CREATES Research Paper 2011-15

A Simple Test for Spurious Regressions

Antonio E. Noriega and Daniel Ventosa-Santaularia

A Simple Test for Spurious Regressions[*,†]

Antonio E. Noriega[‡]   Daniel Ventosa-Santaulària[§]

This version: May, 2, 2011

Abstract

The literature on spurious regressions has found that the t-statistic for testing the null of no relationship between two independent variables diverges asymptotically under a wide variety of nonstationary data generating processes for the dependent and explanatory variables. This paper introduces a simple method which guarantees convergence of this t-statistic to a pivotal limit distribution, when there are drifts in the integrated processes generating the data, thus allowing asymptotic inference. We show that this method can be used to distinguish a genuine relationship from a spurious one among integrated ($I(1)$ and $I(2)$) processes. Simulation experiments show that the test has good size and power properties in small samples. We apply the proposed procedure to several pairs of apparently independent integrated variables (including the marriages and mortality data of Yule, 1926), and find that our procedure, in contrast to standard ordinary least squares regression, does not find (spurious) significant relationships between the variables.

Keywords: Spurious regression, integrated process, detrending, asymptotic theory, Cointegration, Monte Carlo experiments.

JEL Classification: *C12, C15, C22, C46.*

1 Introduction

For many years, the statistics and econometrics literatures have studied the phenomenon of spurious relationships among independent variables under a wide variety of data generating processes (DGPs). One early reference is that of Yule (1926), which finds a correlation above 95% between the proportion of Church of England marriages to all marriages and the mortality rate, for the years 1866-1911. More recently, using computer simulation methods, Granger and Newbold (1974) found a significant t-ratio for the slope parameter in a simple linear regression model, assuming independent driftless random walks for the dependent and explanatory variables. Later on, the asymptotic theory developed by Phillips (1986) provided a theoretical explanation of the results in the experimental study of Granger and Newbold (1974): the t-ratio does not possess a limiting distribution, it rather diverges to infinity as the sample size grows, implying that, asymptotically, the t-ratio would always reject the (true) null of no relationship. The rate at which the statistic diverges is $T^{1/2}$, according to Phillips (1986).[1] When allowing for drifts in the random walk representation for the dependent and explanatory variables in a linear regression model, Entorf (1997) shows that divergence occurs at (a faster) rate T. In a recent paper, Noriega and Ventosa-Santaulària (2007) show that the phenomenon of spurious regression is pervasive under a wide range of combinations of DGPs for both the dependent and explanatory variables.[2]

In search for a convergent t-statistic in spurious regressions, Sun (2004) recognizes that the "divergence of the usual t-statistic seems to be a defining characteristic of a spurious regression" (p. 943). He shows that such divergence arises from an underestimated standard error of the ordinary least square (OLS) estimator, and proposes an alternative estimator, based on the HAC standard error estimator with a bandwidth proportional to the sample size. In this set-up, the resulting t-statistic no longer diverges. Sun (2004) finds, however, that the (convergent) limiting distribution of the t-statistic depends on nuisance parameters; in particular, on the memory parameters of the underlying fractional processes he assumes for the dependent and explanatory variables. He argues that, although of theoretical interest, these results have little practical importance, given that parameters in the DGP are generally unknown, and therefore, inference is not feasible.

In this paper, we take a different route. We propose to filter out nuisance parameters via OLS linear detrending on each variable. Residuals from these regressions are then used to verify the

---

[1]In a subsequent paper, the representation theory developed by Phillips (1998) shows that a trending stochastic (deterministic) process can be represented as an infinite linear combination of trending deterministic (stochastic) functions with random coefficients. In such an asymptotic environment, the regression t-ratios of the fitted coefficients diverge at rate $T^{1/2}$.

[2]Other papers studying the phenomenon of spurious regression under different assumptions are: Marmol (1995, 1996, 1998), Cappuccio and Lubian (1997), Granger et. al. (1998), Tsay and Chung (1999), Hassler (1996, 2000, 2003), Kim, Lee and Newbold (2004), Noriega and Ventosa-Santaulària (2006), and Stewart (2006). A literature review can be found in Ventosa-Santaulària (2009).

significance of the relationship through a rescaled version of the standard t-statistic. An analogous approach can be found in the seminal paper by Granger and Newbold (1974), who argue, in the context of estimating equations in econometrics, that "One method we are currently considering is to build single series models for each variable, using the methods of Box and Jenkins (1970) for example, and then searching for relationships between series by relating the residuals from these single models." (pp. 117-118). They further argue that in building regression models, the quantity to be explained is not the variation in the original series, but the variation in the residual part.

Using asymptotic theory, we show that, when both dependent and explanatory variables follow an integrated process, the proposed t-statistic will not depend on nuisance parameters and will not diverge, except for the case when the dependent and explanatory variables are not independent, effectively eliminating the spurious regression problem. We also compute both finite sample and asymptotic critical values, which can be used to distinguish a genuine relationship from a spurious one. Simulation experiments reveal that this procedure works well in finite samples.

Next section briefly introduces the types of DGPs analysed in the paper, which are widely used in empirical work in econometrics. Section 3 shows how standard inference might be complicated by the presence of nuisance parameters, even for (appropriately rescaled) convergent t-statistics. In order to overcome this problem, the approach outlined above to testing for a statistical relationship in a simple regression model is introduced in Section 4, which results in both convergent and pivotal limit distributions of the t-statistic, thus allowing asymptotic inference. Section 5 presents Monte Carlo experiments which report size and power properties of the proposed testing procedure. Section 6 presents several empirical applications of our procedure. We find that, for instance, the high statistical correlation between marriages and mortality found by Yule (1926) is indeed spurious, once our filtering procedure is applied. Last section concludes.

2 Trending mechanisms in the DGP

We consider the following regression model, estimated by OLS:

$$y_t = \hat{\alpha} + \hat{\delta} x_t + \hat{\eta}_t, \tag{1}$$

used as a vehicle for testing the null hypothesis $H_0 : \delta = 0$. Note that the nature of the trending mechanism in the dependent and explanatory variables is unknown a priory. The following assumption summarizes the DGPs considered below for both the dependent and the explanatory variables in model (1). The DGPs in Table 1 include stochastic trending mechanisms, which are widely used in applied work in economics, to model variables such as nominal and real output, consumption, money, prices, among others.

Table 1

ASSUMPTION. The DGPs for $z = y, x$ are as follows.

DGP	Name	Model
1.	I(1)	$\Delta z_t = u_{zt}$
2.	I(1)+*drift*	$\Delta z_t = \mu_z + u_{zt}$
3.	I(2)	$\Delta^2 z_t = u_{zt}$
4.	CI(1,1)	$y_t = \alpha_1 + \beta_1 x_t + u_{yt}$
5.	CI(2,1)	$y_t = \alpha_2 + \beta_2 x_t + \xi_{yt}$
6.	Corr(1)	$y_t = \alpha_3 + \beta_3 x_t + \xi_{yt}$

In Table 1, u_{yt} and u_{xt} are independent innovations obeying Assumption 1 in Phillips (1986), and $\xi_{yt} = \sum_{i=1}^{t} u_{yi}$, that is ξ_{yt} follows an $I(1)$ process. DGP 1 is a driftless random walk, while DGP 2 is a random walk with drift, μ_z. DGP 3 represents an integrated process with double unit roots, that is, one that has to be differenced twice to make it stationary. Under DGP 4, x_t is assumed to follow DGPs 1 or 2 and, if $\beta_1 \neq 0$, then y_t and x_t are cointegrated $CI(1,1)$, following the notation in Engle and Granger (1987). In this case, even though both variables are $I(1)$, the linear combination produces stationary errors. Under DGP 5 it is assumed that x_t follows an $I(2)$ process and, if $\beta_2 \neq 0$, then y_t and x_t are cointegrated $CI(2,1)$. Here a linear combination of $I(2)$ processes reduces the order of integration to $I(1)$. Finally, DGP 6 corresponds to the case of $x_t \sim I(1)$ with drift, correlated with y_t (assuming $\beta_3 \neq 0$), but not cointegrated, since $\xi_{yt} \sim I(1)$. We call this case Corr(1), meaning that the $I(1)$ variables are (only) correlated.

3 The divergent nature of the t-statistic

Assume that interest centers on testing the null hypothesis of no relationship between two random variables y and x, i.e., $H_0 : \delta = 0$, using as a vehicle regression model (1). Rejection of the null when variables are independent is known as a spurious regression.

