

## Granger revisited: $t$ values and the empirical OLS bias with stationary and non-stationary time series using Monte Carlo simulations

Carlos Guerrero-de-Lizardi<sup>1</sup>

Universidad Nacional Autónoma de México, México

### Abstract

The conduction of a *reliable* statistical analysis is based on the recognition of the statistical features of the time series at stake and on the underlying probabilistic assumptions of the applied model. Our purpose is to illustrate this kind of analysis using Granger's groundbreaking ideas. Our Monte Carlo results show that in the presence of stationary and non-stationary time series, the standard ordinary least squares inference could be misleading. We will graphically address the empirical distribution of the estimator's bias as well as the inconvenience of using standard errors to illustrate how the true variation is underestimated. We recommend following Granger's suggestions, which we highlight with originality using a "measurement in economics" perspective. Our quantitative exercises are replicable to the extent that we fully shared our codes in addition to using an open-access database of the seminal paper written by Nelson and Plosser (1982). Our main conclusion is simple: empirical researchers should be cautious when drawing qualitative findings based on a standard ordinary least squares inference carried out in the context of a regression analysis.

*JEL Classification:* C15, C22, C87

*Keywords:* reliable statistical analysis, standard OLS inference, empirical bias, replicability, measurement in economics

## Granger revisitado: valores $t$ y el sesgo empírico de OLS con series temporales estacionarias y no estacionarias utilizando simulaciones de Monte Carlo

### Resumen

La realización de un análisis estadístico confiable se fundamenta en el reconocimiento de las características estadísticas de las series de tiempo en juego y de los supuestos probabilísticos subyacentes del modelo aplicado. Nuestro propósito es ilustrar este tipo de análisis utilizando algunas ideas de Granger. Nuestros resultados de Monte Carlo muestran que en presencia de series de tiempo estacionarias y no estacionarias, la inferencia basada en los mínimos cuadrados ordinarios puede ser engañosa. Abordamos gráficamente la distribución empírica del sesgo del estimador y el inconveniente del uso de errores estándar que subestiman su verdadera variación. Recomendaremos seguir las sugerencias de Granger, que destacamos con originalidad utilizando una perspectiva desde la "medición en economía". Nuestros ejercicios cuantitativos son replicables en la medida en que compartimos completamente nuestros códigos y utilizamos la base de datos de acceso abierto del documento seminal escrito por Nelson y Plosser (1982). Nuestra conclusión principal es simple: los investigadores empíricos deben ser absolutamente cautelosos al momento de extraer conclusiones cualitativas basadas en una

<sup>1</sup>Thanks are due to Armando J. Guerrero and the two anonymous referees for their suggestions.

\*Sin fuente de financiamiento para el desarrollo de la investigación

## Resumen

inferencia estándar de mínimos cuadrados ordinarios realizada en el contexto de un análisis de regresión.

Clasificación JEL: C15, C22, C87

Palabras clave: análisis estadístico confiable, inferencia estándar basada en los MCO, sesgo empírico, replicabilidad, medición en economía

“I am often asked how the idea of *cointegration* came about; was it the result of logical deduction or a flash of inspiration? In fact, it was rather more prosaic. A colleague, David Hendry, stated that the difference between a pair of integrated series could be stationary. My response was that it could be proved that he was wrong, but in attempting to do so, I showed that he was correct, and generalized it to *cointegration*, and proved the consequences such as the error-correction representation.” Clive W. J. Granger, *Nobel Lecture*.

## 1. Introduction

The most common estimator used in applied econometrics is that of ordinary least squares simply due to its “ideal” theoretical characteristics. However, Clive Granger used nonsense correlations introduced by Yule (1926) and a framework based on balanced and unbalanced equations to illustrate some “non-ideal” outcomes; Granger explored the fact that in the presence of stationary and non-stationary time series, the standard ordinary least squares inference could be misleading.

Our purpose is to illustrate the meaning of the conduction of a *reliable* statistical analysis using some of Granger’s groundbreaking ideas. In section 2, we offer a background to understand some fundamental principles. In sections 3 and 4, we present our simulation results of the empirical bias of OLS. In section 5, we share some final thoughts.

## 2. Background

Granger coined the term *cointegration* to describe a regression in which a pair of variables maintain a genuine long-run relationship instead of a nonsense one, which he referred to as spurious. Looked upon as a long-memory uncovering process, the invention of *cointegration* was preceded by Granger’s concerns about nonsense correlations introduced by Yule (1926) and a framework based on balanced and unbalanced equations. As a result of such endeavors, Granger was awarded the 2003 Nobel prize in Economic Sciences.

