_y > \frac{1}{2}$. Tables B.4, B.5, and B.6, the three cases are

9

studied under regressions involving two, three, and four regressors, respectively.[1]

By and large, the simulation results show that the rejection rates of the t-tests remain fairly stable for $0 < d_{x_i} + d_y < \frac{1}{2}$, they slowly increase for $d_{x_i} + d_y = \frac{1}{2}$, and increase at a faster pace for $\frac{1}{2} < d_{x_i} + d_y < 1$. As for the $\mathcal{F}$ joint significance test statistic, its behavior is analogous to that of the t-ratios albeit dependent on $\overline{d_x} + d_y$. Moreover, our simulations show that, as the sample size increases, the R^2 collapses to zero, whilst the $\mathcal{DW}$ approaches the value predicted in item 9 of the theorem. In particular, when all processes have the same order of integration, for $0 < d < \frac{1}{4}$, the theorem states that the t-ratios do not diverge. This is confirmed by simulations (see Table B.1's results in column 1, when $d = 1/5$), where the rejection rate remains stable around $0.07 - 0.10$ for sample sizes above 500 observations, but the simulations 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. Moreover, simulations also show that the joint significance test exhibits relatively adequate rejection rates (this is, close to the nominal 5%) for small samples (100) and small values of d (see Table 1). The R^2 are rather low even for samples sizes as small as 100 observations. As for Tables B.4, B.5, and B.6, these results allow us to observe that the behavior of the R^2 and $\mathcal{F}$ statistic does indeed depend on the sum $\overline{d_x} + d_y$. When the value of this sum is lower than $\frac{1}{2}$ the R^2 takes lower values than in the other cases and the $\mathcal{F}$ statistic is relatively stable. When the value of this sum is instead greater than or equal to $\frac{1}{2}$, the R^2 takes higher values and the $\mathcal{F}$ appears to grow as the sample size increases. These results are in line with those of the theorem.

---

[1]For these simulations, Table B.3 shows the specific values of d_z for each variable in each case.

Chapter 3

Concluding remarks

We studied the asymptotic and the finite-sample behavior of the OLS-estimated multivariate regression with an arbitrary finite number of regressors and a constant term, where all the variables are independent stationary fractionally integrated processes. Our findings are in line with what is already established in the literature. In particular, the asymptotic behavior of the estimates and their associated t-ratios does not depend on the number of regressors in the specification, but rather on the persistence of the processes that generated the regressand and the particular regressor series. Hence, when the variables behave as stationary long-memory processes, inference drawn from the t-ratios or the $\mathcal{F}$ joint test can be unreliable. Moreover, Monte Carlo simulations confirm our asymptotic results and reveal that the phenomenon of spurious regression becomes more acute as the persistence of the variables rises. Our findings support the conjecture that spurious effects are attributable to persistence rather than nonstationarity.

Appendix A

Proof of the theorem

Since we rely on the results of TC's Lemma 1, we reproduce the relevant ones here as Lemma A1:

Lemma A1: *Let Assumption 1 hold. Then, as $T \to \infty$:*

1. $\sum_{t=1}^{T} z_t = O_p\left(T^{\frac{1}{2}+d_z}\right)$.

2. $\sum_{t=1}^{T} z_t^2 = O_p(T)$.

3. $\sum_{t=1}^{T} x_{i,t} y_t = \begin{cases} O_p\left(T^{\frac{1}{2}}\right) & \text{if } 0 < d_{x_i} + d_y < \frac{1}{2}, \\ O_p\left(\sqrt{T \ln T}\right) & \text{if } d_{x_i} + d_y = \frac{1}{2}, \\ O_p\left(T^{d_{x_i}+d_y}\right) & \text{if } \frac{1}{2} < d_{x_i} + d_y < 1, \end{cases}$

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

4. $\sum_{t=1}^{T} x_{i,t} x_{j,t} = \begin{cases} O_p\left(T^{\frac{1}{2}}\right) & \text{if } 0 < d_{x_i} + d_{x_j} < \frac{1}{2}, \\ O_p\left(\sqrt{T \ln T}\right) & \text{if } d_{x_i} + d_{x_j} = \frac{1}{2}, \\ O_p\left(T^{d_{x_i}+d_{x_j}}\right) & \text{if } \frac{1}{2} < d_{x_i} + d_{x_j} < 1, \end{cases}$

   for $i, j = 1, \ldots, k$ and $i \neq j$.

Here, $z = y, x_1, \ldots, x_k$.

To show items 1 and 2 we make use of the OLS estimator formula:

$$(X'X)\hat{\beta} = X'Y, \tag{A.1}$$

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

Expression (A.1) can be alternatively written as

$$
\begin{bmatrix}
T & \sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{2,t} & \cdots & \sum_{t=1}^{T} x_{k,t} \\
\sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{1,t}^2 & \sum_{t=1}^{T} x_{1,t} x_{2,t} & \cdots & \sum_{t=1}^{T} x_{1,t} x_{k,t} \\
\sum_{t=1}^{T} x_{2,t} & \sum_{t=1}^{T} x_{1,t} x_{2,t} & \sum_{t=1}^{T} x_{2,t}^2 & \cdots & \sum_{t=1}^{T} x_{2,t} x_{k,t} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sum_{t=1}^{T} x_{k,t} & \sum_{t=1}^{T} x_{1,t} x_{k,t} & \sum_{t=1}^{T} x_{2,t} x_{k,t} & \cdots & \sum_{t=1}^{T} x_{k,t}^2
\end{bmatrix}
\begin{bmatrix}
\hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k
\end{bmatrix}
=
\begin{bmatrix}
\sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_{1,t} y_t \\ \sum_{t=1}^{T} x_{2,t} y_t \\ \vdots \\ \sum_{t=1}^{T} x_{k,t} y_t
\end{bmatrix}.
\tag{A.2}
$$

From (A.2), we observe that

$$
\hat{\beta}_0 = T^{-1} \left(\sum_{t=1}^{T} y_t - \sum_{n=1}^{k} \hat{\beta}_n \sum_{t=1}^{T} x_{n,t} \right).
\tag{A.3}
$$