In a recent paper, Noriega and Ventosa-Santaulària (2007, NVS hereafter) showed that the t-statistic ($t_{\hat{\delta}}$) in a spurious regression does not possess an asymptotic distribution under a wide variety of Data Generating Processes, including trend-stationary processes, single and double unit root processes, broken-mean- and broken-trend-stationary processes, and combinations thereof. Instead, the t-statistic diverges to infinity as the sample size grows.[3]

In order to obtain a convergent t-statistic, the latter should be rescaled by T^κ. NVS find that κ is generally 1/2, but in some cases $\kappa = 1$, or $\kappa = 3/2$, depending on the trending behaviour of the

---

[3]These results were obtained by NVS from the calculation of the order in probability of the t-statistic for all combinations of DGPs considered.

dependent and explanatory variables.

Theorem 1 presents the asymptotic behaviour of the t-statistic $t_{\hat{\delta}}$ from (1) for four combinations of $DGPs$ in the Assumption: 1) a driftless random walk on a driftless random walk (called $t_{\hat{\delta}}^1$), 2) a random walk with drift on a random walk with drift ($t_{\hat{\delta}}^2$), 3) a double unit root process on a double unit root process ($t_{\hat{\delta}}^3$), and 4) a double unit root process on a double unit root process under cointegration [CI(2,1)] ($t_{\hat{\delta}}^4$).[4] In the Theorem, convergence in distribution and in probability are denoted as $\overset{D}{\to}$ and $\overset{p}{\to}$, respectively, and W_z is a standard Wiener process, i.e., $W_z(r)$ is normally distributed for every r in $[0,1]$; that is $W_z(r) \sim N(0,r)$. To simplify notation, all integrals are understood to be taken over the interval $[0,1]$, with respect to the Lebesgue measure, i.e., integrals such as $\int W_z, \int W_z^2, \int rW_z$, and $\int W_x W_y$ are short for $\int_0^1 W_z(r)dr$, $\int_0^1 W_z^2(r)dr$, $\int_0^1 rW_z(r)dr$, and $\int_0^1 W_x W_y(r)dr$, respectively. Also, $\overline{W}_z = \int_0^r W_z(s)ds$, for $s \in [0,1]$, $s < r$.

The proof of Theorem 1 is provided in the Appendix.[5]

THEOREM 1. Consider testing the null hypothesis $H_0 : \delta = 0$ in regression model (1).

a) Denote by $\hat{\delta}^1$ and $t_{\hat{\delta}}^1$ the OLS estimate of δ, and the t-statistic for testing H_0, respectively, when both y and x follow an $I(1)$ process. Then as $T \to \infty$,

$$\hat{\delta}^1 \overset{D}{\to} \left(S_{xy} - S_x S_y\right)\left(S_{x2} - S_x^2\right)^{-1}$$
$$T^{-1/2}t_{\hat{\delta}}^1 \overset{D}{\to} \left(S_{xy} - S_y S_x\right)\left(S_1\right)^{-1/2}$$

b) Denote by $\hat{\delta}^2$ and $t_{\hat{\delta}}^2$ the OLS estimate of δ, and the t-statistic for testing H_0, respectively, when both y and x follow an $I(1)$ plus drift process. Then as $T \to \infty$,

$$\hat{\delta}^2 \overset{p}{\to} \frac{\mu_y}{\mu_x}$$
$$T^{-1}t_{\hat{\delta}}^2 \overset{D}{\to} \mu_x \mu_y \left(12 S_2\right)^{-1/2}$$

c) Denote by $\hat{\delta}^3$ and $t_{\hat{\delta}}^3$ the OLS estimate of δ, and the t-statistic for testing H_0, respectively, when both y and x follow an $I(2)$ process. Then as $T \to \infty$,

$$\hat{\delta}^3 \overset{D}{\to} \frac{\sigma_y}{\sigma_x} S_3 S_4^{-1}$$
$$T^{-1/2}t_{\hat{\delta}}^3 \overset{D}{\to} S_3 S_5^{-1/2}$$

d) Denote by $\hat{\delta}^4$ and $t_{\hat{\delta}}^4$ the OLS estimate of δ, and the t-statistic for testing H_0, respectively, when x_t is generated by an $I(2)$ process, and y_t is generated by a $CI(2,1)$ process, as in DGP 5. Then as $T \to \infty$,

$$\hat{\delta}^4 \overset{D}{\to} \beta_{y2} S_4 \left(\int \overline{W}_x\right)^{-2}$$

---

[4]Results for the case $CI(1,1)$ are well known (Stock, 1987) and therefore are not reported here.

[5]Results in parts *a)*, *b)*, and *c)* of Theorem 1 confirm and extend results in Noriega and Ventosa-Santaulària (2007) (some of which had already been obtained in Phillips (1986), Park and Phillips (1989), Marmol (1995) and Entorf (1997)), as NVS only derived the order in probability of the t-statistic, and not the corresponding asymptotic distribution, as is done in Theorem 1.

$$T^{-1/2} t_{\hat{\delta}}^4 \xrightarrow{D} S_4^{1/2} \int \overline{W}_x \left(\int \overline{W}_x^2 \right)^{-1}$$

where S_z, S_{z2}, S_{xy}, St_z, and S_i for $i = 1, 2 \ldots, 5$, are functions of Wiener processes defined in Appendix A.3.

As can be seen in Theorem 1, the slope parameter does not converge to its true value of zero for the first three combinations of DGPs. To confirm the spurious nature of the relationship, note that the t-statistic diverges in all cases, thus indicating that the null hypothesis of no relationship will be rejected in large samples. The $CI(2, 1)$ case of part d shows that the estimate does not converge to its true value, β_2, and its associated t-ratio diverges at rate $T^{1/2}$. Furthermore, note that when variables follow $DGP\ 2\ (I(1) + drift)$, typically the leading case in macroeconomics, the normalized asymptotic distribution is not pivotal: it depends on the deterministic drift parameters. Hence, even after using an appropriate rescaling, inference is not possible in this case due to the presence of nuisance parameters.

4 A simple test for spurious regression

Results from last section make clear that for the case of a unit root process with drift, even knowing the scaling factor needed for the statistic to achieve a well-defined limit, the corresponding asymptotic distribution is not pivotal (not nuisance-parameter-free). We propose below a simple method which filters out the nuisance parameters, thus allowing asymptotic inference.

The procedure starts by linearly detrending each variable through the following OLS regression:

$$z_t = c_z + b_z t + \varepsilon_{zt}, \quad t = 1, 2, ..., T. \tag{2}$$

for $z = x, y$. Residuals are defined as:

$$\hat{\varepsilon}_{zt} = z_t - \hat{c}_z - \hat{b}_z t,$$

which are used to estimate the following equation

$$\hat{\varepsilon}_{yt} = c_f + \beta_f \hat{\varepsilon}_{xt} + \nu_t. \tag{3}$$

As can be seen, equation (3) uses generated variables: residuals obtained from a first round of estimation. Pagan (1984) shows that when regressors are residuals from another model, a two-step regression estimator will be consistent and efficient, and "valid inferences can be made with the

standard errors provided as output from a second stage regression" (p. 242).[6] Additionally, as proven by Frisch and Waugh (1933), identical results for the estimation of β_f and its t-statistic from (3) would be obtained if instead regression model (1) was used with an additional time trend term.[7]

The next theorem provides the asymptotic theory related to the OLS estimator $\hat{\beta}_f$, and a rescaled version of its associated t-statistic, $T^{-1/2}t_{\beta_f}$, which we call τ, in equation (3). It also reports the asymptotic behavior of the R^2 statistic. Note from Theorem 2 that parts *a* and *b* (*c* and *d*) refer to $I(1)$ ($I(2)$) processes. The proof is outlined in the Appendix.

