Unbalanced Regressions and Spurious Inference

Daniel Ventosa-Santaulària

Centro de Investigación y Docencia Económicas, CIDE, Mexico City, Mexico
Email: daniel@ventosa-santaularia.com

ABSTRACT

Spurious regression has been extensively studied in time series econometrics since Granger and Newbold's [1] seminal paper. Recently, it has been advanced that this phenomenon is due to a mistreatment of short-range autocorrelation in the residuals of the regression when at least one of the variables in a bivariate regression is stationary. HAC errors, feasible GLS and Cochrane-Orcutt-type procedures are then proposed to draw correct inference. Such a proposal should be cautiously considered, since nonsense inference might also be due to deterministic trend mechanisms, structural breaks, and long range dependence. In these cases, standard autocorrelation correction procedures would not solve the problem of spurious regression. We aim to make the later argument clear.

Keywords: Spurious Regression; Stationarity; Unbalanced Regression; Unit Root

1. Unbalanced Spurious Regressions

Spurious regression has been extensively studied since Granger and Newbold [1] seminal paper in which independent nonstationary variables are simulated and then used to estimate a simple bivariate regression. Phillips [2] provided the theoretical framework to understand the phenomenon in the simplest case (independent driftless unit root processes). Since then, the spurious regression phenomenon has been identified for many data-generating processes (DGPs), such as unit root with drifts, (broken-) trend stationary and long range, for example[1].

Here, we are concerned with the results presented in Noriega and Ventosa-Santaulària [4] and Stewart [5] pertaining to the spurious regression phenomenon under the following conditions: 1) both variables, y_t and x_t (see Equation (1)), are stationary (*i.e.*, integrated of order 0, $I(0)$), and 2) at least one of the variables (the regressor or the regressand) is integrated of order 1, $I(1)$. The later combinations result in an unbalanced regression and Noriega and Ventosa-Santaulària [4] found that, in a simple regression specification,

$$y_t = \alpha + \beta x_t + u_t, \tag{1}$$

where either y_t, x_t (or both) is $I(0)$[2], the *t*-ratio associated with $\hat{\beta}$, $t_{\hat{\beta}}$, does not diverge as the sample size grows; *i.e.* $t_{\hat{\beta}} = O_p(1)$. Results in Noriega and Ventosa-

Santaulària [4] imply that the asymptotic spurious regression phenomenon does not occur. Nevertheless, nonsense inference cannot be fully discarded. In a recent paper, Stewart [5] argues that, although the *t*-ratio does not diverge, it may not necessarily converge to a standard normal distribution. Furthermore, in the absence of autocorrelation in the DGP's innovations, only when both variables are *iid* $I(0)$ processes, the *t*-ratio behaves—asymptotically—as a standard normal. Other DGP combinations, such as $y_t \sim I(1)$ − $x_t \sim I(1)$ and vice versa, do have asymptotic nonstandard distributed *t*-ratios. Nevertheless, the size distortions are better explained by the presence of autocorrelation in the DGP innovations. This point is illustrated by Stewart [5] throughout a number of finite-sample experiments. The problem comes as no surprise since the estimated residuals behave as an autocorrelated process and size distortion should be expected in that case. Moreover, the use of heteroskedasticity and autocorrelation consistent (HAC) errors considerably reduce size distortions in some cases, as argues Stewart [5]. **Table 1** summarizes the relevant DGPs for both the dependent and the explanatory variables, similar to those used by Noriega and VentosaSantaulària [4] and Stewart [5], to estimate a simple linear specification.

For simplicity, we assume that innovations, e_{zt}, for $z = x, y$, are *iid* white noises. Following Noriega and Ventosa-Santaulària [4] and using the aforementioned DGPs, we present the following corollary:

Corollary *Let* y_t *and* x_t, *be generated by DGPs i and j of* **Table 1**. *Denote* C_{i-j} *as the DGP combination that generated y and x, respectively, and use them to*

---

[1]For a recent survey see Ventosa-Santaulària [3].
[2]In Noriega and Ventosa-Santaulària [4], the other variable may behave as: 1) driftless $I(1)$; 2) $I(1)$ with drift (with a possible drift break); 3) $I(2)$; and 4) (Broken-)Trend stationary.

