ASOCIACIÓN ESPURIA EN EL MODELO DE REGRESIÓN LOGÍSTICA CON SERIES DE TIEMPO

SPURIOUS ASOCIATION IN LOGISTIC TIME SERIES BINARY REGRESSION MODELS

Gustavo **Ramírez-Valverde**[1], J. Carlos **Islas-Monroy**[1], Benito **Ramírez-Valverde**[2]

[1]Estadística. Campus Montecillo. Colegio de Postgraduados. 56230. Montecillo, Estado de México. (gramirez@colpos.mx). [2]Estrategias para el Desarrollo Agrícola Regional. Campus Puebla. Carretera Federal, México Puebla. 72760. Puebla, Puebla.

Resumen

La asociación espuria en el modelo de regresión lineal ocurre cuando la variable independiente contribuye a explicar de manera importante la variabilidad de la variable respuesta de acuerdo con la prueba de hipótesis sobre el parámetro de la variable independiente, a pesar de que ambas variables no tienen ninguna relación. En modelos con variable respuesta categórica la presencia de asociación espuria no está reportado en la literatura. Por tanto, en este estudio se muestra empíricamente la existencia del fenómeno en regresión logística cuando los datos son generados por diferentes procesos de serie de tiempo que involucran series no estacionarias. El análisis de los resultados obtenidos indica que este fenómeno ocurre cuando el mecanismo generador de la variable respuesta y de la variable explicativa es no estacionario.

Palabras clave: modelos de respuesta binaria, regresión con series de tiempo, procesos no estacionarios, regresión logística.

Abstract

Spurious association in the linear regression model occurs when the independent variable makes an important contribution to explaining the variability of the response variable according to the hypothesis test of the parameter of the independent variable, even though the two variables have no relationship. In models with categorical response variable, the presence of spurious association is not reported in the literature. Therefore, in this study the existence of the phenomenon is shown empirically in logistic regression when the data are generated by different processes of time series that involve non-stationary series. The analysis of the results indicates that this phenomenon occurs when the generating mechanism of the response variable and of the explicative variable is non-stationary.

Key words: binary response models, regression with time series, non-stationary processes, logistic regression.

Introducción

El fenómeno de asociación espuria en el modelo de regresión lineal ocurre cuando la variable independiente contribuye a explicar la variabilidad de la variable respuesta, a pesar de que evidentemente las variables no tienen relación de acuerdo con la prueba de hipótesis sobre el parámetro de la variable independiente. Yule (1926) describe un ejemplo de este fenómeno donde hay una correlación de 0.95 entre dos variables independientes. Las variables eran la proporción de matrimonios por la iglesia sobre todos los casamientos en Inglaterra de 1866 a 1911, con la variable tasa de mortalidad estandarizada por

Introduction

The phenomenon of spurious association in the linear regression model occurs when the independent variable contributes to the explanation of the variability of the response variable, even though the variables evidently have no relationship according to the hypothesis test of the parameter of the independent variable. Yule (1926) describes an example of this phenomenon where there is a correlation of 0.95 between two independent variables. The variables were the proportion of church marriages over all marriages in England from 1866 to 1911, with the variable standardized mortality rate for every 1000 persons in the same period, showing how in cases of series generated with similar mechanisms and with deterministic tendency, there is spurious association when short time intervals were

* Autor responsable ❖ Author for correspondence.
Recibido: marzo, 2010. Aprobado: junio, 2011.
Publicado como ARTÍCULO en Agrociencia 45: 583-591. 2011.

cada 1000 personas en el mismo periodo, mostrando como en casos de series generadas con mecanismos semejantes y con tendencia determinística, se presenta asociación espuria cuando se observaban intervalos de tiempo pequeños, y esta asociación disminuía al aumentar el intervalo de tiempo, sugiriendo el mecanismo generador de las observaciones como explicación para que aparezcan asociaciones espurias. Según Granger y Newbold (1974), al simular X_t e Y_t caminatas aleatorias independientes, se presenta asociación espuria en algunos casos, a pesar de que por construcción X e Y son independientes. Phillips (1998) reporta que los mecanismos de tendencia en ambas series es lo que frecuentemente conduce a relaciones de regresión espuria, dando una explicación teórica al fenómeno en regresión lineal simple y múltiple.

A partir de los resultados de Grager y Newbold (1974) se muestra la presencia de regresión espuria en procesos integrados de orden 2 (Haldrup, 1994), en procesos de orden *d* (Marmol, 1995), en casos donde el orden de integración de la variable respuesta es diferente al orden de la variable independiente (Marmol, 1996), en procesos integrados de orden fraccional (Marmol 1998), en casos donde las series temporales tenían una raíz unitaria y presentaban deriva (Entorf, 1997), en series que no involucran un proceso persistente (Granger *et al.*, 2001), y Kim *et al. (*2004) lo hacen usando una serie estacionaria con una tendencia deterministica lineal generada independientemente. Noriega y Ventosa-Santaulària (2006 y 2007) estudiaron el fenómeno de la asociación espuria en series integradas de orden 1 y 2, con tendencias deterministas sujetas a cambios estructurales, deriva y sus combinaciones, mientras que Zaldivar *et al.* (2009) encuentran presencia de asociación espuria en series con especificaciones dinámicas.