THEOREM 2. Consider testing the null hypothesis $H_0 : \beta_f = 0$ in regression model (3). The asymptotic behaviour ($T \to \infty$) of the OLS estimator $\hat{\beta}_f$, its associated $T^{1/2}$-rescaled t-statistic, τ, and the R^2 statistic is as follows:

a) When x_t is generated by an $I(1)$ or an $I(1) + drift$ process, and y_t is generated by a $CI(1,1)$ process, as in DGP 4:

$$\hat{\beta}_f \xrightarrow{p} \beta_1$$
$$\tau = O_p(T^{1/2})$$
$$(1 - R^2) = O_p(T^{-1})$$

b) When x_t and y_t are independent from each other and generated by $I(1)$ or $I(1)+drift$ processes:

$$\hat{\beta}_f = O_p(1)$$
$$\tau \xrightarrow{D} ND^{-1/2}$$
$$(1 - R^2) = O_p(1)$$

c) When x_t is generated by an $I(2)$ process, and y_t is generated by a $CI(2,1)$ process, as in DGP 5:

$$\hat{\beta}_f \xrightarrow{p} \beta_2$$
$$\tau = O_p(T)$$
$$(1 - R^2) = O_p(T^{-1})$$

d) When x_t and y_t are independent and generated by $I(2)$ processes:

$$\hat{\beta}_f \xrightarrow{D} (\sigma_y/\sigma_x)Q_5Q_4^{-1}$$
$$\tau \xrightarrow{D} Q_5Q_6^{-1/2}$$
$$(1 - R^2) = O_p(1)$$

---

[6]Note however, that this result concerns estimators from regression models in which only the regressor is a generated variable.

[7]See also Lovell (2008) or Greene (1997, pp. 246-247).

e) When x_t is generated by an $I(1) + drift$ process, and y_t is generated a $Corr(1)$ process, as in DGP 6:

$$\hat{\beta}_f \overset{D}{\to} Q_1/\sigma_x Q_2$$
$$\tau \overset{D}{\to} Q_1(-\sigma_y^2 Q_3)^{-1/2}$$
$$(1 - R^2) = O_p(1)$$

where N, D, and Q_i for $i = 1, 2, \ldots, 6$ are functions of Wiener processes defined in Appendix A.4.

Theorem 2 provides useful results. First, the estimated slope parameter in the detrended regression model (3) converges to the cointegrating parameter of model 4 in the Assumption, β_1, when variables cointegrate, as shown in part *a*, implying that the cointegrating parameter will be consistently estimated from regression model (3).[8] Furthermore, under cointegration, the rescaled t-statistic diverges, correctly indicating a long-run relationship, as shown also in part *a*. Second, as shown in part *b*, the rescaled t-statistic does not diverge for independent integrated processes, thus avoiding the (asymptotic) spurious regression problem. Furthermore, the t-statistic converges to a pivotal limiting distribution. Note that this holds true with or without a drift in the DGP. Similar conclusions can be reached for $I(2)$ processes, as shown in parts *c* and *d*. Part *e* indicates that when variables are correlated but not cointegrated, the test has no power, since τ does not diverge; instead, it converges to a non-pivotal distribution. Finally, note that the R^2 converges in probability to one, only when there is cointegration among the variables.

Summing up, the t-statistic will diverge only when there is a long-run cointegration relationship between the variables; otherwise it will not grow with the sample size.

Based on the preceding results, we propose a simple test which allows to distinguish a true linear relationship among two integrated random variables, from a spurious one. The test is based on τ, the $T^{1/2}$-rescaled t-statistic of β_f in regression model (3) for testing the null hypothesis $H_0 : \beta_f = 0$. Under the null, the filtered variables are asymptotically linearly independent. A true relationship occurs when the null is rejected.

For the case when the variables are independent and follow any combination of $I(1)$ and $I(1) + drift$ processes, the resulting formulae (Theorem 2, part *b*) show that the asymptotic distribution is pivotal, i.e. free of nuisance parameters. This implies that the above procedure allows inference by means of an appropriately rescaled pivotal statistic, whose distribution can be tabulated.

We simulated the limit expression for τ in Theorem 2 (parts *b* and *d*) and generate asymptotic critical values, which we report in Table 2 in the row indicated by the symbol ∞.[9]

---

[8]Nelson and Kang (1981) argue that the dynamics of econometric models estimated from inappropriately detrended integrated variables, may be an artifact of the trend removal procedure. Note that this phenomenon does not seem to affect the consistency with which the slope parameter is estimated, as shown in parts *a)* and *c)* of Theorem 2.

[9]The number of replications is $10,000$ and the simulation of the Brownian motions follows Perron (1989, p. 375).

Table 2

CRITICAL VALUES FOR THE t-STATISTICS

T	τ [case I(1)-I(1) with or without drift]				τ [case I(2)-I(2)]			
	1%	5%	10%	20%	1%	5%	10%	20%
25	±1.28	±0.92	±0.76	±0.57	±5.47	±3.49	±2.70	±1.92
50	±1.28	±0.92	±0.76	±0.58	±5.95	±3.67	±2.82	±2.00
100	±1.28	±0.92	±0.76	±0.58	±5.87	±3.73	±2.86	±2.01
200	±1.28	±0.92	±0.76	±0.58	±5.90	±3.74	±2.85	±2.03
500	±1.28	±0.92	±0.76	±0.58	±5.95	±3.74	±2.85	±2.03
1,000	±1.28	±0.93	±0.76	±0.58	±6.04	±3.78	±2.90	±2.04
∞	±1.29	±0.93	±0.76	±0.58	±6.04	±3.79	±2.92	±2.06

Table 2 also reports critical values based on simulated data for samples $T = 25, 50, 100, 200,$ $500, 1000$. The left part of Table 2 shows critical values for the (normalized) t-statistic, τ, for the case I(1)-I(1) (whether the variables have a drift or not), while the right part shows critical values for the case I(2)-I(2), also based on Theorem 2 (part d). It is worth noting the closeness of the asymptotic and finite sample critical values.
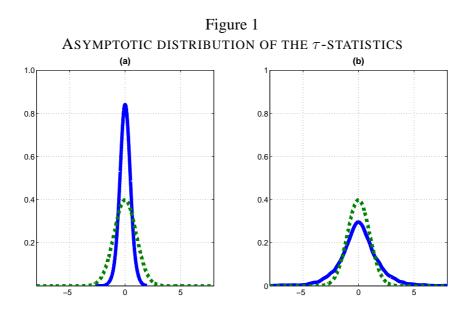
As a guide on the use of critical values in Table 2, assume that unit root tests (such as Dickey-Fuller, or Ng-Perron tests) have led the researcher to the conclusion that both y and x are $I(1)$, with a sample size of $T = 100$. The use of τ, together with critical values provided in Table 2, allows to test for a relationship between these two integrated variables. A (low) value of the statistic which does not reject the null (lower than, say, 0.92, the critical value at the 5% level), will indicate that the variables are two independent random walks. On the other hand, a large value of the statistic (larger than 0.92) will be indicative of the variables being cointegrated.

Figure 1 plots the asymptotic distribution of τ for the I(1)+drift case (left panel) and for the I(2) case (right panel) under the null hypothesis. Clearly, both of the distributions display a marked departure from normality.

4.1 Some extensions

We have also examined the case of spurious regression between two Trend-Stationary (TS) processes (see Kim, Lee and Newbold, 2004) as well as combinations among TS and $I(d)$ processes, for $d = 1, 2$. Results indicate that $t_{\hat{\beta}_f}$ converges to a standard normal distribution under the null hypothesis when innovations in the DGP for both dependent and explanatory variables are *iid*. However, when the innovations are autocorrelated, the limit distribution of $t_{\hat{\beta}_f}$ under the null is

---

A Matlab code is available from the authors upon request.

Figure 1
ASYMPTOTIC DISTRIBUTION OF THE τ-STATISTICS

Panel (a): τ [I(1)+drift-I(1)+drift case]; Panel (b): τ [I(2)-I(2) case]. $\sigma_z^2 = 1$ for $z = x, y$; $T = 1,000$, $R = 10,000$. A standard normal distribution (dash-dotted curve) is depicted for comparison.

not nuisance-parameter-free. These results, which are still preliminary and out of the scope of the present paper, will be reported elsewhere.

5 Finite Sample Properties

5.1 Size and power of the τ test

We computed rejection rates of the proposed τ-statistics for testing the null hypothesis $H_0 : \beta_f = 0$ in equation (3), using critical values from Table 2 with a 5% nominal level. Rejection rates were computed on simulated data for samples of size $T = 50, 100, 200, 300$ and 500, using the models in the Assumption, and $10,000$ replications.

