



# **COLEGIO DE POSTGRADUADOS**

**INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS**

**CAMPUS MONTECILLO**

**SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA**

**ESTADÍSTICA**

**CORRELACIÓN ESPURIA EN MODELOS  
DE REGRESIÓN DE SERIES DE TIEMPO  
CON VARIABLE DE RESPUESTA BINARIA**

**JUAN CARLOS ISLAS MONROY**

**T E S I S**

**PRESENTADA COMO REQUISITO PARCIAL  
PARA OBTENER EL GRADO DE:**

**MAESTRO EN CIENCIAS**

**MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2007**

## **AGRADECIMIENTOS**

Al Colegio de Postgraduados por haberme brindado la oportunidad de progresar profesionalmente.

Al Dr. Gustavo Ramírez Valverde por su acertada dirección y apoyo para la realización de esta tesis.

Al grupo de asesores por su valiosa participación.

A todos los maestros, personal de estadística y demás personas que contribuyeron en mi desarrollo durante mi estancia en el Colegio.

## **DEDICO EL PRESENTE TRABAJO**

### **A mis primeros maestros, mis padres**

Sra. Mericia Monroy Solís

Sr. Juan de Dios Islas Licona

### **A mis hermanos**

Cuauhtémoc, Angélica y Jesús

A Sami e Israel

A todos mis amigos y compañeros con quienes he compartido momentos especiales y han pasado a formar parte de mi vida.

Con toda sinceridad

Juan Carlos

## ÍNDICE

RESUMEN .....	i
ABSTRACT .....	ii
INTRODUCCIÓN .....	1
CAPÍTULO I. CORRELACIÓN ESPURIA EN REGRESIÓN LINEAL.....	4
1.1 Definiciones .....	4
1.2 Regresión lineal .....	6
1.3 Pruebas de Wald .....	9
CAPÍTULO II. MODELOS DE REGRESIÓN PARA VARIABLE DE RESPUESTA BINARIA .....	15
2.1 Modelo de regresión logístico .....	16
2.1.1 Modelo e inferencia .....	16
2.1.2 Problemas de estimación .....	19
2.2 Ecuaciones de estimación generalizadas .....	20
2.2.1 Parametrizando la matriz de correlación de trabajo .....	23
2.2.1.1 Correlación autorregresiva .....	24
2.2.1.2 Correlación estacionaria .....	25
2.2.1.3 Correlación no estacionaria .....	26
2.2.1.4 Correlación no estructurada.....	27
2.2.2 Procedimiento de estimación .....	27
2.2.3 Probando coeficientes en ecuaciones de estimación generalizadas ....	28
2.2.3.1 Pruebas de Wald .....	29
2.3 Modelo de cadena de Markov de primer orden .....	29
CAPÍTULO III. ESTUDIO DE SIMULACIÓN .....	33
3.1 Procesos de serie de tiempo considerados .....	34
3.2 Generación de observaciones .....	35
3.3 Simulación en el modelo de regresión logístico.....	36
3.4 Simulación en el modelo de ecuaciones de estimación generalizadas .....	37
3.4.1 Modelo con una variable explicatoria .....	38
3.4.2 Modelo con una variable explicatoria continua y un tratamiento .....	39
3.5 Simulación en el modelo de cadena de Markov de primer orden .....	40
3.6 Problemas de estimación.....	41
CAPÍTULO IV. RESULTADOS Y DISCUSIÓN.....	43
4.1 Modelo de regresión logístico .....	43
4.1.1 Tamaño de la prueba .....	43
4.1.2 Valor máximo del estadístico t .....	47
4.2 Ecuaciones de estimación generalizadas .....	50
4.2.1 Modelo con una variable explicatoria, matriz de correlación no estructurada .....	50
4.2.2 Modelo con un tratamiento y una variable explicatoria, matriz de correlación no estructurada.....	53
4.2.3 Modelo con una variable explicatoria, matriz de correlación independiente.....	59
4.2.4 Modelo con un tratamiento y una variable explicatoria, matriz de correlación independiente. ....	62

4.3 Modelo de cadena de Markov de primer orden .....	67
4.3.1 Modelo basado en $P_0(t)$ .....	67
4.3.2 Modelo basado en $P_1(t)$ .....	72
CAPÍTULO V. CONCLUSIONES Y ESTUDIO POSTERIOR .....	77
5.1 Conclusiones .....	77
5.2 Estudio posterior.....	77
LITERATURA CITADA.....	79
Anexo I. Tamaño de la prueba y valor máximo del estadístico t de algunos modelos.....	81
Anexo II. Diagramas de flujo utilizados en cada modelo de regresión .....	89

## ÍNDICE DE CUADROS

Cuadro 1. Tamaño de la prueba estimado para la hipótesis $H_0:\beta_1=0$ .....	46
Cuadro 2. Tamaño de la prueba para la hipótesis $H_0:\beta_1=0$ , $\eta_{it} = \beta_0 + \beta_1x_{it}$ .....	51
Cuadro 3. Valores máximos de t para $\beta_1$ , $\eta_{it} = \beta_0 + \beta_1x_{it}$ .....	52
Cuadro 4. Tamaño de la prueba para la hipótesis $H_0:\beta_1=0$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	54
Cuadro 5. Tamaño de la prueba para la hipótesis $H_{0\tau} : \tau = 0$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	55
Cuadro 6. Valores máximos de t para $\beta_1$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	56
Cuadro 7. Valor máximo de t para $\tau$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	57
Cuadro 8. Tamaño de la prueba para la hipótesis $H_0:\beta_1=0$ , $\eta_{it} = \beta_0 + \beta_1x_{it}$ .....	60
Cuadro 9. Valores máximos de t para $\beta_1$ , $\eta_{it} = \beta_0 + \beta_1x_{it}$ .....	61
Cuadro 10. Tamaño de la prueba para la hipótesis $H_0:\beta_1=0$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	63
Cuadro 11. Valores máximos de t para $\beta_1$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	64
Cuadro 12. Tamaño de la prueba para la hipótesis $H_{0\tau} : \tau = 0$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	65
Cuadro 13. Valor máximo del estad t para $\tau$ , $\eta_{it} = \beta_0 + \tau_{it} + \beta_1x_{it}$ .....	66

Cuadro A1. Tamaño de la prueba para $H_0:\beta_1=0$ , modelo de regresión logístico.....	82
Cuadro A2. Valor máximo de t para $\beta_1$ , modelo de regresión logístico.....	83
Cuadro A3. Tamaño de la prueba para $H_{00}: \beta_{01}=0$ , $\eta_t = \beta_0 + \beta_1x_t$ . Modelo de cadena de Markov.....	85
Cuadro A4. Valor máximo de t para el coeficiente $\beta_{01}$ , $\eta_t = \beta_0 + \beta_1x_t$ . Modelo de cadena de Markov.....	86
Cuadro A5. Tamaño de la prueba para la hipótesis $H_{10}: \beta_{11}=0$ , $\eta_t = \beta_0 + \beta_1x_t$ . Modelo de cadena de Markov.....	87
Cuadro A6. Valor máximo de t para el coeficiente $\beta_{11}$ , $\eta_t = \beta_0 + \beta_1x_t$ . Modelo de cadena de Markov.....	88

### ÍNDICE DE FIGURAS

Figura 1. Comportamiento del tamaño de la prueba para $H_0:\beta_1=0^{**}$ .....	45
Figura 2. Comportamiento del valor máximo de t para $\beta_1^{**}$ . ....	47
Figura 3. Comportamiento del tamaño de la prueba para $H_{00}:\beta_{01}=0$ , $\eta_t = \beta_0 + \beta_1x_t^{**}$ . ....	68
Figura 4. Comportamiento del valor máximo de t para el coeficiente $\beta_{01}$ , $\eta_t = \beta_0 + \beta_1x_t^{**}$ . ....	71
Figura 5. Comportamiento del tamaño de la prueba para $H_{10}: \beta_{11}=0$ , $\eta_t = \beta_0 + \beta_1x_t^{**}$ . ....	73
Figura 6. Comportamiento del valor máximo de t para el coeficiente $\beta_{01}$ , $\eta_t = \beta_0 + \beta_1x_t^{**}$ . ....	75
Figura A1. Diagrama de flujo, modelo de regresión logístico.....	89
Figura A2. Diagrama de flujo, ecuaciones de estimación generalizadas.....	90
Figura A3. Diagrama de flujo, modelo de cadena de Markov.....	91

# CORRELACIÓN ESPURIA EN MODELOS DE REGRESIÓN DE SERIES DE TIEMPO CON VARIABLE DE RESPUESTA BINARIA

## RESUMEN

Este trabajo muestra empíricamente la existencia del fenómeno de correlación espuria en modelos de regresión de respuesta binaria cuando los datos son generados por diferentes procesos de serie de tiempo. Los modelos estudiados son; el logístico, para observaciones independientes; un modelo de cadena de Markov de primer orden que considera dependencia entre las observaciones; y el de ecuaciones de estimación generalizadas para mediciones repetidas en estudios longitudinales. Los resultados indican que este fenómeno ocurre en los tres modelos de regresión excepto bajo algunas condiciones. Dada la hipótesis nula  $H_0:\beta_1=0$ ; en el modelo logístico y en el de cadena de Markov, el tamaño de la prueba presenta valores mucho mayores de lo esperado y aumenta con el tamaño de muestra en todos los casos excepto cuando en el modelo interviene un proceso estacionario. El valor máximo del estadístico t se comporta de manera similar. En ecuaciones de estimación generalizadas, al utilizar la matriz de correlación más apropiada, en las dos hipótesis  $H_{0\beta}:\beta_1=0$  y  $H_{0\tau}:\tau=0$  ( $\tau$  es el efecto de un tratamiento) el fenómeno es muy parecido: cuando la variable de respuesta se basa en un proceso estacionario el tamaño de la prueba no crece con el tamaño de muestra, por el contrario, con cualquier otra combinación de procesos de las variables de respuesta y explicatoria el tamaño de la prueba crece rápidamente con el tamaño de muestra.

Palabras clave: Modelos de respuesta binaria, series de tiempo

# SPURIOUS CORRELATION IN BINARY RESPONSE

## MODELS WITH TIME SERIES

### ABSTRACT

This work shows empirically the existence of the phenomenon of spurious correlation in binary regression models when data are generated by different time series processes. The models studied were; the logistic, for independent observations; a first order Markov chain model which considers dependent observations; and the generalized estimating equations for repeated measures in longitudinal studies. The results indicate that this phenomenon occurs in the three regression models except under some conditions. Given the null  $H_0:\beta_1=0$ ; in the logistic model and in the Markov chain model, the size of the test presents values greater than expected and increases with sample size except when the model involves a stationary process. The maximum value of the t-statistic behaves similar. In the generalized estimating equations using the most suitable correlation matrix, the phenomenon is very similar in the two hypothesis  $H_{0\beta}:\beta_1=0$  and  $H_{0\tau}:\tau=0$  ( $\tau$  being a treatment effect): when the response variable is a stationary process the size of the test do not grows with the sample size, on the contrary, with any other combination of the generating process of the response and explanatory variables the size of the test grows rapidly with the sample size.

Keywords: Binary response models, time series



## INTRODUCCIÓN

El fenómeno de correlación 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 ambas variables no tienen ninguna relación, un ejemplo de este fenómeno lo encontramos en Yule (1926) donde es reportada una correlación de 0.95 entre dos variables independientes; una variable es la proporción de casamientos por la iglesia en Inglaterra en el periodo de 1866 a 1911, y la otra variable es la tasa de mortalidad estandarizada por cada 1000 personas en el mismo periodo.

La ocurrencia de este fenómeno en el modelo de regresión lineal ha recibido especial atención en los últimos años, sin embargo, en regresión para variable de respuesta binaria, este problema no ha sido reportado en la literatura, lo cuál no implica que en dichos modelos el fenómeno no ocurra. Un ejemplo de este fenómeno en regresión de respuesta binaria se obtiene al modelar una serie de la ocurrencia del lluvia en una estación meteorológica del país como variable de respuesta y una serie de observaciones del precio del dólar a la venta como variable explicatoria, esto es, al modelar la probabilidad de lluvia en el tiempo  $t$  con el precio del dólar. Obviamente estas variables no tienen relación alguna en la naturaleza, sin embargo, al modelar la probabilidad de ocurrencia de lluvia en función del precio del dólar con el modelo de regresión logístico se obtiene que la variable independiente es significativa. Podría atribuirse esta dependencia significativa obtenida entre ambas variables al hecho de que al usar este modelo no se cumple el supuesto de independencia en las observaciones, ya que en

ambas variables las observaciones están correlacionadas. Teniendo esto en cuenta se utilizaron dos modelos adicionales que toman en cuenta la dependencia de las observaciones, el primero de ellos es el modelo de ecuaciones de estimación generalizadas propuesto por Liang y Zeger (1986) para datos de mediciones repetidas en estudios longitudinales en el que se pueden introducir correlaciones no estacionarias de las observaciones agrupadas en conglomerados, cada conglomerado contiene las observaciones repetidas realizadas en un individuo particular; el coeficiente obtenido fue 21.83 y un p-valor  $< 0.05$ ; el segundo modelo está basado en una cadena de Markov de primer orden, donde se modeló la probabilidad de la ocurrencia de lluvia en el día actual dada la ocurrencia o no de lluvia en el día anterior, utilizando el precio del dólar como variable explicatoria. El coeficiente de la regresión fue 19.37 con valor de  $t = 2.50$ . La dependencia entre las variables es significativa aún en modelos que consideran autocorrelación de las observaciones. Esta dependencia es considerada espuria porque naturalmente la ocurrencia de lluvia y el precio del dólar no tienen ninguna relación.

El presente trabajo tiene por objetivo probar empíricamente la existencia del fenómeno de correlación espuria en modelos de respuesta binaria. La evidencia de este fenómeno se mostrará utilizando el tamaño de la prueba y el valor máximo del estadístico  $t$  obtenido de 1000 repeticiones de un experimento de regresión con datos simulados.

En el capítulo 1 se presenta una revisión del fenómeno de correlación espuria en el modelo de regresión lineal así como algunas de las condiciones bajo las que ocurre.

En el capítulo 2, se describen los modelos de regresión de respuesta binaria ya mencionados. Posteriormente, en el 3 se lleva a cabo el estudio de simulación para la obtención del tamaño de la prueba y el valor máximo del estadístico  $t$  y su comportamiento en cada modelo de regresión en función del tamaño de muestra. En el capítulo 4 se discuten los resultados obtenidos del estudio de simulación obtenidos y finalmente en el capítulo 5 se emiten las conclusiones y se proponen algunos aspectos para ser investigados posteriormente.

## CAPÍTULO I. CORRELACIÓN ESPURIA EN REGRESIÓN LINEAL

Antes de abordar el tema central de la presente unidad definiremos algunos conceptos involucrados en la existencia del fenómeno de correlación espuria en el modelo de regresión lineal.

### 1.1 Definiciones

**Serie de tiempo.** Una serie de tiempo es un conjunto de observaciones generadas secuencialmente en el tiempo. En el presente trabajo denotaremos a una serie de tiempo como  $\{z_t\}_{t=1}^T$  o únicamente como  $z_t$  y comúnmente nos referiremos a la serie de tiempo como proceso.

**Proceso autoregresivo.** Este proceso es una serie de tiempo en la que el valor actual de la serie  $z_t$  puede representarse como una combinación lineal de un número finito  $p$  de valores rezagados  $z_{t-1}, \dots, z_{t-p}$  y un término aleatorio  $u_t$  como en la siguiente expresión

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + u_t, u_t \sim \text{niid}(0, \sigma^2) \quad (1.1)$$

Este proceso se denota como AR(p).

**Proceso estacionario.** Una serie de tiempo  $z_t$  es llamada proceso estacionario si cumple las siguientes condiciones

a)  $E(z_t) = 0$ , para  $t = 1, \dots, T$

$$b) E(z_t^2) = \sigma^2, \text{ para } t=1, \dots, T \quad (1.2)$$

$$c) E(z_t z_s) = 0, \text{ para } t \neq s \text{ y } t, s = 1, \dots, T$$

Estas propiedades indican que la media y la varianza del proceso estacionario es independiente de  $t$  y cada autocovarianza depende únicamente de la distancia entre el tiempo  $t$  y el tiempo  $s$ .

Serie de tiempo no estacionaria. Una serie de tiempo  $z_t$  que no cumple con las propiedades de la expresión (1.2) es no estacionaria.

Caminata aleatoria y proceso integrado. El proceso no estacionario de caminata aleatoria está dado por

$$z_t = z_{t-1} + u_t \quad (1.3)$$

Donde  $u_t$  cumple con las propiedades (1.2).

Cuando  $u_t \sim \text{niid}(0, 1)$  el proceso es llamado caminata aleatoria estandarizada.

Nótese que la diferencia en las observaciones de  $z_t$  es un proceso estacionario

$$z_t - z_{t-1} = u_t$$

Un proceso que necesita ser diferenciado  $d$  veces de esta forma para lograr que sea estacionario es llamado proceso integrado. El número de diferencias  $d$ , es llamado orden de integración y tal proceso se denota como  $I(d)$ . De acuerdo a esto la caminata aleatoria es un proceso integrado de orden 1.

Raíz unitaria. Consideremos el proceso (1.1) con  $p=1$ :

$$z_t = \phi_1 z_{t-1} + u_t, u_t \sim \text{niid}(0, \sigma^2) \quad (1.4)$$

Si  $\phi_1=1$  se obtiene el proceso de caminata aleatoria y la diferencia  $z_t - z_{t-1} = u_t$  puede expresarse como  $z_t(1-B) = u_t$ , donde  $B$  es el operador de retraso,  $B(z_t) = z_{t-1}$ . En este caso la raíz del coeficiente  $1-B$  es 1 y decimos que este proceso tiene raíz unitaria. Cuando  $|\phi_1| < 1$  el proceso es estacionario y la raíz de  $(1-B)=0$  es  $B=1/|\phi_1|$  y cae fuera del círculo unitario.

Probabilidad límite. Un vector aleatorio  $a(y_T)$  de tamaño  $T$  converge en probabilidad al vector  $a_T$  si para toda  $\varepsilon > 0$ ,

$$\lim_{T \rightarrow \infty} P(\|a(y_T) - a_T\| < \varepsilon) = 1 \quad (1.5)$$

Donde la doble barra indica la norma euclídeana.

Cuando se cumple (1.5) para  $a(y_T)$  se dice que la probabilidad límite (plim) de este vector es el vector  $a_T$  y  $a_T$  puede ser un vector aleatorio o no aleatorio de escalares.

## 1.2 Regresión lineal

Consideremos el modelo de regresión lineal con  $p$  variables

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_p x_{pt} + u_t, u_t \sim \text{niid}(0, \sigma^2), t=1, \dots, T \quad (1.6)$$

Donde  $y_t$  es la t-ésima observación de la variable dependiente;  $x_{it}$  ( $1 \leq i \leq p$ ), es la t-ésima observación de la i-ésima variable independiente;  $u_t$  es un término de error y  $\beta_0, \beta_1, \dots, \beta_p$  son parámetros desconocidos.

En notación matricial escribimos este modelo como

$$y = X\beta + u \quad (1.7)$$

Los supuestos básicos en el modelo de regresión lineal son

- a) Las variables explicatorias son fijas
- b) El rango de X es  $p+1$
- c) Los términos de error  $u_t$  no están correlacionados, tienen media cero y varianza constante, lo que se conoce como ruido blanco.

En la expresión (1.7)  $u = (u_1, u_2, \dots, u_T)'$ , por lo tanto el supuesto c) se expresa como

$$E(u) = 0, V(u) = E(uu') = \sigma^2 I \quad (1.8)$$

Donde 0 es un vector de ceros, e I es la matriz identidad de tamaño T.

Los parámetros desconocidos  $\beta_0, \beta_1, \dots, \beta_p$  son estimados por la expresión

$$\hat{\beta} = (X'X)^{-1} X'y \quad (1.9)$$

Dos propiedades deseables de un estimador es que sea insesgado y consistente.

La primera de ellas implica que el valor esperado del estimador es igual al valor

verdadero del parámetro. La segunda propiedad implica que el valor del parámetro estimado tiende al valor verdadero conforme el tamaño de muestra tiende a infinito.

Para mostrar que el estimador de  $\beta$  dado por (1.9) es insesgado suponemos que el modelo (1.7) está correctamente especificado y que el proceso generador de los datos ocurre con  $\beta = \beta_r$ , entonces sustituimos el vector  $y$  en función de  $\beta_r$  en (1.9) para obtener

$$\hat{\beta} = (X'X)^{-1} X'(X\beta_r + u) = \beta_r + (X'X)^{-1} X'u \quad (1.10)$$

Tomando la esperanza de  $\hat{\beta}$  bajo los supuestos a) y c) de (1.7) se obtiene que  $\hat{\beta}$  es insesgado.

Para verificar que el estimador de  $\beta$  es consistente nos basamos en la expresión (1.10) en la que debemos probar que  $(X'X)^{-1}X'u$  tiene probabilidad límite cero. Esta expresión es el producto de  $(X'X)^{-1}$  y  $X'u$ , pero ni  $X'X$  ni  $X'u$  tiene probabilidad límite, sin embargo podemos dividir estas dos expresiones por  $n$  sin cambiar el valor de  $(X'X)^{-1}X'u$  y de esa manera se transforma a  $(X'X)^{-1}$  y  $X'u$  en dos cantidades que bajo supuestos razonables tienen probabilidad límite no estocástica. Entonces para el segundo término de (1.10) tenemos



$$\left( \text{plim}_{T \rightarrow \infty} \frac{1}{T} X'X \right)^{-1} \text{plim}_{T \rightarrow \infty} \frac{1}{T} X'u = S^{-1} \text{plim}_{T \rightarrow \infty} \frac{1}{T} X'u = 0 \quad (1.11)$$

Sonde S es una matriz no estocástica de rango completo.