The path to the Nobel was not an easy one, in fact it was quite the opposite. In the following quote Granger (2010, p. 3) explains the difficulties he faced publishing his findings: “*Econometrica* rejected the paper for various reasons, such as wanting a deeper theory and some discussion of the testing question and an application. As I knew little about testing I was very happy to accept Rob’s offer of help with the revision. I re-did the representation theorem and he produced the test and application, giving a paper by Granger and Engle which evolved into a paper by Engle and Granger, whilst I was away for six months leave in Oxford and Canberra. This new paper was submitted to *Econometrica* but it was also rejected for not being sufficiently original. I was anxious to submit it to David Hendry’s new econometrics journal but Rob wanted to explore other possibilities. These were being considered when the editor of *Econometrica* asked us to re-submit because they were getting so many submissions on this topic that they needed our paper as some kind of base reference.”

To introduce his viewpoint, Yule (1926) proposed the correlation coefficient between mortality per 1,000 persons in England and Wales and the proportion of Church of England Marriages per 1,000 of all marriages

as an example. The correlation obtained was 0.9512! Yule (1926) conducted a *reliable* statistical analysis to the extent: 1) he was aware that the correlation between two variables is a meaningful measure of its linear relationship only if its means are constant over time; 2) he identified the high autocorrelation of the selected variables, and 3) he concluded that the underlying probabilistic assumption of the correlation coefficient was violated and, therefore, the obtained measure was likely misleading.<sup>2</sup> In this regard, Johnston and DiNardo (1997, p. 10) added: “no British politician proposed closing down the Church of England to confer immortality on the electorate.”

As an extension of his work on spurious regressions during the early 1980s, Granger investigated what he named “balanced models”, defined as follows (Granger, 1993, p. 12): “An equation which has  $X_t$  having the same dominant features as  $Y_t$  will be called balanced. Thus, if  $X_t$  is I(0),  $Y_t$  is I(1), the equation will not be balanced. It is a necessary condition that a specified model be balanced for the model to be satisfactory.” As a rule of thumb in applied economics, we point out that the level of a variable is I(1), and its first difference is I(0). The specification of the Granger’s Ordinary Least Squares (OLS) bivariate regression is the following:

$$Y_t = \alpha_1 + \alpha_1 X_t + U_t \quad (1)$$

Having in mind “the balanced equation law”, we would expect that  $U_t$  has the same property of  $Y_t$ , and so the third assumption of the classical linear regressions model will not be obeyed, meaning we would expect some low Durbin-Watson statistics, a well-known symptom of spurious regression.

To explain a nonsense regression, Granger used to quote a book written long ago by a coauthor of Yule, M. G. Kendall. We have two AR(1) processes:

$$Z1_t = \beta_1 Z1_{t-1} + U_t \quad (2)$$

$$Z2_t = \beta_2 Z2_{t-1} + U_t \quad (3)$$

with  $\beta_1 = \beta_2 = \beta$ , that is  $Z1_t$  and  $Z2_t$  both obeyed the same autoregressive model, and  $U_t$  are “white noise” innovations independent of each other at all pairs of time. The sample correlation ( $R$ ) between  $n$  consecutive terms of  $Z1_t$  and  $Z2_t$  has the following variance (Kendall, 1954, p. 113):

$$var (R) = \frac{1 (1 + \beta^2)}{n (1 - \beta^2)} \quad (4)$$

where  $n$  is the sample size. According to Granger (2003, p. 558), if  $\beta$  “is near one and  $n$  not very large, then the  $var(R)$  will be quite big, which can only be achieved if the distribution of  $R$  values has large weights near the extreme values of -1 and 1, which will correspond to” a “significant” coefficient value in a bivariate regression of  $Z1_t$  on  $Z2_t$ .

The idea of *cointegration* between a pair of variables I(1), which generates a linear combination I(0), was a natural extension of Granger’s concerns on spurious regression and balanced models. Suppose  $Y_t$  and  $W_t$  are integrated of order one, and we run the following regression:

$$Y_t = \gamma_0 + \gamma_1 W_t + U_t \quad (5)$$

$Y_t$  and  $W_t$  are cointegrated if  $Y_t - \gamma_1 W_t$  eliminates its common stochastic trend. In this particular case,  $\gamma_1$  is called the “cointegrating” parameter, and  $U_t$  is I(0) –implying a stable and unique long-run equilibrium relationship between  $Y_t$  and  $W_t$ .

---

<sup>2</sup>It is possible that Yule (1926) constitutes an undeclared critique to Hooker (1901), who reported a correlation between the moving average (which he called trend) of the marriage rate and of the trade per head, from 1861 to 1895, equal to 0.80.

Lastly, a permanent concern of Professor Granger was the generation of spurious results. Posthumously, he sent us an elegant warning (White and Granger, 2011, p. 15): “When looking for causal relations between stochastic trends, it is important to recognize that Reichenbach’s principle of common cause (Reichenbach, 1956) does not necessarily apply. Briefly, Reichenbach’s principle holds that if we see correlation between two series, then one series must cause the other, or there must be an underlying common cause. With stochastic trends, however, observed correlations may be spurious, as pointed out long ago by Yule (1926).” Their message should guide our empirical work: a high correlation between a pair of variables should not lead us to assume a cause-effect relationship or the existence of a common cause, but rather to recognize it as a result of expressing the statistical properties of the variables at stake in addition to the sensitivity of the statistic used.