En economía aumenta el uso de modelos cuando la respuesta es binaria; por ejemplo, la aplicación de técnicas de predicción del cambio de signo de los retornos de mercado es un tema de creciente interés para la comunidad financiera. Algunos estudios que muestran el uso de esto modelos son los siguientes: Wu y Zhang (1997) sugieren que las estrategias de transacción basadas en la estimación de la dirección del cambio en el nivel de precios son más efectivas y pueden generar beneficios más altos que aquellas basadas en una predicción puntual del nivel de precios de los instrumentos financieros; Lo y MacKinlay (1988)

observed, and this association decreased when the time interval was increased, suggesting generating mechanisms of the observations as explanation for the appearance of spurious associations. According to Granger and Newbold (1974), when simulating X_t and Y_t independent random walks, spurious association appears in some cases, although by construction X and Y are independent. Phillips (1998) reports that the mechanisms of tendency in both series is what frequently leads to relationships of spurious regression, giving a theoretical explanation to the phenomenon in simple and multiple linear regression.

From the results of Grager and Newbold (1974), the presence of linear regression is shown in integrated process of order 2 (Haldrup, 1994), in process of order *d* (Marmol, 1995), in cases where the order of integration of the response variable is different from the order of the independent variable (Marmol, 1996), in integrated processes of fractional order (Marmol, 1998), in cases where the temporal series had a unitary root and presented drift (Entorf, 1992) in series that do not involve a persistent process. Granger *et al.* (2001) and Kim *et al.* (2004) do this using a stationary series with a deterministic linear tendency that is independently generated. Noriega and Ventosa-Santaulària (2006 and 2007) studied the phenomenon of spurious association in integrated series of order 1 and 2, with deterministic tendencies subject to structural changes, drift and its combinations, whereas Zaldivar *et al.* (2009) find presence of spurious association in series with dynamic specifications.

In economics there is an increase in the use of models when the response is binary. For example, the application of techniques of prediction of the change in sign of the market returns is a topic of growing interest for the financial community. Some studies that show the use of these models are as follows: Wu and Zhang (1997) suggest that the strategies of transaction based on the estimation of the direction of change in the price level are more effective and can generate higher benefits than those based on a punctual prediction of the price level of the financial instruments; Lo and MacKinlay (1988) report the weekly returns for a variety of indices and medium sized portfolios with data of the U.S., Western Europe and Japan for the period of 1962 to 1985; and Conrad and Kaul (1988) point out the predictability

reportan los retornos semanales para una variedad de índices y portafolios de tamaño medio con datos de los EE.UU., Europa Occidental y Japón para el período de 1962 a 1985; y Conrad y Kaul (1988) señalan la predictibilidad de los retornos en el corto plazo usando datos semanales para el período de 1962 a 1985.

En regresión logística el problema de asociación espuria no se encuentra reportado en la literatura, por lo que se realizó un estudio de simulación basado en distintos procesos generadores de los datos, donde la variable numérica X fue generada independientemente de la variable respuesta binaria Y.

Materiales y Métodos

Modelo de regresión logística

Sean y_1, y_2,..., y_T, T observaciones de una variable de respuesta binaria con función de distribución de probabilidades Bernoulli (π_t) (t=1,…,T). Se quiere modelar la probabilidad de éxito π_t en función de una variable explicativa fija x_t. El modelo de regresión logístico se expresa como:

$$\pi_t = \frac{\exp(\beta_0 + \beta_1 x_t)}{1 + \exp(\beta_0 + \beta_1 x_t)}$$

La función de log-verosimilitud l, de las observaciones es:

$$l(\beta; y) = \sum_{t=1}^{T} \{ y_t(\beta_0 + \beta_1 x_t) - \ln[1 + \exp(\beta_0 + \beta_1 x_t)] \}$$

De la cual se obtiene el vector de derivadas parciales $\frac{\partial l}{\partial \beta}$, que al evaluarse en β e igualando a cero cada uno de sus elementos resulta un sistema de dos ecuaciones no lineales en los parámetros desconocidos β_0 y β_1. Este sistema no tiene solución cerrada y puede resolverse numéricamente. La generalización de este modelo al caso de p variables explicativas es directa.

Para inferir sobre los parámetros se utiliza el estadístico de Wald que bajo Ho tiene distribución asintótica Chi-cuadrada con p grados de libertad $\left(x_p^2\right)$, donde p es el número de parámetros en el vector β y $\hat{\beta}$ es el vector de parámetros estimados (Dobson, 1990). Para probar la hipótesis nula H_0: β_j=0 para alguna j:1≤j≤p se usa el estadístico $t = \hat{\beta}_j / e.e.\left(\hat{\beta}_j\right)$, donde

of the returns in short term using weekly data for the period of 1962 to 1985.

In logistic regression, the problem of spurious association is not reported in the literature, therefore a simulation study was made based on different generating processes of the data, where the numerical value X was generated independently of the binary response variable Y.