As for $\hat{\beta}_n$, for $n = 1, \ldots, k$, by Cramer's rule we have that

$$
\hat{\beta}_n = \frac{\Delta_n}{\Delta},
\tag{A.4}
$$

where

$$
\Delta = \left| X'X \right|
$$

$$
=
\begin{vmatrix}
T & \sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{2,t} & \cdots & \sum_{t=1}^{T} x_{k,t} \\
\sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{1,t}^2 & \sum_{t=1}^{T} x_{1,t} x_{2,t} & \cdots & \sum_{t=1}^{T} x_{1,t} x_{k,t} \\
\sum_{t=1}^{T} x_{2,t} & \sum_{t=1}^{T} x_{1,t} x_{2,t} & \sum_{t=1}^{T} x_{2,t}^2 & \cdots & \sum_{t=1}^{T} x_{2,t} x_{k,t} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sum_{t=1}^{T} x_{k,t} & \sum_{t=1}^{T} x_{1,t} x_{k,t} & \sum_{t=1}^{T} x_{2,t} x_{k,t} & \cdots & \sum_{t=1}^{T} x_{k,t}^2
\end{vmatrix},
\tag{A.5}
$$

and

$$\Delta_n = \begin{vmatrix} T & \sum_{t=1}^{T} x_{1,t} & \cdots & \sum_{t=1}^{T} x_{n-1,t} & \sum_{t=1}^{T} y_t & \sum_{t=1}^{T} x_{n+1,t} & \cdots & \sum_{t=1}^{T} x_{k,t} \\ \sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{1,t}^2 & \cdots & \sum_{t=1}^{T} x_{1,t}x_{n-1,t} & \sum_{t=1}^{T} x_{1,t}y_t & \sum_{t=1}^{T} x_{1,t}x_{n+1,t} & \cdots & \sum_{t=1}^{T} x_{1,t}x_{k,t} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ \sum_{t=1}^{T} x_{n-1,t} & \sum_{t=1}^{T} x_{1,t}x_{n-1,t} & \cdots & \sum_{t=1}^{T} x_{n-1,t}^2 & \sum_{t=1}^{T} x_{n-1,t}y_t & \sum_{t=1}^{T} x_{n-1,t}x_{n+1,t} & \cdots & \sum_{t=1}^{T} x_{n-1,t}x_{k,t} \\ \sum_{t=1}^{T} x_{n,t} & \sum_{t=1}^{T} x_{1,t}x_{n,t} & \cdots & \sum_{t=1}^{T} x_{n-1,t}x_{n,t} & \sum_{t=1}^{T} x_{n,t}y_t & \sum_{t=1}^{T} x_{n,t}x_{n+1,t} & \cdots & \sum_{t=1}^{T} x_{n,t}x_{k,t} \\ \sum_{t=1}^{T} x_{n+1,t} & \sum_{t=1}^{T} x_{1,t}x_{n+1,t} & \cdots & \sum_{t=1}^{T} x_{n-1,t}x_{n+1,t} & \sum_{t=1}^{T} x_{n+1,t}y_t & \sum_{t=1}^{T} x_{n+1,t}^2 & \cdots & \sum_{t=1}^{T} x_{n+1,t}x_{k,t} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^{T} x_{k,t} & \sum_{t=1}^{T} x_{1,t}x_{k,t} & \cdots & \sum_{t=1}^{T} x_{n-1,t}x_{k,t} & \sum_{t=1}^{T} x_{k,t}y_t & \sum_{t=1}^{T} x_{n+1,t}x_{k,t} & \cdots & \sum_{t=1}^{T} x_{k,t}^2 \end{vmatrix}.$$

(A.6)

To find the order in probability of $\hat{\beta}_n$ we shall first determine the order in probability of Δ and Δ_n. To do so, we triangulate the matrices whose determinant is equal to Δ and Δ_n, as the determinant of any triangular matrix is merely the product of the elements along the main diagonal. Hence, we seek to find the order in probability of said elements after triangulation. We find that the order in probability of the elements along the main diagonal after triangulation is the same as that prior to triangulation for both Δ and Δ_n. Note that the operations required to carry out triangulation of the matrices, the addition of rows multiplied by a scalar to other rows, leave the determinant of the matrix unchanged.

Let us first look at (A.5), the denominator in expression (A.4). If we add the first row multiplied by scalar $-\frac{\sum_{t=1}^{T} x_{i-1,t}}{T}$ to the i-th row, for $i = 2, \ldots, k+1$, we arrive at the following:

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

To continue with triangulation of the matrix whose determinant is equal to Δ, we would

14

next add the second row multiplied by scalar $-\frac{T\sum_{t=1}^{T} x_{1,t}x_{i-1,t} - \sum_{t=1}^{T} x_{1,t} \sum_{t=1}^{T} x_{i-1,t}}{T\sum_{t=1}^{T} x_{1,t}^2 - \left(\sum_{t=1}^{T} x_{1,t}\right)^2}$ to the i-th row, for $i = 3, \ldots, k$. This would render all terms in the second column, starting from the third row, 0. Should we continue this process for a total of $k+1$ steps analogous to the two described above,[1] 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 $m = 1, \ldots, k+1$,

$$\Delta[i,j,m] = \begin{cases} \Delta[i,j,m-1] - \frac{\Delta[i,m,m-1]}{\Delta[m,m,m-1]}\Delta[m,j,m-1] & \text{if } m \le i-1, \\ \Delta[i,j,m-1] & \text{if } m > i-1, \end{cases} \tag{A.7}$$

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

From (A.7) we deduce that

$$\Delta[i,j,m] = \sum_{t=1}^{T} x_{i-1,t}x_{j-1,t} - \sum_{r=1}^{\min\{m,i-1\}} \frac{\Delta[i,r,r-1]\Delta[r,j,r-1]}{\Delta[r,r,r-1]}. \tag{A.8}$$

More specifically, after triangulation,

$$\Delta[i,j,k+1] = \begin{cases} \sum_{t=1}^{T} x_{i-1,t}x_{j-1,t} - \sum_{r=1}^{i-1} \frac{\Delta[i,r,r-1]\Delta[r,j,r-1]}{\Delta[r,r,r-1]} & \text{if } j \ge i, \\ 0 & \text{if } j < i. \end{cases} \tag{A.9}$$

At every step in the the triangulation process and, consequently, after triangulation of the matrix whose determinant Δ is equal to, the elements along the main diagonal are $O_p(T)$ as $T \to \infty$:

Lemma A2: *Let Assumption 1 hold. Then, as $T \to \infty$, for $a = 1, \ldots, k+1$ and all m:*

$$\Delta[a,a,m] = O_p(T).$$

---

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

Proof: Note from Lemma A1 that prior to triangulation the elements along the main diagonal are $O_p(T)$. As triangulation consists solely in the addition of other terms to the original terms, it follows that after triangulation (i.e., once we are left with an upper triangular matrix), all elements of the main diagonal must be an order in probability greater than or equal to $O_p(T)$. If all terms added throughout the triangulation process are an order in probability lower than or equal to $O_p(T)$, the element will remain $O_p(T)$ after triangulation. If, on the other hand, any of the terms added is an order in probability strictly greater than $O_p(T)$, then after triangulation the element will too be an order in probability strictly greater than $O_p(T)$. Note further from Lemma A1 that, before triangulation, all terms outside of the main diagonal are an order in probability strictly lower than $O_p(T)$.

From (A.8) observe that in each term added successively throughout the triangulation process, $-\frac{\Delta[i,r,r-1]\Delta[r,j,r-1]}{\Delta[r,r,r-1]}$ for $r = 1, \ldots, i-1$, the terms in the numerator $\Delta[i,r,r-1]$ and $\Delta[r,j,r-1]$ are always elements from outside the main diagonal,[2] whereas the term in the denominator $\Delta[r,r,r-1]$ is always an element from the main diagonal.