Table 3 shows rejection rates of the τ test for different combinations of parameter values, assuming x_t and y_t have been generated as independent $I(1)$ processes (Panel *a*), or as cointegrated processes (Panel *b*). In order to study the effect on size and power of autocorrelation in the processes' disturbances u_{zt}, we allow $u_{zt} = \rho_z u_{zt-1} + \eta_{zt}$, for $\rho_z = 0.0, 0.5$ with $z = y, x$, and $\eta \sim iid\mathcal{N}(0, 1)$.[10] As can be seen, size tends to be conservative across T and parameter values (with the exception of $T = 50$ and $\rho_y = \rho_x = 0.5$, in which case size is 0.07). Turning to Panel *b* in Table 3, we find that power is generally high, except for samples sizes smaller than $T = 200$ and small parameter values.

---

[10] We assume $\sigma_y^2 = \sigma_x^2 = 1$ in all simulations.

Table 3

REJECTION RATES OF THE TEST STATISTICS τ.

Panel (a)

Relationship	Parameters					Sample Size				
	μ_x	μ_y	β_{y1}	$\rho_{x,1}$	$\rho_{y,1}$	50	100	200	300	500
	0.00	0.00	−	0.00	0.00	0.02	0.02	0.02	0.03	0.02
			−	0.50	0.50	0.07	0.04	0.04	0.03	0.03
	0.50	0.30	−	0.00	0.00	0.02	0.03	0.02	0.02	0.02
			−	0.50	0.50	0.07	0.05	0.04	0.03	0.03
		1.25	−	0.00	0.00	0.02	0.02	0.02	0.02	0.02
			−	0.50	0.50	0.07	0.04	0.04	0.03	0.03
		1.50	−	0.00	0.00	0.02	0.02	0.02	0.02	0.02
			−	0.50	0.50	0.07	0.05	0.04	0.03	0.03
	1.50	0.30	−	0.00	0.00	0.02	0.02	0.03	0.03	0.02
			−	0.50	0.50	0.07	0.05	0.03	0.03	0.03
		0.50	−	0.00	0.00	0.02	0.02	0.02	0.02	0.02
			−	0.50	0.50	0.07	0.04	0.03	0.03	0.03
		1.25	−	0.00	0.00	0.03	0.02	0.03	0.02	0.03
			−	0.50	0.50	0.07	0.04	0.03	0.03	0.03

Independent DGPs (row group label, Panel (a))

Panel (b)

	μ_x	α_y	β_{y1}	$\rho_{x,1}$	$\rho_{y,1}$	50	100	200	300	500
	0.00	0.30	0.30	0.00	0.00	0.01	0.10	0.41	0.65	0.91
				0.50	0.50	0.03	0.62	0.92	0.99	1.00
	0.30	0.50	0.30	0.00	0.00	0.02	0.10	0.40	0.65	0.91
				0.50	0.50	0.28	0.61	0.93	0.99	1.00
			0.50	0.00	0.00	0.24	0.62	0.94	0.99	1.00
				0.50	0.50	0.74	0.96	1.00	1.00	1.00
			1.25	0.00	0.00	0.99	1.00	1.00	1.00	1.00
				0.50	0.50	1.00	1.00	1.00	1.00	1.00
			1.50	0.00	0.00	1.00	1.00	1.00	1.00	1.00
				0.50	0.50	1.00	1.00	1.00	1.00	1.00
			1.75	0.00	0.00	1.00	1.00	1.00	1.00	1.00
				0.50	0.50	1.00	1.00	1.00	1.00	1.00
			2.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
				0.50	0.50	1.00	1.00	1.00	1.00	1.00

Cointegrated (row group label, Panel (b))

Panel (a): x_t and y_t independently generated by DGP (2); panel (b): x_t and y_t generated by DGP (2), and (4), respectively ($x_t, y_t \sim CI(1,1)$).

Table 4 shows size and power results under the assumption that variables are generated by an $I(2)$ process. Panel *a* shows that the test has the correct size for all parameter values and sample sizes. Panel *b* corresponds to the case when the variables cointegrate in such a way that a linear combination of the two $I(2)$ variables is $I(1)$, that is, variables are $CI(2,1)$. As can be seen, power is generally very high.[11]

Note that our proposed test is based on prior statistical inference used to determine whether there is a unit root or not in each variable. Since this pre-testing may induce size distortions, we applied a Bonferroni correction. Additional Monte Carlo experiments were carried out where the τ test depends on inference drawn from Dickey-Fuller tests applied to the individual series. Results

---

[11]We also studied the power of the test assuming variables are $I(2)$ and $CI(2,2)$, that is, a linear combination of two $I(2)$ variables is $I(0)$. Again, power is generally high.

Table 4

REJECTION RATES OF τ.

Panel (a)

Relationship	Parameters					Sample Size				
	σ^2_{ux}	σ^2_{uy}	β_{y2}	$\rho_{x,1}$	$\rho_{y,1}$	50	100	200	300	500
Independent DGPs	0.50	0.30	–	0.00	0.00	0.05	0.05	0.05	0.05	0.05
			–	0.50	0.50	0.06	0.06	0.04	0.06	0.05
		1.25	–	0.00	0.00	0.04	0.04	0.05	0.05	0.05
			–	0.50	0.50	0.05	0.05	0.05	0.05	0.04
		1.50	–	0.00	0.00	0.05	0.05	0.05	0.05	0.04
			–	0.50	0.50	0.06	0.05	0.05	0.05	0.05
	1.50	0.30	–	0.00	0.00	0.05	0.05	0.05	0.05	0.06
			–	0.50	0.50	0.06	0.05	0.05	0.05	0.05
		0.50	–	0.00	0.00	0.05	0.05	0.05	0.05	0.05
			–	0.50	0.50	0.06	0.05	0.04	0.04	0.05
		1.25	–	0.00	0.00	0.05	0.05	0.05	0.05	0.05
			–	0.50	0.50	0.06	0.05	0.05	0.05	0.06

Panel (b)

Relationship	σ^2_{ux}	σ^2_{uy}	β_{y2}	$\rho_{x,1}$	$\rho_{y,1}$	50	100	200	300	500
Cointegrated	1.00	1.00	0.30	0.00	0.00	0.32	0.68	0.92	0.98	0.99
				0.50	0.50	0.42	0.72	0.94	0.98	0.99
			0.50	0.00	0.00	0.58	0.88	0.98	0.99	0.99
				0.50	0.50	0.66	0.90	0.99	0.99	1.00
			1.25	0.00	0.00	0.93	0.99	1.00	1.00	1.00
				0.50	0.50	0.95	0.99	1.00	1.00	1.00
			1.50	0.00	0.00	0.96	0.99	1.00	1.00	1.00
				0.50	0.50	0.97	0.99	1.00	1.00	1.00
			1.75	0.00	0.00	0.98	0.99	1.00	1.00	1.00
				0.50	0.50	0.98	0.99	1.00	1.00	1.00
			2.00	0.00	0.00	0.99	0.99	1.00	1.00	1.00
				0.50	0.50	0.99	0.99	1.00	1.00	1.00

Panel (a): x_t and y_t independently generated by DGP (3); panel (b): x_t and y_t generated by DGP (3), and (5), respectively ($x_t, y_t \sim CI(2,1)$).

(not reported but available from the authors upon request) reveal that a Bonferroni correction does not seem to be necessary, as the size properties of the test are nearly identical to the ones reported in Tables 3 and 4.

5.2 Spurious regression and cointegration

We have shown that the proposed test has power for distinguishing among independent and cointegrated processes. In order to investigate this issue in more depth, we designed a Monte Carlo experiment, through which we compare the small sample performance of our τ statistic with that of standard cointegration tests, such as the residual based test of Engle and Granger (1987), and the Johansen (1988) test. For this experiment, we generated 10,000 samples of sizes $T = 50, 100, 200, 300$, and 500 of integrated processes (with drift) for y_t and x_t, under two hypotheses:

the variables are independent of each other, and the variables cointegrate. We then calculate the proportion of times the null hypothesis $H_0 : \beta_f = 0$ in regression model (3) is rejected at the nominal size of 5%, out of 10,000 replications, under each hypothesis.

As can be seen from column 3 in Table 5, the τ test does a very good job in discerning independent processes from cointegrated ones: under the hypothesis of independence, the rejection rate equals nominal size (5%), while power is nearly 80%, with a sample as small as fifty observations. In the Table, EG, Tr and Eig stand for the Engle-Granger (1987) test, the trace test of Johansen (1988), and the eigenvalue test of Johansen (1988), respectively. Note that the performance of the four tests is very similar for sample sizes above 100 observations. When innovations u_{zt} in the DGP are correlated, column 4 of Table 5 shows some size distortions (the rejection rate reaches 10% for $T = 50$) and lower power (52%). Note, however, that these problems quickly disappear as the sample size grows.