### 3. The empirical bias of OLS in the presence of non-stationary time series

In their seminal paper, Granger and Newbold (1974) performed a mini Monte Carlo as a way “to find evidence of spurious regressions” (Granger, 2003, p. 558).<sup>3</sup> They generated pairs of independent random walks without drift, each of length 50 ( $T = 50$ ), and run 100 bivariate regressions. Considering the knowledge at the beginning of the seventies, in their simulation it was expected that roughly 95 percent of  $|t|$  values on the estimated parameter would be less than 1.96, but it was the case only in 23 occasions.<sup>4</sup> Since the random walks were unrelated by design, the t values were misleading, and so, using their simulation results Granger and Newbold (1974, p. 115) suggested a new critical value of 11.2 when assessing the significance of the coefficient at the 5% level. At present we know that this suggestion is inconvenient, in the sense that Phillips (1986, p. 318) demonstrated that it has not asymptotic sense to apply the correction based on the standardized statistic, that is,  $\left(\frac{|t|}{\sqrt{T}}\right)$ .

Inspired by Granger and Newbold (1974) we designed a Monte Carlo simulation. In our case, each random walk has 100 terms -approximately the same length of the variables contained in the database of Nelson and Plosser (1982) that we will use later-, and the number of replications was the nowadays standard 10,000. Box 1 shows our (EViews) program.

---

<sup>3</sup>According to Davidson and MacKinnon (1993, p. 732), “The term is reported to have originated with Metropolis and Ulam (1949). If it had been coined a little later, it might have been called the ‘Las Vegas method’ instead of the ‘Monte Carlo method.’”

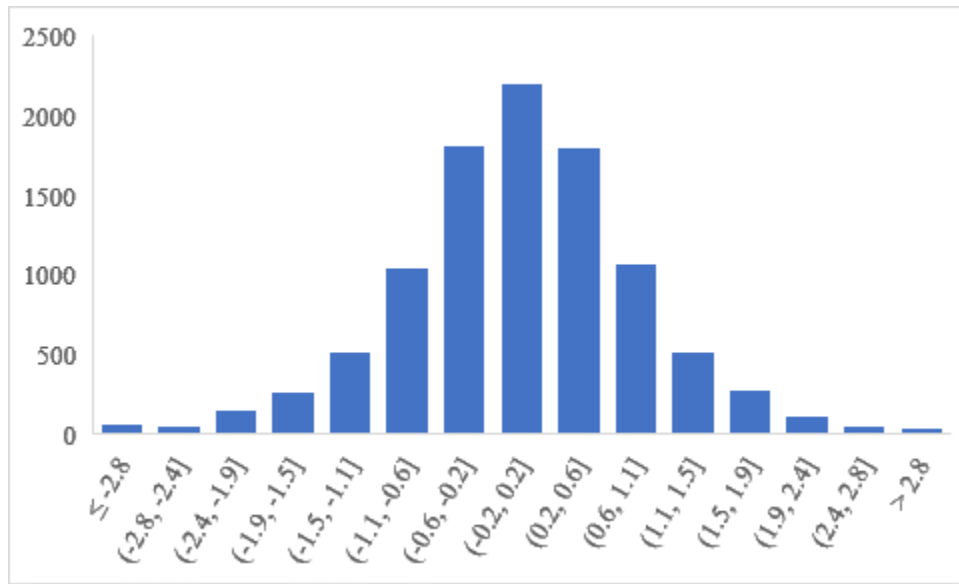
<sup>4</sup>Student’s t appeared in 1908. The pseudonym was used by William S. Gosset in 19 times because his employer, Arthur Guinness, required that his identity “be shielded from competitors”; Gosset “was a master brewer and rose in fact to the top of the top of the brewing industry: Head Brewer of Guinness” (Ziliak, 2008, p. 199 and p. 201). Student (1908, p. 13) design a Monte Carlo to empirically analyzed the probable error of a mean: “The material used was a correlation table containing the height and left middle finger measurements of 3000 criminals, from a paper by W. R. Macdonell (Biometrika, Vol. I. p. 219). The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book which thus contains the measurements of 3000 criminals in a random order. Finally each consecutive set of 4 was taken as a sample -750 in all- and the mean, standard deviation, and correlation of each sample determined. The difference between the mean of each sample and the mean of the population was then divided by the standard deviation of the sample, giving us the z of Section III.” By the way, Sir Francis Galton, a Charles Darwin’s cousin, coined the term “regression”. His work had great influence in Ronald A. Fisher and Karl Pearson, who were sponsors of Gosset studies. We would say that this great scientific community, of which G. Yule and R. Hooker were part, applied a naive Darwinian analysis in their research.