Materials and Methods

Logistic regression model

Let y_1, y_2,…, y_T, T be observations of a binary response variable with function of distribution of Bernoulli probabilities (π_t) (t = 1,…, T). The probability of success π_t is to be modeled as a function of a fixed explicative variable x_t. The logistic regression model is expressed as:

$$\pi_t = \frac{\exp(\beta_0 + \beta_1 x_t)}{1 + \exp(\beta_0 + \beta_1 x_t)}$$

The function of log-likelihood l of the observations is as follows:

$$l(\beta; y) = \sum_{t=1}^{T} \{ y_t(\beta_0 + \beta_1 x_t) - \ln[1 + \exp(\beta_0 + \beta_1 x_t)] \}$$

From which the vector of partial derivates $\frac{\partial l}{\partial \beta}$ is obtained, which when evaluated in β and equaling to zero each one of its elements results in a system of two non-linear equations in the unknown parameters β_0 and β_0. This system does not have a closed solution and can be resolved numerically. The generalization of this model to the case of p explicative variables is direct.

To infer about the parameters the Wald statistic is used, which under Ho has chi-squared asymptotic distribution with p degrees of freedom $\left(x_p^2\right)$, where p is the number of parameters in the vector β and $\hat{\beta}$ is the vector of estimated parameters (Dobson, 1990). To test the null hypothesis H_0: β_j=0 for any j:1≤j≤p the statistic $t = \hat{\beta}_j / e.e.\left(\hat{\beta}_j\right)$, is used where $e.e.\left(\hat{\beta}_i\right)$ is the standard error of the estimator $\hat{\beta}_i$, and the asymptotic distribution of t under H_0 is standard normal.

Time series processes considered

Noriega and Ventosa-Santaulària (2007) used experiments of simulation and an asymptotic analysis to show the existence of

$e.e.(\hat{\beta}_i)$ es el error estándar del estimador $\hat{\beta}_i$, y la distribución asintótica de t bajo H_0 es normal estándar.

Procesos de serie de tiempo considerados

Noriega y Ventosa-Santaulària (2007) usaron experimentos de simulación y un análisis asintótico para mostrar la existencia del fenómeno de asociación espuria en el modelo de regresión clásico con diferentes combinaciones de procesos generadores de las variables de respuesta y explicativa. En el presente estudio se utilizaron procesos semejantes a los usados por Noriega y Ventosa-Santaulària (2007), para conocer si los resultados en modelos de regresión con respuesta binaria tenían comportamiento semejante al de regresión lineal, y los seis procesos considerados se muestran en el Cuadro 1.

En el Cuadro 1 $u_{zt} \sim$ NIID(0, σ_z^2), DU_{zt} es una variable indicadora que afecta a los periodos posteriores a la fecha del cambio estructural (Tb_z) en una magnitud θ_z, esto es, DU_{zt} = 1 si t > Tb_z y DU_{zt} = 0 si t≤ Tb_z; DT_{zt} es una variable indicadora que afecta la pendiente de la serie en una magnitud ϕ_z en los periodos posteriores a la fecha del cambio estructural, esto es, DT_{zt} = (t− Tb_z) si t > Tb_z y DT_{zt} =0 si t≤ Tb_z; I(·) muestra el orden de integración del proceso, *br* indica corte estructural, *dr* la tendencia lineal y *TS* el proceso estacionario en tendencias.

Estudio de simulación

El objetivo fue mostrar la existencia del fenómeno de asociación espuria en el modelo de regresión logística, mediante simulación. En cada modelo se simularon 36 situaciones resultantes de combinar los seis procesos del Cuadro 1 con $z_t = x_t, y_t$ como generadores de las observaciones de la variable explicativa x_t y de la variable de respuesta y_t. Las variables x_t y y_t se generaron de forma independiente. La simulación se realizó en el paquete R usando los generadores de números aleatorios propios del lenguaje R.

Generación de las observaciones

Los componentes estocásticos del Cuadro 1 se obtuvieron con un generador de números aleatorios suponiendo σ_z^2=1 y $Tb_z \sim$ U (1, T), donde U denota la función de distribución de probabilidades uniforme discreta. La generación de las observaciones de la variable explicativa x_t, fue directa del Cuadro 1, ya que estas observaciones corresponden a variables numéricas.

La variable de respuesta binaria y_t se obtuvo con un proceso Bernoulli como generador de los datos con $P(Y_t=1) = \pi_t$, donde π_t depende de una variable subyacente w_t generada como alguno de los procesos del Cuadro 1. Además, la relación entre $P(Y_t=1) = \pi_t$ y w_t sigue un modelo de regresión logística;

Cuadro1. Procesos utilizados en el estudio de simulación.
Table 1. Processes used in the simulation study.