Para escribir la primera igualdad se supuso que  $\text{plim}_{T \rightarrow \infty} \frac{1}{T} X'X = S$ , donde S es una matriz finita, lo cual no es del todo irrazonable dado que cada elemento de la matriz  $\frac{1}{T} X'X$  es un promedio de T números:

$$\left( \frac{1}{T} X'X \right)_{ij} = \frac{1}{T} \sum_{t=1}^T X_{ti} X_{tj} \quad 1 \leq i, j \leq p+1$$

La segunda igualdad se obtiene bajo el supuesto de que  $E(u_t | X_t) = 0$  y aplicando la Ley de los Números Grandes<sup>†</sup> tenemos

$$\text{plim}_{T \rightarrow \infty} \frac{1}{T} X'u = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T X_t u_t = 0$$

Por lo tanto, por (1.11) el estimador de  $\beta$  dado por (1.9) es consistente.

### 1.3 Pruebas de Wald

Supongamos que se tiene un modelo general de regresión con k variables explicatorias. Un modelo reducido basado en el modelo general se obtiene al imponer un conjunto de m restricciones, tales restricciones pueden especificarse

<sup>†</sup> Esta ley establece que para dos números pequeños  $\epsilon$  y  $\delta$ ,  $\epsilon > 0$  y  $0 < \delta < 1$ , existe un entero T tal que si una muestra aleatoria de tamaño T o mayor se obtiene de una población con función de densidad  $f_X(x)$ , entonces la probabilidad de que la media muestral se desvíe de la media poblacional por una cantidad menor que  $\epsilon$  es mayor que  $1 - \delta$ , tan cercana a 1 como se desee (Mood, *et al.*, 1974).

mediante la expresión:

$$R\beta=r$$

Donde  $R$  es una matriz de rango completo de tamaño  $m \times T$ ,  $r$  es un  $m$  vector y  $\beta$  es el vector de parámetros definido en (1.7). Los valores tomados por los elementos de  $R$  y  $r$  se especifican bajo la hipótesis nula bajo la cual los elementos del vector  $R\beta-r$  deben ser cercanos a cero, siendo  $\hat{\beta}$  es el estimador de máxima verosimilitud de  $\beta$ . Se quiere considerar algún criterio para determinar cuando  $R\hat{\beta}-r$  es suficientemente cercano a cero y por tanto consistente con la hipótesis nula. Para ello consideramos que  $\sqrt{T}(\hat{\beta}-\beta)$  tiene distribución límite normal multivariada con vector de medias cero y matriz de varianzas y covarianzas  $\Gamma$ , donde  $\Gamma$  es una matriz positiva definida tal que

$$\Gamma = \lim T^{-1}I$$

$I$  es la matriz información,  $I = E(UU')$ , y  $U$  es el vector de primeras derivadas parciales de la función de log-verosimilitud de las observaciones.

Por lo tanto se tiene que  $\sqrt{TR}(\hat{\beta}-\beta) = \sqrt{T}(R\hat{\beta}-r)$  tiene distribución límite normal multivariada con vector de medias cero y matriz de varianzas y covarianzas dada por  $R\Gamma^{-1}R'$ , de donde el estadístico

$$T(R\hat{\beta}-r)'[R\Gamma^{-1}R']^{-1}(R\hat{\beta}-r) \sim \chi_m^2$$

Donde  $\chi$  denota la distribución Ji-Cuadrada. La distribución de este estadístico se mantiene si la matriz  $T\Gamma$  es reemplazada por la matriz  $\Gamma^*$ , la cuál tiene la propiedad de que  $\text{plim } T^{-1}\Gamma^*(\hat{\beta}) = \Gamma(\beta)$  siendo  $\hat{\beta}$  un estimador consistente de  $\beta$ . Igualando  $\hat{\beta}$  con  $\hat{\beta}'$  se obtiene la definición común del estadístico de Wald (Harvey, 1990):

$$W = (R\hat{\beta}' - r)' [R\Gamma^* - 1(\hat{\beta}'R)']^{-1} (R\hat{\beta}' - r)$$

Cuando la hipótesis nula es falsa,  $W$  tiende a ser muy grande y una elección natural de la región crítica es  $W > \kappa$ . Si la distribución de  $W$  no puede ser determinada con exactitud, un valor adecuado de  $\kappa$  se puede obtener de las tablas de la distribución Ji-Cuadrada.

### 1.3 Correlación espuria

Como el ejemplo mencionado en la introducción del presente trabajo, existen muchos otros que muestran la posibilidad de encontrar variables en la naturaleza aparentemente relacionadas cuando dichas variables son enteramente independientes. Hendry (1980) reporta un ejemplo econométrico en el que hay un coeficiente de correlación cercano a 1 entre el nivel de precios y la precipitación acumulada en el Reino Unido.

Phillips (1998) menciona que la generalidad de los mecanismos de tendencia en los datos es lo que conduce a correlaciones espuria. Particularmente se ha mostrado la existencia de este fenómeno en diferentes combinaciones de procesos generadores de los datos de la variable dependiente y la variable

explicatoria. En un principio Granger y Newbold (1974) mostraron mediante simulación que al usar caminatas aleatorias independientes como variable de respuesta y explicatoria el fenómeno se presenta. Noriega y Ventosa-Santaulária (2006) mostraron la ocurrencia de este fenómeno cuando las variables de respuesta y explicatoria que intervienen en el modelo son generadas en base a diferentes procesos como son procesos estacionarios con cambio estructural, con tendencia determinística y con tendencia estocástica.

Cuando las variables  $y_t$  y  $x_t$  son no estacionarias y son independientes una de la otra, el procedimiento de mínimos cuadrados ordinarios aplicado al modelo de regresión

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

Tiene las siguientes implicaciones (Noriega y Ventosa-Santaulária, 2005):

- a) El estimador de  $\beta_1$  no converge a su verdadero valor cero.
- b) El valor t para probar la hipótesis nula  $H_0: \beta_1=0$  diverge.

Lo que indica una relación espuria entre  $y_t$  y  $x_t$  aún asintóticamente.

Un caso especial nos muestra las desventajas de los estimadores de mínimos cuadrados en presencia de procesos no estacionarios: cuando una variable con una raíz unitaria es usada como regresora, el supuesto para obtener consistencia en el estimador de los parámetros no se cumple, se supone que la probabilidad límite de la matriz  $T^{-1}X'X$  es la matriz finita S, pero esto es falso cuando una o

mas de las series regresoras tengan una raíz unitaria. Por ejemplo, supongamos que una variable regresora es la caminata aleatoria estandarizada

$$z_t = z_{t-1} + u_t, u_t \sim \text{niid}(0,1)$$

En este caso uno de los elementos de la matriz  $X'X$  es  $\sum_{t=1}^T z_t^2$ , que equivale a:

$$\sum_{t=1}^T \left( \sum_{r=1}^t \sum_{s=1}^t u_r u_s \right) \quad (1.12)$$

Debido a la forma de  $u_t$ , únicamente los términos con  $r=s$  contribuyen a la esperanza de la expresión (1.12). La cual, como  $E(u_r^2) = 1$ , esta dada por:

$$\sum_{t=1}^T \sum_{r=1}^t E(u_r^2) = \sum_{t=1}^T t = \frac{1}{2} T(T+1) \quad (1.13)$$

Ahora, si  $z$  denota el vector de tamaño  $T$  con elementos  $z_t$ , entonces la esperanza de  $T^{-1}z'z$ , por (1.13) resulta ser  $\frac{1}{2}(T+1)$  que evidentemente es de orden estocástico  $T$ , y por lo tanto hace imposible que  $T^{-1}z'z$  tenga una probabilidad límite finita. Este hecho implica que ninguno de los resultados sobre consistencia y normalidad asintótica aplican en modelos donde una o más de las variables regresoras tienen una raíz unitaria.

Cuando en el modelo intervienen procesos AR estacionarios tanto en la variable explicatoria como en las variables de respuesta y son independientes, podemos ver que ocurre considerando el modelo de regresión

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \varepsilon_t \sim \text{niid}(0, \sigma^2)$$

Donde  $y_t$  y  $x_t$  son generadas por el proceso AR(1):

$$z_t = \varphi_1 z_{t-1} + u_t, |\varphi_1| < 1, u_t \sim \text{niid}(0, \sigma^2)$$

En este caso la matriz  $T^{-1}X'X$  tiene una probabilidad límite finita, sin embargo, debido a que ni la constante, ni  $x_t$  explican a  $y_t$  el término de error para la observación  $t$  es  $\varepsilon_t = y_t$  que no es ruido blanco sino un proceso AR(1).

Lo anterior muestra diversas situaciones y los problemas en regresión lineal cuando se presenta el fenómeno de correlación espuria, estas situaciones están basadas obviamente en que las variables dependiente y explicatoria son independientes. Cuando estas variables realmente están relacionadas pero son procesos no estacionarios integrados de orden 1 se obtiene un modelo de regresión mal especificado. Para evitar resultados espurios en ambos casos, se recomienda el uso de técnicas de cointegración mediante las cuales es posible encontrar la verdadera relación entre las variables utilizando las distribuciones correctas para probar la significancia de tales relaciones (Murray, 1994).

## CAPÍTULO II. MODELOS DE REGRESIÓN PARA VARIABLE DE RESPUESTA BINARIA

En la naturaleza es común encontrar fenómenos donde la variable de respuesta toma uno de dos posibles valores tales como presencia o ausencia de una característica. Entonces estas variables se pueden representar por la variable binaria  $y_i$  tal que

$$y_i = \begin{cases} 1 & \text{si la característica está presente} \\ 0 & \text{en caso contrario} \end{cases}$$

Se acostumbra llamar éxito cuando  $y_i = 1$ . En este tipo de variables el interés se centra en modelar la probabilidad de éxito en función de una o más covariables. Un ejemplo simple de este tipo de variable es el medio de transporte elegido por un individuo o conjunto de individuos para trasladarse de su casa al trabajo, el valor tomado por la variable puede codificarse como 1 cuando el individuo elige transporte público o 0 cuando elige viajar en automóvil. El tipo de modelo utilizado para estimar la probabilidad en cuestión depende de la estructura de los datos.

En este trabajo se consideran tres modelos que permiten estimar la probabilidad arriba mencionada en función de un conjunto de  $k$  variables explicatorias: a) regresión logística; b) un modelo de cadena de Markov de primer orden y c) ecuaciones de estimación generalizadas. El primero permite modelar la probabilidad de éxito en función de  $k$  variables explicatorias cuando las observaciones son hechas transversalmente en el tiempo; el modelo de cadena de Markov permite modelar esta dependencia cuando observaciones sobre un

individuo son hechas a través del tiempo o tienen correlación natural y el de ecuaciones de estimación generalizadas permite modelar la misma dependencia cuando las observaciones son hechas sobre uno o mas individuos a través del tiempo. En las siguientes secciones se describe de manera más amplia cada uno de estos modelos.

## 2.1 Modelo de regresión logístico

### 2.1.1 Modelo e inferencia

Sean  $y_1, y_2, \dots, y_T$ ,  $T$  observaciones de una variable aleatoria con función de distribución de probabilidades Bernoulli( $\pi_t$ ). Se quiere modelar la dependencia de la probabilidad de éxito  $\pi_t$ ,  $t = 1, 2, \dots, T$  en función de un conjunto de  $k$  variables explicatorias  $x_{1t}, x_{2t}, \dots, x_{kt}$ . El modelo de regresión logístico de la dependencia de  $\pi_t$  en esas  $k$  variables explicatorias está dada por

$$\pi_t = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)} \quad (2.1)$$

O bien

$$\ln\left(\frac{\pi_t}{1 - \pi_t}\right) = \eta_t$$

Donde  $\eta_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt}$ , es llamado predictor lineal.



La función de verosimilitud para la muestra  $y_1, y_2, \dots, y_T$  con probabilidades correspondientes  $\pi_t$  puede ser escrita como  $L = \prod_{t=1}^T \pi_t^{y_t} (1-\pi_t)^{1-y_t}$ , y la función de log-verosimilitud resulta

$$l = \sum_{t=1}^T [y_t \ln(\pi_t) + (1-y_t) \ln(1-\pi_t)] \quad (2.2)$$

Las derivadas parciales de  $l$  con respecto a los parámetros desconocidos  $\beta_0, \beta_1, \dots, \beta_k$  son

$$\frac{\partial l}{\partial \beta_j} = \sum_{t=1}^T y_t x_{jt} - \sum_{t=1}^T x_{jt} \exp(\eta_t) (1 + \exp(\eta_t))^{-1}, \quad 0 \leq j \leq k$$

Que al igualarse a cero se obtiene un conjunto de  $k+1$  ecuaciones no lineales en los parámetros desconocidos que únicamente pueden resolverse de manera numérica (Collet, 1991).

Un método de solución es el método de Scoring de Fisher, en el cuál la solución está dada por

$$\hat{\beta}^{(m)} = \hat{\beta}^{(m-1)} + [I^{(m-1)}]^{-1} U^{(m-1)} \quad (2.3)$$

Donde  $\hat{\beta}^{(m)}$ , es el vector de parámetros estimados en la  $m$ -ésima iteración;  $I^{(m-1)}$  denota la matriz información evaluada en el vector de parámetros estimados  $\hat{\beta}^{(m-1)}$  y  $U$  es el vector de scores de primeras derivadas parciales de  $l$  evaluado en  $\hat{\beta}^{(m-1)}$ .

Habiendo obtenido el valor de los parámetros desconocidos puede estimarse el valor del predictor lineal como  $\hat{\eta}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + L + \hat{\beta}_k x_{kt}$  y a su vez estimar las probabilidades  $\pi_t$ :  $\hat{\pi}_t = \exp(\hat{\eta}_t) [1 + \exp(\hat{\eta}_t)]^{-1}$ .

Para inferir sobre los parámetros se utiliza el estadístico de Wald,  $W = (\hat{\beta} - \beta) I(\hat{\beta} - \beta)$ , que asintóticamente tiene distribución  $\chi_p^2$ , donde  $p = k+1$  es el número de parámetros en el vector  $\beta$  (Dobson, 1990). Para probar la hipótesis nula de que un coeficiente particular es cero se usa el estadístico t,  $t = (\hat{\beta}) / e.e(\hat{\beta})$  (e.e.=error estándar) cuya distribución asintótica es normal estándar bajo  $H_0$  (Myers, *et al.*, 2002).

La devianza es una buena medida del grado en el cual el modelo actual se desvía del modelo completo o saturado. Esta medida está basada en la razón de verosimilitud maximizada en el espacio paramétrico restringido por el modelo estimado actualmente,  $\hat{L}_c$ , y la verosimilitud maximizada del modelo para el cuál los valores estimados de las probabilidades coinciden con las proporciones observadas (espacio paramétrico irrestricto),  $\hat{L}_f$ . La devianza se obtiene con la expresión

$$D = -2 \ln(\hat{L}_c / \hat{L}_f) = -2 (\ln \hat{L}_c - \ln \hat{L}_f)$$

Y tiene distribución aproximada  $\chi_{T-p}^2$ . Se obtienen valores grandes de la devianza cuando  $\hat{L}_c$  es pequeño con respecto a  $\hat{L}_f$ , lo que indica que el modelo actual es

pobre. Se obtienen valores pequeños de la devianza cuando  $\hat{L}_c$  y  $\hat{L}_f$  son similares, lo que indica que el modelo actual es bueno.

Debido a que el valor esperado de una variable aleatoria con distribución  $\chi_v^2$  es  $v$ , al estimar un modelo de regresión logístico una regla útil de aceptación es cuando la devianza es aproximadamente igual a sus grados de libertad, o bien, si la devianza media (definida como el cociente  $D/v$ ), tiene un valor cercano a 1.

### 2.1.2 Problemas de estimación

Un problema que puede ocurrir al estimar un modelo es que las ecuaciones de estimación no tengan una solución única, en otras palabras, no existe un único estimador de máxima verosimilitud de los parámetros. La unicidad de los estimadores de máxima verosimilitud depende de la configuración de los puntos muestrales en el espacio de observación. Existen tres categorías mutuamente excluyentes basadas en la configuración de los puntos muestrales: completamente separados, cuasi-completamente separados y sobrepuestos. Las primeras dos configuraciones dan estimados infinitos no únicos mientras que la tercera configuración provee estimadores finitos y únicos (Albert y Anderson, 1984).

Sea  $x$  un vector de  $k$  covariables, la separación completa implica la existencia de un subvector  $x_s$  contenido en  $x$  para el cuál todas las observaciones pueden ser categorizadas como  $y_t = 0$  o  $y_t = 1$ . La presencia de este subvector causa monotonía en la función de log-verosimilitud y resultan estimadores de máxima verosimilitud  $\hat{\beta}_s$  para las variables en  $x_s$  que son infinitos negativa o positivamente

y errores estándar estimados que también son infinitos. La separación cuasi completa denota el caso cuando tal categorización ocurre únicamente para un subconjunto de observaciones.

Con respecto a la separación, una alternativa para tratar este problema es eliminar la variable o variables que causen el problema lo cuál no es muy atractivo porque la variable o variables problemáticas pueden tener una gran influencia en el fenómeno de interés (Zorn, 2005).

En estudios de simulación, cuando este problema no es de interés primario puede considerarse la alternativa de omitir las réplicas del experimento en las que los datos presenten este problema.

## 2.2 Ecuaciones de estimación generalizadas

Al estimar un modelo de regresión logístico con datos binomiales, si se tiene la certeza de que el modelo estimado es correcto, pero la devianza media es mayor que uno, puede ocurrir que el supuesto de la variabilidad binomial no se cumple, entonces se dice que los datos exhiben sobredispersión o heterogeneidad. Las causas aparentes de sobredispersión son: el componente sistemático del modelo es inadecuado en alguna forma (omisión de variables explicatorias relevantes u omisión de interacciones), la presencia de uno o más outliers en los datos, y la elección de una función de liga incorrecta. Una vez que estas causas posibles de devianza media inadecuada han sido eliminadas, la sobredispersión puede ser explicada por la ocurrencia de variación entre las probabilidades de respuesta o correlación entre las respuestas binarias (Collet, 1991).

La correlación entre observaciones binarias puede surgir principalmente cuando las observaciones son hechas sobre una o varias unidades experimentales a través del tiempo y a los datos así obtenidos son llamados datos longitudinales o datos de panel, en este caso cada panel representa una unidad experimental (Hardin y Hilbe, 2003).

Para el análisis de datos longitudinales Liang y Zeger (1986) presentaron una extensión de los modelos lineales generalizados. Esta extensión es una clase de ecuaciones de estimación generalizadas que proporcionan estimadores consistentes de los parámetros de la regresión y de sus varianzas.

El término ecuaciones de estimación generalizadas, indica que se introduce un componente de varianza de segundo orden directamente en la ecuación de estimación conjunta. Nos referimos a estimación conjunta cuando simplemente las observaciones son agrupadas ignorando a cual panel pertenecen naturalmente.

Suponiendo que tenemos  $n$  paneles, las ecuaciones de estimación generalizadas para modelos lineales generalizados son de la forma

$$\Psi(\beta) = \left[ \left\{ \sum_{i=1}^n \sum_{t=1}^{n_i} \frac{y_{it} - \mu_{it}}{a(\varphi)V(\mu_{it})} \left( \frac{\partial \mu}{\partial \eta} \right)_{it} x_{jit} \right\}_{j=1,2,\dots,p} \right]_{p \times 1} = [0]_{p \times 1} \quad (2.4)$$

Donde  $y_{it}$  es la  $t$ -ésima observación en el  $i$ -ésimo individuo o unidad experimental;  $\mu_{it}$  es la media de la variable de respuesta;  $a(\varphi)$  es un parámetro de escala en la función de distribución de la variable de respuesta y  $\eta$  es el predictor lineal.

Reescribiendo esta expresión en términos de los paneles tenemos

$$\Psi(\beta) = \left[ \left\{ \sum_{i=1}^n \mathbf{x}'_{ji} D \left( \frac{\partial \mu}{\partial \eta} \right) [V(\mu_i)]^{-1} \begin{pmatrix} y_i - \mu_i \\ \mathbf{a}(\varphi) \end{pmatrix} \right\}_{j=1,2,\dots,p} \right]_{p \times 1} = [0]_{p \times 1} \quad (2.5)$$

Donde  $D()$  denota una matriz diagonal,  $V(\mu_i)$  es una matriz diagonal que puede descomponerse en la forma

$$V(\mu_i) = \left[ D(V(\mu_{it}))^{1/2} I_{(n_i \times n_i)} D(V(\mu_{it}))^{1/2} \right] \quad (2.6)$$

En esta expresión observamos que estas ecuaciones consideran a las observaciones dentro de los paneles como independientes.