**Box 1.** Code to run a Monte Carlo simulation using random walks

```
'Create a workfile undated, range 1 to 10,000
!reps = 10000
for !i=1 to !reps
  genr innovationrw1{!i}=@nrnd
  genr innovationrw2{!i}=@nrnd
  smpl 1 1
  genr rw1{!i}=0
  genr rw2{!i}=0
  smpl 2 100
  genr rw1{!i}=rw1{!i}(-1)+innovationrw1{!i}
  genr rw2{!i}=rw2{!i}(-1)+innovationrw2{!i}
  smpl 1 100
  matrix(!reps,2) results
  equation eq{!i}.ls rw1{!i}=c(1)*rw2{!i}
  results(!i,1)=eq{!i}.@coefs(1)
  results(!i,2)=eq{!i}.@tstats(1)
  d innovationrw1{!i}
  d innovationrw2{!i}
  d rw1{!i}
  d rw2{!i}
  d eq{!i}
next
'Copy and paste "results" to Excel
'In EViews type: data NONSENSE TsNONSENSE
'Copy and paste from Excel to EViews NONSENSE TsNONSENSE
```

It is necessary to specify that the “!reps” command sets the number of replicas; in the following lines, we create (“genr”) a pair of stationary time series that contain pseudo-random draws from a standard normal distribution. Initially, we focused on the first observations (“smpl 1 1”) to fix the initial conditions of the two random walks. We then generated the rest of its values by stacking the innovations generated. To create a new object, specifically speaking a matrix with two columns, we invoke the command “matrix”, we estimate the regression using OLS (“ls”), and we deposit the values of the estimated coefficients and its corresponding  $t$  values in “results”. It is necessary to erase the inputs of the exercise for obvious reasons (“d” from delete), and to use the “next” command to end the code. Finally, with the last three suggestion lines, we change the format of the estimated coefficients and  $t$  values as columns of a matrix to variables (data), using the spreadsheet as a “scale” software.

The following figure shows statistics about the values of the NONSENSE estimated coefficients (y-axis frequency and x-axis coefficient values) and the table its  $t$  values tabulation.



**Figure 1.** Histogram (y-axis frequency and x-axis values) of the estimated coefficients.

**Table 1.** Tabulation of *t* values.

Value	Count	Percent
$[-62, -2)$	4,285	42.85
$[-2, 0)$	748	7.48
$[0, 2)$	742	7.42
$[2, 68)$	4,225	42.25
Total	10,000	100

In figure 1, the horizontal axis constitutes the bias in estimating the slope coefficient -to the extent the true value is zero- and the vertical axis is its relative frequency. In other words, we would expect coefficients with a value around zero and non-significant *t* values in the 95 percent of the regressions since, by design, drift-free independent random walks were used. But the results obtained were not as expected: in 7,800 cases the values of the coefficients exceeded the value of  $|0,2|$  and we barely obtained in 1,490 occasions *t* values less than  $|2|$ . Hence, there is a problem if we apply the “standard OLS inference” (Granger et al. 2001, p. 900). In our days we know that the heart of the matter is the use of standard errors which underestimate the true variation of the OLS estimators; unfortunately, there is not a practical solution, that is, an alternative mechanical estimator of the standard errors useful to construct an adjusted *t* statistic (Sun, 2004).

#### 4. The empirical bias of OLS in the presence of stationary time series

According to Phillips (2003, p. 2), “a primary limitation on empirical knowledge is that the true model for any given data is unknown and, in all practical cases, unknowable. Even if a formulated model were correct, it would still depend on parameters that need to be estimated from data”.

Using a particular type of stationary time series Greene (2003, pp. 44-5) proposed the following Monte Carlo experiment: 1) Generate two stationary random variables,  $W_t$  and  $X_t$ ; 2) Generate  $U_t = 0,5W_t$ , and  $Y_t = 0,5X_t + U_t$ ; 3) Run 10,000 regressions, and 4) Randomly select 500 samples of 100 observation and make a graph. As we read in other books frequently, his conclusion was simple (Greene, 2003, p. 45): “note that the distribution of slopes has a mean roughly equal to the ‘true values’ of 0.5”. It is convenient to note the

following:

1. Our preference for an unbiased estimator stems from the “hope” that a particular estimate will be close to the mean of the estimator’s sampling distribution; but “it is possible to have an ‘unlucky’ sample and thus a bad estimate” (Kennedy, 2003, p. 16 and p. 31). This handicap of empirical sciences explains the origin of the cruel story of the three econometricians who go duck hunting. The first shoots about a foot in front of the duck, the second about a foot behind, and the third yells, “We got him!”
2. We knew the true model, so we avoid the “alias o bias matrix” (Draper and Smith, 1998, chapter 10). But since this ideal situation is never present in empirical research, we should take advantage of the following axiom (Granger, 1993, p. 2): “any model will be only an approximation to the generating mechanism. . . It follows that several models can all be equally good approximations. How can one judge the quality of the approximation means further discussion.”
3. Professor Greene explored a very extreme time series in the sense that  $W_t$ ,  $X_t$ , and  $Y_t$  can be written as first order autoregressive processes with a coefficient equal to zero or, in other words, because the correlation coefficients between  $W_t$  and  $W_{t-k}$  is equal to zero for all  $k \neq 0$ , and similarly for  $X_t$ , and  $Y_t$ . Indeed, it is impossible to imagine economic variables with a similar data generating process. In this sense, we recommend never lose sight of the statistical properties of the dependent variable and the regressors because (Patterson, 2000, p. 317) “the properties of the OLS estimator of the coefficients and the distribution of this estimator depend crucially upon these properties.”