Identificación	Proceso	Descripción
1	*I*(0)	$z_t = \mu_z + u_{zt}$
2	*I*(0)+*br*	$z_t = \mu_t + \theta_z DU_{zt} + u_{zt}$
3	*TS*	$z_t = \mu_z + \beta_z t + u_{zt}$
4	*TS*+*br*	$z_t = \mu_z + \theta_z DU_{zt} + \beta_z t + \phi_z DT_{zt} + u_{zt}$
5	*I*(1)	$\Delta z_t = u_{zt}$
6	*I*(1)+*dr*	$\Delta z_t = \mu_z + u_{zt}$

the phenomenon of spurious association in the classic regression model with different combinations of generating processes of the response and explicative variables. Processes similar to those used by Noriega and Ventosa-Santaulària (2007) were employed in the present study to know if the results in regression models with binary response had behavior similar to that of linear regression, and the six processes considered are shown in Table 1.

In Table 1, $u_{zt} \sim$ NIID(0, σ_z^2), DU_{zt} is an indicative variable that affects the periods after the date of the structural change (Tb_z) in a magnitude of θ_z, that is, DU_{zt} = 1 if t > Tb_z and DU_{zt} = 0 if t≤ Tb_z; DT_{zt} is an indicative variable that affects the slope of the series in a magnitude of ϕ_z in the periods after the date of the structural change, that is, DT_{zt} = (t− Tb_z) if t > Tb_z and DT_{zt} =0 if t≤ Tb_z; I(·) shows the order of integration of the process, *br* indicates structural cut, *dr* the linear tendency, and *TS* the stationary process in tendencies.

Simulation study

The objective was to show the existence of the phenomenon of spurious association in the logistic regression model through simulation. In each model, 36 situations were simulated resulting from combining the six processes of Table 1 with $z_t = x_t, y_t$ as generators of the observations of the explicative variable x_t and of the response variable y_t. The variables x_t and y_t were generated independently. The simulation was carried out in the R package using the generators of proper random numbers of the R language.

Generation of the observations

The stochastic components of Table 1 were obtained with a generator of random numbers assuming σ_z^2=1 and $Tb_z \sim$ U (1, T), where *U* denotes the uniform discrete distribution function of probabilities. The generation of the observations of the explicative variable x_t was direct from Table 1, given that these observations correspond to numerical variables.

esto es, $\pi_t = \exp(\beta_0 + \beta_1 w_t)[1 + \exp(\beta_0 + \beta_1 w_t)]^{-1}$, ($\beta_0$ y β_1 son constantes arbitrarias distintas de cero) y una vez obtenido el valor de w_t se calculaba el valor de π_t con $\pi_t = \exp(\beta_0 + \beta_1 w_t)[1 + \exp(\beta_0 + \beta_1 w_t)]^{-1}$ y la observación y_t se obtuvo usando el método de la transformada inversa (Ross, 1999).

Simulación en el modelo de regresión logística

El experimento consistió en modelar la dependencia de $P(Y_t=1|x_t)$ en x_t (t=1,…,T) a través del modelo logístico donde el proceso generador de los datos de Y_t fue independiente del proceso generador de x_t, esto es, la hipótesis H_O: $\beta_1=0$ es cierta. Este experimento se realizó con 1000 repeticiones para cada uno de los niveles de los factores estudiados (tipo de proceso de la variable respuesta y tipo de proceso de la variable independiente) y se registró el rechazo (1) o no rechazo (0) de la hipótesis nula H_O: $\beta_1=0$ en los 1000 experimentos simulados. Se estudiaron los tamaños de muestra: 25, 50, 75, 100, 250, 500, 750, 1000, 2500 y 5000.

El tamaño de la prueba se estimó con la proporción de rechazos obtenida con el cociente nr/r, donde nr es el número de rechazos y r el número de repeticiones del experimento.

Resultados y Discusión

Se realizaron las simulaciones de los 36 casos resultantes de utilizar como x_t variable explicativa a cada uno de los seis procesos estudiados (Cuadro 1) y ser combinados con estos procesos como variable respuesta y_t. En las Figuras aparece un recuadro en la parte inferior con cuatro números compuestos de dos dígitos y cada uno de estos números representa una combinación de los procesos del Cuadro 1: el primer dígito indica el proceso que generó y_t, y el segundo el proceso que generó x_t.

En la Figura 1 se observa que cuando la variable respuesta y_t fue un proceso estacionario (proceso 1) y la variable explicativa x_t fue alguno de los seis procesos del Cuadro 1, no se presentó la regresión espuria y el tamaño de la prueba se mantiene en un valor cercano al nominal α=0.05. Este resultado coincide con el reportado por Noriega y Ventosa-Santaulària (2006 y 2007) en regresión lineal.

La Figura 2 muestra los casos en que la variable repuesta y_t no era estacionario, pero la falta de estacionariedad se debió a un cambio estructural

The binary response variable y_t was obtained with a Bernoulli process as generator of the data with $P(Y_t = 1) = \pi_t$, where π_t depends on a subjacent variable w_t generated as one of the process of Table 1. Furthermore, the relationship between $P(Y_t=1) = \pi_t$ and w_t follows a logistic regression model; that is, $\pi_t = \exp(\beta_0 + \beta_1 w_t)[1 + \exp(\beta_0 + \beta_1 w_t)]^{-1}$, β_0 and β_1 are arbitrary constants different from zero) and once the value of w_t was obtained, the value of π_t was calculated with $\pi_t = \exp(\beta_0 + \beta_1 w_t)[1 + \exp(\beta_0 + \beta_1 w_t)]^{-1}$ and the observation y_t was obtained using the inverse transform method (Ross, 1999).