Las ecuaciones de estimación generalizadas propuestas por Liang y Zeger (1986) se basan en la distribución marginal de la variable de respuesta, para la cual el valor esperado y las funciones de varianza son promediadas sobre los paneles, entonces la matriz identidad en (2.6) es la matriz de correlación dentro de los paneles y puede ser reemplazada por estructuras de correlación diferentes de manera que la forma de la matriz de varianza resulte

$$V(\mu_i) = \left[ D(V(\mu_{it}))^{1/2} R(\alpha)_{n_i \times n_i} D(V(\mu_{it}))^{1/2} \right]_{n_i \times n_i} \quad (2.7)$$

Donde  $R(\alpha)$  es una matriz simétrica que cumple con los requerimientos de una matriz de correlación y es llamada matriz de correlación de trabajo,  $\alpha$  es un  $n_i$  vector que caracteriza completamente a  $R(\alpha)$ .

Cuando  $R(\alpha)$  en (2.7) es la matriz identidad como en (2.6), el modelo de estimación es llamado modelo de independencia.

La expresión (2.7) será igual a  $Cov(Y_i)$  si  $R(\alpha)$  es la verdadera matriz de correlación para las  $Y_i$ 's (Hardin y Hilbe, 2003).

### 2.2.1 Parametrizando la matriz de correlación de trabajo

Cuando se piensa que las observaciones dentro de los paneles no tienen un orden específico y que están igualmente correlacionadas, únicamente un parámetro adicional debe ser estimado. Alternativamente, podemos hipotetizar una estructura más complicada bajo el supuesto de que las observaciones dentro de los paneles tienen un orden específico. En este caso, requerimos de un vector de parámetros adicionales de hasta  $\max_i n_i - 1$  parámetros, o una matriz entera de parámetros de hasta  $\binom{\max_i n_i}{2} - n_i$  parámetros adicionales (Hardin y Hilbe, 2003).

Formalmente podemos escribir el estimador para los parámetros de asociación auxiliares como la ecuación de estimación

$$\Psi(\alpha) = \sum_{i=1}^n \left( \frac{\partial \xi_i}{\partial \alpha} \right) H_i^{-1} (W_i - \xi_i) = [0]_{qx1} \quad (2.8)$$

Donde

$$W_i = (r_{i1}r_{i2}, r_{i1}r_{i3}, \dots, r_{i(n_i-1)}r_{in_i})'_{qx1} \quad (2.9)$$

$$H_i = D(V(W_{ij}))_{q \times q} \quad (2.10)$$

$$\xi_i = E(W_i)_{q \times 1} \quad (2.11)$$

En estas ecuaciones  $r_{ij}$  es el  $ij$ -ésimo residual de Pearson,  $H_i$  es una matriz diagonal y  $q = \binom{n_i}{2}$ . La matriz de correlación entra a través de la ecuación (2.11).

Las siguientes secciones presentan planteamientos para especificar la estructura estimada de correlación dentro de los paneles.

### 2.2.1.1 Correlación autorregresiva

Cuando las observaciones dentro de los paneles tienen un orden natural, es razonable considerar que la asociación entre ellas depende del tiempo y una parametrización autorregresiva de la matriz de correlación es adecuada.

En este caso  $\alpha$  es un vector que se estima con la expresión (Hardin y Hilbe, 2003)

$$\alpha = \frac{1}{\phi} \left[ \sum_{i=1}^n \left( \frac{\sum_{t=1}^{n_i-0} \{f_{i,t} f_{i,t+0}\}}{n_i}, \dots, \frac{\sum_{t=1}^{n_i-k} \{f_{i,t} f_{i,t+k}\}}{n_i} \right) \right] \quad (2.12)$$

Donde  $\phi^{-1} = \sum_i \sum_{t=1}^{n_i} \{f_{it}^2\} / n-p$ ,  $\{f_{it}^2\}$  son los residuales de Pearson obtenidos al estimar el

modelo y se calculan con la expresión

$$\{f_{it}\} = (y_{it} - \hat{\mu}_{it}) / \sqrt{V(\hat{\mu}_{it})}$$



Otra forma de estimar  $\alpha$  es con la pendiente de la regresión de  $\ln(\frac{t'}{t})$  sobre  $\ln(|t-t'|)$  (Liang y Zeger, 1986).

### 2.2.1.2 Correlación estacionaria

En esta estructura se puede hipotetizar que las correlaciones existen para un número pequeño de unidades de tiempo. Aquí, se especifica una diferencia en tiempo máxima para la cual las observaciones podrían estar correlacionadas, de esta forma la matriz de correlación resulta en forma de banda diagonal. En este caso  $\alpha$  es un vector de correlaciones para  $k$  rezagos especificados por el analista (Hardin y Hilbe, 2003).

Usando los residuales de Pearson, la estimación de  $\alpha$  se hace en la misma forma que la correlación autorregresiva y usamos la expresión (2.12).

La matriz de correlación hipotetizada contiene unos en la diagonal principal,  $\alpha_1$  en las primeras diagonales adyacentes a la principal,  $\alpha_2$  en las segundas diagonales adyacentes a la principal y así sucesivamente. Los elementos de esta matriz se pueden describir como

$$R_{uv} = \begin{cases} \alpha_{|u-v|} & \text{si } |u-v| \leq k \\ 0 & \text{de otro modo} \end{cases}$$

### 2.2.1.3 Correlación no estacionaria

En esta estructura de correlación cada elemento es estimado de la información disponible. Mediante esta estructura se especifica una matriz de parámetros  $\alpha$  dada por la siguiente expresión (Hardin y Hilbe, 2003)

$$\alpha = G \left( \sum_{i=1}^n n_i \right) \left( \sum_{i=1}^n \sum_{t=1}^{n_i} f_{i,t}^2 / n_i \right)^{-1} \quad (2.13)$$

$$G = \begin{bmatrix} g_{1,1} \left. \vphantom{g_{1,1}} \right\}^2 & g_{1,2} \left. \vphantom{g_{1,2}} \right\}^2 & L & g_{1,n_i} \left. \vphantom{g_{1,n_i}} \right\}^2 \\ g_{2,1} \left. \vphantom{g_{2,1}} \right\}^2 & g_{2,2} \left. \vphantom{g_{2,2}} \right\}^2 & L & g_{2,n_i} \left. \vphantom{g_{2,n_i}} \right\}^2 \\ M & M & O & M \\ g_{n_i,1} \left. \vphantom{g_{n_i,1}} \right\}^2 & g_{n_i,2} \left. \vphantom{g_{n_i,2}} \right\}^2 & K & g_{n_i,n_i} \left. \vphantom{g_{n_i,n_i}} \right\}^2 \end{bmatrix}$$

$$g_{u,v} = \left( \sum_{i=1}^n I(i,u,v) \right)^{-1}, \quad I(i,u,v) = \begin{cases} 1 & \text{si el panel } i \text{ tiene observaciones en } (u,v) \\ 0 & \text{de otro modo} \end{cases}$$

Al igual que en la estructura estacionaria, la matriz de correlación no estacionaria usa las correlaciones estimadas para un número  $g$  determinado de bandas. La matriz de correlación de trabajo se especifica como

$$R_{uv} = \begin{cases} 1 & \text{si } u=v \\ \alpha_{uv} & \text{si } 0 < |u-v| \leq g \\ 0 & \text{de otro modo} \end{cases}$$

La matriz de correlación no estacionaria no necesariamente es invertible y al calcularse pueden encontrarse problemas numéricos, especialmente para conjuntos de datos desbalanceados.

#### 2.2.1.4 Correlación no estructurada

La matriz de correlación no estructurada es el caso más general de correlación y es igual a la matriz de correlación no estacionaria para el máximo rezago. La matriz de correlación es especificada como  $R=\alpha$ , donde  $\alpha$  está especificada en la ecuación (2.13). Al igual que en la estructura de correlación no estacionaria, la matriz de correlación estimada no necesariamente es invertible y en el proceso de estimación pueden encontrarse problemas numéricos (Hardin y Hilbe, 2003).

Existen otras estructuras de correlación que pueden considerarse, por ejemplo; la correlación intercambiable, en la que se supone que las observaciones entre paneles tienen una correlación común, sin embargo esta estructura es apropiada cuando los datos no tienen una dependencia en el tiempo.

Estas ecuaciones de estimación enfatizan en la dependencia entre las observaciones de los sujetos para estimar consistentemente la varianza de los estimadores propuestos aún cuando la estructura de correlación propuesta sea incorrecta como frecuentemente puede ocurrir (Liang y Zeger, 1986).

#### 2.2.2 Procedimiento de estimación

Combinando las ecuaciones de estimación (2.4) para los parámetros de la regresión y (2.8) para los parámetros auxiliares, las ecuaciones de estimación completas están dadas por

$$\psi(\beta, \alpha) = (\psi_\beta(\beta, \alpha), \psi_\alpha(\beta, \alpha)) = \begin{pmatrix} \sum_{i=1}^n \mathbf{x}'_{ji} D\left(\frac{\partial \mu_i}{\partial \eta}\right) (V(\mu_i))^{-1} \left(\frac{y_i - \mu_i}{a(\varphi)}\right) \\ \sum_{i=1}^n \left(\frac{\partial \xi_i}{\partial \alpha}\right)' H_i^{-1}(W_i - \xi_i) \end{pmatrix} \quad (2.14)$$

$$V(\mu_i) = D(V(\mu_{it}))^{1/2} R(\alpha) D(V(\mu_{it}))^{1/2} \quad (2.15)$$

En cada paso del algoritmo de estimación, se estima  $R(\alpha)$  y posteriormente es usada para estimar  $\beta$ . El algoritmo se detiene cuando un cambio en los parámetros estimados en dos pasos sucesivos es menor al valor propuesto por algún criterio de convergencia.

### 2.2.3 Probando coeficientes en ecuaciones de estimación generalizadas

La prueba de hipótesis de interés está basada en la hipótesis nula  $H_0: \beta_r = \beta_0$  donde  $\beta_r$  es un subvector de parámetros y  $\beta_0$  es el valor hipotetizado de  $\beta_r$ .

Sea  $\hat{\beta}$  el vector de coeficientes estimados y  $\hat{\beta}_1$  el vector de coeficientes estimados basado en el modelo de independencia. El estadístico para la prueba basada en  $\hat{\beta}$  es construido usando el estimador de varianza conjunto, el cuál ignora la correlación dentro de los paneles, y asumiendo que el estadístico tiene distribución  $\chi_r^2$ . El planteamiento basado en  $\hat{\beta}_1$  se caracteriza por construir el estadístico de prueba usando un estimado de la varianza robusta, en este caso también se asume que dicho estadístico tienen distribución  $\chi_r^2$  (Hardin y Hilbe, 2003).

### 2.2.3.1 Pruebas de Wald

Los estadísticos de prueba típicamente son calculados usando pruebas de Wald. Usando el estimador sándwich de la varianza se obtiene un estadístico de prueba de Wald generalizado

$$T_w = n(\hat{\beta}_r - \beta_0)' V_{MS}^{-1}(\hat{\beta}_r - \beta_0) \quad (2.16)$$

Donde  $n$  es el número de individuos; MS indica sándwich modificado. En este estadístico generalmente son usados  $r$  grados de libertad equivalentes al número de parámetros en  $\hat{\beta}$ . Al usar el estimador sándwich de la varianza para construir el estadístico de prueba debemos asegurarnos que hay menos covariables que individuos, ya que de otra forma el estimador sándwich de varianza es singular (Hardin y Hilbe, 2003).

### 2.3 Modelo de cadena de Markov de primer orden

Consideremos ahora el proceso estocástico  $\{y_t, t=0,1,2..\}$  donde  $y_t$  toma únicamente dos valores: 0 ó 1. Si  $y_t=u$  se dice que el proceso está en el estado  $u$  al tiempo  $t$ . Suponemos que siempre que el proceso esté en el estado  $u$  al tiempo  $t$ , hay una probabilidad fija  $P_{uv}$  de que posteriormente esté en el estado  $v$ . Es decir

$$P_{uv} = P[y_{t+1} = v | y_t = u, y_{t-1} = u_{t-1}, \dots, y_1 = u_1], \quad t=1, \dots, T \quad (2.17)$$

Para todos los estados  $u_1, \dots, u_{t-1}, u, v$  y  $t \geq 1$ . Tal proceso estocástico es conocido como Cadena de Markov (Ross, 1996).

La expresión (2.17) nos dice que la distribución condicional de cualquier estado futuro  $y_{t+1}$  dados los estados pasados  $y_1, \dots, y_{t-1}$  y el actual  $y_t$ , es independiente de los estados pasados y depende únicamente del actual, a lo que se le llama propiedad Markoviana y así el proceso  $\{y_t, t=0,1,2,\dots\}$  es un proceso de cadena de Markov de primer orden y (2.17) puede expresarse como

$$P_{uv} = P[y_t = v | y_{t-1} = u], t = 1, \dots, T$$

Dado que el proceso debe hacer una transición a algún estado tenemos

$$P_{uv} \geq 0, u, v \geq 0; \sum_{v=0}^{\infty} P_{uv} = 1, u=0,1,\dots$$

Y la matriz de probabilidades de transición es

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} \text{ y } P_{uv} = P(y_{t+1}=v | y_t=u).$$

Calculando la función de log-verosimilitud  $l$  de las probabilidades  $P_{uv}$  usando la propiedad Markoviana resulta

$$l = l(y_1) + l(y_2 | y_1) + \dots + l(y_t | y_{t-1})$$

Todos los términos a excepción del primero están dados por la matriz  $P$ . El primer término será considerado como fijo de forma que contribuye como un factor constante en  $l$ . De este modo la log-verosimilitud resulta

$$l = \sum r_{uv} \ln P_{uv}$$

Donde  $r_{uv}$  es la proporción observada de transiciones del estado u al v.

Esta log-verosimilitud es la misma del muestreo de una distribución multinomial y los estimadores de máxima verosimilitud de los parámetros individuales  $P_{uv}$  son las proporciones correspondientes.

Cuando se asume estacionaridad en el análisis de cadena de Markov se asume que  $P_{uv}(t) = P_{uv}$ . En muchas situaciones este supuesto no es apropiado de manera que  $P_{uv}$  no será constante a través del tiempo. Una forma de aproximar estas probabilidades es usando la log-verosimilitud que resulta

$$l = \sum_{t=0}^T \sum_u [n_{u1}(t) \ln P_u(t) + n_{u0}(t) \ln (1 - P_u(t))]$$

Donde  $P_u(t) = P[y(t)=1 | y(t-1)=u]$ ;  $n_{uv}(t)$ , es el número de periodos en los que el proceso está en el estado v.

Esta log-verosimilitud es la misma que el Kernel de la log-verosimilitud para un modelo multinomial de una tabla de contingencia de  $2 \times 2 \times T$  o para  $2T$  probabilidades binomiales, donde  $n_{u1}(t) \sim B(n_{u+}(t), P_u(t))$ , el signo “+” indica suma sobre el subíndice.

Stern y Coe (1984) proponen este procedimiento de aproximación mediante el modelo

$$P_u(t) = g(\eta_u(t))$$

Donde  $g$  es la función de liga logística y  $\eta$  es el predictor lineal en función del tiempo únicamente.

Si nuestro interés es modelar la probabilidad de que el estado actual de la cadena sea 1 dado el estado anterior en función de  $k$  variables explicatorias  $x_1, \dots, x_k$ , entonces la probabilidad  $P_u(t) = P[y_t = 1 | y_{t-1} = u, x_{1t}, \dots, x_{kt}]$  es igual a la probabilidad  $P_{z_u}(t) = P[z_{ut} = 1 | x_{1t}, \dots, x_{kt}]$  donde  $z_{ut}$  es el  $t$ -ésimo elemento de un vector  $z_u$  que toma en cuenta las transiciones de la cadena  $\{y_t\}_{t=1}^T$  del estado  $u$  en el tiempo  $t-1$  al estado 1 en el tiempo  $t$ . El elemento  $z_{0t}$  del vector  $z_0$  es 1 si  $y_t=1$  y  $y_{t-1}=0$  y es 0 si  $y_t=0$  y  $y_{t-1}=0$  ó 1. El elemento  $z_{1t}$  del vector  $z_1$  es 1 si  $y_t=1$  y  $y_{t-1}=1$  y es 0 si  $y_t=0$  y  $y_{t-1}=0$  ó 1.

De este modo se tendrán dos modelos de probabilidad  $P_u(t)$ , uno para cada valor de  $u$  y por lo tanto dos vectores de parámetros.

Para modelar la dependencia de  $P_{u_1}(t)$  en las  $k$  variables explicatorias, se usa el modelo de regresión logístico para el cuál esta dependencia tomará la forma

$$P_u(t) = \frac{\exp(\beta_0 + \beta_1 x_{1t} + L + \beta_k x_{kt})}{1 + \exp(\beta_0 + \beta_1 x_{1t} + L + \beta_k x_{kt})} \quad (2.18)$$

Cuya estimación e inferencia se describió en la sección 2.1.



### CAPÍTULO III. ESTUDIO DE SIMULACIÓN

Se realizó un estudio de simulación para mostrar la existencia del fenómeno de correlación espuria en modelos de regresión de respuesta binaria cuando las observaciones son generadas por procesos de serie de tiempo. Los modelos de regresión utilizados son los mismos que se describieron en el capítulo 2, como se indicó; el modelo de cadena de Markov y el de ecuaciones de estimación generalizadas consideran observaciones serialmente correlacionadas.

En cada modelo simulado se observó:

- a) El tamaño de prueba dada la hipótesis  $H_0:\beta_1=0$ , donde  $\beta_1$  es el coeficiente de una variable explicatoria  $x_t$ ; se espera que bajo  $H_0$  el tamaño de la prueba sea menor o igual al nivel de significancia empleado, cosa que no ocurre en presencia de correlación espuria.
- b) El valor máximo del estadístico t para el coeficiente  $\beta_1$ ; bajo la hipótesis nula se espera que el estadístico t tenga una distribución normal y que sus valores no estén muy alejados de la media de la distribución.

En el modelo de ecuaciones de estimación generalizadas se estudian dos situaciones: en una de ellas se considera únicamente a la variable  $x_t$  como variable explicatoria y en la otra además de  $x_t$  se considera el efecto de un tratamiento  $\tau$  en la mitad de los individuos; por lo tanto, en el primer caso la hipótesis de interés es  $H_0:\beta_1=0$  y el valor del estadístico t para  $\beta_1$ ; en el segundo

caso las hipótesis de interés son  $H_{0\tau}: \tau=0$  y  $H_{0\beta_1}: \beta_1 = 0$  y se obtiene el valor del estadístico para  $\tau$  y para  $\beta_1$ .

En cada modelo se simularon 49 situaciones resultantes de combinar 7 procesos de serie de tiempo como generadores de las observaciones de la variable explicatoria  $x_t$  y de la variable de respuesta  $y_t$ . Las variables  $x_t$  y  $y_t$  se generaron de manera independiente, por lo tanto las hipótesis  $H_0:\beta_1=0$ ,  $H_{0\tau}: \tau=0$  y  $H_{0\beta_1}: \beta_1 = 0$  son verdaderas.

### 3.1 Procesos de serie de tiempo considerados

En la expresión (3.1) se muestran los procesos de serie de tiempo considerados en el presente estudio como generadores de la variable explicatoria  $x_t$  y la variable de respuesta  $y_t$ ; los procesos con cambio estructural son casos especiales de los procesos considerados en Noriega y Ventosa-Santaulária (2006).

$$\begin{array}{lll}
 1) & I(0) & z_t = \mu_z + u_{zt} \\
 2) & I(0)+br & z_t = \mu_z + \theta_z DU_{zt} + u_{zt} \\
 3) & TS & z_t = \mu_z + \beta_z t + u_{zt} \\
 4) & TS+br & z_t = \mu_z + \theta_z DU_{zt} + \beta_z t + \phi_z DT_{zt} + u_{zt} \\
 5) & I(1) & z_t = z_{t-1} + u_{zt} \\
 6) & I(1)+dr & z_t = \mu_z + z_{t-1} + u_{zt} \\
 7) & I(1)+dr+br & z_t = \mu_z + z_{t-1} + \theta_z DU_{zt} + u_{zt}
 \end{array} \tag{3.1}$$

Donde  $u_{zt} \sim \text{niid}(0, \sigma_z^2)$ ,  $DU_{zt} = 1(t > Tb_z)$  es una variable indicadora que afecta a los periodos de tiempo posteriores a la fecha del cambio estructural,  $Tb_z$ , en una

$$\text{magnitud } \theta_z, \text{ es decir } DU_{zt} = \begin{cases} 1 & t > Tb_z \\ 0 & t \leq Tb_z \end{cases}.$$

$DT_{zt} = (t - Tb_z)1(t > Tb_z)$  es una variable indicadora que afecta la pendiente de la serie en una magnitud  $\varphi_z$  en los periodos posteriores a la fecha del cambio estructural.  $I(\cdot)$  indica el orden de integración del proceso; br corte estructural; dr tendencia lineal y TS tendencia estacionaria.

De acuerdo a las definiciones del capítulo 1, el proceso 1 de la expresión (3.1) es un proceso estacionario; el 2 también es estacionario pero con un cambio estructural o cambio en el nivel de las observaciones; el 3 es un proceso con tendencia lineal en función del tiempo; el 4 es un proceso con tendencia lineal pero con un cambio estructural; el 5 es una caminata aleatoria; el 6 es una caminata aleatoria con tendencia lineal; y el 7 es una caminata aleatoria con tendencia y cambio estructural.

### 3.2 Generación de observaciones

Los componentes estocásticos en la expresión (3.1) se obtuvieron con un generador de números aleatorios asumiendo que la varianza del término aleatorio  $u_{zt}$  es igual a uno ( $\sigma_z^2 = 1$ ) y que la fecha de cambio estructural se distribuye de mane uniforme en  $(1, T)$ . La generación de las series de observaciones de la

variable explicatoria  $x_t$ , fue directa de (3.1) ya que estas observaciones son continuas.