In Granger et al. (2001) and Granger (2003, p. 560), it was shown that spurious regressions can also occur, “although less clearly”, with stationary time series. They generated two AR(1) processes as in our equations (2) and (3) with  $0 < \beta_1 = \beta_2 = \beta \leq 1$ , and run regressions following equation (1) with sample sizes varying between 100 and 10,000 -to suggest that the nonsense regression issue is not a small sample property. Table 2 is a summary of their results:

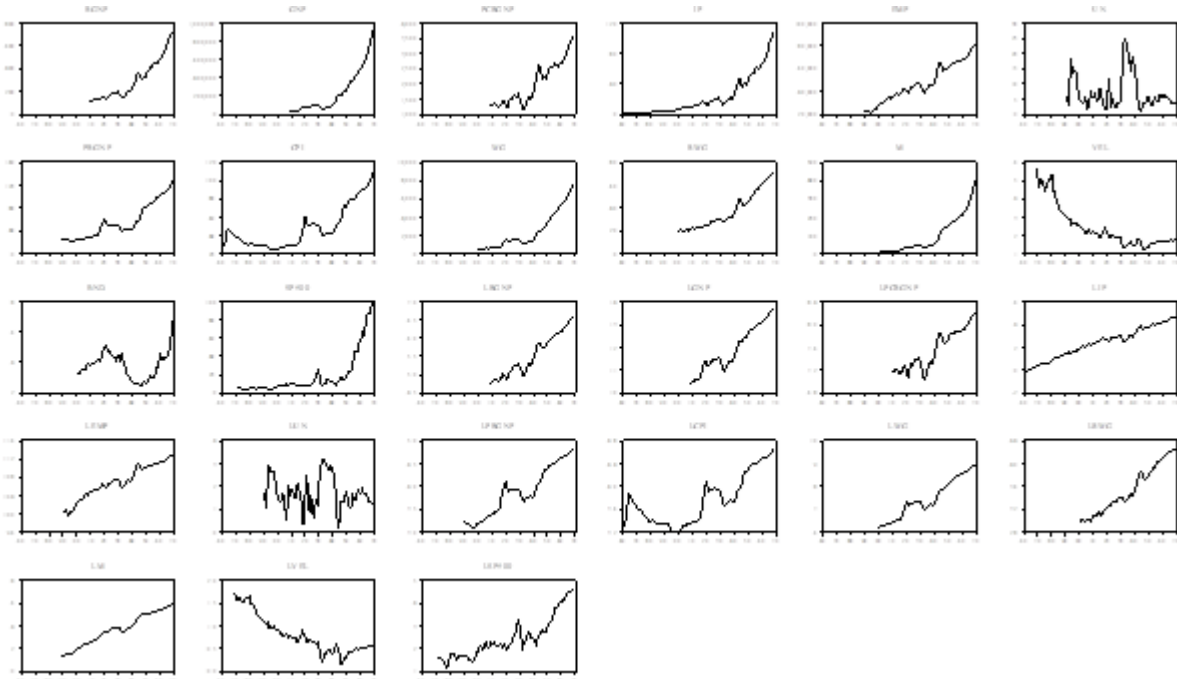
**Table 2.** Regressions between independent AR(1) series ( $\beta_1 = \beta_2 = \beta$  percentage of  $|t| > 2$ )

Sample series	$\beta = 0$	$\beta = 0.25$	$\beta = 0.5$	$\beta = 0.75$	$\beta = 0.9$	$\beta = 1.0$
100	4.9	5.6	13.0	29.9	51.9	89.1
500	5.5	7.5	16.1	31.6	51.1	93.7
2,000	5.6	7.1	13.6	29.1	52.9	96.2
10,000	4.1	6.4	12.3	30.5	52.0	98.3

Source: Granger (2003, p. 560).

According to Granger (2003, p. 557), Yule (1926) is a much-cited paper but not sufficiently understood. The content of Table 2 supports this point of view in the sense that, as in Yule (1926), the clue is the degree of the autocorrelation of involved variables in a regression. It is not necessary to have random walks to generate nonsense exercises. It is just great, with AR(1) processes with coefficients between 0.25 and 0.9, and distinct sample sizes, Granger et al. (2001) were in a position to put all the applied econometricians in check.