Simulation in the logistic regression model

The experiment consisted of modeling the dependence of $P(Y_t=1|x_t)$ in x_t (t=1,…,T) by means of the logistic model where the generating process of the data of Y_t was independent from the generating process of x_t, that is, the hypothesis H_O: $\beta_1=0$ is correct. This experiment was carried out with 1000 replicates for each one of the levels of the factors studied (type of process of the response variable and type of process of the independent variable) registering rejection (1) or non-rejection (0) of the null hypothesis H_O: $\beta_1=0$ in the 1000 simulated experiments. The following sample sizes were studied: 25, 50, 75, 100, 250, 500, 750, 1000, 2500 and 5000.

The size of the test was estimated with the proportion of rejections obtained with the quotient nr/r, where nr is the number of rejections and r the number of replicates of the experiment.

Results and Discussion

Simulations were made of the 36 cases resulting from using x_t as explicative variable to each one of the six processes studied (Table 1) and combined with these processes as response variable y_t. In the Figures there is an inset in the lower portion with four numbers comprised of two digits and each one of these numbers represents one combination of the process of Table 1: the first digit indicates the process that generated y_t, and the second the process that generated x_t.

In Figure 1 it is observed that when the response variable y_t was a stationary process (process 1) and the explicative variable x_t was one of the six process of Table 1, spurious regression did not occur and the size of the test is maintained in a value close to the nominal α=0.05. This result coincides with what was reported by Noriega and Ventosa-Santaulària (2006 and 2007) in linear regression.

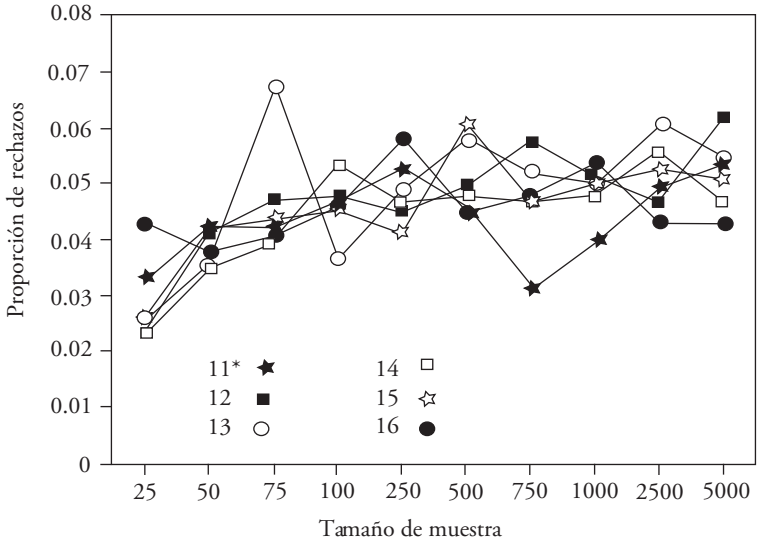

Figura 1. Tamaño de la prueba con la variable y_t estacionaria (proceso 1) con cada uno de los seis procesos como x_t.
*** El primer dígito indica el proceso que generó y_t, el segundo el proceso que generó x_t.**

Figure 1. Size of the test with the stationary y_t variable (process 1) with each one of the six processes as x_t.
***The first digit indicates the process that generated y_t, the second the process that generated x_t.**

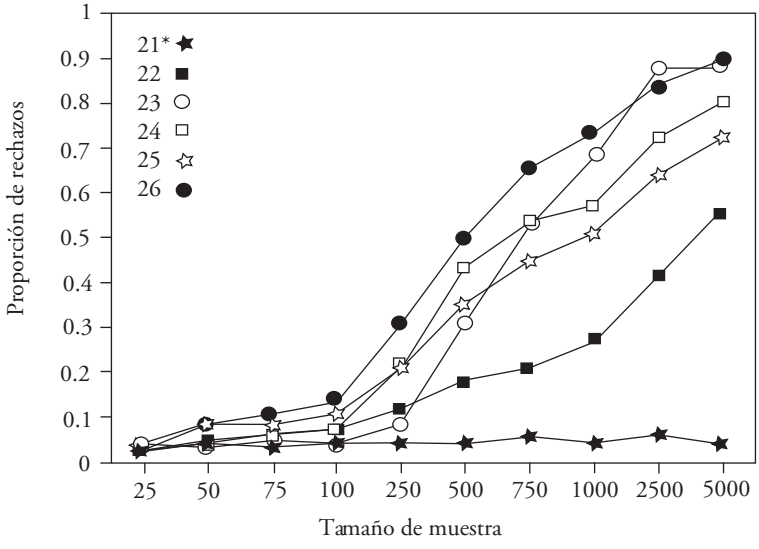

Figura 2. Tamaño de la prueba con la variable y_t no estacionaria (no estacionariedad por un cambio estructural - proceso 2) con cada uno de los seis procesos como x_t.
*** El primer dígito indica el proceso que generó y_t, el segundo el proceso que generó x_t.**