Para generar las observaciones binarias de la variable de respuesta  $y_t$ , se consideró que  $y_t \sim \text{Bernoulli}(\pi_t)$ . Generamos una secuencia de observaciones  $\{w_t\}_{t=1}^T$  en base a un proceso de serie de tiempo de la expresión (3.1) que llamaremos  $w_t$  y suponemos que las probabilidades  $\pi_t$  se relacionan con  $w_t$  a través de la función logística  $\pi_t = \exp(\beta_0 + \beta_1 w_t) [1 + \exp(\beta_0 + \beta_1 w_t)]^{-1}$ , donde  $\beta_0$  y  $\beta_1$  son constantes arbitrarias distintas de cero, específicamente se asignó  $\beta_0 = -1$  y  $\beta_1 = 0.5$ . Una vez que se tiene  $\pi_t$  obtenemos el valor de  $y_t$  con el método de la transformada inversa (véase Ross, 1999).

Habiendo generado las observaciones de cada variable, de respuesta y explicatoria, se procedió a la simulación en los modelos de regresión como a continuación se describe.

### 3.3 Simulación en el modelo de regresión logístico

En este apartado la simulación consistió en modelar la dependencia de  $P(y_t=1)$  en  $x_t$  ( $t=1, \dots, T$ ) a través del modelo de regresión logístico aunque el proceso generador de los datos de  $y_t$  fue independiente del proceso generador de  $x_t$ ; esto es la hipótesis  $H_0: \beta_1=0$  es cierta.

En cada regresión registramos el rechazo (1) o no rechazo (0) de la hipótesis nula  $H_0: \beta_1=0$  y el valor del estadístico  $t$ . Este experimento se realizó con 1000

repeticiones para cada uno de los diferentes tamaños de muestra. En cada tamaño de muestra se obtuvo el valor máximo del estadístico  $t$ . Finalmente el tamaño de la prueba es estimado con el cociente  $nr / r$  donde  $nr$  es el número de rechazos y  $r$  el número de repeticiones del experimento.

Los tamaños de muestra fueron 25, 50, 75, 100, 250, 500, 750, 1000, 2500 y 5000.

En el Anexo II Figura A1, se encuentra el diagrama de flujo utilizado para la simulación en este modelo.

#### 3.4 Simulación en el modelo de ecuaciones de estimación generalizadas

El estudio de simulación en este modelo comprende dos escenarios que se diferencian por las covariables presentes en el predictor lineal de la expresión (2.4), en un escenario existe solo una variable explicatoria continua y en otro existe una variable explicatoria continua y otra variable explicatoria binaria que representa la presencia (o ausencia) del efecto de un tratamiento.

Cada escenario se simuló con dos estructuras de correlación; una de ellas es la independiente y la otra es no estructurada. La estructura de correlación independiente se considera más apropiada cuando los procesos 1 y 2 de (3.1) generan las observaciones de la variable de respuesta y explicatoria ya que estos procesos generan observaciones no correlacionadas. Cuando en el modelo intervienen los procesos 3 a 7 la forma de correlación no estructurada se

considera más apropiada ya que esta matriz es una matriz autoregresiva del máximo rezago que puede existir en los datos.

La forma de las observaciones en la aplicación de este modelo es un poco distinta de la forma de los datos en el modelo de regresión logístico y en el de cadena de Markov de primer orden. En este modelo consideramos que tenemos  $n$  individuos (o paneles) sobre los que se realiza un conjunto de  $n_i$  observaciones en el tiempo ( $1 \leq i \leq n$ ), en nuestro caso  $n_i$  es la misma para toda  $i$ . Se quiere modelar la probabilidad  $P(y_{it} = 1)$  ( $1 \leq t \leq n_i$ ) en función de una o dos covariables para lo cual utilizamos el modelo logístico.

#### 3.4.1 Modelo con una variable explicatoria

El modelo de regresión logístico para modelar la dependencia de  $P(y_{it} = 1)$  en una variable explicatoria es de la forma

$$P(y_{it} = 1 | x_{it}) = \frac{\exp(\beta_0 + \beta_1 x_{it})}{1 + \exp(\beta_0 + \beta_1 x_{it})}, \quad t = 1, \dots, n_i \quad (3.2)$$

Las secuencias de observaciones  $\{y_t\}_{t=1}^{n_i}$  y  $\{x_t\}_{t=1}^{n_i}$  son generadas de manera independiente por algún proceso de la expresión (3.1) de acuerdo a alguna de las 49 combinaciones de estos procesos.

El experimento consistió en estimar el modelo (3.2) con 6 individuos (i.e.  $1 \leq i \leq 6$ ) registrando el rechazo (1) de la hipótesis  $H_0: \beta_1 = 0$  y el valor del estadístico  $t$ .

El experimento se realizó con 1000 repeticiones de las cuales se obtuvo el valor máximo del estadístico  $t$  y tamaño de la prueba para diferentes tamaños de muestra  $n_i$ , como se describió con anterioridad. Los tamaños de muestra considerados son 3, 4, 5, 6, 7 y 8.

### 3.4.2 Modelo con una variable explicatoria continua y un tratamiento

En este caso la expresión (3.2) es ligeramente distinta ya que deseamos modelar la probabilidad  $P(y_{it} = 1)$  en función del efecto de un tratamiento y de una variable explicatoria continua, al considerar el tratamiento la expresión (3.2) resulta

$$P(y_{it} = 1 | \tau_{it}, x_{it}) = \frac{\exp(\beta_0 + \tau_{it} + \beta_1 x_{it})}{1 + \exp(\beta_0 + \tau_{it} + \beta_1 x_{it})}, \quad t = 1, \dots, n_i \quad (3.3)$$

Las secuencias de observaciones  $\{y_t\}_{t=1}^{n_i}$  y  $\{x_t\}_{t=1}^{n_i}$  nuevamente se generan de manera independiente por algún proceso de la expresión (3.1) de acuerdo a alguna de las 49 combinaciones de estos procesos; el tratamiento  $\tau$  es aplicado a la mitad de los individuos de tal forma que  $\tau_{it} = 0$  para  $n/2$  individuos y  $\tau_{it} = 1$  para el grupo restante.

El experimento consistió en estimar el modelo (3.3) con 6 individuos (i.e.  $1 \leq i \leq 6$ ) registrando el rechazo (1) de las hipótesis  $H_{0\beta} : \beta_1 = 0$  y  $H_{0\tau} : \tau = 0$  y el valor del estadístico para  $\tau$  y  $\beta_1$ . Bajo la hipótesis nula  $H_{0\tau} : \tau = 0$  no hay efecto del tratamiento  $\tau$ .

Con 1000 repeticiones de este experimento para cada tamaño de muestra se obtuvo una estimación del tamaño de la prueba de las hipótesis  $H_{0\beta}$  y  $H_{0\tau}$  y el valor máximo del estadístico  $t$  para  $\tau$  y  $\beta_1$ . Los tamaños de muestra  $n_i$  considerados son 3, 4, 5, 6, 7 y 8.

En el Anexo II Figura A2, se encuentra el diagrama de flujo general utilizado para la simulación en este modelo.

### 3.5 Simulación en el modelo de cadena de Markov de primer orden

Ahora suponemos que la secuencia de observaciones  $\{y_t\}_{t=1}^T$  es un proceso estocástico de cadena de Markov de primer orden con dos estados: 0 y 1, nuestro interés es modelar la probabilidad de que el estado de la cadena en el periodo actual sea el estado 1 dado el estado anterior en función de una variable explicatoria  $x_t$  cuando  $y_t$  y  $x_t$  son generadas de manera independiente. El modelo de regresión utilizado es el modelo de regresión logístico, entonces la expresión (2.18) resulta

$$P_u(t) = \frac{\exp(\beta_0 + \beta_1 x_t)}{1 + \exp(\beta_0 + \beta_1 x_t)} \quad (3.4)$$

Donde  $P_u(t) = P[y_t = 1 | y_{t-1} = u, x_t]$ .

Pero  $P_u(t)$  es equivalente a la probabilidad  $P(z_{ut}=1)$  donde la secuencia  $\{z_{ut}\}_{t=1}^T$  es la secuencia de transiciones de la variable binaria  $y_t$  del estado  $u$  al estado 1 ( $u=0,1$ ); se tendrán dos secuencias de este tipo y dos modelos, uno basado en  $u=0$  y



otro basado en  $u=1$ . En cada uno de estos modelos se registró el rechazo (1) de la hipótesis nula  $H_{u0}: \beta_{u1}=0$  y el valor del estadístico  $t$ . El subíndice  $u$  en  $H$  y en  $\beta$  se utilizó para diferenciar a que modelo pertenece la hipótesis nula y el coeficiente en función del valor de  $u$ .

El experimento se realizó con 1000 repeticiones para cada tamaño de muestra. Se obtuvo el valor máximo del estadístico  $t$  y la estimación del tamaño de la prueba como se describió con anterioridad. Los tamaños de muestra son 25, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900 y 1000.

En el Anexo II Figura A3, se encuentra el diagrama de flujo utilizado para la simulación en este modelo.

### 3.6 Problemas de estimación

Al utilizar el modelo de regresión logístico y el modelo de cadena de Markov para modelar un conjunto de datos, es probable encontrar problemas de estimación basados en la configuración de los puntos muestrales u observaciones como se mencionó en la sección 2.1.2.

Para mostrar consistentemente la existencia del fenómeno de correlación espuria, durante la simulación en el modelo de regresión logístico y en el modelo de cadena de Markov se eliminaron las series de observaciones que presentaron problemas de convergencia.

Para detectar el problema de cuasiseparación se utilizó un procedimiento del mismo paquete que identifica el número de puntos muestrales que se encuentran

sobrepuestos, cuando únicamente un punto muestral se encontró sobrepuesto se procedió a eliminar esa repetición del experimento.

En el modelo de ecuaciones de estimación generalizadas, un problema de estimación está relacionado con la forma de la matriz de correlación hipotetizada, mencionamos que cuando la matriz hipotetizada es no estructurada la matriz estimada puede no ser invertible y en consecuencia encontrarse problemas numéricos. En este caso el paquete estadístico indica cuando ocurre este problema y en tal situación la repetición del experimento fue omitida.

## CAPÍTULO IV. RESULTADOS Y DISCUSIÓN

Los resultados obtenidos del estudio de simulación se discuten en las siguientes secciones. Los resultados del modelo de regresión logístico y del modelo de cadena de Markov se presentan en forma numérica y en forma gráfica. Los resultados del modelo de ecuaciones de estimación generalizadas son reportados únicamente en forma numérica.

### 4.1 Modelo de regresión logístico

Presentamos los resultados obtenidos del comportamiento del tamaño de la prueba estimado y el valor máximo del estadístico  $t$  en función del tamaño de muestra.

#### 4.1.1 Tamaño de la prueba

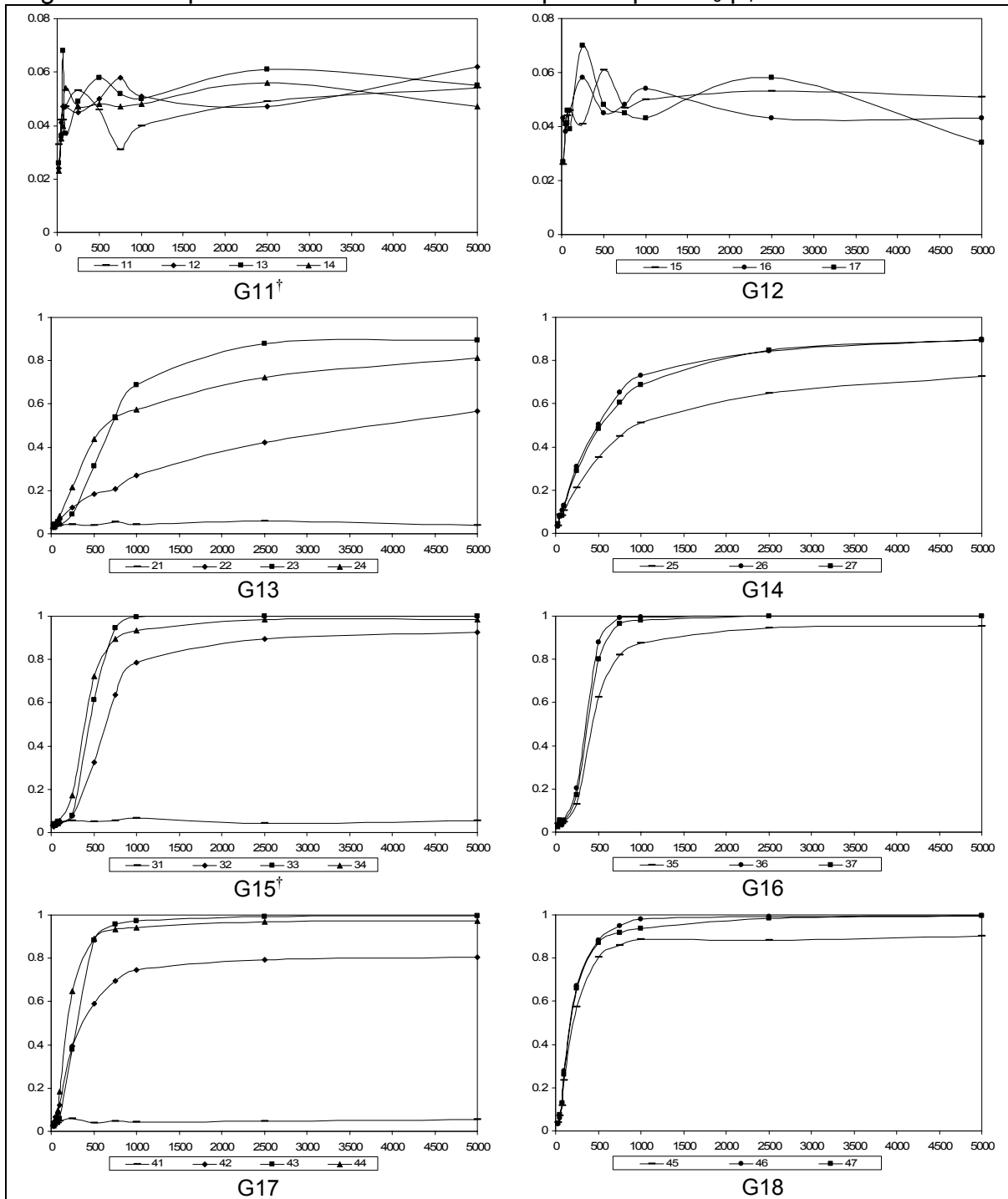
El comportamiento del tamaño de la prueba se presenta en la Figura 1 (en el Anexo I se reportan los valores numéricos de estas estimaciones, Cuadro A1), primeramente observamos las combinaciones donde el proceso 1 interviene como generador de  $y_t$  y los procesos 1 a 7 intervienen como generadores de  $x_t$  (Gráficas G11 y G12). En estas combinaciones el tamaño de la prueba se mantiene alrededor de 0.05. También observamos en las gráficas G13, G15, G17, G19, G111 y G113 en las que el proceso 1 aparece como generador de  $x_t$ , el tamaño de la prueba toma valores relativamente pequeños en cada tamaño de muestra. Estos resultados son similares a lo observado en el modelo de regresión lineal

cuando el proceso 1 está presente en el modelo (Ver Noriega y Ventosa-Santaulária, 2006).

En las gráficas G13 a G114 se observa que el tamaño de la prueba en las combinaciones donde no está presente el proceso 1, tiende a uno a medida que aumenta el tamaño de muestra aunque la tasa a la que incrementa es distinta en cada combinación, esto implica que la probabilidad de ocurrencia de correlación espuria aumenta con el tamaño de muestra; por ejemplo cuando el proceso 2 genera a  $y_t$  o a  $x_t$  o a ambos la velocidad de crecimiento del tamaño de la prueba es menor que en los demás casos. En algunas combinaciones, por ejemplo 57, 66, 67, 75, 76 y 77 ocurre este fenómeno aún para el tamaño de muestra mas pequeño que se utilizó (25) ya que el tamaño de la prueba en estos casos es mayor que 0.1 (Véase Cuadro 1).

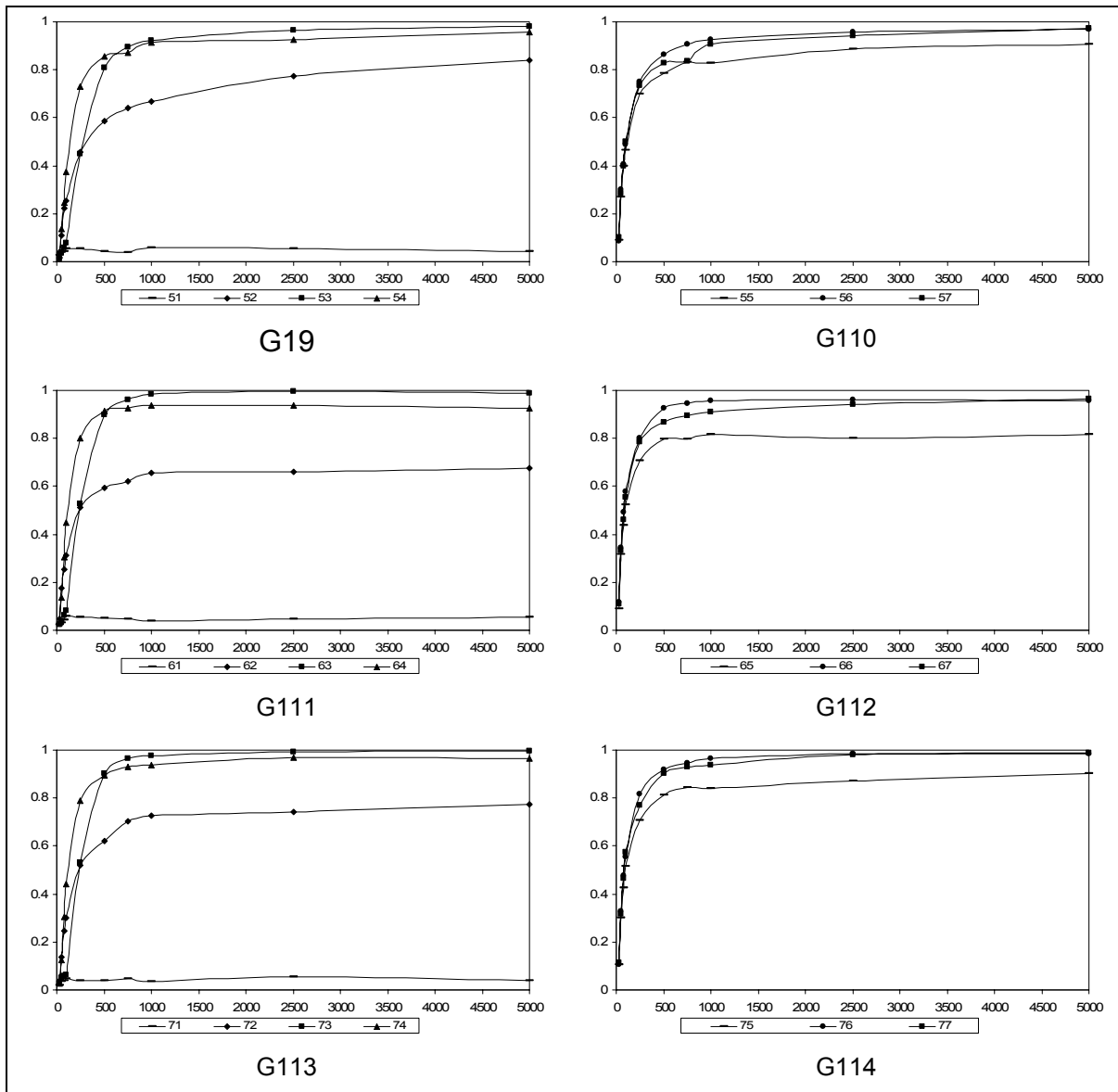
En regresión lineal en las combinaciones donde no está presente el proceso 1 ocurre algo similar a lo observado en este modelo, el tamaño de la prueba tiene a uno a medida que aumenta el tamaño de muestra (Véase Noriega y Ventosa-Santaulária, 2006).

Figura 1. Comportamiento del tamaño de la prueba para  $H_0: \beta_1=0^{**}$



\*\*En la leyenda de cada gráfico aparecen cuatro (o tres) números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

Figura 1. Comportamiento del tamaño de la prueba para  $H_0:\beta_1=0^{**}$ . Continuación



\*\*En la leyenda de cada gráfico aparecen cuatro (o tres) números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . † Número de gráfica.

Cuadro 1. Tamaño de la prueba estimado para la hipótesis  $H_0:\beta_1=0$ .