In their seminal (Campbell and Perron, 1991, p. 147) paper, Nelson and Plosser (1982, p. 147) analyzed 27 variables in “natural logs except for the bond yield”: real GNP, nominal GNP, real per capita GNP, industrial production, employment, unemployment rate, GDP deflator, consumer prices, wages, real wages, money stock, velocity, bond yields, and common stock prices. Theoretically speaking, there is a mixture of stationary and non-stationary time series. So as not to enter into an unnecessary discussion, the next figure shows all the variables and Table 3 its autocorrelation coefficients.



**Figure 2.** Nelson and Plosser (1982), annual dataset, 1860-1970, in “natural logs except for the bond yield”.

Source: Nelson and Plosser (undated).

**Table 3.** Autocorrelation of the variables used by Nelson and Plosser (1982)

rgnp	0.942	rwg	0.960	lemp	0.956
gnp	0.922	m	0.935	lun	0.754
pcrgnp	0.944	vel	0.947	lprgnp	0.965
ip	0.950	bnd	0.839	lcpi	0.963
emp	0.954	sp500	0.950	lwg	0.956
un	0.856	lrgnp	0.951	lrwg	0.962
prgnp	0.950	lgnp	0.947	lm	0.963
cpi	0.950	lprgnp	0.947	lvel	0.958
wg	0.940	lip	0.969	lsp500	0.955

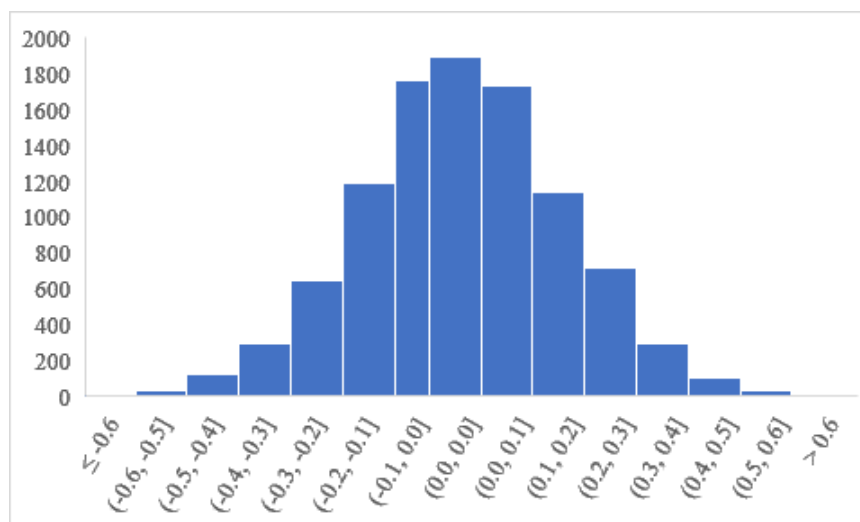
To run four simulations, we use as a point of reference the minor values of the autocorrelations contained in Table 3. In each case, for example, we run the following general program:



**Box 2.** Code to run a Monte Carlo simulation using  $I(0)$  time series

```
'Create a workfile undated, range 1 to 10,000
!reps = 10000
for !i=1 to !reps
  gener innovationx1{!i}=@nrnd
  gener innovationx2{!i}=@nrnd
  smpl 1 1
  gener x1{!i}=0
  gener x2{!i}=0
  smpl 2 100
  gener x1{!i}=(0.754)*x1{!i}(-1)+innovationx1{!i}
  gener x2{!i}=(0.754)*x2{!i}(-1)+innovationx2{!i}
  smpl 1 100
  matrix(!reps,2) results
  equation eq{!i}.ls x1{!i}=c(1)*x2{!i}
  results(!i,1)=eq{!i}.@coefs(1)
  results(!i,2)=eq{!i}.@tstats(1)
  d innovationx1{!i}
  d innovationx2{!i}
  d x1{!i}
  d x2{!i}
  d eq{!i}
next
'Export "results" to Excel
'In EViews: data NONSENSE tsNONSENSE
'Copy and paste from Excel to EViews NONSENSE tsNONSENSE
```

The user should only change the autoregressive values in the program (lines 10 and 11). The following figure and table show statistics about the values of the estimated coefficients and its  $t$  values (y-axis frequency and x-axis values). Following Patterson (2000, pp. 317-26), we improved its titles.



**Figure 3.** The empirical distribution of the bias of OLS estimator of the slope.

**Table 4.** Simulations results using database of Nelson and Plosser (1982)

$\beta =$	0.754	0.839	0.856	0.922
Percentage of $ t  \leq 2$	71.14	60.86	56.23	42.98

Before commenting on our results, we want to focus on a critical aspect of the dissemination of sciences. In 2011 Crocker and Cooper, editors of the prestige Science, wrote a note entitled “Addressing scientific fraud”, showing evidence that fraud is a common practice. In this regard, as the first line of defense, Crocker and Cooper (2011, p. 1182) ask for “greater transparency with data”.