Table 5
REJECTION RATES OF THE τ, ENGLE-GRANGER, AND JOHANSEN TESTS.

T	Case	τ	τ^a	EG	EG^a	Tr.	Tr.^{a}	Eig.	Eig.^{a}
						1 lag; no det. part			
50	I	0.05	0.10	0.05	0.02	0.07	0.09	0.07	0.09
	CI	0.78	0.52	1.00	0.33	0.99	0.28	0.99	0.26
100	I	0.05	0.09	0.05	0.02	0.06	0.07	0.06	0.07
	CI	0.98	0.77	1.00	0.89	1.00	0.77	1.00	0.78
200	I	0.05	0.07	0.05	0.02	0.06	0.06	0.06	0.06
	CI	0.99	0.96	1.00	0.99	1.00	0.99	1.00	0.99
300	I	0.05	0.06	0.04	0.02	0.07	0.06	0.06	0.06
	CI	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
500	I	0.05	0.06	0.05	0.02	0.08	0.06	0.07	0.06
	CI	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00

EG: Engle-Granger test; Tr/Eig: Johansen's Trace and Eigenvalue tests; ^{a} Innovations follow a stationary $AR(1)$ process: $u_{zt} = 0.75u_{zt-1} + \epsilon_{zt}$; $\epsilon_{zt} \sim \mathcal{N}(0,1)$; $\mu_x = 0.03$, $\mu_y = 0.04$, $\beta_y = 0.7$ (for cointegrated relationships), and $R = 1,000$.

Hence, the τ test could be used not only to distinguish a genuine relationship from a spurious one, but also to distinguish independent from cointegrated processes. There may be gains from using the τ test as a cointegration test, given its relative simplicity: on the one hand, the Engle-Granger test must control for autocorrelation by means of augmentation terms; on the other, the Johansen test requires the specification of the initial VAR model, as well as decisions regarding the inclusion of deterministic components.

Overall, we believe that the practitioner could benefit from applying both tests, given their different nature: the EG procedure is based on the properties of the residuals whilst ours is based

on the parameter estimate. If both tests find evidence of a genuine relationship, the practitioner should be more confident about the validity of such inference. On the contrary, if the results of the tests are incompatible, then the practitioner could consider this as evidence of potential misleading inference and should therefore revise the empirical exercise. The use of the τ statistic could be therefore considered as a companion test in a cointegration analysis, capable of confirming the inference drawn from other tests or to cast doubts about its validity.

To illustrate these arguments, we performed an additional Monte Carlo experiment. We generated 10,000 samples of two independent $I(1)$ processes of sizes reported in the first column on Table 6, in order to study the joint behaviour of the τ test and the EG one. Because the processes are generated independently of each other, we expect that the tests do not reject their null hypotheses, namely, no relationship, and no cointegration, respectively. To verify this, we counted the percentage of times this occurs, and present these percentages in column 2 of Table 6. At a nominal level of 5%, and assuming the tests are independent, we should expect to get something close to the following probability:

$$\Pr(\tau \text{ does not reject } H_0 \cap \text{EG does not reject } H_0 \mid H_0 \text{ is true}) = (1-\alpha)^2 = (0.95)^2 = 0.9025$$

Table 6

TYPE-I ERROR REJECTION RATES OF THE τ AND EG TESTS.

T	Both accept	Only τ accepts	Only EG accepts	Both reject
50	0.9061	0.0478	0.0418	0.0043
100	0.9110	0.0446	0.0405	0.0039
150	0.9047	0.0420	0.0479	0.0054
200	0.9065	0.0446	0.0441	0.0048
300	0.9048	0.0419	0.0480	0.0053
500	0.9072	0.0438	0.0431	0.0059
1,000	0.9058	0.0424	0.0473	0.0045

Level of the tests: $\alpha = 0.05$. Innovations are $\sim iid\mathcal{N}(0,1)$; $\mu_x = 0.03$, $\mu_y = 0.04$, and $R = 10,000$.

As can be seen from the values reported in column 2 in Table 6, it seems the tests are indeed independent.Columns 3 and 4 in Table 6 report estimates of the following probability:

$$\Pr(\text{one test does not reject } H_0 \cap \text{ the other one does reject } H_0 \mid H_0 \text{ is true}) = (1-\alpha)\alpha$$
$$= (0.95)0.05$$
$$= 0.0475$$

From the reported values, we can conclude that at least one of the tests does not reject the (true)

null is around 95% (calculated as the sum of columns 2 and 3, or 2 and 4). Finally, column 5 estimates the following probability:

$$\Pr(\tau \text{ rejects } H_0 \cap \text{EG rejects } H_0 \mid H_0 \text{ is true}) = \alpha^2 = 0.0025$$

which means that it is very unlikely that both tests get the wrong outcome of rejecting the true null.

Turning to power issues, we present Monte Carlo experiments that allow the presence of auto-correlation in the innovations: in Table 7 innovations follow an $AR(1)$ process with autoregressive parameter equal to 0.75. In this case, the EG test must include one lag of the dependent variable as an additional regressor in the auxiliary regression. We included such a lag in the EG test and compare the rejection rates of the joint application of the τ test and the EG one. From Table 7, which is constructed under the alternative hypothesis of cointegration, when the practitioner runs both tests with a sample of size $T = 100$, there is a 93% chance of rejecting the null with at least one of the tests ($0.613 + 0.148 + 0.167$), which is better than 76% ($0.613 + 0.148$), or 78% ($0.613 + 0.167$), the chances of rejecting the null using only the τ statistic, or the EG test. It appears that, for relatively small samples (below 100 observations), the application of both tests ensures higher power. Hence, there seems to be potential benefits in the use of both tests instead of only one.

Table 7
TYPE-II ERROR REJECTION RATES OF THE τ AND EG TESTS.

T	Both reject	Only τ rejects	Only EG rejects	Both accept
50	0.1662	0.3593	0.0977	0.3768
100	0.6131	0.1480	0.1665	0.0724
150	0.8911	0.0116	0.0951	0.0022
200	0.9627	0.0001	0.0372	0.0000
300	0.9952	0.0000	0.0048	0.0000
500	1.0000	0.0000	0.0000	0.0000
1,000	1.0000	0.0000	0.0000	0.0000

Level of the tests: $\alpha = 0.05$. Innovations are $\sim AR(1)$ with $\rho = 0.75$; $\mu_x = 0.03$, $\mu_y = 0.04$, $\beta_y = 0.7$ (for cointegrated relationship), and $R = 10,000$.

6   Empirical illustrations

In a frequently cited but insufficiently read paper (as Granger (2001), p.557 argues), Yule (1926) first discussed the nonsense correlations that can be found "between quantities varying with the time, to which we cannot attach any physical significance whatever,…"(p. 2). He illustrated this using annual data for the years 1866-1911 on the proportion of Church of England marriages to all

marriages (per 1,000 persons) and the mortality rate (per 1,000 persons) in England and Wales.

Yule (1926) found a correlation coefficient between these two variables of 0.9512, and argued that even though it could be possible that the spread of scientific thinking and the progress of science might be behind the fall in marriages and mortality, respectively, and hence a common factor influences both series, it is nevertheless clear that "the correlation is simply sheer nonsense; that it has no meaning whatever; that it is absurd to suppose that the two variables in question are in any sort of way, however indirect, causally related to one another" (p.2).

The purpose of this section is twofold. First, we use real data to illustrate the possibility of finding spurious statistical relationships between variables which, on *a priori* grounds, should bear no relationship to each other. The variables we study are the following:

1. Annual data (1866-1911) on the proportion of Church of England marriages to all marriages (per 1,000 persons) in England and Wales (*marriages*, henceforth).[12]

2. Annual data (1866-1911) on the mortality rate (per 1,000 persons) in England and Wales (*mortality*, henceforth).[13]

3. Monthly data (1991:1-2005:12) on total number of vehicle sales in the US (*cars* henceforth).[14]

4. Monthly data (1991:1-2005:12) on number of murders in the US (*murders* henceforth).[15]

5. Monthly data (1991:1-2005:12) of the inflation rate in Brazil (*infbrazil*, henceforth).[16]

6. Monthly data (1991:1-2005:12) on cash in vault of commercial and development banks in Mexico (*BCbanks*, henceforth).[17]

Second, after showing that the usual OLS regression techniques indicate the presence of linear (spurious) relationships between combinations of the variables, we show that, once our proposed procedure is implemented, the statistical relationship vanishes, leading to what we believe should be, on a priori grounds, correct inference.