Then, observe from (A.9) that for the a-th element along the main diagonal to be an order in probability strictly greater than $O_p(T)$ after triangulation, it it must be the case that $\frac{\Delta[a,r_1,r_1-1]\Delta[r_1,a,r_1-1]}{\Delta[r_1,r_1,r_1-1]}$ be an order in probability strictly greater than $O_p(T)$ for some $r_1 \in \{1, \ldots, a-1\}$. Given that $\Delta[r_1,r_1,r_1-1]$ is known to be an order in probability greater than or equal to $O_p(T)$, a necessary, though not sufficient, condition for this to occur is that either $\Delta[a,r_1,r_1-1]$ or $\Delta[r_1,a,r_1-1]$, both elements from outside the main diagonal distinct from 0, be an order in probability greater than $O_p(T)$. Likewise, observe from (A.8) that for either of these terms to be an order in probability greater than $O_p(T)$, a necessary condition, in turn, is that either $\Delta[a,r_2,r_2-1]$, $\Delta[r_2,r_1,r_2-1]$, $\Delta[r_1,r_2,r_2-1]$ or $\Delta[r_2,a,r_2-1]$ be an order in probability greater than $O_p(T)$ for some $r_2 \in \{1, \ldots, r_1-1\}$. Note that $r_2 < r_1$; this is, the necessary condition for the elements outside the main diagonal to be an order in probability greater than $O_p(T)$ at a certain step in the triangulation process is that certain elements, also from outside the main

---

[2]In particular, these are always distinct from 0 when added to elements (i,j) for which $j \geq i$.

16

diagonal, be themselves greater than $O_p(T)$ at a preceding step in the triangulation process.

Should we continue the same argument iteratively, we would eventually arrive at the following: for any element of the main diagonal (a,a) to be an order in probability strictly greater than $O_p(T)$ as $T \to \infty$ after triangulation, at least one element from outside the main diagonal (i,j) such that

$$(i,j) \in \{(i,1) \mid i \in \{2,\ldots,a\}\} \bigcup \{(1,j) \mid j \in \{2,\ldots,a\}\}$$

must have originally (i.e., at $m = 0$) been an order in probability strictly greater than $O_p(T)$. Yet we know that originally all elements outside the diagonal were strictly less than $O_p(T)$.

We have thus shown that after triangulation all elements of the main diagonal are $O_p(T)$. Given that the determinant of any triangular matrix is equal to the product of the elements along the main diagonal, it follows from Lemma A2 that, as $T \to \infty$,

$$\Delta = O_p\left(T^{k+1}\right). \tag{A.10}$$

We now turn our attention to (A.6), the numerator in expression (A.4). To simplify the argument put forward, we assume without loss of generality that $n = k$. As before, we seek to turn the matrix whose determinant is equal to Δ_k into an upper triangular matrix. Note that Δ_k differs from Δ in that the original elements of the last column are not $\sum_{t=1}^T x_{i-1,t} x_{k,t}$ but rather $\sum_{t=1}^T x_{i-1,t} y_t$ for each row $i = 1,\ldots,k+1$. Note also from Lemma A1 that the last element along the main diagonal is originally an order in probability strictly lower than $O_p(T)$ as $T \to \infty$. Thus, the argument outlined previously no longer applies in its entirety.

Proceeding as we did before, we let $\Delta_k[i,j,m]$ denote the element at the i-th row and j-th column at step m in the triangulation process of the matrix whose determinant is equal to Δ_k. Then, for $m = 1,\ldots,k+1$,

$$\Delta_k[i,j,m] = \begin{cases} \Delta_k[i,j,m-1] - \frac{\Delta_k[i,m,m-1]}{\Delta_k[m,m,m-1]}\Delta_k[m,j,m-1] & \text{if } m \leq i-1, \\ \Delta_k[i,j,m-1] & \text{if } i-1 < m, \end{cases} \tag{A.11}$$

17

where

$$\Delta_k[i,j,0] = \sum_{t=1}^{T} x_{i-1,t} z_{j-1,t},$$

and

$$z_{j-1} = \begin{cases} x_{j-1} & \text{if } j \neq k+1, \\ y_t & \text{if } j = k+1. \end{cases}$$

As before, from (A.11) it follows that

$$\Delta_k[i,j,m] = \sum_{t=1}^{T} x_{i-1,t} x_{j-1,t} - \sum_{r=1}^{\min\{m,i-1\}} \frac{\Delta_k[i,r,r-1]\Delta_k[r,j,r-1]}{\Delta_k[r,r,r-1]}. \tag{A.12}$$

Recall that we seek to find the order in probability of the elements along the main diagonal after triangulation. In view of the argument sketched previously for Δ, the first k elements of the main diagonal are always (i.e., at all steps in the triangulation process) $O_p(T)$:

Lemma A3: *Let Assumption 1 hold. Then, as $T \to \infty$, for $a = 1,\ldots,k$ and all m:*

$$\Delta_k[a,a,m] = O_p(T).$$

Therefore, the only order in probability we require to find is that of the last element of the main diagonal. As regards this element, we have that once triangulation is complete,

$$\Delta[k+1,k+1,k+1] = \sum_{t=1}^{T} x_{k,t} y_t - \sum_{r=1}^{k} \frac{\Delta[k+1,r,r-1]\Delta[r,k+1,r-1]}{\Delta[r,r,r-1]}. \tag{A.13}$$

Observe from (A.13) that the denominator of the terms added to element $(k+1,k+1)$ throughout the triangulation process is always one of the first k elements of the main diagonal, which are known to be $O_p(T)$ at all steps. The numerator of the added terms is comprised of elements from outside the main diagonal distinct from 0 at differing steps in the triangulation

process. To these elements, in turn, are added terms composed of the product of other two elements from outside of the main diagonal divided by an element from the main diagonal, as seen in (A.12). The order in probability of the elements outside of the main diagonal when added to element $(k+1, k+1)$ is established in the following lemma, in which we posit that all terms added throughout the triangulation process are an order in probability equal to or lower than that of the original term, $\sum_{t=1}^{T} x_{i-1,t} z_{j-1,t}$:

Lemma A4: *Let Assumption 1 hold. Then, for all $i \neq j$:*

$$
\Delta[i, j, \min\{i-1, j-1\}] = \begin{cases} O_p(T^{\frac{1}{2}}) & \text{for } 0 < d_{x_k} + d_{z_{j-1}} < \frac{1}{2}, \\ O_p\left[(T \ln T)^{\frac{1}{2}}\right] & \text{for } d_{x_k} + d_{z_{j-1}} = \frac{1}{2}, \\ O_p(T^{d_{x_k} + d_{z_{j-1}}}) & \text{for } \frac{1}{2} < d_{x_k} + d_{z_{j-1}} < 1. \end{cases}
$$

Proof: The above occurs if and only if at every step in the triangulation process up to step $m = \min\{i-1, j-1\}$ the term added is an order in probability lower than or equal to the order in probability of the original term, which we shall prove by induction.