Figure 2. Size of the test with the non-stationary variable y_t (non-stationarity due to a structural change – process 2) with each one of the six process as x_t.
***The first digit indicates the process that generated y_t, the second the process that generated x_t.**

(proceso 2). Se puede observar que cuando la variable explicativa x_t fue alguno de los procesos no estacionarios del Cuadro 1 (procesos del 2 al 6), los valores de el tamaño de la prueba tienen una clara tendencia a aumentar con el tamaño de muestra dando evidencias de regresión espuria, y la proporción de rechazos en lugar de disminuir aumenta fuertemente al grado de obtener entre 60 y 90 % de rechazos en tamaños de muestra de 5000.

La Figura 2 muestra un cierto ordenamiento respecto a la magnitud del problema y fue más grave en los procesos no estacionarios con orden de integración 1 (procesos 5 y 6) y los no estacionarios con tendencia determinística (procesos 3 y 4). El proceso

Figure 2 shows the cases in which the response variable y_t was not stationary, but the lack of stationarity was due to a structural change (process 2). It can be observed that when the explicative variable x_t was one of the non-stationary processes of Table 1 (processes 2 to 6), the values of the size of the test have a clear tendency to increase with the size of the sample, showing evidence of spurious regression and the proportion of rejections, instead of decreasing, increases strongly, to the degree of obtaining between 60 and 90 % of rejections in sample sizes of 5000.

Figure 2 shows a certain ordering with respect to the magnitude of the problem, and it was more serious in the non-stationary processes with integration

no estacionario con menos problemas fue aquel con falta de estacionariedad debida exclusivamente a un cambio estructural (proceso 2).

Los resultados de los procesos que tuvieron como variable respuesta los procesos 3, 4, 5 y 6 del Cuadro 1 (Figura 3, 4, 5 y 6) tuvieron un comportamiento muy similar al observado cuando la variable repuesta y_t no era estacionario, pero la falta de estacionariedad se debió a un cambio estructural (Figura 2), aunque con efectos más marcados. Estos resultados coinciden con lo reportado en regresión lineal por Phillips (1998), Noriega y Ventosa-Santaulària (2006 y 2007), Marmol (1995, 1996 y 1998), Kim *et al*. (1994) y Haldrup (1994).

La similitud en los resultados de la regresión logística y la regresión lineal podrían explicarse si en la regresión logística se supone una variable latente no observable que determina la probabilidad de que la variable respuesta sea igual a la categoría uno, y que las relaciones expresadas en la literatura para dos variables continuas (presencia de asociación espuria) se mantienen con la variable latente continua, por lo que finalmente se expresa en la aparición de asociación espuria en modelos de regresión logística.

Conclusiones

El fenómeno de asociación espuria se presentó en el modelo de regresión logística y al igual que en regresión lineal, la ocurrencia de este fenómeno depende principalmente de la no estacionariedad.

order of 1 (processes 5 and 6) and the non-stationary processes due to deterministic tendencies (processes 3 and 4). The non-stationary process with fewest problems was the one with lack of stationarity due exclusively to a structural change (process 2).

The results of the process that had as response variable the processes 2, 4, 5 and 6 of Table 1 (Figures 3, 4, 5 and 6) exhibited a behavior very similar to that observed when the response variable y_t was not stationary, but the lack of stationarity was due to a structural change (Figure 2), although with more marked effects. These results coincide with what was reported in linear regression by Phillips (1998), Noriega and Ventosa-Santaulària (2006 and 2007), Marmol (1995, 1996 and 1998), Kim *et al.* (1994) and Haldrup (1994).

The similarity in the results of the logistic regression and the linear regression could be explained if in the logistic regression it is assumed that there is a non-observable latent variable that determines the probability that the response variable is equal to category one, and that the relationships expressed in the literature for two continuous variables (presence of spurious association) are maintained with the continuous latent variable, thus finally it is expressed in the appearance of spurious association in logistic regression models.

Conclusions

The phenomenon of spurious association appeared in the logistic regression model, and just as in linear