---

[12]Data for *Church marriages* are from Office for National Statistics, U.K. (http://www.statistics.gov.uk/cci/nscl.asp?ID=7537).

[13]Data for the *mortality rate* series come from Mitchell (1988).

[14]The source is Bureau of Economic Analysis: Auto and Truck Sales, Production, Exports and Inventories (thousands), from www.FreeLunch.com - http://www.economy.com/freelunch. The sample period (1991:1-2005:12) comprises 180 observations.

[15]The source is FBI: Crime in the United States; BOC: County City Data Book, from www.FreeLunch.com - http://www.economy.com/freelunch.

[16]Calculated as the percentage change in the consumer price index. Central Bank of Brazil (Banco Central do Brasil). https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries

[17]Thousands of Pesos, nominal stocks. The source is Bank of Mexico (Banco de México). http://www.banxico.org.mx/sitioingles/billetesymonedas/estadisticas/banknotescoins/banknotescoinsincircu.htm

We start our empirical investigation by uncovering the order of integration of the variables. As a first step, we follow Dickey and Pantula (1987), who observed empirically that the probability of rejecting the null hypothesis of one unit root (denoted H_1) against the alternative of stationarity (H_0) increases with the number of unit roots present. In order to overcome this possibility, we use the methodology suggested by Pantula (1989), which consists of an asymptotically consistent sequential procedure for testing the null hypothesis H_r : exactly r unit roots, against the alternative H_{r-1} : exactly $(r-1)$ unit roots, with $r = m, ..., d+1, d$, where m is an assumed maximum number of unit roots present in the data, and d the true number of unit roots present in the data. Pantula suggests that the hypotheses must be tested sequentially in the order $H_m, H_{m-1}, ..., H_d$. We assume that it is known a priori that the maximum possible number of unit roots present in the data is 2. Based on Pantula's results, we perform unit root tests downwards, starting with a test of the null hypothesis H_2: exactly two unit roots (or a unit root in the first differences of the data). If the null H_2 is rejected, then we test the null H_1: one unit root, against the alternative of stationarity, otherwise, we infer there are two unit roots in the series.

This procedure is implemented by using seven tests: the Augmented Dickey-Fuller unit root test (see Said and Dickey (1984)), the KPSS stationarity test[18] (see Kwiatkowski, et.al (1992)), the ERS point optimal unit root test (see Elliott, et.al. (1996)) and the four unit root tests with good size and power properties of Ng-Perron (see Ng and Perron (2001)). Table 8 summarizes the time series properties of the variables.

As can be deducted from this Table, most variables seem to follow a unit root process, with the exceptions of *murders*, which seems to be an $I(2)$ process, and *marriages* which is described as $I(1)$ by some tests, or $I(2)$ by others. Inference from the various tests is summarized in Table 9.

Using several combinations of these integrated series, we examine regression results under two approaches: simple linear ordinary least squares on (1), and the procedure proposed in section 4, which uses $\tau = T^{-1/2}t_{\beta_f}$ from OLS estimation of equation (3) as the test statistic. Table 10 collects the results.

---

[18]In the case of the *KPSS* stationarity test, we start by testing the null H_1 against H_2, that is, the null of stationarity in the first differences of the data, against the alternative of two unit roots. If the null H_1 is rejected, then we stop and conclude that the series has two unit roots. If the null is not rejected, we proceed to test the null of stationarity, H_0, against the alternative of a unit root, H_1.

Table 8

RESULTS OF THE UNIT ROOT TESTS

Series	ADF	KPSS[†]	ERS[†]	Ng-Perron			
				MZa	**MZt**	**MSB**	**MPT**
$\Delta marriage$	-2.03	0.83***	3.61*	-5.68	-1.67*	0.29	4.36*
$marriage$	——	——	15.10	——	-1.60	——	16.11
$\Delta mortality$	-9.48***	0.02	3.30*	-14.91***	-2.61***	0.17**	2.11**
$mortality$	-1.19	0.43***	12.02	-7.49	-1.93	0.26	12.16
$\Delta cars$	-19.41***	0.013	0.48***	-122.74***	-7.80***	0.06***	0.25***
$cars$	-1.40	75.90***	319.00	-0.15	-0.20	1.35	338.37
$\Delta murders$	-0.81	147.42***	38.05	-1.07	-0.72	0.68	22.69
$murders$	——	——	——	——	——	——	——
$\Delta Infbrazil$	-12.92***	0.048	0.28***	-149.93***	-8.66***	0.06***	0.16***
$Infbrazil$	-2.48	11.86***	7.68	-11.93	-2.42	0.20	7.75
$\Delta BCbanks$	-19.12***	0.01	0.342***	-122.39***	-7.81***	0.06***	0.22***
$BCbanks$	0.22	119.45***	132.67	-1.77	-0.93	0.53	50.86

***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

[†] The statistics reported are those which estimate the residual spectrum at frequency zero by OLS (using OLS-detrended methods does not change the results).

We let the maximum value of lag length at $k_{max} = int(12(T/100)^{1/4})$, see Ng and Perron (2001). The lag length is selected by MBIC, MAIC and MHQ; except in the Ng-Perron test where the lag is selected like in Perron and Qu (2007).

Table 9

ORDER OF INTEGRATION FROM UNIT ROOT TESTS.

Series	ADF	KPSS	ERS	Ng-Perron			
				MZa	MZt	MSB	MPT
$marriage$	$I(2)$	$I(2)$	$I(1)$	$I(2)$	$I(1)$	$I(2)$	$I(1)$
$mortality$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$
$cars$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$
$murders$	$I(2)$	$I(2)$	$I(2)$	$I(2)$	$I(2)$	$I(2)$	$I(2)$
$Infbrazil$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$
$BCbanks$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$	$I(1)$

Table 10

y_t	x_t	$t_{\hat{\delta}}$	τ
$mortality$	$marriages$	23.542***	0.477
$cars$	$Infbrazil$	-6.390***	-0.062
$cars$	$BCbanks$	10.108***	-0.006
$Infbrazil$	$BCbanks$	-11.258***	-0.118
$mortality$	$\Delta marriages$	0.688	0.010
$cars$	$\Delta murders$	2.047**	-0.068
$Infbrazil$	$\Delta murders$	-0.393	0.258
$BCbanks$	$\Delta murders$	4.424***	0.063

***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

Results from Table 10 indicate that, in almost all cases, simple OLS regression among apparently independent variables will result in rejection of the null of no relationship, leading one to conclude in favour of spurious relationships. The case of $mortality$ on $\Delta marriages$ suggests that $marriages$ is indeed $I(1)$, not $I(2)$. If this was the case, the application of the difference operator would eliminate the stochastic trend in this variable, i.e. $\Delta marriages \sim I(0)$. As shown in Noriega and Ventosa-Santaulària (2007), the t-statistic does not diverge when one of the variables is trendless. Hence, in this case we conclude that standard inference through $t_{\hat{\delta}}$ in regression model (1) for these variables would indicate a spurious rejection of the null.

The last column of Table 10 shows that the proposed procedure indicates, as one would expect, that the variables are not statistically related, since the τ statistic is not significant at conventional levels, using critical values from the left panel of Table 2.

7   Conclusions

This paper has proposed a simple procedure to overcome the spurious regression problem in a simple linear regression model when the variables are integrated processes. We study two cases, one in which both dependent and explanatory variables are integrated processes of order one (with and without drift), the leading case in many empirical studies in macroeconomics, and one in which the variables are integrated of order two.

In the context of a simple linear regression model, it is well known that, when both dependent and explanatory variables follow an $I(1)$ plus drift process, the t-statistic of the slope parameter diverges, while the corresponding rescaled statistic converges to a well defined distribution, expressed in terms of Wiener processes, but dependent on nuisance parameters (the drift parameters). We introduce a simple approach based on linear filtering of the data, which results in a t-statistic with a well defined asymptotic distribution free of nuisance parameters. We tabulated both the asymptotic distribution of this statistic and its finite sample counterpart, and report critical values for various samples sizes and significance levels.