Table 1. The DGPs for $z_t = x_t, y_t$.

Case	Name	Model
1	$I(0)$	$z_t = \mu_z + e_{zt}$
2	I(1)	$z_t = z_{t-1} + e_{zt}$

estimate specification (1) *by ordinary least square. The asymptotic distribution of the t-ratio associated with* $\hat{\beta}$, $t_{\hat{\beta}}$, *is*:

- $C_{1-2}: t_{\hat{\beta}} \xrightarrow{d} \dfrac{\int_0^1 \omega_x \, \mathrm{d}\omega_y(r) - \omega_y(1)\int_0^1 \omega_x(r)\,\mathrm{d}r}{\left[\int_0^1 \omega_x^2(r)\,\mathrm{d}r - \left(\int_0^1 \omega_{x(r)}\,\mathrm{d}r\right)^2\right]^{1/2}}$

- $C_{2-1}: t_{\hat{\beta}} \xrightarrow{d} \dfrac{\int_0^1 \omega_y \, \mathrm{d}\omega_x(r) - \omega_x(1)\int_0^1 \omega_y(r)\,\mathrm{d}r}{\left[\int_0^1 \omega_y^2(r)\,\mathrm{d}r - \left(\int_0^1 \omega_y(r)\,\mathrm{d}r\right)^2\right]^{1/2}}$

where $\omega_z(.)$, *for* $z = x, y,$, *is a standard Brownian motion and* $\xrightarrow{d}$ *denotes convergence in distribution.*

Proof: See Noriega and Ventosa-Santaulària [4][3].

Results in the corollary reveal that the asymptotic distribution of the t-ratio is nonstandard when the regression is unbalanced. However, a simple simulation of the asymptotic distribution shows a striking resemblance of this distribution with a standard normal (insets (*a*) and (*b*) in **Figure 1**). Such resemblance fades out in the presence of autocorrelation (insets (*c*) and (*d*) in **Figure 1**). We, therefore, confirm that the size distortions pointed out by Stewart [5] are due to autocorrelation; the latter happens to be an important source of spurious regression when at least one of the variables is $I(0)$ and confirms the results of Granger Hyung and Jeon [6] and Mikosch and Vries [7] results. Nevertheless, short range autocorrelation should not be considered as the sole *source* of spurious inference. It is well documented that deterministic trends, structural breaks, and long range dependence, also generate nonsense inference (see Perron [8] and Tsay and Chung [9]). It is important to note that the latter cannot be prevented by using Cochrane-Orcutt or Feasible GLS.
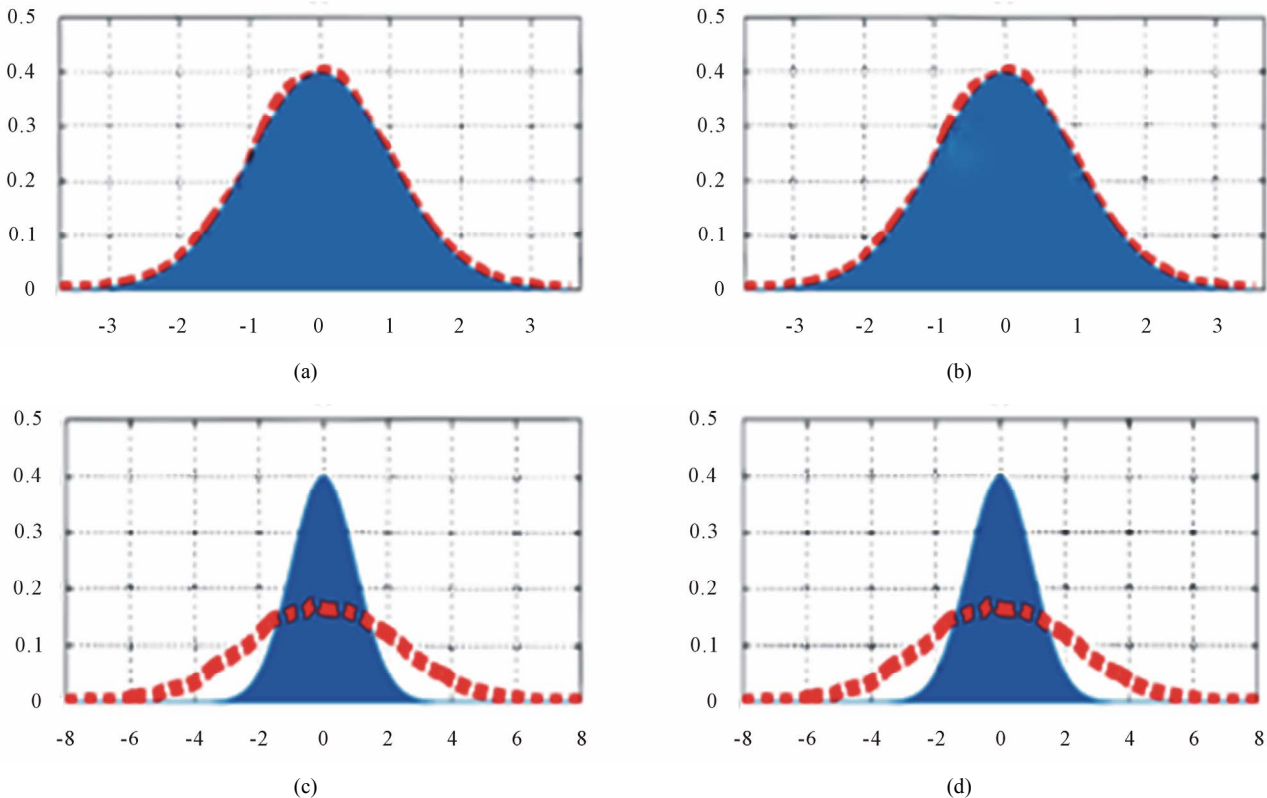
Using standard correction procedures to deal with the