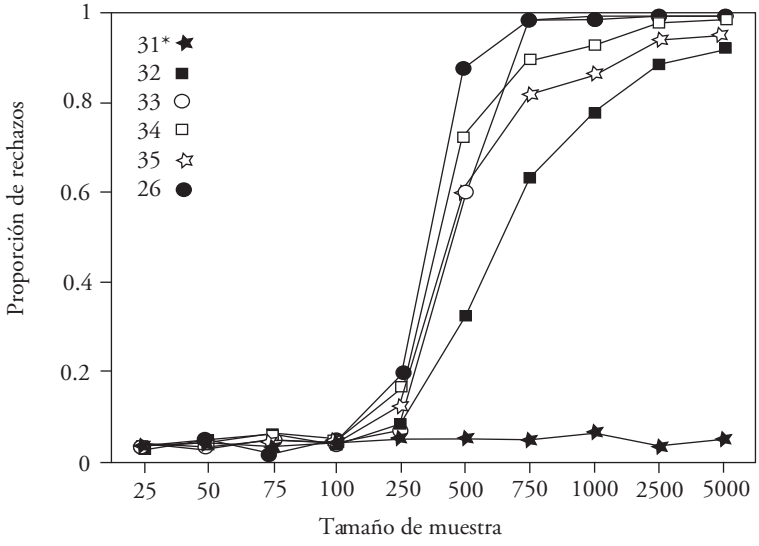

Figura 3. Tamaño de la prueba con la variable y_t no estacionaria (no estacionariedad por una tendencia determinística - proceso 3) con cada uno de los seis procesos como x_t. * El primer dígito indica el proceso que generó y_t, el segundo el proceso que generó x_t.

Figure 3. Size of the test with the non-stationary y_t variable (non-stationarity due to a deterministic tendency – process 3) with each one of the six processes as x_t. *The first digit indicates the process that generated y_t, the second the process that generated x_t.

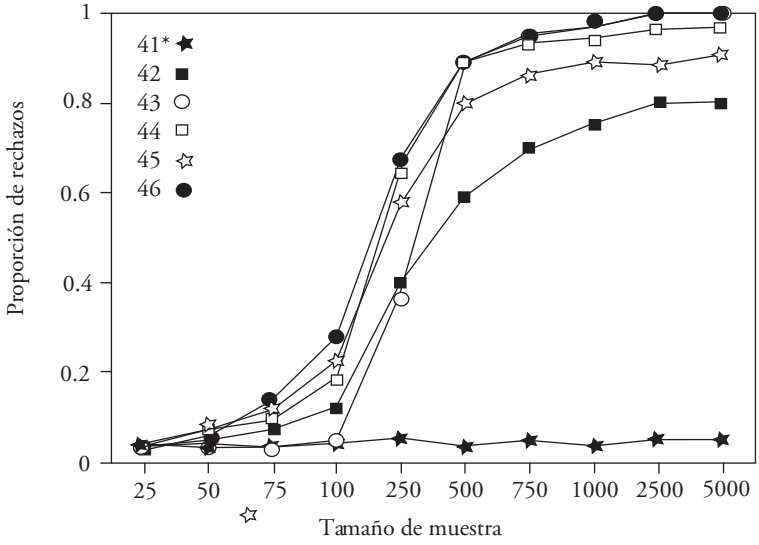

Figura 4. Tamaño de la prueba con la variable y_t no estacionaria (no estacionariedad por una tendencia determinística y por un cambio estructural - proceso 4) con cada uno de los procesos como x_t.
*** El primer dígito indica el proceso que generó y_t, el segundo el proceso que generó x_t.**

Figure 4. Size of the test with the non-stationary y_t variable (non-stationarity due to a deterministic tendency and a structural change – process 4) with each one of the six processes as x_t.
***The first digit indicates the process that generated y_t, the second the process that generated x_t.**

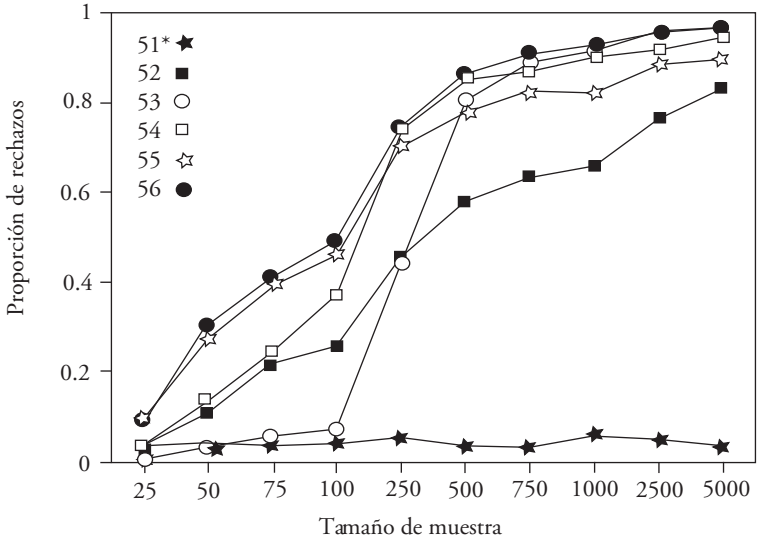

Figura 5. Tamaño de la prueba con la variable y_t no estacionaria, es una caminata aleatoria con orden de integración 1 (proceso 5) con cada uno de los procesos en la expresión 1 como x_t.
*** El primer dígito indica el proceso que generó y_t, el segundo el proceso que generó x_t.**

Figure 5. Size of the test with the non-stationary y_t variable, in a random walk with integration order 1 (process 5) with each one of the processes in expression 1 as x_t.
***The first digit indicates the process that generated y_t, the second the process that generated x_t.**