Comb*	Tamaños de muestra, T									
	25	50	75	100	250	500	750	1000	2500	5000
57	0.102	0.285	0.400	0.501	0.733	0.827	0.836	0.907	0.942	0.972
66	0.117	0.343	0.492	0.578	0.802	0.924	0.944	0.958	0.96	0.957
67	0.110	0.331	0.462	0.553	0.786	0.867	0.894	0.910	0.942	0.963
75	0.104	0.302	0.427	0.516	0.708	0.812	0.842	0.841	0.873	0.901
76	0.107	0.328	0.478	0.556	0.817	0.918	0.946	0.964	0.983	0.983
77	0.113	0.316	0.464	0.575	0.770	0.904	0.929	0.936	0.981	0.987

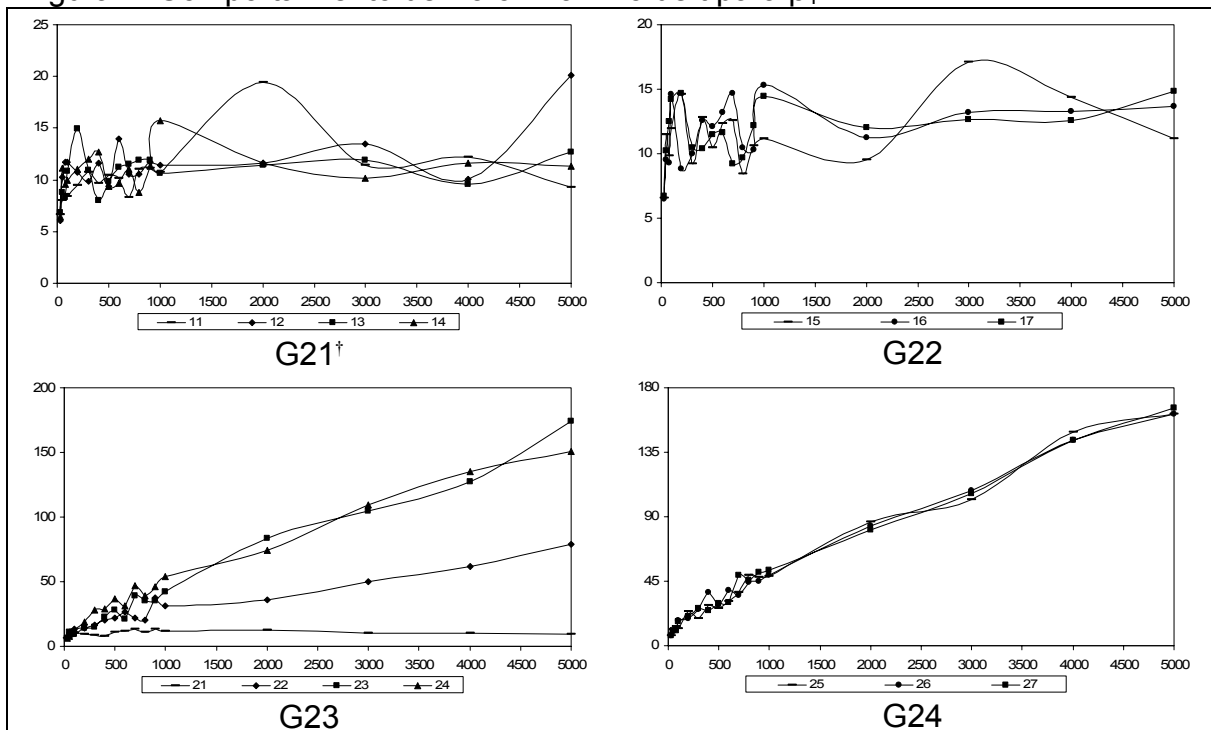
\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

#### 4.1.2 Valor máximo del estadístico t

Ahora observamos el comportamiento del valor máximo del estadístico t en la Figura 2 (en el Anexo I se reportan los valores numéricos de estas estimaciones, Cuadro A2). En las gráficas G21 y G22 donde el proceso 1 genera a  $y_t$ , el valor máximo del estadístico t se mantiene relativamente pequeño en todos los tamaños de muestra. También observamos que cuando el proceso 1 aparece como generador de  $x_t$  el valor máximo del estadístico t también toma valores pequeños (gráficas G23, G25, G27, G29, G211 y G213).

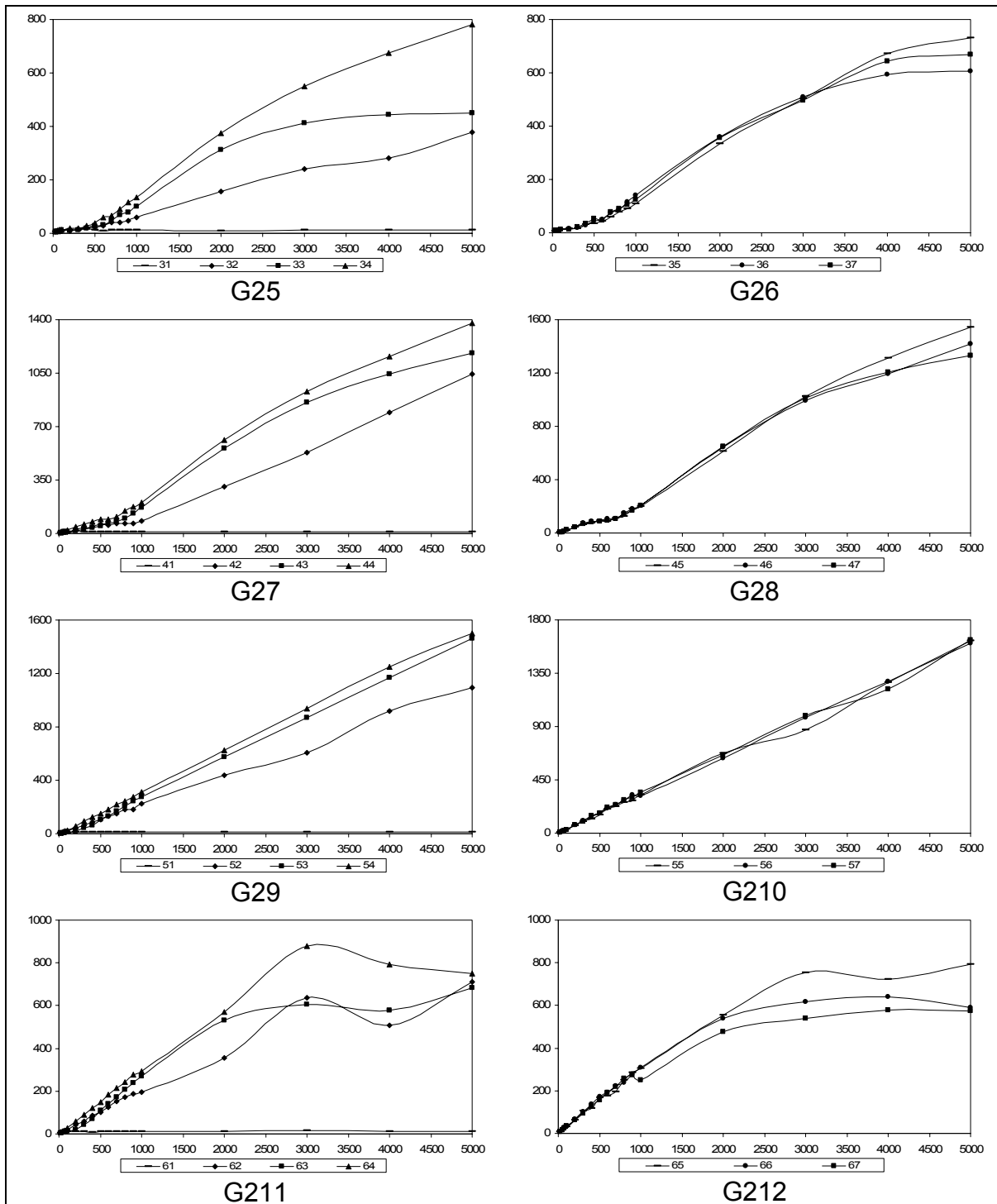
Cuando el proceso 1 no interviene en el modelo el valor máximo del estadístico t aumenta con el tamaño de muestra (ver gráficas G23 a G214).

Figura 2. Comportamiento del valor máximo de t para  $\beta_1^{**}$ .



\*\*En la leyenda de cada gráfico aparecen cuatro (o tres) números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

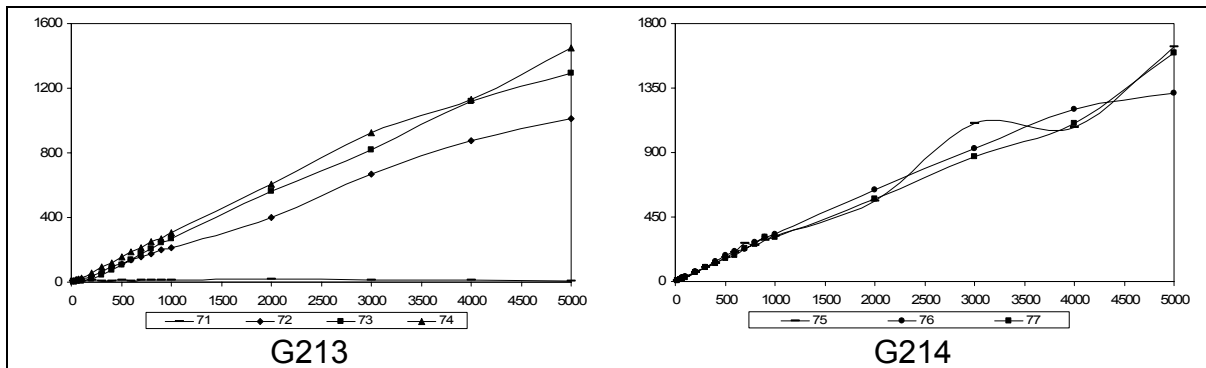
Figura 2. Comportamiento del valor máximo de  $t$  para  $\beta_1^{**}$ . Continuación



\*\*En la leyenda de cada gráfico aparecen cuatro (o tres) números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.



Figura 2. Comportamiento del valor máximo de  $t$  para  $\beta_1^{**}$ . Continuación



\*\*En la leyenda de cada gráfico aparecen cuatro (o tres) números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

En general en este modelo se observa que al tener tamaños de muestra grandes el tamaño de la prueba tiende a 1 y el valor de  $t$  tiende a infinito y por lo tanto la probabilidad de ocurrencia de correlación espuria es mayor conforme aumenta el tamaño de muestra.

Los resultados obtenidos en este modelo son parecidos a lo que se ha observado en el modelo de regresión lineal utilizando los mismos procesos generadores de los datos, es decir, con la presencia del proceso 1 que es estacionario y genera observaciones no correlacionadas, no ocurre el fenómeno de correlación espuria y en los demás casos la probabilidad de ocurrencia se incrementa con el tamaño de muestra. Con respecto al valor del estadístico  $t$ , aunque en este trabajo no se derivó la tasa de divergencia de este estadístico como se hizo en Noriega y Ventosa-Santaulária (2006), el comportamiento observado del su valor máximo nos da evidencia de que al no estar presente el proceso 1 en el modelo, tiende a crecer con el tamaño de muestra.

Es importante recalcar que este modelo no toma en cuenta la correlación de las observaciones y por tanto, no se cumple el supuesto de observaciones independientes, sin embargo, el uso de este tipo de regresión con observaciones correlacionadas conduce a obtener una correlación espuria entre  $y_t$  y  $x_t$ .

## 4.2 Ecuaciones de estimación generalizadas

Primeramente se discutirán los resultados de este modelo al suponer una matriz de correlación no estructurada en las ecuaciones de estimación y posteriormente se discutirán los resultados obtenidos al utilizar la matriz de independencia. Los tamaños de muestra, equivalentes al número de observaciones por individuo en todos los casos fueron 3, 4, 5, 6, 7 y 8.

### 4.2.1 Modelo con una variable explicatoria, matriz de correlación no estructurada,

Tamaño de la prueba y valor máximo de t

En el Cuadro 2 aparecen los valores del tamaño de la prueba para la hipótesis  $H_0: \beta_1=0$  estando presente únicamente la variable  $x_t$  como variable explicatoria. El valor máximo del estadístico t se presenta en la Cuadro 3.

Como se observa en el Cuadro 2, el comportamiento del tamaño de la prueba es muy similar en todas las combinaciones y tiende a crecer con el número de observaciones por individuo. El valor máximo del estadístico t es considerablemente grande en la mayoría de los casos aún cuando el número de observaciones por individuo es pequeño.

Cuadro 2. Tamaño de la prueba para la hipótesis  $H_0: \beta_1=0$ ,  $\eta_{it} = \beta_0 + \beta_1 x_{it}$ .

Comb*	Tamaños de muestra, $n_i$					
	3	4	5	6	7	8
11	0.13	0.22	0.32	0.38	0.46	0.51
12	0.15	0.23	0.32	0.40	0.46	0.52
13	0.14	0.23	0.32	0.39	0.47	0.53
14	0.15	0.21	0.30	0.38	0.44	0.50
15	0.19	0.27	0.36	0.46	0.51	0.54
16	0.19	0.26	0.38	0.46	0.49	0.57
17	0.16	0.26	0.36	0.45	0.47	0.57
21	0.18	0.27	0.34	0.42	0.47	0.54
22	0.14	0.23	0.34	0.39	0.48	0.53
23	0.15	0.24	0.30	0.38	0.46	0.54
24	0.14	0.23	0.32	0.39	0.45	0.53
25	0.17	0.31	0.40	0.46	0.54	0.59
26	0.17	0.27	0.39	0.48	0.51	0.54
27	0.18	0.29	0.37	0.44	0.51	0.59
31	0.14	0.22	0.30	0.39	0.47	0.51
32	0.13	0.20	0.31	0.40	0.45	0.51
34	0.14	0.22	0.30	0.40	0.43	0.53
35	0.14	0.24	0.32	0.37	0.43	0.53
36	0.16	0.28	0.39	0.46	0.50	0.54
37	0.19	0.29	0.39	0.42	0.47	0.56
41	0.16	0.28	0.37	0.42	0.51	0.58
42	0.13	0.23	0.29	0.39	0.43	0.51
43	0.14	0.22	0.33	0.36	0.43	0.51
44	0.17	0.20	0.32	0.40	0.44	0.52
45	0.14	0.22	0.31	0.37	0.43	0.52
46	0.16	0.24	0.34	0.42	0.51	0.52
47	0.17	0.26	0.37	0.41	0.48	0.55
51	0.17	0.25	0.32	0.41	0.47	0.56
52	0.15	0.23	0.30	0.38	0.43	0.52
53	0.14	0.22	0.34	0.38	0.44	0.53
54	0.15	0.22	0.30	0.39	0.42	0.49
55	0.14	0.23	0.31	0.39	0.43	0.50
56	0.18	0.27	0.38	0.43	0.48	0.60
57	0.17	0.25	0.34	0.39	0.51	0.58
61	0.16	0.26	0.38	0.42	0.48	0.54
62	0.14	0.22	0.28	0.38	0.45	0.54
63	0.16	0.21	0.33	0.37	0.45	0.54
64	0.16	0.24	0.31	0.38	0.44	0.49
65	0.15	0.23	0.33	0.37	0.43	0.50
66	0.16	0.27	0.37	0.44	0.51	0.57
67	0.16	0.27	0.37	0.47	0.51	0.56
71	0.20	0.29	0.38	0.43	0.49	0.58
72	0.14	0.24	0.31	0.36	0.45	0.52
73	0.14	0.22	0.32	0.38	0.45	0.51
74	0.14	0.22	0.31	0.39	0.43	0.51
75	0.16	0.23	0.30	0.40	0.45	0.49
76	0.16	0.26	0.36	0.44	0.50	0.55
77	0.17	0.26	0.35	0.42	0.52	0.57

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo dígito indica el proceso que generó a  $x_t$ .

Cuadro 3. Valores máximos de t para  $\beta_1$ ,  $\eta_{it} = \beta_0 + \beta_1 x_{it}$ .

Comb*	Tamaños de muestra, $n_i$					
	3	4	5	6	7	8
11	27962082.3	4.25E+27	6.04E+31	Inf	5.59E+31	3.13E+33
12	1.69E+32	6.63E+28	2.94E+30	Inf	1.77E+31	4.16E+30
13	Inf	8.08E+28	2.92E+32	4.66E+33	9.40E+31	1.10E+33
14	1.46E+30	3884829.62	Inf	2.00E+31	1.81E+32	7.55E+30
15	5.6191E+10	1.60E+31	7.03E+29	2.67E+30	1.17E+29	1.97E+32
16	4145.11172	Inf	4.82E+32	7.16E+30	Inf	1.92E+30
17	Inf	521781.763	Inf	Inf	Inf	7.12E+31
21	7.83E+30	1.53E+30	2.02E+28	4.03E+31	4.77E+29	3.83E+31
22	3.44E+30	2583.92871	4.36E+30	5.88E+33	1.02E+30	1.72E+29
23	Inf	9.99E+27	2.29E+35	2.51E+31	6.86E+30	Inf
24	1.53E+30	9.38E+31	2.14E+30	4.59E+29	2.25E+31	2.39E+31
25	Inf	1.97E+32	Inf	Inf	6.10E+31	5.13E+32
26	584.602982	4.32E+26	2.19E+31	5.85E+31	Inf	2.68E+31
27	1.04E+27	1.51E+29	5.14E+31	1.24E+20	Inf	3.38E+31
31	8.20E+28	3321.01094	3.66E+36	5.66E+30	2.33E+31	7.85E+30
32	4.45E+29	9.70E+29	5.03E+30	5.72E+36	1.61E+31	9.20E+30
34	75.6411478	2.12E+28	Inf	4.22E+31	1.96E+31	8.71E+31
35	Inf	8.28E+30	2.45E+28	3.04E+32	2.54E+31	Inf
36	1.37E+27	4.79E+28	4.68E+31	1.64E+31	Inf	5.53E+31
37	Inf	Inf	1.59E+30	5.24E+31	2.22E+32	Inf
41	4.98E+30	Inf	1.46E+32	4.78E+31	7.20E+30	6.14E+30
42	184.992199	8.69E+29	1.98E+30	2.34E+31	3.03E+31	3.27E+30
43	2.00E+31	Inf	2.80E+31	1.11E+30	1.41E+32	1.78E+33
44	3.19E+30	Inf	3.60E+32	5.78E+33	5.24E+31	3.35E+33
45	5.85E+30	3.63E+28	1.47E+31	4.84E+31	7.12E+33	2.28E+32
46	Inf	Inf	Inf	5.52E+30	5.94E+31	Inf
47	1.77E+30	Inf	2.23E+29	2.61E+33	Inf	2.10E+29
51	Inf	2.49E+31	Inf	4.44E+29	Inf	Inf
52	8.44E+30	3.29E+29	Inf	4.76E+29	Inf	3.49E+33
53	Inf	5.01E+30	3.66E+30	7.50E+31	2.86E+31	1.33E+31
54	Inf	1.71E+30	1.84E+30	Inf	4.77E+30	Inf
55	Inf	1.29E+30	3.19E+30	1.32E+31	7.34E+32	8.49E+33
56	3.25E+30	Inf	Inf	Inf	Inf	7.59E+31
57	1.99E+30	Inf	3.16E+32	3.31E+31	Inf	Inf
61	1.53E+30	16258.7423	Inf	Inf	Inf	Inf
62	1.02E+29	Inf	1.01E+32	1.72E+31	3.15E+33	3.38E+30
63	2.06E+30	1.88E+31	Inf	Inf	6.40E+33	Inf
64	Inf	2.59E+30	4.61E+30	5.60E+33	Inf	Inf
65	1.38E+33	9.47E+27	1.19E+29	Inf	Inf	Inf
66	Inf	1.92E+27	6.45E+30	Inf	Inf	4.45E+32
67	9.72E+30	Inf	3.28E+30	Inf	Inf	5.15E+33
71	1.37E+29	Inf	1.19E+29	1.64E+31	Inf	Inf
72	1.73E+30	1.81E+32	1.75E+34	1.09E+30	2.87E+33	Inf
73	Inf	9513.68744	Inf	6.43E+31	4.23E+34	Inf
74	4.71E+27	Inf	1.04E+31	3.06E+31	9.01E+31	7.77E+32
75	3.26E+30	2.68E+29	4.58E+31	8.76E+30	1.17E+31	4.60E+32
76	590.628759	2.80E+30	Inf	Inf	Inf	Inf
77	Inf	8.42E+28	2.52E+30	Inf	Inf	Inf

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

En algunas combinaciones de procesos generadores de los datos intervienen procesos que generan observaciones no correlacionadas como es el caso de los procesos 1 y 2, en esta situación se puede argumentar que los parámetros estimados, tanto del tamaño de la prueba como del valor máximo del estadístico t, no son eficientes, ya que la matriz de correlación más apropiada es la matriz de independencia. No obstante, en las demás combinaciones las observaciones están correlacionadas y los estimadores son eficientes ya que en el proceso de estimación se consideró una matriz de correlación no estructurada en la que las correlaciones son autorregresivas. Por lo tanto, no estando el proceso 1 o 2 en el modelo, ocurre correlación espuria entre las variables  $y_t$  y  $x_t$  y la probabilidad de ocurrencia aumenta con el número de observaciones por individuo.

4.2.2 Modelo con un tratamiento y una variable explicatoria, matriz de correlación no estructurada.

Ahora consideramos el caso donde interviene en el modelo un tratamiento y una variable explicatoria. El tamaño de la prueba se obtuvo por separado, es decir un tamaño de prueba para cada hipótesis:  $H_{0\beta}: \beta_1=0$  y  $H_{0\tau}: \tau=0$ . El valor máximo del estadístico t se obtuvo para  $\tau$  y para  $\beta_1$ .

En las Cuadros 4 y 5 observamos los valores del tamaño de la prueba para las hipótesis  $H_{0\beta}: \beta_1=0$  y  $H_{0\tau}: \tau=0$ . Estos resultados son parecidos al caso en el que únicamente la variable  $x_t$  se utilizó como explicatoria; el tamaño de la prueba tiende a crecer con el número de observaciones por individuo alcanzando valores cercanos a 0.6 cuando se tienen 6 observaciones por individuo.