Figure 1. *t*-ratio asymptotic distribution for unbalanced regressions: insets (a) and (c) *I*(0) vs *I*(1); insets (b) and (d) *I*(1) vs *I*(0); insets (a) and (b) *iid* innovations; insets (c) and (d) *AR*(1) innovations ($\varphi_x = 0.4$ and $\varphi_y = 0.7$). $\sigma_e^2 = 1$ Number of replications: 10,000. The blue area corresponds to the standard normal distribution whilst the red dashed line depicts the asymptotic distribution of the *t*-ratio.

---

[3]Noriega and Ventosa-Santaulària [4] only provide the order in convergence of the *t*-ratio. However, by following the instructions in the appendix, the asymptotic expressions can also be obtained, as we demonstrated in this paper. The *Mathematica* code is available upon request.

spurious regression phenomenon is tempting, even if such procedures cannot always provide correct inference (see Stewart [5] and McCallum [10], for example). Sun [11] proposed a convergent *t*-statistic using modified HAC errors with a bandwidth proportional to the sample size when the variables are highly persistent. The author acknowledges, however, that such a procedure cannot be used in empirical applications, since the limit distribution of the test depends on the memory parameter under the null hypothesis and critical values cannot, therefore, be tabulated. McCallum [10] and Kolev [12] also advocate classical correction procedures to deal with spurious regressions, such as the Cochrane-Orcutt procedure and Feasible Generalized Least Squares. They argue that using them reduces size distortions of the *t*-test. However, Martínez-Rivera and Ventosa-Santaulària [13] proved that such methods are not always effective and remain highly dependent on the DGP of the series[4].

2. Concluding Remarks

There is finite-sample evidence showing that spurious inference in unbalanced regressions mostly occurs when the innovations of the DGPs are not *iid*. In that sense, standard autocorrelation-correction procedures, such as HAC errors, Feasible GLS and Cochrane-Orcutt estimates, have been advanced to eliminate/reduce the size distortions and, thus, spurious inference. This approach should, nevertheless, be reconsidered. First, there is evidence that spurious regression using stationary series cannot always be interpreted as a short range autocorrelation phenomenon: long range dependence and structural breaks (level shifts, for example) also cause spurious inference; spurious regression cannot, therefore, be always corrected using classical procedures. Second, an unbalanced regression (in which the order of integrations of the involved series is not the same) is an empirical situation which remains to be proved relevant. The estimation of an unbalanced regression is not intuitive, although there are cases such as in the predictive equation in the finance literature, in which the market returns (usually, found to be stationary) is regressed against dividend yield (stationary but highly persistent). Spurious regression cannot be simply considered as a short-memory autocorrelation phenomenon and cannot, therefore, be treated using standard procedures. The main conclusion is, therefore, twofold: 1) practitioners should interpret cautiously their results whenever they find evidence of autocorrelation, since the inference could be spurious; 2) they should, however, be aware that spurious regressions arise for many diverse reasons, autocorrelation being only one of them; standard autocorrelation correction procedures are not to be considered as the sole solution to prevent spurious inference; on the contrary: parameter stability, long memory and cointegration tests should always be also considered.