At least in economics, Crocker and Cooper (2011) hit the nail on the head. To cite but one terrible example, Duvendack, Palmer-Jones, and Reed (2015, pp. 181-182) uncovered the following: “What can we learn from our analysis of replication studies? Most importantly, and perhaps not too surprisingly, the main takeaway is that, conditional on the replication having been published, there is a high rate of disconfirmation. Over the full set of replication studies, approximately two out of every three studies were unable to confirm the original findings. Another 12 percent disconfirmed at least one major finding of the original study, while confirming others (Mixed?). In other words, nearly 80 percent of replication studies have found major flaws in the original research.”

Hence, we used an open-access database of a seminal paper and explained our computational code step by step. We can now begin to analyze our results.

In Figure 3, the horizontal axis measures the bias in estimating the slope coefficient, and the vertical axis is its relative frequency. Once again, in our simulations, we obtained unexpected coefficient values and t values. For Granger (2003, p. 560), the implication of these results is that “applied econometricians should not worry about spurious regressions only when dealing with I(1), unit root, processes.” But his caveat has been practically ignored by a large portion of the specialized literature. Excellent econometric literature -both textbooks and journal articles- exposes OLS without looking at the statistical properties of the variables we are dealing with, visualizes non-stationary time series in some moments as dominated by a deterministic trend and in others by a stochastic one, and last but not least, limits the problem of spurious regressions to the case of integrated variables.<sup>5</sup>

## 5. Final messages from “measurement in economics” and to perform a “reliable statistical analysis”

Measurement in economics is not a unified field but fragmented in subfields such as econometrics, index theory, and national accounts (Boumans, 2007, p. 3). It is not an exaggeration to point out that currently, econometric analysis constitutes a central measuring instrument used in our science. It is worth mentioning that, as other measuring instruments, initially functioned as an “artifact of measurement”, but rapidly became an “analytical device” (Klein, 2001; and Morgan 2001). In this regard, the guru of measurement in economics warned us that (Boumans 2005, p. 121): “a relevant problem of instruments used to make unobservables visible is how to distinguish between the facts about the phenomenon and the artifacts created by the instrument.”

Therefore, it is always necessary to keep in mind that the output of the econometric analysis shapes our scientific study of objects. Granger’s path-breaking ideas already showed us the dangers of carrying out

<sup>5</sup>A short list would be the following: Baltagi (2011), Chipman (2011), Gujarati and Porter (2008), Heij et al. (2004), Maddala and Kim (2002), Maddala and Lahiri (2009), Verbeek (2017), and Wooldridge (2012). Incidentally, Granger (2009, pp. 258-9) suggested the following classification: “Class A: Basically Dogmatic. 1. Jeff Wooldridge, *Introductory Econometrics*. . . 2. Russell Davidson and James MacKinnon, *Estimation and Inference in Econometrics*. . . Class B: Mixture of Dogmatic and Pragmatic. 1. Fumio Hayashi, *Econometrics*. . . 2. James Davidson, *Econometric Theory*. . . Class C: Largely Pragmatic. Michio Hatanaka, *Time Series Based Econometrics*. . . 2. James Stock and Mark Watson, *Introduction to Econometrics*. . . It should be noted that these texts have been ranked only on their attitude towards pragmatism and may not agree with rankings on other qualities.”

a regression that does not take into account the statistical characteristics of the analyzed variables and of ignoring the fact that the properties of the OLS estimator depend crucially upon these characteristics. We should be cautious when drawing qualitative conclusions based on a standard OLS inference carried out in the context of a regression analysis. Indeed, it is urgent that the specialized literature completely assimilates Granger's contributions regarding the risk of obtaining spurious results in economics.