The asymptotic theory behind our proposed procedure implies that, when variables cointegrate, the test will reject the null of no correlation. On the other hand, when variables are independent, then the test will not reject asymptotically. A small Monte Carlo experiment reveals that our proposed test statistic does a very good job in discerning independent from cointegrated variables and could be therefore considered as a "companion test" in a cointegration analysis.

Finally, we applied the proposed procedure to the famous Yule (1926) data set on marriages and mortality rates, and found that under our method we no longer find a (spurious) significant relationship between these two variables. Some additional empirical exercises confirm that our procedure seems to work in practice.

References

Box, G. and G. Jenkins (1970), *Time Series Analysis, Forecasting, and Control*, San Francisco: Holden Day.

Cappuccio, N. and D. Lubian (1997), "Spurious Regressions Between $I(1)$ Processes with Long Memory Errors", *Journal of Time Series Analysis*, 18, 341-354.

Dickey, D. A. and S. G. Pantula (1987), "Determining the Order of Differencing in Autoregressive Processes", *Journal of Business and Economic Statistics*, 5(4), 455-461.

Entorf, H. (1997), "Random Walks with Drifts: Nonsense Regression and Spurious Fixed-Effect Estimation", *Journal of Econometrics*, 80, 287-296.

Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation,

Estimation and Testing", *Econometrica*, 55, 251-76.

Elliott, G., T.J. Rothenberg and J.H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root", *Econometrica* 64, 813-836.

Frisch, R. and F. V. Waugh (1933), "Partial Time Regressions as Compared with Individual Trends", *Econometrica*, 1(4), 387-401.

Granger, C.W.J. and P. Newbold (1974), "Spurious Regression in Econometrics", *Journal of Econometrics*, 2, 111-120.

Granger, C.W.J., N. Hyung, and Y. Jeon (1998), "Spurious Regressions with Stationary Series", Mimeo.

Granger, C.W.J (2001), "Spurious Regressions in Econometrics", Chapter 26 in *A Companion to Theoretical Econometrics*, edited by Badi H. Baltagi, Blackwell Publishing.

Greene, W. (1997), *Econometric Analysis*, third edition, Prentice Hall.

Hassler, U. (1996), "Spurious Regressions when Stationary Regressors are Included", *Economics Letters*, 50, 25-31.

Hassler, U. (2000), "Simple Regressions with Linear Time Trends", *Journal of Time Series Analysis*, 21, 27-33.

Hassler, U. (2003), "Nonsense Regressions Due to Neglected Time-Varying Means", *Statistical Papers*, 44, 169-182.

Johansen, S. (1988), "Statistical Analysis of Cointegrated Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

Kim, T.-H., Y.-S. Lee and P. Newbold (2004), "Spurious Regressions with Stationary Processes Around Linear Trends", *Economics Letters*, 83, 257-262.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159-178.

Lovell, M. C. (2008), "A Simple Proof of the FWL (Frisch-Waugh-Lovell) Theorem", *The Journal of Economic Education*, 39(1), 88-91.

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

Marmol, F. (1996), "Nonsense Regressions Between Integrated Processes of Different Orders", *Oxford Bulletin of Economics and Statistics*, 58(3), 525-36.

Marmol, F. (1998), "Spurious Regression Theory with Nonstationary Fractionally Integrated Processes", *Journal of Econometrics*, 84, 233-50.

Mitchell B.R. (1988), *British Historical Statistic*s, Cambridge University Press.

Nelson, C. R., and H. Kang (1981), "Spurious Periodicity in Inappropriately Detrended Time Series", *Econometrica*, 49(3), 741-751.

Ng, S. and P. Perron (2001), "Lag Length Selection and the Construction of Unit Root Tests

with Good Size and Power", *Econometrica* 69(6), 1519-54.

Noriega, A. and D. Ventosa-Santaulària (2006), "Spurious Regression Under Broken-Trend Stationarity", *Journal of Time Series Analysis*, 27(5), 671-684.

Noriega, A. and D. Ventosa-Santaulària (2007), "Spurious Regression and Trending Variables", *Oxford Bulletin of Economics and Statistics*, 69(3), 439-444.

Pagan, A. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, 25(1), 221-247.

Pantula, S. G. (1989), "Testing for Unit Roots in Time Series Data" , *Econometric Theory*, 5, 256-271.

Park, J. Y. and P. C. B. Phillips (1989), "Statistical Inference in Regressions with Integrated Processes: Part 2", *Econometric Theory*, 5, 95-131.

Perron, P. and Z. Qu (2007), "A Simple Modification to Improve the Finite Sample Properties of Ng and Perron's Unit Root Tests", *Economics Letters*, 94, 12-19.

Phillips, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics*, 33, 311-340.

Phillips, P.C.B. (1998), "New Tools for Understanding Spurious Regressions", *Econometrica*, 66(6), 1299-1325.

Said, S. E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order", *Biometrika*, 71, 599-607.

Sargan, J. D. (1980), "Some Tests of Dynamic Specification for a Single Equation", *Econometrica*, 48, 879-897.

Stewart, C. (2006), "Spurious Correlation of $I(0)$ Regressors in Models With an $I(1)$ Dependent Variable", *Economics Letters*, 91(2), 184-189.

Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55(5), 1035–1056.

Sun, Y. (2004), "A Convergent t-statistic in Spurious Regressions", *Econometric Theory*, 20, 943-962.

Tsay, W. J. and C. F. Chung (2000), "The Spurious Regression of Fractionally Integrated Processes", *Journal of Econometrics*, 96(1), 155-82.

Ventosa-Santaulària, D. (2009), "Spurious Regression", *Journal of Probability and Statistics*, article ID 802975, 27 pages, doi:10.1155/2009/802975.

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), pp. 1-63.

APPENDIX A.1

Equation (1) can be written in matrix form $y = X\beta + \varepsilon$, with y a $T \times 1$ vector of y_t data, X a $T \times 2$ matrix comprising a constant term and data on x_t, and ε a $T \times 1$ vector of zero mean disturbances. The vector of OLS estimators is defined as:

$$\widehat{\beta} = \left[\begin{array}{c} \widehat{\alpha} \\ \widehat{\delta} \end{array} \right] = (X'X)^{-1}X'y,$$

where (all sums run from $t = 1$ to T) $X'X = \left[\begin{array}{cc} T & \Sigma x_t \\ \Sigma x_t & \Sigma x_t^2 \end{array} \right]$, and $X'y = \left[\begin{array}{c} \Sigma y_t \\ \Sigma x_t y_t \end{array} \right]$. The t-statistic is defined by

$$t_{\widehat{\delta}} = \widehat{\delta} \left[\widehat{\sigma}_\varepsilon^2 (X'X)_{22}^{-1} \right]^{-1/2},$$

where $\widehat{\sigma}_\varepsilon^2$ is the estimated regression variance,

$$\widehat{\sigma}_\varepsilon^2 = \frac{\Sigma(y_t - \widehat{\alpha} - \widehat{\delta}x_t)^2}{T}$$

and $(X'X)_{22}^{-1}$ denotes the 2nd diagonal element of $(X'X)^{-1}$. With the aid of a *Mathematica 7.0* code, for each combination of DGPs in the Assumption for y and x, we compute the order of magnitude of $\widehat{\delta}$, $\widehat{\sigma}_\varepsilon^2$, and $(X'X)_{22}^{-1}$, and therefore we derive the order of magnitude of $t_{\widehat{\delta}}$. The code also allows us to derive the expression for the asymptotic distribution of $t_{\widehat{\delta}}$. This code is available at *http://dl.dropbox.com/u/1307356/SpRegTest/NVSSpRegTest1.zip*