---

[4]There are other approaches worth mentioning. In Davidson and Mac-Kinnon [14], for example, the authors consider that the spurious regression phenomenon is, at least partially due to a misspecification of the model. The authors argue that instead of Equation (1), practitioners should estimate $y_t = \alpha + \beta x_t + \gamma y_{t-1} + \mu_y$. Using this specification makes the null hypothesis of the *t*-test valid. Simulations presented in Davidson and MacKinnon [14] reveal, however, that even a correct specification is unable to provide an adequate size of the *t*-test under the null hypothesis when the variables are nonstationary.

REFERENCES

[1] C. W. J. Granger and P. Newbold, "Spurious Regressions in Econometrics," *Journal of Econometrics*, Vol. 2 No. 2, 1974, pp. 111-120. doi:10.1016/0304-4076(74)90034-7

[2] P. C. B. Phillips, "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, Vol. 33, No. 3, 1986, pp. 311-340. doi:10.1016/0304-4076(86)90001-1

[3] D. Ventosa-Santaulària, "Spurious Regression," *Journal of Probability and Statistics*, Vol. 2009, No. 1, 2009, pp. 155-182. doi:10.1155/2009/802975

[4] A. E. Noriega and D. Ventosa-Santaulària, "Spurious Regression and Trending Variables," *Oxford Bulletin of Economics and Statistics*, Vol. 69, No. 3, 2007, pp. 439-444. doi:10.1111/j.1468-0084.2007.00481.x

[5] C. Stewart, "A Note on Spurious Significance in Regressions Involving I(0) and I(1) Variables," *Empirical Economics*, Vol. 41, No. 3, 2011, pp. 565-571. doi:10.1007/s00181-010-0404-5

[6] C. W. J. Granger IV, N. Hyung and Y. Jeon, "Spurious Regressions with Stationary Series," *Applied Economics*, Vol. 33, No. 7, 2001, pp. 899-904.

[7] T. Mikosch and C. G. Vries, "Tail Probabilities for Regression Estimators," *Tinbergen Institute Discussion Paper*s, TI 2006-085/2, 2006.

[8] P. Perron, "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, Vol. 57, No. 6, 1989, pp. 1631-1401. doi:10.2307/1913712

[9] W. J. Tsay and C. F. Chung, "The Spurious Regression of Fractionally Integrated Processes," *Journal of Econometrics*, Vol. 96, No. 1, 2000, pp. 155-182. doi:10.1016/S0304-4076(99)00056-1

[10] B. T. McCallum, "Is the Spurious Regression Problem Spurious?" *Economics Letters*, Vol. 107, No. 3, 2010, pp. 321-323. doi:10.1016/j.econlet.2010.02.004

[11] Y. Sun, "A Convergent T-Statistic in Spurious Regressions," *Econometric Theory*, Vol. 20, No. 5, 2004, pp. 943-962. doi:10.1017/S0266466604205072

[12] G. I. Kolev, "The 'Spurious Regression Problem' in the Classical Regression Model Framework," *Economics Bulletin*, Vol. 31, No. 1, 2011, pp. 925-937.

[13] B. Martínez-Rivera and D. Ventosa-Santaulària, "A Comment on 'Is the Spurious Regression Problem Spurious?'" *Economics Letters*, Vol. 115, No. 2, 2012, pp. 229-231.

[14] D. Davidson and J. G. MacKinnon, "Econometric Theory and Methods," Oxford University Press, New York, 2004.