Cuadro 4. Tamaño de la prueba para la hipótesis  $H_0:\beta_1=0$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Tamaños de muestra, $n_i$					
	3	4	5	6	7	8
11	0.14	0.22	0.27	0.31	0.37	0.44
12	0.16	0.23	0.27	0.30	0.38	0.51
13	0.17	0.23	0.27	0.30	0.38	0.46
14	0.16	0.20	0.25	0.28	0.38	0.46
15	0.20	0.29	0.39	0.42	0.47	0.57
16	0.21	0.29	0.35	0.45	0.48	0.57
17	0.19	0.27	0.39	0.44	0.50	0.60
21	0.17	0.24	0.29	0.32	0.41	0.48
22	0.14	0.22	0.31	0.30	0.41	0.50
23	0.16	0.23	0.27	0.33	0.40	0.46
24	0.16	0.23	0.28	0.35	0.40	0.48
25	0.21	0.32	0.40	0.47	0.53	0.62
26	0.22	0.28	0.40	0.44	0.52	0.60
27	0.21	0.31	0.37	0.45	0.52	0.59
31	0.17	0.21	0.27	0.32	0.37	0.49
32	0.16	0.19	0.28	0.32	0.38	0.46
34	0.15	0.23	0.28	0.30	0.36	0.48
35	0.19	0.32	0.39	0.44	0.51	0.58
36	0.22	0.31	0.38	0.43	0.48	0.58
37	0.20	0.30	0.36	0.43	0.50	0.58
41	0.15	0.22	0.27	0.28	0.36	0.45
42	0.14	0.21	0.28	0.28	0.39	0.46
43	0.17	0.20	0.27	0.31	0.36	0.45
44	0.18	0.23	0.27	0.30	0.35	0.43
45	0.18	0.26	0.32	0.39	0.49	0.54
46	0.20	0.28	0.36	0.43	0.46	0.54
47	0.21	0.29	0.33	0.40	0.47	0.53
51	0.15	0.24	0.26	0.28	0.35	0.45
52	0.17	0.22	0.26	0.27	0.37	0.45
53	0.17	0.24	0.25	0.28	0.32	0.38
54	0.15	0.21	0.23	0.28	0.36	0.43
55	0.22	0.27	0.33	0.37	0.46	0.54
56	0.20	0.27	0.34	0.37	0.47	0.51
57	0.19	0.27	0.36	0.38	0.42	0.54
61	0.16	0.21	0.25	0.28	0.35	0.43
62	0.18	0.21	0.26	0.29	0.34	0.45
63	0.18	0.23	0.28	0.31	0.36	0.40
64	0.16	0.22	0.25	0.29	0.35	0.44
65	0.20	0.29	0.37	0.40	0.47	0.53
66	0.20	0.26	0.36	0.42	0.47	0.54
67	0.23	0.28	0.36	0.40	0.49	0.53
71	0.16	0.23	0.25	0.27	0.31	0.44
72	0.14	0.21	0.26	0.28	0.36	0.44
73	0.17	0.22	0.26	0.28	0.35	0.42
74	0.14	0.22	0.26	0.31	0.36	0.41
75	0.20	0.29	0.38	0.38	0.48	0.53
76	0.19	0.29	0.35	0.40	0.46	0.54
77	0.19	0.27	0.36	0.36	0.50	0.56

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_i$ , el segundo el proceso que generó a  $x_i$ .

Cuadro 5. Tamaño de la prueba para la hipótesis  $H_{0\tau} : \tau = 0$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Tamaños de muestra, $n_i$					
	3	4	5	6	7	8
11	0.11	0.20	0.26	0.32	0.41	0.49
12	0.11	0.21	0.27	0.32	0.42	0.47
13	0.12	0.22	0.29	0.32	0.40	0.46
14	0.10	0.20	0.27	0.37	0.42	0.47
15	0.16	0.27	0.34	0.42	0.46	0.58
16	0.17	0.24	0.33	0.42	0.47	0.54
17	0.16	0.25	0.36	0.41	0.50	0.57
21	0.13	0.22	0.30	0.38	0.42	0.50
22	0.11	0.22	0.27	0.37	0.44	0.53
23	0.15	0.20	0.29	0.34	0.41	0.50
24	0.15	0.21	0.29	0.38	0.43	0.51
25	0.16	0.28	0.37	0.47	0.48	0.59
26	0.18	0.27	0.38	0.44	0.50	0.57
27	0.19	0.27	0.36	0.46	0.48	0.59
31	0.11	0.21	0.28	0.36	0.42	0.48
32	0.12	0.23	0.25	0.34	0.42	0.48
34	0.12	0.20	0.28	0.35	0.40	0.47
35	0.17	0.27	0.35	0.44	0.51	0.57
36	0.14	0.27	0.33	0.41	0.49	0.55
37	0.14	0.26	0.34	0.43	0.49	0.57
41	0.11	0.19	0.24	0.31	0.37	0.46
42	0.13	0.18	0.22	0.32	0.35	0.46
43	0.10	0.18	0.28	0.32	0.37	0.47
44	0.10	0.18	0.26	0.31	0.38	0.45
45	0.13	0.24	0.34	0.36	0.48	0.53
46	0.17	0.24	0.32	0.40	0.46	0.53
47	0.16	0.23	0.31	0.39	0.46	0.49
51	0.11	0.18	0.24	0.29	0.38	0.46
52	0.14	0.17	0.25	0.32	0.37	0.44
53	0.13	0.18	0.27	0.30	0.38	0.45
54	0.11	0.18	0.26	0.32	0.35	0.43
55	0.15	0.23	0.30	0.36	0.46	0.52
56	0.12	0.25	0.31	0.35	0.43	0.52
57	0.13	0.24	0.31	0.36	0.43	0.53
61	0.11	0.17	0.30	0.34	0.38	0.49
62	0.12	0.20	0.28	0.33	0.39	0.47
63	0.11	0.19	0.25	0.33	0.41	0.47
64	0.12	0.22	0.25	0.32	0.43	0.47
65	0.16	0.25	0.34	0.40	0.48	0.49
66	0.16	0.24	0.34	0.41	0.45	0.56
67	0.16	0.25	0.34	0.42	0.46	0.52
71	0.14	0.20	0.23	0.30	0.40	0.46
72	0.13	0.21	0.27	0.32	0.38	0.50
73	0.13	0.18	0.25	0.31	0.39	0.46
74	0.12	0.20	0.27	0.32	0.39	0.47
75	0.14	0.25	0.35	0.39	0.45	0.51
76	0.16	0.25	0.33	0.40	0.42	0.52
77	0.16	0.23	0.33	0.39	0.46	0.51

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

En esta situación se esperaban valores del tamaño de la prueba cercanos a 0.05 (el nivel de significancia de la regresión) al menos en los casos donde intervienen procesos con observaciones correlacionadas, ya que las ecuaciones de estimación toman en cuenta dichas correlaciones, sin embargo, en los cuadros anteriores se observa que la probabilidad de ocurrencia de este fenómeno es alta aún para un número pequeño de observaciones. Los valores máximos del estadístico t (Cuadros 6 y 7) refuerzan esta afirmación porque tienden a crecer conforme aumenta el número de observaciones por individuo alcanzando valores mas altos con 7 u 8 observaciones.

Cuadro 6. Valores máximos de t para  $\beta_1$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$

Comb*	Tamaños de muestra, $n_i$					
	3	4	5	6	7	8
11	118.48	442.36	753.05	965.06	2194.21	1554.02
12	73.24	451.33	258.66	698.13	6223.27	43461.96
13	90.92	439.35	367.40	820.07	15904.99	591.56
14	39.96	164.97	1979.12	1485.63	568.92	8842.42
15	193.58	2272.25	2698.29	4721.95	53703.88	16444.44
16	225.35	1589.56	1337.24	2474.49	22774.49	125411.84
17	114.36	1030.55	4926.75	4687.32	313273.44	56749.37
21	168.33	1178.59	396.68	442.20	3576.01	3422.31
22	55.74	202.86	2452.82	1175.05	87356.82	32130.16
23	1095.94	560.59	591.10	546.72	53462.78	10185.29
24	863.50	1050.57	339.86	1802.17	2431.83	13768.55
25	339.91	1735.49	1456.90	2112.81	325142.97	19015.57
26	219.11	695.55	2622.97	12168.57	99370.37	107026.92
27	270.41	3778.18	4243.18	4672.07	58809.00	61059.61
31	48.21	117.29	201.31	460.85	42961.13	9405.61
32	199.25	593.57	688.36	687.12	348.49	3481.46
34	50.19	253.05	270.02	224.11	31099.44	8381.78
35	302.64	663.51	1352.68	2596.17	14891.84	15884.51
36	203.82	394.19	2601.17	2798.22	5974.46	12668.75
37	72.41	332.31	5020.08	20359.28	83011.49	25251.46
41	32.48	176.14	3201.19	252.10	2352.04	36999.72
42	290.20	699.89	451.82	302.87	7061.62	3979.34
43	63.87	715.61	57418.55	790.17	1068.32	18758.92
44	33.25	251.26	267.45	247.56	1103.43	67501.98
45	144.21	1509.38	1214.93	4597.13	79354.48	33061.95
46	836.06	738.50	1081.43	14436.57	27965.69	15195.73
47	311.24	100.91	1052.02	1430.12	257537.26	169674.73

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .



Cuadro 6. Valores máximos de t para  $\beta_1$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ , continuación

51	233.52	835.07	637.47	672.40	401.73	25474.96
52	86.54	628.77	239.46	702.80	2634.56	686.82
53	147.34	134.96	456.47	198.64	2298.27	1970.09
54	31.82	441.59	185.06	10459.73	410.11	1624.09
55	335.46	9731.86	306.13	21759.14	10534.70	7395.08
56	161.13	845.14	1242.88	1351.73	8495.55	1898.80
57	4067.61	248.44	1400.79	1603.42	15443.56	354135.21
61	105.87	753.91	651.16	981.81	2300.81	9763.48
62	65.78	355.92	265.94	1455.10	795.28	3524.83
63	133.67	122.81	228.09	3501.91	967.57	11867.91
64	40.14	234.53	970.51	596.95	862.63	31404.89
65	99.65	381.98	460.40	3272.57	1654.57	4571.04
66	162.32	259.97	2324.10	1402.97	10869.98	7639.84
67	172.29	378.28	1879.64	6147.49	16643.58	23968.34
71	42.62	326.69	68.10	71976.43	3525.73	2943.65
72	37.63	203.73	369.66	2057.03	2593.23	32562.66
73	39.64	353.27	254.73	154.24	5292.64	78142.99
74	34.54	181.96	529.41	991.27	2626.95	1783.29
75	103.71	204.18	6133.62	2681.88	204775.98	5127.01
76	157.66	130.86	605.34	5112.64	2496.46	3753.41
77	287.17	766.68	3181.38	4045.73	11525.32	2026.29

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro 7. Valor máximo de t para  $\tau$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Tamaños de muestra, $n_i$					
	3	4	5	6	7	8
11	4.90E+11	2.03E+10	8.99E+10	5.33E+08	4.97E+09	1.06E+09
12	2.05E+02	6.90E+10	2.66E+09	2.81E+09	1.36E+10	1.25E+08
13	3.54E+02	7.03E+02	5.95E+09	2.42E+09	1.09E+09	9.75E+08
14	1.34E+13	2.29E+09	4.13E+08	6.10E+08	2.88E+09	1.29E+10
15	1.86E+02	6.03E+08	7.82E+10	7.12E+08	4.13E+08	6.68E+08
16	3.14E+08	7.27E+08	9.52E+07	2.53E+09	9.04E+08	3.17E+09
17	2.44E+02	4.59E+09	7.10E+10	1.98E+09	4.43E+08	1.25E+09
21	4.89E+11	1.31E+11	1.22E+10	4.06E+08	1.36E+09	1.16E+09
22	7.33E+09	1.05E+11	5.03E+09	5.75E+10	3.22E+10	1.75E+10
23	5.66E+09	5.50E+08	4.77E+09	2.42E+09	1.30E+09	3.03E+09
24	1.08E+02	1.69E+02	2.12E+05	3.29E+11	4.30E+09	5.65E+09
25	2.47E+02	2.21E+03	2.57E+03	1.09E+09	9.37E+09	1.17E+09
26	2.63E+09	4.21E+10	2.22E+10	1.13E+09	8.49E+09	5.09E+09
27	2.54E+09	5.95E+02	3.83E+12	1.04E+10	4.21E+09	1.08E+09
31	4.53E+01	1.53E+02	7.86E+08	2.03E+08	7.71E+08	1.72E+09
32	1.02E+02	7.56E+09	5.06E+10	3.68E+10	1.13E+09	1.56E+09
34	7.52E+01	1.33E+11	3.91E+09	1.36E+11	2.24E+11	5.17E+08
35	1.44E+02	1.60E+10	1.07E+09	7.15E+08	1.17E+09	7.62E+09
36	1.07E+11	1.94E+08	4.04E+08	2.30E+08	4.58E+08	1.20E+09
37	1.30E+02	7.85E+09	5.23E+10	1.46E+09	1.02E+10	4.07E+10
41	1.49E+10	4.82E+09	4.15E+08	3.44E+09	1.66E+10	1.52E+11
42	8.25E+09	2.14E+02	3.79E+08	3.30E+09	1.52E+09	2.91E+09
43	4.61E+01	2.74E+02	1.69E+03	1.06E+10	1.47E+10	2.44E+09
44	1.02E+02	2.17E+11	1.66E+12	5.99E+08	3.51E+08	8.38E+08

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro 7 Valor máximo de t para  $\tau$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ , continuación

45	1.71E+10	1.83E+10	2.64E+08	1.89E+08	4.36E+08	4.37E+09
46	1.69E+12	2.75E+08	2.48E+10	6.98E+08	1.22E+09	4.25E+09
47	1.59E+10	3.95E+08	3.31E+08	4.56E+08	8.32E+09	1.07E+09
51	6.84E+01	2.21E+09	5.51E+08	3.05E+02	5.12E+10	7.68E+08
52	3.70E+01	1.12E+12	4.10E+09	2.44E+10	1.23E+10	7.24E+09
53	5.61E+09	3.22E+09	2.78E+09	5.38E+10	4.95E+10	1.64E+09
54	1.30E+02	9.65E+09	1.07E+10	2.87E+11	8.43E+09	5.62E+09
55	6.18E+02	3.17E+07	9.96E+09	5.67E+08	2.35E+09	3.61E+07
56	9.40E+01	5.33E+09	9.20E+08	3.03E+08	3.48E+09	5.61E+09
57	1.51E+02	2.97E+07	3.38E+10	2.01E+09	2.22E+09	4.59E+08
61	6.45E+10	2.26E+02	3.84E+09	5.34E+08	2.79E+08	1.22E+09
62	1.25E+10	6.48E+10	3.94E+08	9.96E+08	1.28E+10	1.59E+09
63	2.99E+10	2.29E+02	7.86E+09	6.39E+09	3.50E+09	2.24E+08
64	5.51E+02	2.56E+09	6.09E+07	7.06E+08	1.10E+11	1.23E+09
65	5.31E+01	1.76E+09	2.58E+07	7.40E+09	4.69E+08	1.66E+10
66	4.09E+13	1.87E+08	7.35E+09	1.21E+09	1.17E+10	2.60E+10
67	8.28E+01	7.12E+08	2.83E+10	4.20E+08	9.71E+09	2.99E+08
71	3.13E+02	2.48E+10	1.86E+09	1.06E+10	4.38E+09	9.88E+10
72	2.54E+01	1.56E+09	3.87E+09	2.86E+08	5.56E+09	2.24E+09
73	6.43E+01	1.79E+08	2.70E+10	1.54E+10	9.46E+09	2.32E+09
74	1.62E+02	3.35E+08	1.65E+09	9.52E+08	1.53E+08	2.82E+09
75	1.73E+02	1.59E+03	7.17E+09	1.21E+08	5.48E+12	3.33E+09
76	9.09E+10	4.09E+10	1.41E+09	2.19E+03	1.68E+09	4.01E+09
77	7.19E+08	1.46E+03	7.94E+08	1.65E+10	2.38E+10	9.21E+08

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

El comportamiento observado del tamaño de la prueba y del valor máximo del estadístico t indican que la variable  $x_t$  y el efecto del tratamiento  $\tau$  tienen una influencia significativa en la probabilidad de éxito de la variable de respuesta cuando el número de observaciones por individuo aumenta. Este resultado tiene validez cuando los procesos 3 a 7 intervienen en el modelo, ya que en estos procesos las observaciones están correlacionadas y de acuerdo con Liang y Zeger (1986) con la estructura de correlación hipotetizada que hemos usado los parámetros estimados son eficientes.

#### 4.2.3 Modelo con una variable explicatoria, matriz de correlación independiente.

A continuación presentamos los resultados obtenidos del modelo de ecuaciones de estimación generalizadas al suponer una matriz de independencia en las ecuaciones de estimación. En este caso los parámetros más eficientes se obtienen cuando las observaciones no están correlacionadas como las generadas por los procesos 1 y 2 y por ello fijaremos nuestra atención en las combinaciones que involucran a estos procesos. Con respecto a las combinaciones donde no intervienen los procesos 1 y 2, no podemos obtener alguna conclusión con respecto a la ocurrencia o no de este fenómeno, ya que aunque el tamaño de la prueba y el valor máximo del estadístico  $t$  son relativamente pequeños, los parámetros obtenidos no son eficientes.

En la Cuadro 8 se presentan los valores estimados del tamaño de la prueba y observamos que en las combinaciones 11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 41, 42, 51, 52, 61, 62, 71 y 72, el tamaño de la prueba aunque no aumenta con el número de observaciones por individuo, oscila en 0.1 (véanse líneas correspondientes en la Cuadro 8), un valor ligeramente mayor de lo esperado; en las combinaciones 15, 16, 17, 25, 26 y 27 el tamaño de la prueba tiende a aumentar conforme aumenta el número de observaciones por individuo, lo que indica que la presencia de procesos derivados de una caminata aleatoria (5, 6 y 7 de la expresión 3.1) como variable explicatoria inducen el incremento del tamaño de la prueba al aumentar el número de observaciones por individuo y por tanto inducen la ocurrencia de correlación espuria.

Cuadro 8. Tamaño de la prueba para la hipótesis  $H_0: \beta_1=0$ ,  $\eta_{it} = \beta_0 + \beta_1 x_{it}$ .

Comb*	Observaciones					
	3	4	5	6	7	8
11	0.111	0.091	0.099	0.106	0.101	0.119
12	0.095	0.103	0.096	0.099	0.117	0.089
13	0.089	0.104	0.080	0.117	0.094	0.102
14	0.103	0.102	0.093	0.108	0.101	0.098
15	0.110	0.107	0.145	0.117	0.154	0.161
16	0.118	0.118	0.138	0.118	0.128	0.159
17	0.114	0.120	0.130	0.122	0.129	0.144
21	0.081	0.095	0.100	0.117	0.115	0.109
22	0.085	0.103	0.093	0.097	0.112	0.107
23	0.097	0.092	0.101	0.112	0.108	0.108
24	0.090	0.088	0.126	0.098	0.107	0.103
25	0.118	0.122	0.131	0.138	0.126	0.150
26	0.108	0.114	0.143	0.130	0.130	0.148
27	0.097	0.114	0.148	0.153	0.152	0.160
31	0.086	0.096	0.103	0.102	0.096	0.101
32	0.102	0.081	0.103	0.113	0.094	0.106
33	0.094	0.092	0.100	0.108	0.096	0.100
34	0.076	0.096	0.111	0.103	0.082	0.111
35	0.112	0.121	0.107	0.137	0.142	0.130
36	0.104	0.139	0.138	0.127	0.168	0.150
37	0.106	0.143	0.120	0.141	0.162	0.143
41	0.114	0.091	0.094	0.091	0.101	0.106
42	0.091	0.102	0.095	0.124	0.104	0.122
43	0.088	0.108	0.084	0.111	0.091	0.114
44	0.086	0.100	0.097	0.107	0.088	0.113
45	0.099	0.134	0.113	0.141	0.146	0.147
46	0.101	0.122	0.137	0.148	0.146	0.138
47	0.105	0.130	0.124	0.131	0.152	0.147
51	0.085	0.107	0.111	0.083	0.121	0.108
52	0.093	0.104	0.100	0.100	0.089	0.125
53	0.104	0.099	0.111	0.097	0.104	0.133
54	0.092	0.102	0.106	0.088	0.122	0.104
55	0.123	0.147	0.145	0.145	0.148	0.159
56	0.101	0.142	0.138	0.147	0.156	0.179
57	0.118	0.133	0.159	0.128	0.149	0.166
61	0.089	0.103	0.094	0.104	0.095	0.100
62	0.094	0.095	0.115	0.129	0.084	0.120
63	0.094	0.096	0.112	0.108	0.115	0.118
64	0.096	0.106	0.110	0.115	0.118	0.107
65	0.120	0.130	0.151	0.136	0.161	0.136
66	0.120	0.133	0.127	0.125	0.151	0.176
67	0.120	0.122	0.150	0.167	0.145	0.141
71	0.080	0.117	0.110	0.091	0.102	0.122
72	0.080	0.071	0.106	0.113	0.104	0.103
73	0.105	0.094	0.122	0.137	0.135	0.118
74	0.099	0.116	0.093	0.113	0.119	0.109
75	0.107	0.118	0.141	0.138	0.157	0.140
76	0.112	0.104	0.119	0.113	0.144	0.154
77	0.098	0.130	0.125	0.145	0.132	0.146

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo dígito indica el proceso que generó a  $x_t$ .