At the first step in the triangulation process, elements (i, j) for which $i > 1$ and $j > 1$ are

$$
\Delta_k[i, j, 1] = \sum_{t=1}^{T} x_{i-1,t} z_{j-1,t} - \frac{\sum_{t=1}^{T} x_{i-1,t} \sum_{t=1}^{T} z_{j-1,t}}{T}.
$$

From Lemma A1 is plain to see that the second term, added at step 1 in the triangulation process, is not an order in probability greater than that of the original term.

At the second step in the triangulation process, elements for which $i > 2$ and $j > 2$ are

$$
\Delta_k[i, j, 2] = \Delta_k[i, j, 1] - \frac{\left(T \sum_{t=1}^{T} x_{1,t} x_{i-1,t} - \sum_{t=1}^{T} x_{1,t} \sum_{t=1}^{T} x_{i-1,t}\right)\left(T \sum_{t=1}^{T} x_{1,t} z_{j-1,t} - \sum_{t=1}^{T} x_{1,t} \sum_{t=1}^{T} z_{j-1,t}\right)}{T\left[T \sum_{t=1}^{T} x_{1,t}^2 - \left(\sum_{t=1}^{T} x_{1,t}\right)^2\right]}.
$$

Observe from Lemma A1 that the term added at step 2 takes its highest order in probability when $\frac{1}{2} < d_{x_1} + d_{x_i}$ and $\frac{1}{2} < d_{x_1} + d_{z_j}$, in which case it is $O_p\left(T^{2d_{x_1} + d_{x_i} + d_{z_j} - 1}\right)$. Conversely, the original term takes its lowest order in probability when $d_{x_{i-1}} + d_{z_{j-1}} < \frac{1}{2}$, in which case it is $O_p\left(T^{\frac{1}{2}}\right)$. Given that, by assumption, $d_{x_{i-1}} + d_{z_{j-1}} < \frac{1}{2}$ and $d_{x_1} < \frac{1}{2}$, it is clear that $2d_{x_1} + d_{x_{i-1}} + d_{z_{j-1}} - 1 < \frac{1}{2}$.

In all other possible cases, either the order in probability of the original term must rise or the order in probability of the added term must fall. Hence, once again, the term added is not an order in probability greater than the original term.

At the m-th step we have that, for $i > m$ and $j > m$,

$$\Delta_k[i,j,m] = \Delta_k[i,j,m] - \frac{\Delta_k[i,m,m-1]\Delta_k[m,j,m-1]}{\Delta_k[m,m,m-1]}. \tag{A.14}$$

Suppose now that, for $i > m-1$ and $j > m-1$,

$$\Delta_k[i,j,m-1] = \begin{cases} O_p(T^{\frac{1}{2}}) & \text{for } 0 < d_{x_{i-1}} + d_{z_{j-1}} < \frac{1}{2}, \\ O_p\left[(T\ln T)^{\frac{1}{2}}\right] & \text{for } d_{x_{i-1}} + d_{z_{j-1}} = \frac{1}{2}, \\ O_p\left(T^{d_{x_{i-1}}+d_{z_{j-1}}}\right) & \text{for } \frac{1}{2} < d_{z_{i-1}} + d_{z_{j-1}} < 1, \end{cases} \tag{A.15}$$

This is, the original term is an order in probability greater than or equal to those added successively throughout the triangulation process up to step $m-1$. Then, substituting (A.15) in (A.14) it can be shown in a similar manner to what we did at the second step that the term added at the m-th step is an order in probability strictly lower than the order in probability of the original term in element (i,j).

We have thus shown that, if at the preceding step in the triangulation process, the proposition holds,[3] then the proposition holds at the current step. Additionally, we had already shown that the proposition holds at the second step. Therefore, we have effectively shown that it holds at every step.

As for the last element of the main diagonal, which is $\sum_{t=1}^{T} x_{k,t} y_t$ prior to triangulation, the following lemma establishes its order in probability after triangulation:

---

[3]The proposition being that the original term is an order in probability equal to or greater than that of the term added at the step.

Lemma A5: *Let Assumption 1 hold. Then,*

$$
\Delta_k[k+1,k+1,k+1] = \begin{cases} O_p(T^{\frac{1}{2}}) & \text{for } 0 < d_{x_k} + d_y < \frac{1}{2}, \\ O_p\left[(T\ln T)^{\frac{1}{2}}\right] & \text{for } d_{x_k} + d_y = \frac{1}{2}, \\ O_p(T^{d_{x_k}+d_y}) & \text{for } \frac{1}{2} < d_{x_k} + d_y < 1. \end{cases}
$$

The proof is straightforward substituting the results of Lemmas A3 and A4 in (A.13).

Hence, given that the determinant of any triangular matrix is the product of the elements of the main diagonal, it follows from Lemmas A3 and A5 that, as $T \to \infty$,

$$
\Delta_n = \begin{cases} O_p(T^{k+\frac{1}{2}}) & \text{for } 0 < d_{x_n} + d_y < \frac{1}{2}, \\ O_p\left[(T^{k+1}\ln T)^{\frac{1}{2}}\right] & \text{for } d_{x_n} + d_y = \frac{1}{2}, \\ O_p(T^{d_{x_n}+d_y+k}) & \text{for } \frac{1}{2} < d_{x_n} + d_y < 1. \end{cases} \tag{A.16}
$$

Substituting results (A.10) and (A.16) in (A.4) concludes the proof of item 1.

As for $\hat{\beta}_0$, note that now all the terms in eq. (A.3) have known orders in probability. Denote by n_1, n_2, and n_3 the subsets of $\{1,\ldots,k\}$ for which $d_{x_n} + d_y < \frac{1}{2}$, $d_{x_n} + d_y = \frac{1}{2}$, and $\frac{1}{2} < d_{x_n} + d_y$, respectively. Then, using Lemma A1, (A.3) can be asymptotically reduced to

$$
\hat{\beta}_0 = T^{-1}\left[O_p\left(T^{d_y+\frac{1}{2}}\right) - \sum_{n\in n_1} O_p\left(T^{d_{x_n}}\right) - \sum_{n\in n_2} O_p\left(T^{d_{x_n}}(\ln T)^{\frac{1}{2}}\right) - \sum_{n\in n_3} O_p\left(T^{2d_{x_n}+d_y-\frac{1}{2}}\right)\right].
$$

Given that $d_y > 0$ and $d_{x_n} < \frac{1}{2}$ for all n, it is clear that the first term dominates all others, which concludes the proof of item 2.