APPENDIX A.2

As in Appendix A.1, Equation (2) can be written in matrix form $z = X\beta + \varepsilon$, with z a $T \times 1$ vector of data ($z_t = y_t, x_t$), X a $T \times 2$ matrix comprising a constant term and a linear trend t, and ε a $T \times 1$ vector of zero mean disturbances. The vector of OLS estimators is defined as:

$$\widehat{\beta} = \left[\begin{array}{c} \widehat{c}_z \\ \widehat{b}_z \end{array} \right] = (X'X)^{-1}X'z,$$

where $X'X = \left[\begin{array}{cc} T & \Sigma t \\ \Sigma t & \Sigma t^2 \end{array} \right]$, and $X'z = \left[\begin{array}{c} \Sigma z_t \\ \Sigma z_t t \end{array} \right]$. With the aid of a *Mathematica 7.0* code (*available at http://dl.dropbox.com/u/1307356/SpRegTest/NVSSpRegTest2.zip*), we are able to compute analytic expressions for $\widehat{\beta}$. The resulting residuals, $\widehat{\varepsilon}_{yt}$ and $\widehat{\varepsilon}_{xt}$ are then used to estimate by OLS

regression model (3):

$$\widehat{\beta}_f = \Sigma \hat{\varepsilon}_{yt} \hat{\varepsilon}_{xt} \left(\Sigma \hat{\varepsilon}_{xt}^2 \right)^{-1}$$

Note that in this case $\Sigma \hat{\varepsilon}_{zt} = \widehat{c}_f = 0$, by construction. The *Mathematica 7.0* code does the rest, deriving a limiting expression for $\widehat{\beta}_f$ (and also a limiting expression for t_{β_f}). The behaviour of the R^2 is obtained by studying the asymptotics of the residual sum of squares and the total sum of squares from $R^2 = 1 - RSS/TSS$, where $RSS = \sum \hat{v}_t^2 = T\hat{\sigma}_\nu^2$ and $TSS = \sum \hat{\varepsilon}_{yt}^2 = T\hat{\sigma}_{\varepsilon_y}^2$, both of which can also be obtained from the *Mathematica 7.0* code.

APPENDIX A.3

The definitions for the expressions used in Theorem 1 are as follows. For $z = x, y$:

$S_z = \sigma_z \int W_z$, $S_{z2} = \sigma_z^2 \int W_z^2$, $S_{xy} = \sigma_x \sigma_y \int W_x W_y$, $St_z = \sigma_z \int r W_z$, and

$S_1 = S_{x2} S_{y2} - S_{y2} S_x^2 - S_{x2} S_y^2 + 2 S_x S_y S_{xy} - S_y S_x^2$

$S_2 = (\mu_x^2)^{-1} \left[\mu_y^2 A_x + \mu_x^2 A_y + 2\mu_x \mu_y (12 St_x St_y - 6 St_y S_x - 6 St_x S_y + 4 S_x S_y - 6 S_{xy}) \right]$

$S_3 = \int \overline{W}_x \int \overline{W}_y - \int \overline{W}_x \overline{W}_y$, $S_4 = \left(\int \overline{W}_x \right)^2 - \int \overline{W}_x^2$

$S_5 = 2 \int \overline{W}_x \int \overline{W}_y \int \overline{W}_x \overline{W}_y - \left(\int \overline{W}_x \overline{W}_y \right)^2 - \int \overline{W}_x^2 \left(\int \overline{W}_y \right)^2 - \left(\int \overline{W}_x \right)^2 \int \overline{W}_y^2 + \int \overline{W}_x^2 \int \overline{W}_y^2$

with $A_z = S_{z2} - 4(3 St_z^2 - 3 St_z S_z + S_z^2)$, and $\sigma_z^2 = \lim_{T \to \infty} (T^{-1} \sum_{t=1}^T u_{zt}^2)$.

APPENDIX A.4

The definitions for the expressions used in Theorem 2 are as follows.

$N = 6 \int r W_x \left(\int W_y - 2 \int r W_y \right) + \int W_x \left(6 \int r W_y - 4 \int W_y \right) + \int W_x W_y$

$D = 4 \left(\int W_x \right)^2 \int W_y^2 - 12 \left(\int r W_x \right)^2 \left[\left(\int W_y \right)^2 - \int W_y^2 \right] + 12 \left(\int W_x \right)^2 \left(\int r W_y \right)^2$

$\quad - 12 \int W_x \int r W_x \left[\left(\int W_y \right)^2 - 2 \int W_y \int r W_y \right] + w_{y1} \int W_x^2 + 4 (w_2 + w_3) \int W_x W_y + \left(\int W_x W_y \right)^2$

$Q_1 = \beta_{y3} \sigma_x w_{x1} + \sigma_y \left[2(w_2 - w_3) - \int W_x W_y \right]$

$Q_2 = 4 \left(\int W_x \right)^2 - \int W_x^2 - 12 \int r W_x \left(\int W_x - \int r W_x \right)$

$Q_3 = 12 \int r W_y \left[2 \int W_x \int r W_x \int W_y - \int W_x^2 \int W_y - \left(\int W_x \right)^2 \int W_y + \int W_x^2 \int r W_y \right]$

$\quad + \left(\int W_y \right)^2 \left[4 \int W_x^2 - 12 \left(\int r W_x \right)^2 \right] + w_{x1} \int W_y^2 + \int W_x W_y \left[4(w_2 + w_3) + \int W_x W_y \right]$

$Q_4 = 4 \left(\int \overline{W}_x \right)^2 - 12 \int \overline{W}_x \int r \overline{W}_x + 12 \left(\int r \overline{W}_x \right)^2 - \int \overline{W}_x^2$

$$Q_5 = 4 \int \overline{W}_x \int \overline{W}_y - 6 \int r\overline{W}_x \int \overline{W}_y - 6 \int \overline{W}_x \int r\overline{W}_y + 12 \int r\overline{W}_x \int r\overline{W}_y - \int \overline{W}_x\overline{W}_y$$

$$Q_6 = \left(\int \overline{W}_x\overline{W}_y\right)^2 - 4\left(3 \int r\overline{W}_x^{\,2} - \int \overline{W}_x^2\right)\left(\int \overline{W}_y\right)^2 - 12\left(\int \overline{W}_x^2 - 2\int \overline{W}_x \int r\overline{W}_x\right)\int \overline{W}_y \int r\overline{W}_y$$
$$+ 12\left[\left(\int \overline{W}_x^2 - \left(\int \overline{W}_x\right)^2\right)\right]\left(\int r\overline{W}_y\right)^2 + Q_4 \int \overline{W}_y^2$$
$$- 4\int \overline{W}_x\overline{W}_y \left(2 \int \overline{W}_x \int \overline{W}_y - 3 \int r\overline{W}_x \int \overline{W}_y - 3 \int \overline{W}_x \int r\overline{W}_y + 6 \int r\overline{W}_x \int r\overline{W}_y\right)$$

$$w_{z1} = 4\left(\int W_z\right)^2 - 12 \int W_z \int rW_z + 12\left(\int rW_z\right)^2 - \int W_z^2$$

$$w_2 = 3 \int rW_x \int W_y - 2 \int W_x \int W_y$$

$$w_3 = 3 \int W_x \int rW_y - 6 \int rW_x \int rW_y$$

Research Papers
2011

2011-01:   Cristina Amado and Timo Teräsvirta: Modelling Volatility by Variance Decomposition

2011-02:   Timo Teräsvirta: Nonlinear models for autoregressive conditional heteroskedasticity

2011-03:   Roxana Halbleib and Valeri Voev: Forecasting Covariance Matrices: A Mixed Frequency Approach

2011-04:   Mark Podolskij and Mathieu Rosenbaum: Testing the local volatility assumption: a statistical approach

2011-05:   Michael Sørensen: Prediction-based estimating functions: review and new developments

2011-06:   Søren Johansen: An extension of cointegration to fractional autoregressive processes

2011-07:   Tom Engsted and Stig V. Møller: Cross-sectional consumption-based asset pricing: The importance of consumption timing and the inclusion of severe crises

2011-08:   Tommaso Proietti and Stefano Grassi: Bayesian stochastic model specification search for seasonal and calendar effects

2011-09:   Matt P. Dziubinski: Option valuation with the simplified component GARCH model

2011-10:   Tim A. Kroencke, Felix Schindler and Andreas Schrimpf: International Diversification Benefits with Foreign Exchange Investment Styles

2011-11:   Eduardo Rossi and Paolo Santucci de Magistris: Estimation of long memory in integrated variance

2011-12:   Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Generalized Jackknife Estimators of Weighted Average Derivatives

2011-13:   Dennis Kristensen: Nonparametric Detection and Estimation of Structural Change

2011-14:   Stefano Grassi and Paolo Santucci de Magistris: When Long Memory Meets the Kalman Filter: A Comparative Study

2011-15:   Antonio E. Noriega and Daniel Ventosa-Santaularia: A Simple Test for Spurious Regressions