Cuadro 9. Valores máximos de t para  $\beta_1$ ,  $\eta_{it} = \beta_0 + \beta_1 x_{it}$ .

Comb*	Observaciones					
	3	4	5	6	7	8
11	22.02	43.60	33.86	119.76	50.61	45.89
12	22.10	37.11	62.28	27.47	69.91	38.41
13	41.34	20.31	27.00	33.35	96.70	49.23
14	27.40	93.26	48.77	40.44	129.53	55.97
15	53.41	31.11	36.09	126.69	44.91	96.07
16	97.76	87.24	99.68	67.91	110.89	105.64
17	52.38	41.19	235.35	66.19	48.99	120.32
21	53.96	31.84	39.53	93.81	53.63	89.94
22	73.23	28.70	370.50	106.28	50.17	48.71
23	55.31	38.69	96.89	64.39	53.20	39.17
24	35.17	110.12	77.92	71.19	68.46	78.22
25	51.23	33.20	40.70	76.38	81.47	71.16
26	40.95	48.61	53.81	65.53	35.90	151.52
27	63.12	90.42	450.69	59.38	93.45	59.40
31	49.64	26.08	77.14	96.23	68.42	25.48
32	34.04	35.84	33.25	106.65	56.37	87.95
33	45.84	149.25	46.57	33.83	83.69	90.18
34	25.27	27.82	36.05	83.05	39.63	77.95
35	23.91	38.74	57.68	253.32	268.09	116.31
36	31.58	183.25	77.64	1654.04	74.49	180.13
37	28.09	26.04	74.00	100.66	49.46	87.74
41	47.62	29.68	309.13	33.53	65.73	57.78
42	24.85	43.68	63.47	50.02	26.95	35.56
43	37.19	68.74	30.29	93.44	56.61	69.23
44	26.58	40.21	37.77	35.74	56.68	89.83
45	24.13	28.29	39.72	46.35	54.75	68.96
46	22.60	27.44	83.94	67.42	167.34	52.17
47	70.00	30.50	41.96	305.61	120.97	295.78
51	48.46	21.80	71.06	22.35	100.65	67.25
52	23.63	68.94	149.75	63.92	103.37	87.94
53	28.26	75.97	180.18	25.06	30.26	53.15
54	27.02	40.78	22.19	40.72	80.27	84.47
55	29.58	35.77	104.29	71.34	178.58	119.75
56	39.91	36.74	61.72	61.51	42.22	81.59
57	28.29	128.89	114.24	38.31	32.34	95.05
61	39.82	44.37	47.31	36.00	82.78	84.34
62	28.16	78.81	48.20	56.02	25.87	81.51
63	30.75	31.81	48.53	259.55	113.67	28.81
64	56.19	31.58	121.76	40.32	398.13	65.75
65	30.56	48.68	79.14	80.89	31.10	36.25
66	148.86	45.49	59.16	45.80	98.12	174.62
67	66.68	58.24	71.61	56.47	51.93	147.59
71	85.17	68.47	855.95	98.26	70.80	34.58
72	42.76	29.45	49.08	46.64	43.02	107.19
73	34.32	73.15	77.47	150.39	36.41	37.35
74	28.41	52.91	100.59	83.47	77.23	40.12
75	31.08	52.62	64.29	69.09	43.10	192.66
76	36.70	34.57	50.76	67.79	42.24	62.33
77	152.72	39.74	51.38	57.00	88.14	51.77

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

El valor máximo del estadístico t (Cuadro 9) en todas las combinaciones tiene un comportamiento aleatorio, y a diferencia del modelo de regresión logístico, en ningún caso se observa una tendencia en el comportamiento de este valor en función del tamaño de muestra.

#### 4.2.4 Modelo con un tratamiento y una variable explicatoria, matriz de correlación independiente.

Se obtuvo una estimación del tamaño de la prueba para cada hipótesis:  $H_{0\beta}: \beta_1=0$  y  $H_{0\tau}: \tau=0$  y el valor máximo del estadístico t para  $\tau$  y  $\beta_1$ , nuevamente nos centraremos en las combinaciones que involucran los procesos 1 y 2.

El tamaño de la prueba para  $\beta_1$  (Cuadro 10) se comporta de manera muy similar al caso anterior en la que no existía un tratamiento, es decir, en presencia de procesos derivados de una caminata aleatoria la probabilidad de que ocurra correlación espuria aumenta a medida que aumenta el número de observaciones por individuo. El valor máximo del estadístico t en las mismas combinaciones (Cuadro 11) nuevamente tiene un comportamiento aleatorio y por consiguiente en general no se observa una tendencia en función del número de observaciones por individuo.

Cuadro 10. Tamaño de la prueba para la hipótesis  $H_0: \beta_1=0$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Observaciones					
	3	4	5	6	7	8
11	0.104	0.096	0.096	0.115	0.096	0.110
12	0.107	0.108	0.106	0.100	0.109	0.103
13	0.109	0.107	0.094	0.113	0.107	0.119
14	0.110	0.106	0.101	0.095	0.106	0.102
15	0.109	0.131	0.145	0.115	0.162	0.169
16	0.104	0.131	0.155	0.130	0.144	0.158
17	0.126	0.125	0.122	0.122	0.136	0.159
21	0.081	0.103	0.098	0.111	0.109	0.106
22	0.095	0.111	0.098	0.095	0.117	0.116
23	0.111	0.104	0.102	0.112	0.100	0.114
24	0.084	0.092	0.120	0.101	0.099	0.101
25	0.124	0.119	0.147	0.141	0.147	0.140
26	0.107	0.114	0.153	0.150	0.145	0.150
27	0.106	0.119	0.138	0.160	0.159	0.156
31	0.082	0.097	0.095	0.115	0.100	0.107
32	0.110	0.080	0.094	0.120	0.091	0.107
33	0.088	0.094	0.099	0.116	0.092	0.098
34	0.087	0.101	0.109	0.109	0.081	0.102
35	0.123	0.128	0.126	0.133	0.143	0.137
36	0.130	0.122	0.146	0.138	0.167	0.173
37	0.123	0.144	0.138	0.131	0.160	0.161
41	0.127	0.097	0.096	0.086	0.107	0.110
42	0.101	0.123	0.097	0.120	0.106	0.116
43	0.092	0.097	0.087	0.108	0.091	0.113
44	0.090	0.111	0.095	0.113	0.086	0.107
45	0.122	0.144	0.113	0.137	0.146	0.145
46	0.121	0.133	0.145	0.155	0.148	0.150
47	0.115	0.148	0.124	0.135	0.156	0.157
51	0.089	0.113	0.114	0.094	0.126	0.104
52	0.100	0.119	0.106	0.090	0.112	0.135
53	0.103	0.108	0.103	0.108	0.098	0.135
54	0.094	0.100	0.114	0.100	0.122	0.097
55	0.138	0.137	0.144	0.166	0.161	0.167
56	0.126	0.145	0.132	0.153	0.175	0.197
57	0.124	0.127	0.163	0.134	0.179	0.189
61	0.101	0.109	0.089	0.119	0.092	0.114
62	0.109	0.106	0.110	0.126	0.091	0.125
63	0.105	0.092	0.113	0.104	0.113	0.121
64	0.099	0.118	0.120	0.121	0.118	0.114
65	0.118	0.132	0.147	0.147	0.176	0.147
66	0.138	0.145	0.150	0.149	0.168	0.180
67	0.120	0.149	0.171	0.168	0.148	0.174
71	0.094	0.122	0.126	0.106	0.105	0.133
72	0.107	0.096	0.109	0.108	0.113	0.106
73	0.122	0.090	0.130	0.123	0.139	0.125
74	0.117	0.124	0.102	0.113	0.115	0.117
75	0.113	0.136	0.155	0.156	0.164	0.151
76	0.114	0.129	0.139	0.134	0.148	0.161
77	0.113	0.147	0.136	0.157	0.163	0.178

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro 11. Valores máximos de t para  $\beta_1$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Observaciones					
	3	4	5	6	7	8
11	31.92	100.67	35.51	53.30	26.47	66.78
12	31.84	33.02	44.93	48.52	52.71	53.09
13	39.50	30.61	25.36	32.87	38.05	57.04
14	25.89	61.57	205.69	29.06	69.30	60.97
15	141.60	35.06	47.49	98.60	68.21	77.87
16	47.08	59.58	46.39	102.54	117.71	80.22
17	34.42	24.07	46.68	88.43	231.84	64.51
21	35.68	93.51	36.06	51.45	81.35	69.46
22	25.02	33.60	193.51	77.55	89.38	64.69
23	41.35	30.60	72.71	58.05	42.06	27.35
24	33.22	48.87	35.05	67.94	72.93	111.38
25	25.64	111.43	199.02	112.95	72.74	67.07
26	26.50	67.79	31.94	77.75	34.30	102.49
27	48.65	55.42	70.36	33.12	91.12	48.74
31	33.44	22.65	45.85	40.12	48.05	87.08
32	22.89	23.62	23.33	73.13	43.46	32.61
33	32.31	39.50	31.69	34.92	56.92	78.20
34	19.43	64.32	37.06	58.76	45.01	31.58
35	35.61	28.83	44.21	52.58	140.98	62.17
36	43.61	30.61	40.40	82.10	80.47	1490.55
37	31.81	120.44	96.54	33.62	57.13	100.74
41	41.18	57.22	55.86	41.96	32.13	48.28
42	102.19	42.25	60.45	76.28	88.48	45.78
43	45.30	52.81	32.45	93.55	80.65	102.11
44	53.25	29.97	49.00	47.61	67.43	202.48
45	56.09	173.16	141.88	80.03	95.68	87.33
46	45.40	44.11	38.39	118.43	205.45	35.02
47	19.61	52.27	54.76	161.19	142.76	117.94
51	52.29	99.81	83.50	34.25	37.06	62.08
52	22.49	51.05	78.34	35.34	50.93	96.76
53	22.60	101.35	158.82	38.04	35.15	47.32
54	72.93	42.29	43.46	32.45	71.65	88.80
55	78.93	61.05	48.46	54.99	65.40	128.33
56	45.82	60.68	57.90	45.35	111.32	88.76
57	28.96	52.17	150.30	120.61	71.63	76.45
61	25.32	48.48	40.58	34.94	19.35	59.31
62	52.80	180.63	44.42	131.84	32.42	82.21
63	32.58	34.54	31.29	37.28	69.44	44.17
64	54.09	41.77	55.63	44.15	459.64	93.06
65	52.27	48.34	50.83	80.72	157.58	49.72
66	32.99	34.59	74.88	57.93	51.75	150.66
67	39.51	71.37	54.59	83.09	186.24	63.81
71	75.83	57.22	66.59	57.40	246.83	32.37
72	69.26	34.61	54.68	47.49	54.19	52.04
73	22.83	35.31	29.54	204.85	61.21	370.98
74	19.55	44.29	66.37	155.17	37.23	62.46
75	23.69	32.60	39.46	48.84	53.50	174.04
76	45.53	40.98	74.33	138.26	53.03	46.30
77	55.43	34.57	77.44	33.80	109.11	62.30

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .



En la Cuadro 12 se presentan los valores estimados del tamaño de la prueba para la hipótesis nula  $H_{0\tau} : \tau = 0$ . En este cuadro observamos que en todas las combinaciones, aún las que involucran observaciones no correlacionadas, el tamaño de la prueba aumenta conforme aumenta el número de observaciones por individuo, esto es, se ha obtenido que el efecto de un tratamiento es significativo para la probabilidad de éxito en la variable de respuesta cuando ambas variables se generaron en forma independiente. El valor máximo del estadístico t (Cuadro 13) se comporta de forma aleatoria y por ello en ninguna de las combinaciones se observa alguna tendencia de este valor pero el comportamiento del tamaño de la prueba para el efecto del tratamiento es una evidencia de que ocurre correlación espuria para esta covariable.

Cuadro 12. Tamaño de la prueba para la hipótesis  $H_{0\tau} : \tau = 0$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Observaciones					
	3	4	5	6	7	8
11	0.095	0.124	0.124	0.132	0.143	0.159
12	0.096	0.116	0.144	0.144	0.137	0.161
13	0.104	0.135	0.124	0.144	0.148	0.141
14	0.105	0.127	0.135	0.126	0.139	0.158
15	0.105	0.123	0.142	0.159	0.161	0.178
16	0.098	0.128	0.148	0.166	0.169	0.145
17	0.119	0.144	0.136	0.156	0.188	0.180
21	0.102	0.116	0.151	0.131	0.177	0.168
22	0.111	0.147	0.142	0.157	0.164	0.149
23	0.107	0.121	0.156	0.147	0.166	0.162
24	0.124	0.124	0.125	0.143	0.146	0.141
25	0.150	0.141	0.141	0.149	0.142	0.171
26	0.124	0.154	0.167	0.175	0.178	0.158
27	0.114	0.137	0.156	0.156	0.158	0.162
31	0.092	0.122	0.141	0.146	0.156	0.150
32	0.098	0.123	0.127	0.122	0.152	0.138
33	0.096	0.119	0.133	0.135	0.141	0.150
34	0.088	0.126	0.119	0.133	0.137	0.155
35	0.104	0.130	0.156	0.139	0.166	0.160
36	0.115	0.139	0.145	0.170	0.154	0.173
37	0.124	0.134	0.142	0.161	0.169	0.146

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro 12. Tamaño de la prueba para la hipótesis  $H_{0\tau} : \tau = 0$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ , continuación

41	0.105	0.118	0.141	0.133	0.133	0.170
42	0.072	0.117	0.147	0.124	0.145	0.143
43	0.094	0.115	0.123	0.136	0.138	0.146
44	0.084	0.104	0.139	0.145	0.141	0.158
45	0.110	0.131	0.153	0.136	0.136	0.172
46	0.108	0.134	0.137	0.153	0.149	0.152
47	0.108	0.150	0.135	0.158	0.159	0.145
51	0.103	0.122	0.144	0.147	0.134	0.135
52	0.099	0.117	0.133	0.133	0.146	0.140
53	0.084	0.124	0.122	0.148	0.149	0.143
54	0.088	0.138	0.131	0.144	0.160	0.146
55	0.117	0.114	0.147	0.171	0.164	0.158
56	0.121	0.133	0.152	0.180	0.172	0.152
57	0.108	0.130	0.159	0.167	0.161	0.163
61	0.097	0.113	0.154	0.153	0.138	0.154
62	0.122	0.108	0.130	0.138	0.141	0.158
63	0.093	0.122	0.144	0.146	0.142	0.152
64	0.121	0.118	0.133	0.140	0.154	0.154
65	0.101	0.145	0.152	0.160	0.181	0.168
66	0.112	0.161	0.165	0.149	0.154	0.171
67	0.134	0.157	0.154	0.174	0.166	0.174
71	0.113	0.115	0.132	0.118	0.144	0.147
72	0.120	0.146	0.139	0.142	0.144	0.137
73	0.088	0.124	0.149	0.147	0.145	0.140
74	0.105	0.116	0.125	0.147	0.142	0.156
75	0.105	0.154	0.164	0.173	0.151	0.173
76	0.122	0.154	0.155	0.182	0.159	0.187
77	0.099	0.124	0.163	0.165	0.180	0.175

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro 13. Valor máximo del estad t para  $\tau$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ .

Comb*	Observaciones					
	3	4	5	6	7	8
11	45.59	590.28	290.03	160.15	113.42	552.15
12	88.89	39.09	444.46	90.88	33.64	137.01
13	106.26	135.14	921.21	368.38	445.76	86892.48
14	254.26	52.91	299.25	1868.74	48.47	71.27
15	23.85	126.96	36.34	112.61	67.94	87.15
16	655.55	64.52	239.81	62.36	1975.31	141.31
17	87.96	100.87	37.41	91.51	233.82	1812.53
21	33.65	24.42	805.89	227.26	182.84	170.84
22	33.30	401.81	132.46	955.39	88.50	90.20
23	29.2	132.38	117.45	702.07	92.88	212.94
24	485.91	72.77	92.28	67.84	96.8	50.9
25	653.77	49.53	207.2	62.02	96.14	116.94
26	53.15	75.38	79.79	219.27	206.72	70.89
27	130.65	397.87	1002.49	63.47	1030.47	86.57
31	38.32	50.79	56.51	48.88	249.28	76.34
32	68.1	50.14	44.77	118	117.47	58.62
33	30.87	53.98	162.1	70.65	338.92	69.66

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro 13. Valor máximo del estadístico t para  $\tau$ ,  $\eta_{it} = \beta_0 + \tau_{it} + \beta_1 x_{it}$ , continuación

34	29.4	64.86	97.62	602.1	314.39	90.5
35	43.09	175.04	61.84	1093.12	119.87	338.42
36	163.33	102.58	58.83	165.44	55.25	461.49
37	275.86	529.87	56.59	192.58	210.7	74.37
41	116.46	60.46	125.08	59.35	172.74	59.74
42	14.79	86.77	58.8	90.56	191.74	185.11
43	24.67	71.29	504.19	74.11	132.22	2089.97
44	37.8	88.94	100.04	72.79	93.17	117.45
45	2098.35	79.48	246.43	327.74	68.33	49.01
46	15.32	58.31	100.74	134.16	119.88	145.13
47	38.91	78.94	60.49	126.75	155.48	85.81
51	66.51	48.59	66.3	50.28	107.1	150.8
52	272.06	431.51	85.1	92.1	129.88	861.95
53	1850.57	43.97	102.73	116.93	113.4	70.68
54	39.69	41.77	436.24	80.48	401.4	143.62
55	190.06	44.68	172.76	242.56	168.43	73.04
56	24.84	89.5	127.24	82.07	70	89.29
57	1498.6	149.91	66.73	90.52	119.27	89.85
61	45.13	58.14	95.26	47.09	69.18	343.53
62	62.2	80.23	144.69	80.59	82.06	49.81
63	20.28	49.36	156.44	81.23	113.65	100.27
64	36.05	57.54	52.96	340.3	114.27	76.01
65	40.11	356.56	249.14	54.95	160.77	87.76
66	39.58	55.19	28.68	130.19	47.79	72.7
67	53.5	128.1	89.38	146.4	83.35	49.61
71	29.87	163.65	78.42	234.32	57.17	184.47
72	246.08	171.3	66.31	108.43	279.37	57.58
73	24.83	204.16	58.94	49.34	74.09	160241.18
74	22.83	26.72	104.8	111.99	46.31	115.68
75	158.8	67.3	155.15	86.34	57.39	76.77
76	30.2	826.75	65.6	65.56	113.52	122.71
77	45.41	101.77	73.23	221.54	731.78	81.78

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

### 4.3 Modelo de cadena de Markov de primer orden

Los resultados obtenidos de la simulación con este modelo serán discutidos en forma separada para cada probabilidad modelada:  $P_0(t)$  y  $P_1(t)$ . Los datos con los que se generaron las gráficas se reportan en el anexo.

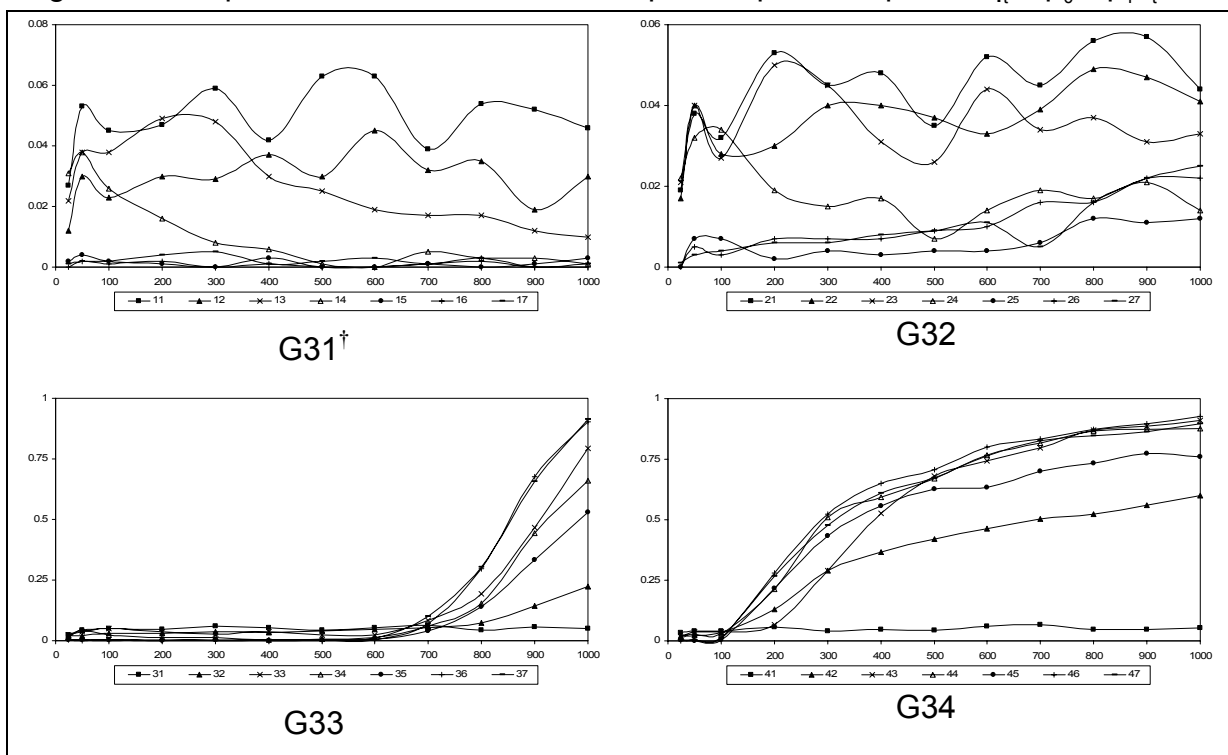
#### 4.3.1 Modelo basado en $P_0(t)$

Tamaño de la prueba para  $H_{00}:\beta_{01}=0$

El comportamiento del tamaño de la prueba para la hipótesis  $H_{00}:\beta_{01}=0$  se presenta en las gráficas de la Figura 3 (En el Anexo I Cuadro A3 se presentan los valores numéricos).