To prove item 3 the following formula is used:

$$s^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t} - \ldots - \hat{\beta}_k x_{k,t} \right)^2$$

$$= \frac{1}{T} \left(\sum_{t=1}^{T} y_t^2 - 2\hat{\beta}_0 \sum_{t=1}^{T} y_t + T\hat{\beta}_0^2 - 2 \sum_{n=1}^{k} \hat{\beta}_n \sum_{t=1}^{T} x_{n,t} y_t + \sum_{n=1}^{k} \hat{\beta}_n^2 \sum_{t=1}^{T} x_{n,t}^2 + \right.$$

$$\left. 2\hat{\beta}_0 \sum_{n=1}^{k} \hat{\beta}_n \sum_{t=1}^{T} x_{n,t} + 2 \sum_{n=1}^{k} \sum_{m>n} \hat{\beta}_n \hat{\beta}_m \sum_{t=1}^{T} x_{n,t} x_{m,t} \right).$$

Observe from Lemma A1 and items 1 and 2 of the theorem that term $\sum_{t=1}^{T} y_t^2$ is $O_p(T)$, in contrast to all other terms, which are an order in probability strictly lower. Thus, the following is true as $T \to \infty$:

$$s^2 = \frac{1}{T} \sum_{t=1}^{T} y_t^2$$

$$\xrightarrow{P} \gamma_y(0).$$

It also follows from the above proof that

$$\sum_{t=1}^{T} \hat{u}_t^2 = O_p(T), \tag{A.17}$$

a result we shall use later.

To show item 4 recall the formula for the estimator of the variance-covariance matrix of the estimators:

$$\widehat{\text{Var}} \left(\hat{\beta} \right) = s^2 \left(X'X \right)^{-1}, \tag{A.18}$$

22

from which it follows that

$$T \widehat{\mathrm{Var}}\left(\hat{\beta}\right)\left(\frac{X'X}{T}\right) = s^2 I.$$

Note that, in the limit, matrix $\left(\frac{X'X}{T}\right)$ is diagonal since all elements along the main diagonal of $X'X$ are $O_p(T)$ and all elements outside of the main diagonal are an order in probability strictly lower (Lemma A1). Further, from (A.18) we have that

$$T \widehat{\mathrm{Var}}\left(\hat{\beta}\right) = s^2 \left(\frac{X'X}{T}\right)^{-1}.$$

Given that the inverse of a diagonal matrix is also diagonal itself, it must be the case that matrix $T \widehat{Var}\left(\hat{\beta}\right)$ is asymptotically diagonal as well. Hence, it is true that, in the limit,

$$\mathrm{diag}\left[T \widehat{\mathrm{Var}}\left(\hat{\beta}\right)\right] \bullet \mathrm{diag}\left(\frac{X'X}{T}\right) = \mathrm{diag}\left\{s^2 I\right\},$$

where "$\bullet$" denotes element-by-element multiplication. In other words,

$$\left[T \begin{bmatrix} s^2_{\hat{\beta}_0} \\ s^2_{\hat{\beta}_1} \\ \vdots \\ s^2_{\hat{\beta}_k} \end{bmatrix}\right] \bullet \left[\frac{1}{T} \begin{bmatrix} T \\ \sum_{t=1}^{T} x^2_{1t} \\ \vdots \\ \sum_{t=1}^{T} x^2_{kt} \end{bmatrix}\right] = \begin{bmatrix} s^2 \\ s^2 \\ \vdots \\ s^2 \end{bmatrix},$$

which, using item 3 of the theorem, becomes

$$\begin{bmatrix} s^2_{\hat{\beta}_0} \\ s^2_{\hat{\beta}_1} \\ \vdots \\ s^2_{\hat{\beta}_k} \end{bmatrix} \bullet \begin{bmatrix} T \\ T\gamma_{x_1}(0) \\ \vdots \\ T\gamma_{x_k}(0) \end{bmatrix} = \begin{bmatrix} \gamma_y(0) \\ \gamma_y(0) \\ \vdots \\ \gamma_y(0) \end{bmatrix}.$$

It follows that

$$T s_{\hat{\beta}_0}^2 \xrightarrow{P} \gamma_y(0),$$

and

$$T s_{\hat{\beta}_n}^2 \xrightarrow{P} \frac{\gamma_y(0)}{\gamma_{x_n}(0)}, \quad \text{for } n = 1, \ldots, k,$$

which concludes the proof of item 4.

Proof of items 5 and 6 is straightforward using the results of items (1-4) and the formula $t_{\beta_n} = \frac{\hat{\beta}_n}{s_{\hat{\beta}_n}}$ for $n = 0, 1, \ldots, k$.

We prove item 7 using the following formula:

$$R^2 = \frac{\sum_{t=1}^{T}(y_t - \bar{y})^2 - \sum_{t=1}^{T} \hat{u}_t^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}. \tag{A.19}$$

Using (A.3), the numerator of (A.19) can be reduced to the following in the limit:

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 - \sum_{t=1}^{T} \hat{u}_t^2 = \frac{1}{T}\left(\sum_{n=1}^{k} \hat{\beta}_n \sum_{t=1}^{T} x_{n,t}\right)^2 + 2\sum_{n=1}^{k} \hat{\beta}_n \sum_{t=1}^{T} x_{n,t} y_t - \sum_{n=1}^{k} \hat{\beta}_n^2 \sum_{t=1}^{T} x_{n,t}^2 -$$
$$2\frac{1}{T}\sum_{t=1}^{T} y_t \sum_{n=1}^{k} \hat{\beta}_n \sum_{t=1}^{T} x_{n,t} - 2\sum_{n=1}^{k}\sum_{m>n} \hat{\beta}_n \hat{\beta}_m \sum_{t=1}^{T} x_{n,t} x_{m,t}. \tag{A.20}$$

From (A.20) and Lemma A1, it follows that, as $T \to \infty$,

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 - \sum_{t=1}^{T} \hat{u}_t^2 = \begin{cases} O_p(1) & \text{for } \bar{d}_x + d_y < \frac{1}{2}, \\ O_p(\ln T) & \text{for } \bar{d}_x + d_y = \frac{1}{2}, \\ O_p(T^{2\bar{d}_x + 2d_y - 1}) & \text{for } \bar{d}_x + d_y > \frac{1}{2}, \end{cases} \tag{A.21}$$

where $\bar{d}_x = \max\{d_{x_1}, d_{x_2}, \ldots, d_{x_k}\}$.

As for the denominator in (A.19), it can shown rather simply using the results of Lemma A1

24

that

$$\sum_{t=1}^{T} (y_t - \bar{y})^2 = O_p(T). \tag{A.22}$$

Substituting (A.22) and (A.21) in (A.19) concludes the proof of item 7.

To show item 8 recall first the formula for the Wald $\mathcal{F}$ statistic associated to the joint significance test:

$$
\begin{aligned}
\mathcal{F} &= \frac{\left[\sum_{t=1}^{T} (y_t - \bar{y})^2 - \sum_{t=1}^{T} \hat{u}_t^2\right]/k}{\sum_{t=1}^{T} \hat{u}_t^2 / [T - (k+1)]} \\
&= \frac{[T - (k+1)] R^2}{k \sum_{t=1}^{T} \hat{u}_t^2 / \sum_{t=1}^{T} (y_t - \bar{y})^2}.
\end{aligned} \tag{A.23}
$$

Substituting (A.17) and (A.22) along with item 7 in (A.23) concludes the proof of item 8.