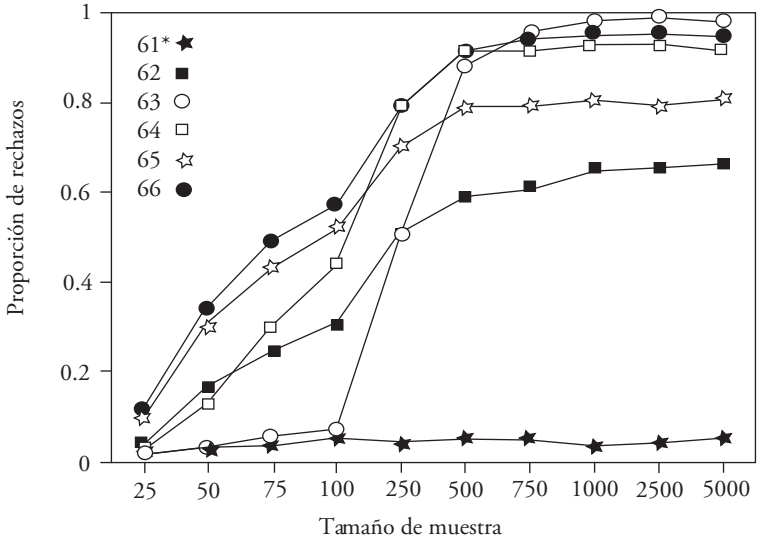

Figura 6. Tamaño de la prueba con la variable y_t no estacionaria con tendencia lineal, (proceso 6) con cada uno de los procesos en la expresión 1 como x_t.
*** El primer dígito indica el proceso que generó y_t, el segundo el proceso que generó x_t.**

Figure 6. Size of the test with the non-stationary y_t variable with linear tendency (process 6) with each one of the processes in expression 1 as x_t.
***The first digit indicates the process that generated y_t, the second the process that generated x_t.**

En todos los casos en que la variable respuesta y la variable explicativa fueron no estacionarios se presentó regresión espuria, sin importar si la no estacionariedad fue de tipo determinístico.

Las 36 combinaciones de procesos generadores de los datos consideradas son un grupo, si no extenso, medianamente amplio de situaciones en las que se puede incurrir en modelos de respuesta binaria erróneos al tratar con observaciones de serie de tiempo.

Literatura Citada

Conrad, J., and G. Kaul. 1988. Time-variation in expected returns. J. Business 61: 409-425.

Dobson, A. J. 1990. An Introduction to Generalized Linear Models. First edition. Ed. Chapman and Hall. London. 174 p.

Entorf, H. 1997. Random walks with drifts: Nonsense regression and spurious fixed-effect estimation. J. Econometrics 80: 287-296.

Granger, C.W.J., N. Hyung, and Y. Jeon. 2001. Spurious regressions with stationary series. Appl. Econ. 33: 899-904.

Granger, C. W. J., and P. Newbold. 1974. Spurious regression in econometrics. J. Econometrics 2:111-120.

Haldrup, N. 1994. The asymptotics of single-equation cointegration regressions with I(1) and I(2) variables. J. Econometrics 63:153-181.

Kim, T.H., Y.S. Lee, and P. Newbold. 2004. Spurious regressions with stationary processes around linear trends. Econ. Lett. 83: 257-262.

Lo, A., and C. MacKinley. 1988. Stock market price do not follow random walk: Evidence from a simple specification test. Rev. Financial Studies 1:41-66.

Marmol, F. 1995. Spurious regressions for I(d) processes. J. Time Series Analysis 16: 313-321.

Marmol, F. 1996. Nonsense regressions between integrated processes of different orders. Oxford Bull. Econ. and Statistics 58(3): 525-36.

Marmol, F. 1998. Spurious regression theory with nonstationary fractionally integrated processes. J. Econometrics 84: 233-50.

regression, the occurrence of this phenomenon depends principally on non-stationarity.

In all of the cases in which the response variable and the explicative variable were non-stationary, spurious regression occurred, regardless of whether the non-stationarity was deterministic.

The 36 conditions of the generating processes of the data considered are a group, if not extense, moderately ample of situations in which can be incurred in erroneous binary response models when dealing with observations of time series.

—End of the English version—

Noriega, E. A., and D. Ventosa-Santaulària. 2006. Spurious regression under broken-trend stationarity. J. Time Series Analysis 27(5): 671-684.

Noriega, E. A., and D. Ventosa-Santaulària. 2007. Spurious regression and trending variables. Oxford Bull. Econ. and Statistics 69(3): 439-444.

Phillips, P. C. B. 1998. New tools for understanding spurious regressions. Econometrica 66 (6): 1299-1325.

Ross, S. M. 1999. Simulación. Segunda edición. Ed. Prentice Hall Hispanoamericana. S. A. México. 282 p.

Wu, Y., and H. Zhang. 1997. Forward premiums as unbiased predictors of future currency depreciation: A non-parametric analysis. J. Int. Money and Finance 16: 609-623.

Yule, G. U. 1926. Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. J. Royal Statistical Soc. 89(1):1-63.

Zaldivar M, G., M. Castro O., y D. Ventosa-Santaulària. 2009. Regresión espuria en especificaciones dinámicas. Ensayos 28(1): 1-20.