## Referencias

- [1] Baltagi, B.H. (2011). *Econometrics*, 5E, Springer.
- [2] Boumans, M. (2005). *How Economists Model the World into Numbers*, London and New York, Routledge.
- [3] Boumans, M. (2007). "Introduction", in *Measurement in Economics: A Handbook*, edited by M. Boumans, Amsterdam, Academic Press, 3-18.
- [4] Campbell, J.Y. and Perron, P. (1991). "Pitfalls and opportunities: What macroeconomists should know about unit roots", in *NBER Macroeconomics Annual*, 6, 141-201.
- [5] Chipman, J.S. (2011). *Advanced Econometric Theory*, Routledge, London and New York.
- [6] Crocker, J. and Cooper, M.L. (2011). "Addressing scientific fraud", *Science*, 334(6060), 1182 (DOI: 10.1126/science.1216775).
- [7] Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press.
- [8] Draper N.R. and Smith, H. (1998). *Applied Regression Analysis*, 3E, Wiley Series in Probability and Statistics, New York.
- [9] Duvendack, M., Palmer-Jones, R., and Reed, W. (2015). "Replications in Economics: A progress report", *Econ Journal Watch*, 12(2), 164-191 (<https://econjwatch.org/issues/volume-12-issue-2-may-2015>).
- [10] Granger, C.W.J. (1993). "General introduction", in *Modelling Economic Series: Readings in Econometric Methodology*, edited by C.W.J. Granger, Oxford University Press, 1-23.
- [11] Granger, C.W.J. (2003). "Spurious regressions in econometrics", *A Companion to Theoretical Econometrics*, edited by Badi H. Baltagi, Blackwell Publishing, 557-61.
- [12] Granger, C.W.J. (2009). "In praise of pragmatics in econometrics", in *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, edited by J.L. Castle and N. Shephard, Oxford University Press, 255-67.
- [13] Granger, C.W.J. (2010). "Some thoughts on the development of *cointegration*", *Journal of Econometrics*, 158, 3-6 (<https://doi.org/10.1016/j.jeconom.2010.03.002>).
- [14] Granger, C.W.J. and Newbold, P. (1974). "Spurious regressions in economics", *Journal of Econometrics*, 2, 111-20 ([https://doi.org/10.1016/0304-4076\(74\)90034-7](https://doi.org/10.1016/0304-4076(74)90034-7)).
- [15] Granger, C.W.J., Hyung, N., and Jeon, Y. (2001). "Spurious regression with stationary series", *Applied Economics*, 33(7), 899-904 (DOI: 10.1080/00036840121734).
- [16] Greene, W. (2003). *Econometric Analysis*, 5E, Prentice Hall.
- [17] Gujarati, D.N. and Porter, D.C. (2008). *Basic Econometrics*, 5E, McGraw-Hill Education.
- [18] Heij, C. et al. (2004). *Econometric Methods with Applications in Business and Economics*, Oxford University Press.
- [19] Hooker, R. (1901). "Correlation of the marriage-rate with trade", *Journal of the Royal Statistical Society*, 64(3), 485-92.
- [20] Johnston, J. and DiNardo, J. (1997). *Econometric Methods*, 4E, McGraw-Hill, Singapore.
- [21] Kendall, M. (1954). *Exercises in Theoretical Statistics*, Charles Griffin and Company Limited, London.

- [22] Kennedy, P. (2003). *A Guide to Econometrics*, 5E, Wiley-Blackwell.
- [23] Klein, J.L. (2001). "Reflections from the age of economic measurement", in *The Age of Economic Measurement*, edited by J.L. Klein and M.M. Morgan, Durham, Duke University Press, 111-36.
- [24] Morgan, M.M. (2001). "Making measuring instruments", in *The Age of Economic Measurement*, edited by J.L. Klein and M.M. Morgan, Durham, Duke University Press, 235-51.
- [25] Maddala, G.S. and Kim, I-M. (2002). *Unit Roots, cointegration and Structural Change*, Cambridge University Press.
- [26] Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4E, Wiley.
- [27] Nelson, Ch. and Plosser, I. (1982). "Trends and random walks in macroeconomic time series. Some evidence and applications", *Journal of Monetary Economics*, 10(2), 139-62 ([https://doi.org/10.1016/0304-3932\(82\)90012-5](https://doi.org/10.1016/0304-3932(82)90012-5)).
- [28] Nelson, Ch. and Plosser, I. (undated). "Nelson\_Plosser", Instructional Stata datasets for econometrics, Boston College Department of Economics.
- [29] Patterson, K. (2000). *An Introduction to Applied Econometrics: A Time Series Approach*, Palgrave.
- [30] Phillips, P.C.B. (1986). "Understanding spurious regressions in econometrics", *Journal of Econometrics*, 33(3), 311-40 ([https://doi.org/10.1016/0304-4076\(86\)90001-1](https://doi.org/10.1016/0304-4076(86)90001-1)).
- [31] Phillips, P.C.B. (2003). "Law and limits of econometrics", Cowles Foundation Discussion Paper, No. 1397.
- [32] Student (1908). "The probable error of a mean", *Biometrika*, 6(1), 1-24.
- [33] Sun, Y. (2004). "A convergent t-statistic in spurious regressions", *Econometric Theory*, 20(5), 943-62 (DOI: 10.1017/S0266466604205072).
- [34] Verbeek, M. (2017). *A Guide to Modern Econometrics*, 5E, John Wiley & Sons Inc.
- [35] White, H. and Granger, C.W.J. (2011). "Consideration of trends in time series", *Journal of Time Series Econometrics*, 3(1), Article 2 (10.2202/1941-1928.1092).
- [36] Wooldridge, J.M. (2012). *Introductory Econometrics: A Modern Approach*, 5E, South-Western.
- [37] Yule, G. (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, 1-64.
- [38] Ziliak, S. (2008). "Guinnessometrics: The economic foundation of 'Student's' t", *Journal of Economic Perspectives*, 22(4), 199-216.