Proof of item 9 comes from the fact that

$$
\begin{aligned}
\mathcal{DW} &= \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2} \\
&= \frac{\sum_{t=2}^{T} \hat{u}_t^2 + \sum_{t=2}^{T} \hat{u}_{t-1}^2 - 2 \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2}.
\end{aligned}
$$

Adding expression $\frac{\hat{u}_1^2 + \hat{u}_T^2}{\sum_{t=1}^{T} \hat{u}_t^2}$, which is negligible as $T \to \infty$, we obtain

$$\mathcal{DW} \approx 2 - 2 \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2}. \tag{A.24}$$

We have that

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

Using Lemma A1 and items 1 and 2 of the theorem, we discard terms inside the brackets known to be $o_p(T)$. Further, given that the underlying processes are independent of one another by Assumption 1, expressions $\frac{1}{T}\sum_{t=2}^{T}x_{n,t}y_{t-1}$, $\frac{1}{T}\sum_{t=2}^{T}x_{n,t-1}y_t$, $\frac{1}{T}\sum_{t=2}^{T}x_{n,t}x_{m,t-1}$, and $\frac{1}{T}\sum_{t=2}^{T}x_{n,t-1}x_{m,t}$ all converge in probability to 0. The only remaining term is $\sum_{t=2}^{T}y_t y_{t-1}$. Therefore,

$$\frac{1}{T}\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}\xrightarrow{P}\gamma_y(1). \tag{A.25}$$

Substituting (A.25) and item 3 of the theorem in (A.24) we find

$$\mathcal{DW}\rightarrow 2-2\rho_y(1). \tag{A.26}$$

Substituting (2.1) in (A.26) concludes the proof of item 9.

Appendix B

Tables

Table B.1: Rejection rate (RR) of t-statistics, Average Value (AV) of the $\mathcal{DW}$, RR of the $\mathcal{F}$ joint test, and AV of the R^2, within an OLS-estimated regression: equal value of d for each series.

T	Spec.	β_i	d 1/5				1/4				3/10			
			RR_t*	$\mathcal{DW}$	R^2	$RR_{\mathcal{F}}$**	RR_t*	$\mathcal{DW}$	R^2	$RR_{\mathcal{F}}$**	RR_t*	$\mathcal{DW}$	R^2	$RR_{\mathcal{F}}$**
T=100	1	β_1	0.0680	1.6030	0.0221	0.0650	0.0880	1.4959	0.0253	0.0980	0.1200	1.4003	0.0290	0.1260
		β_2	0.0480				0.0680				0.1060			
	2	β_1	0.0670				0.0860				0.1080			
		β_2	0.0490	1.6153	0.0336	0.0690	0.0710	1.5114	0.0381	0.1070	0.0990	1.4196	0.0428	0.1390
		β_3	0.0680				0.0930				0.0990			
	3	β_1	0.0650				0.0800				0.1110			
		β_2	0.0500	1.6259	0.0448	0.0760	0.0750	1.5270	0.0513	0.1110	0.1000	1.4408	0.0574	0.1610
		β_3	0.0680				0.0860				0.0970			
		β_4	0.0720				0.0900				0.1160			
T=500	1	β_1	0.0940	1.5341	0.0052	0.0960	0.1390	1.3926	0.0070	0.1890	0.1930	1.2519	0.0092	0.2690
		β_2	0.0730				0.1550				0.1870			
	2	β_1	0.0940				0.1400				0.1890			
		β_2	0.0690	1.5372	0.0076	0.1040	0.1540	1.3976	0.0102	0.2220	0.1830	1.2604	0.0141	0.3450
		β_3	0.0790				0.1200				0.2060			
	3	β_1	0.0950				0.1320				0.1860			
		β_2	0.0690	1.5405	0.0102	0.1220	0.1500	1.4028	0.0135	0.2440	0.1870	1.2680	0.0185	0.4030
		β_3	0.0830				0.1160				0.1980			
		β_4	0.0810				0.1250				0.1850			
T=1,000	1	β_1	0.0840	1.5214	0.0025	0.0870	0.1120	1.3707	0.0034	0.1650	0.2380	1.2127	0.0052	0.3180
		β_2	0.0750				0.1410				0.2230			
	2	β_1	0.0840				0.1170				0.2400			
		β_2	0.0760	1.5231	0.0038	0.1050	0.1400	1.3738	0.0054	0.2340	0.2200	1.2172	0.0076	0.3910
		β_3	0.0840				0.1550				0.2190			
	3	β_1	0.0860				0.1160				0.2310			
		β_2	0.0750	1.5248	0.0052	0.1200	0.1450	1.3767	0.0071	0.2600	0.2200	1.2213	0.0100	0.4260
		β_3	0.0860				0.1550				0.2190			
		β_4	0.1050				0.1420				0.2140			
T=2,000	1	β_1	0.0910	1.5117	0.0014	0.1290	0.1400	1.3621	0.0017	0.1780	0.2130	1.1980	0.0027	0.3290
		β_2	0.1040				0.1440				0.2400			
	2	β_1	0.0930				0.1410				0.2120			
		β_2	0.1050	1.5126	0.0021	0.1450	0.1420	1.3637	0.0027	0.2270	0.2390	1.2007	0.0042	0.4140
		β_3	0.1000				0.1430				0.2470			
	3	β_1	0.0940				0.1390				0.2120			
		β_2	0.1040	1.5135	0.0028	0.1640	0.1400	1.3652	0.0036	0.2470	0.2390	1.2030	0.0055	0.4810
		β_3	0.1000				0.1430				0.2520			
		β_4	0.0980				0.1350				0.2340			

Here, *,**.- RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests, respectively. The parameter $0 < d < 1/2$ is the same for all the series. The nominal size of all tests is 5%. The number of replications is 10,000.

Table B.2: Rejection rate (RR) of t-statistics, Average Value (AV) of the $\mathcal{DW}$, RR of the $\mathcal{F}$ joint test, and AV of the R^2, within an OLS-estimated regression: equal value of d for each series.

T	Spec.	β_i	d: 7/20				d: 2/5				d: 9/20			
			$RR_t{*}$	$\mathcal{DW}$	R^2	$RR_{\mathcal{F}}{**}$	$RR_t{*}$	$\mathcal{DW}$	R^2	$RR_{\mathcal{F}}{**}$	$RR_t{*}$	$\mathcal{DW}$	R^2	$RR_{\mathcal{F}}{**}$
T=100	1	β_1	0.1450	1.2760	0.0354	0.1810	0.1990	1.1713	0.0460	0.2880	0.2540	1.0580	0.0566	0.3470
		β_2	0.1440				0.2140				0.2450			
	2	β_1	0.1380	1.3022	0.0525	0.2180	0.1980	1.2055	0.0662	0.3120	0.2370	1.0970	0.0806	0.4380
		β_2	0.1360				0.2070				0.2220			
		β_3	0.1350				0.1860				0.2330			
	3	β_1	0.1380	1.3293	0.0697	0.2490	0.1690	1.2397	0.0866	0.3690	0.2400	1.1346	0.1049	0.4850
		β_2	0.1280				0.1960				0.2200			
		β_3	0.1310				0.1790				0.2150			
		β_4	0.1410				0.1880				0.2260			
T=500	1	β_1	0.2820	1.0936	0.0135	0.4190	0.3960	0.9418	0.0214	0.5610	0.4900	0.7948	0.0318	0.6710
		β_2	0.3010				0.3690				0.4810			
	2	β_1	0.2740	1.1051	0.0201	0.5130	0.3910	0.9575	0.0313	0.6880	0.4680	0.8135	0.0449	0.7960
		β_2	0.2930				0.3650				0.4700			
		β_3	0.2900				0.3740				0.4560			
	3	β_1	0.2700	1.1160	0.0263	0.5810	0.3770	0.9729	0.0409	0.7520	0.4570	0.8341	0.0587	0.8550
		β_2	0.2860				0.3650				0.4690			
		β_3	0.2810				0.3750				0.4470			
		β_4	0.2590				0.3720				0.4450			
T=1,000	1	β_1	0.3120	1.0556	0.0077	0.4710	0.4500	0.8919	0.0124	0.6100	0.5250	0.7278	0.0189	0.7370
		β_2	0.3250				0.4120				0.5270			
	2	β_1	0.3090	1.0622	0.0115	0.5550	0.4480	0.9004	0.0179	0.7250	0.5120	0.7412	0.0289	0.8560
		β_2	0.3250				0.4030				0.5100			
		β_3	0.3110				0.3950				0.5510			
	3	β_1	0.3120	1.0689	0.0153	0.6500	0.4330	0.9102	0.0240	0.8000	0.5050	0.7524	0.0371	0.9160
		β_2	0.3180				0.4000				0.4970			
		β_3	0.3060				0.3960				0.5400			
		β_4	0.3270				0.4160				0.4980			
T=2,000	1	β_1	0.3410	1.0290	0.0043	0.4890	0.4420	0.8608	0.0069	0.6460	0.5780	0.6935	0.0111	0.7800
		β_2	0.3350				0.4690				0.5830			
	2	β_1	0.3420	1.0327	0.0064	0.5790	0.4450	0.8663	0.0104	0.7790	0.5750	0.7004	0.0162	0.8840
		β_2	0.3260				0.4680				0.5780			
		β_3	0.3240				0.4750				0.5560			
	3	β_1	0.3430	1.0363	0.0085	0.6740	0.4420	0.8712	0.0135	0.8530	0.5770	0.7072	0.0214	0.9340
		β_2	0.3350				0.4660				0.5660			
		β_3	0.3210				0.4670				0.5570			
		β_4	0.3460				0.4310				0.5510			

Here, *,**.- RR_t and $RR_{\mathcal{F}}$ account for rejection rate of the t-ratio and the $\mathcal{F}$ tests, respectively. The parameter $0 < d < 1/2$ is the same for all the series. The nominal size of all tests is 5%. The number of replications is 10,000.

Table B.3: Parametric setting of the Monte-Carlo experiments

Case	Variable	σ^2 under different variance	d
$\overline{d_x}+d_y<\frac{1}{2}$	y	2.00	0.25
	x_1	1.00	0.20
	x_2	0.75	0.15
	x_3	1.25	0.10
	x_4	1.50	0.07
$\overline{d_x}+d_y=\frac{1}{2}$	y	2.00	0.25
	x_1	1.00	0.25
	x_2	0.75	0.20
	x_3	1.25	0.15
	x_4	1.50	0.10
$\overline{d_x}+d_y>\frac{1}{2}$	y	2.00	0.25
	x_1	1.00	0.30
	x_2	0.75	0.20
	x_3	1.25	0.15
	x_4	1.50	0.10

Table B.4: Rejection rate (RR) of t-statistics, Average Value (AV) of the $\mathcal{DW}$, RR of the $\mathcal{F}$ joint test, and AV of the R^2, within a OLS-estimated regression: two regressors and different different values of d for each series.

DGP		Same variance ($\sigma^2=2$)				Different variance			
		Sample size (T)				Sample size (T)			
		100	500	1,000	2,000	100	500	1,000	2,000
$\overline{d_x}+d_y<\frac{1}{2}$	$RR_{t_{\beta_1}}$	0.0808	0.1000	0.1082	0.1182	0.0810	0.0949	0.1038	0.1159
	$RR_{t_{\beta_2}}$	0.0722	0.0796	0.0888	0.0924	0.0763	0.0812	0.0899	0.0924
	R^2	0.0241	0.0053	0.0028	0.0015	0.0245	0.0053	0.0028	0.0015
	$RR_{\mathcal{F}}$	0.0828	0.1104	0.1215	0.1239	0.0852	0.1007	0.1167	0.1302
	$\mathcal{DW}$	1.4987	1.4001	1.3792	1.3661	1.5012	1.4011	1.3795	1.3645
$\overline{d_x}+d_y=\frac{1}{2}$	$RR_{t_{\beta_1}}$	0.0906	0.1189	0.1367	0.1530	0.0890	0.1228	0.1357	0.1525
	$RR_{t_{\beta_2}}$	0.0805	0.0947	0.1091	0.1161	0.0841	0.1023	0.1070	0.1155
	R^2	0.0256	0.0060	0.0032	0.0017	0.0256	0.0061	0.0032	0.0017
	$RR_{\mathcal{F}}$	0.0927	0.1338	0.1564	0.1799	0.0980	0.1408	0.1580	0.1741
	$\mathcal{DW}$	1.4992	1.4030	1.3802	1.3647	1.5024	1.4028	1.3803	1.3659
$\overline{d_x}+d_y>\frac{1}{2}$	$RR_{t_{\beta_1}}$	0.1016	0.1482	0.1731	0.1908	0.0977	0.1447	0.1602	0.1908
	$RR_{t_{\beta_2}}$	0.0849	0.0977	0.1059	0.1157	0.0771	0.0968	0.1040	0.1163
	R^2	0.0266	0.0064	0.0036	0.0019	0.0260	0.0064	0.0034	0.0019
	$RR_{\mathcal{F}}$	0.1063	0.1561	0.1844	0.2041	0.1009	0.1541	0.1732	0.2070
	$\mathcal{DW}$	1.5104	1.4047	1.3809	1.3663	1.5090	1.4014	1.3821	1.3661

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

Table B.5: Rejection rate (RR) of t-statistics, Average Value (AV) of the $\mathcal{DW}$, RR of the $\mathcal{F}$ joint test, and AV of the R^2, within a OLS-estimated regression: three regressors and different different values of d for each series.

DGP		Same variance ($\sigma^2 = 2$) Sample size (T)				Different variance Sample size (T)			
		100	500	1,000	2,000	100	500	1000	2,000
$\overline{d_x} + d_y < \frac{1}{2}$	$RR_{t_{\beta_1}}$	0.0798	0.0993	0.1072	0.1113	0.0784	0.0983	0.1119	0.1157
	$RR_{t_{\beta_2}}$	0.0775	0.0851	0.0876	0.0890	0.0695	0.0863	0.0886	0.0925
	$RR_{t_{\beta_3}}$	0.0642	0.0706	0.0686	0.0742	0.0683	0.0701	0.0713	0.0760
	R^2	0.0350	0.0077	0.0039	0.0020	0.0352	0.0077	0.0040	0.0021
	$RR_{\mathcal{F}}$	0.0804	0.1074	0.1102	0.1238	0.0809	0.1047	0.1203	0.1289
	$\mathcal{DW}$	1.5087	1.4027	1.3793	1.3658	1.5027	1.4043	1.3797	1.3644
$\overline{d_x} + d_y = \frac{1}{2}$	$RR_{t_{\beta_1}}$	0.0936	0.1186	0.1332	0.1464	0.0892	0.1138	0.1292	0.1466
	$RR_{t_{\beta_2}}$	0.0838	0.0981	0.1092	0.1197	0.0827	0.0977	0.1090	0.1115
	$RR_{t_{\beta_3}}$	0.0714	0.0820	0.0866	0.0909	0.0703	0.0816	0.0908	0.0923
	R^2	0.0374	0.0085	0.0045	0.0024	0.0372	0.0084	0.0045	0.0024
	$RR_{\mathcal{F}}$	0.0936	0.1381	0.1579	0.1766	0.0982	0.1366	0.1610	0.1719
	$\mathcal{DW}$	1.5125	1.4056	1.3823	1.3653	1.5132	1.4061	1.3823	1.3665
$\overline{d_x} + d_y > \frac{1}{2}$	$RR_{t_{\beta_1}}$	0.0984	0.1438	0.1693	0.1879	0.0967	0.1437	0.1646	0.1908
	$RR_{t_{\beta_2}}$	0.0781	0.0992	0.1059	0.1181	0.0813	0.0974	0.1107	0.1170
	$RR_{t_{\beta_3}}$	0.0722	0.0794	0.0881	0.0946	0.0731	0.0812	0.0885	0.0927
	R^2	0.0375	0.0090	0.0049	0.0026	0.0378	0.0089	0.0048	0.0026
	$RR_{\mathcal{F}}$	0.1021	0.1582	0.1851	0.2112	0.1014	0.1583	0.1815	0.2108
	$\mathcal{DW}$	1.5163	1.4077	1.3839	1.3672	1.5159	1.4078	1.3828	1.3668

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

Table B.6: Rejection rate (RR) of t-statistics, Average Value (AV) of the $\mathcal{DW}$, RR of the $\mathcal{F}$ joint test, and AV of the R^2, within a OLS-estimated regression: three regressors and different different values of d for each series.

DGP		Same variance ($\sigma^2 = 2$) Sample size (T)				Different variance Sample size (T)			
		100	500	1,000	2,000	100	500	1,000	2,000
	$RR_{t_{\beta_1}}$	0.0817	0.0968	0.1032	0.1171	0.0780	0.0956	0.1140	0.1155
	$RR_{t_{\beta_2}}$	0.0701	0.0824	0.0835	0.0931	0.0694	0.0850	0.0889	0.0987
	$RR_{t_{\beta_3}}$	0.0644	0.0669	0.0703	0.0695	0.0650	0.0672	0.0719	0.0767
$\overline{d_x} + d_y < \frac{1}{2}$	$RR_{t_{\beta_4}}$	0.0617	0.0633	0.0638	0.0662	0.0576	0.0620	0.0590	0.0627
	R^2	0.0456	0.0099	0.0050	0.0026	0.0453	0.0098	0.0051	0.0026
	$RR_{\mathcal{F}}$	0.0811	0.1023	0.1119	0.1263	0.0738	0.0994	0.1143	0.1270
	$\mathcal{DW}$	1.5124	1.4040	1.3811	1.3664	1.5107	1.4060	1.3825	1.3656
	$RR_{t_{\beta_1}}$	0.0867	0.1176	0.1308	0.1439	0.0927	0.1144	0.1353	0.1496
	$RR_{t_{\beta_2}}$	0.0763	0.0958	0.1049	0.1116	0.0839	0.1013	0.1135	0.1192
	$RR_{t_{\beta_3}}$	0.0691	0.0805	0.0895	0.0876	0.0721	0.0822	0.0814	0.0879
$\overline{d_x} + d_y = \frac{1}{2}$	$RR_{t_{\beta_4}}$	0.0646	0.0689	0.0713	0.0753	0.0631	0.0723	0.0698	0.0742
	R^2	0.0477	0.0107	0.0057	0.0030	0.0484	0.0109	0.0057	0.0030
	$RR_{\mathcal{F}}$	0.0895	0.1332	0.1533	0.1685	0.0951	0.1344	0.1574	0.1740
	$\mathcal{DW}$	1.5228	1.4073	1.3832	1.3674	1.5207	1.4060	1.3846	1.3666
	$RR_{t_{\beta_1}}$	0.1019	0.1447	0.1630	0.1880	0.0988	0.1456	0.1661	0.1851
	$RR_{t_{\beta_2}}$	0.0789	0.0990	0.1095	0.1197	0.0774	0.0962	0.1096	0.1138
	$RR_{t_{\beta_3}}$	0.0687	0.0827	0.0888	0.0860	0.0727	0.0835	0.0880	0.0940
$\overline{d_x} + d_y > \frac{1}{2}$	$RR_{t_{\beta_4}}$	0.0648	0.0669	0.0696	0.0757	0.0630	0.0677	0.0716	0.0704
	R^2	0.0491	0.0113	0.0060	0.0032	0.0487	0.0113	0.0060	0.0032
	$RR_{\mathcal{F}}$	0.1028	0.1543	0.1709	0.1937	0.0994	0.1510	0.1777	0.1959
	$\mathcal{DW}$	1.5247	1.4085	1.3844	1.3685	1.5217	1.4089	1.3857	1.3677

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

References

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

Box, G., and Jenkins, G. (1976). *Time series analysis: Forecasting and control.* Holden-Day.

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

Durlauf, S., and Phillips, P. (1988). "Trends versus Random Walks in Time Series Analysis." *Econometrica*, *56*(6), 1333–1354.

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

Granger, C., 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., and Newbold, P. (1974). "Spurious regressions in econometrics." *Journal of Econometrics*, *2*, 11-20.

Hosking, J. (1981). "Fractional differencing." *Biometrika*, *68*, 165-76.

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

Phillips, P. (1986). "Understanding spurious regressions in econometrics." *Journal of Economet-rics*, *33*, 311-340.

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