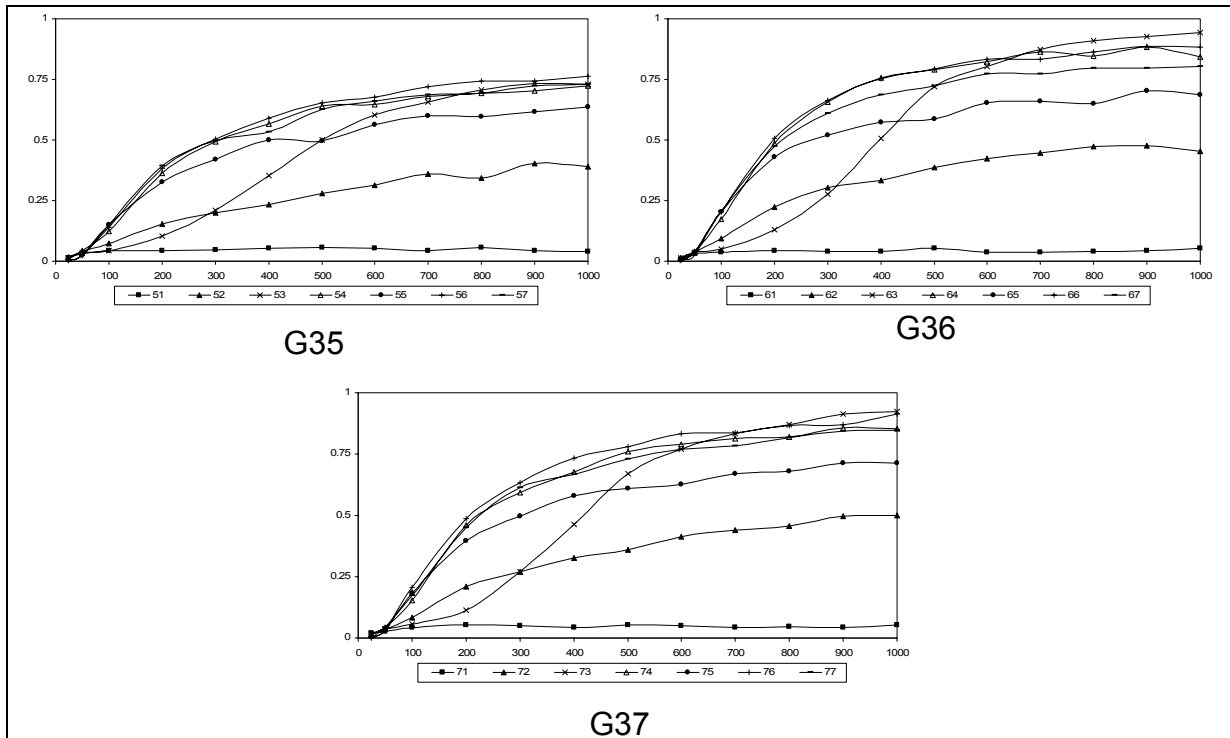
Primeramente notamos en las gráficas de la Figura 3 que cuando el proceso estacionario (1) genera alguna de las variables ya sea de respuesta o explicatoria, el tamaño de la prueba es muy pequeño y en algunos casos tiende a cero a medida que aumenta el tamaño de muestra. En la gráfica G32 el tamaño de la prueba aunque no tiende a cero en ninguna combinación pero se mantiene menor que 0.06 en todos los casos.

Figura 3. Comportamiento del tamaño de la prueba para  $H_{00}:\beta_{01}=0$ ,  $\eta_t = \beta_0 + \beta_1 x_t$  \*\*.



\*\*En la leyenda de cada gráfico aparecen siete números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

Figura 3. Comportamiento del tamaño de la prueba para  $H_{00}:\beta_{01}=0$ ,  $\eta_t = \beta_0 + \beta_1 x_t$  \*\*, continuación.



\*\*En la leyenda de cada gráfico aparecen siete números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

Cuando el proceso 3 genera a  $y_t$  (gráfica G33), a diferencia de las situaciones donde los procesos 4, 5, 6 o 7 generan a  $y_t$  el tamaño de la prueba empieza a tomar valores grandes pero a partir de tamaños de muestra alrededor de 700.

Valor máximo del estadístico t para  $\beta_{01}$

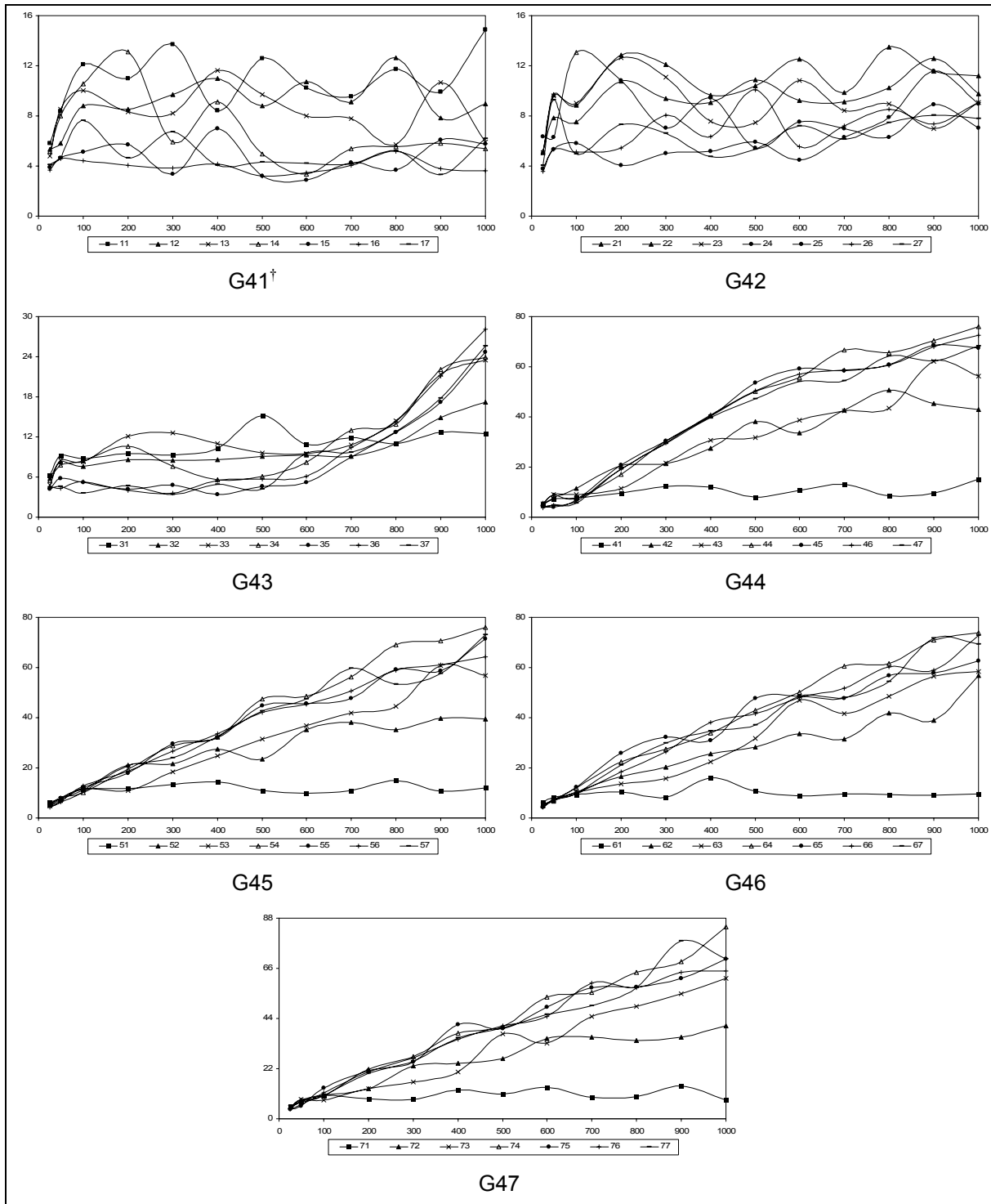
En la Figura 4 se muestra el comportamiento del valor máximo del estadístico t para los distintos tamaños de muestra (En el Anexo I Cuadro A4 se presentan los valores numéricos). En las combinaciones donde los procesos 1 y 2 están presentes como generadores de la variable de respuesta este valor es pequeño y no aumenta con el tamaño de la muestra. Por el contrario, cuando estos procesos no intervienen en el modelo como generadores de  $y_t$  (o de  $x_t$  en el caso del

proceso 1) el valor máximo de  $t$  tiende a crecer indefinidamente pero a una razón menor que como sucedía en el modelo de regresión logístico (véanse gráficas G173 a G177).

El comportamiento del tamaño de la prueba y del valor máximo del estadístico  $t$  nuevamente dan evidencia de la ocurrencia del fenómeno de correlación espuria en este modelo, cosa que no se esperaba dado que este modelo considera que las observaciones no son independientes. Generalmente el fenómeno se presenta a partir de tamaño de muestra 100 cuando la variable de respuesta es generada por un proceso no estacionario y la variable explicatoria es generada por procesos distintos del proceso 1.

En esta situación particular, en que se modela la probabilidad de éxito en el periodo actual dado que en el periodo anterior ocurrió fracaso, a diferencia de lo que se observó en el modelo logístico, el fenómeno de correlación espuria no se presenta cuando el proceso 2 genera a la variable de respuesta.

Figura 4. Comportamiento del valor máximo de  $t$  para el coeficiente  $\beta_{01}$ ,  $\eta_t = \beta_0 + \beta_1 x_t^{**}$ .



\*\*En la leyenda de cada gráfico aparecen siete números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

#### 4.3.2 Modelo basado en $P_1(t)$

Tamaño de la prueba para  $H_{10}:\beta_{11}=0$

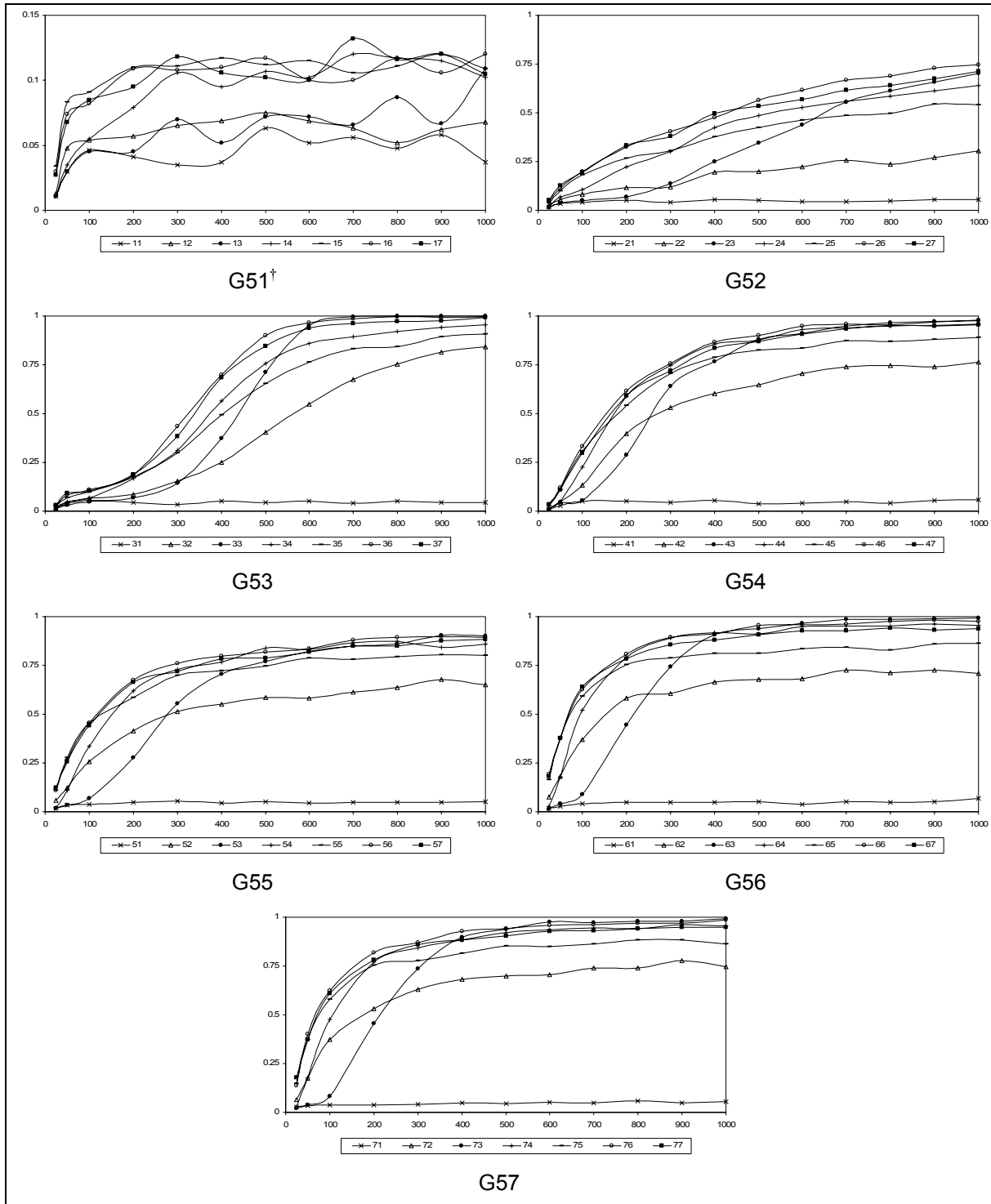
En la Figura 5 se presenta el comportamiento del tamaño de la prueba para  $H_{10}:\beta_{11}=0$ , las gráficas se generaron con los valores correspondientes reportados en el Anexo I (Cuadro A5).

En la gráfica G51 observamos que cuando el proceso 1 genera a  $y_t$  y  $x_t$  el tamaño de la prueba oscila alrededor de 0.05. En los demás casos oscila alrededor de 0.05 para tamaños de muestra menores que 100 pero asciende a valores alrededor de 0.1 para tamaños de muestra mayores. Cuando  $y_t$  es generada por un proceso distinto del proceso 1 y  $x_t$  es generada por el proceso 1 el tamaño de la prueba oscila en 0.05 (véanse gráficas G53 a G57).

En las gráficas G52 a G57 notamos que cuando el proceso 1 no interviene en el modelo, el tamaño de prueba tiende a crecer a medida que aumenta el tamaño de muestra y a partir de tamaño de muestra 500 alcanza valores próximos a 1, excepto en los casos de la gráfica G52 donde la tasa de incremento es menor.



Figura 5. Comportamiento del tamaño de la prueba para  $H_{10}: \beta_{11}=0$ ,  
 $\eta_t = \beta_0 + \beta_1 x_t$  \*\*.



\*\*En la leyenda de cada gráfico aparecen siete números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . †Número de gráfica.

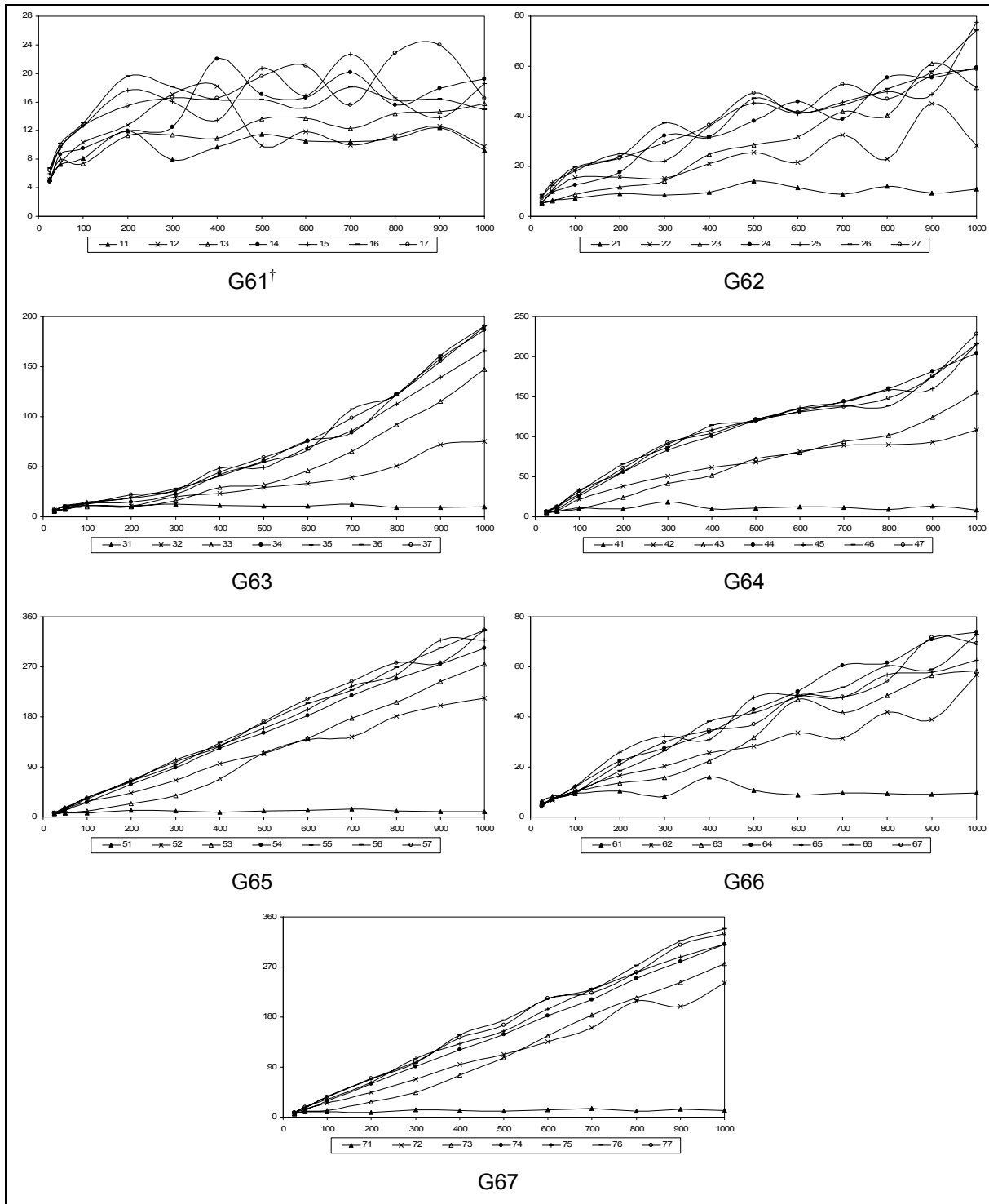
Valor máximo del estadístico t para  $\beta_{11}$

El comportamiento del valor máximo del estadístico t para los distintos tamaños de muestra se presentan en la Figura 6, (En el Anexo I Cuadro A6 se presentan los valores numéricos).

El valor máximo de t cuando interviene el proceso 1 es pequeño y no crece con el tamaño de muestra lo cual concuerda con lo observado con respecto al tamaño de la prueba. En las combinaciones donde el proceso 1 no interviene, el comportamiento es distinto en cada caso pero siempre aumenta con el tamaño de muestra.

En las combinaciones de procesos 55, 56, 57, 65, 66, 67, 75, 76 y 77 el tamaño de la prueba es alto desde tamaño de muestra 25 y observamos en las gráficas para los mismos casos que el valor máximo de t (Figura 6) crece desde tamaños de muestra muy pequeños, lo que indica que en estas combinaciones es inevitable obtener correlación espuria.

Figura 6. Comportamiento del valor máximo de  $t$  para el coeficiente  $\beta_{01}$ ,  $\eta_t = \beta_0 + \beta_1 x_t^{**}$ .



\*\*En la leyenda de cada gráfico aparecen siete números con dos dígitos que corresponden a las combinaciones de los procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ . <sup>†</sup>Número de gráfica.

Concluimos que al no estar presente el proceso 1 en este modelo de regresión se tienen resultados espuria salvo en algunos casos donde el tamaño de muestra es pequeño. A diferencia del caso anterior, la presencia del proceso 2 en la variable de respuesta no evita que ocurra este fenómeno y a excepción de esto notamos que las condiciones en las que se presenta este fenómeno en el modelo de cadena de Markov son muy similares a las condiciones en las que se presenta en el modelo de regresión logístico y aún en el modelo de regresión para variable de respuesta continua.

## CAPÍTULO V. CONCLUSIONES Y ESTUDIO POSTERIOR

### 5.1 Conclusiones

Con base en el estudio de simulación realizado en el presente trabajo empíricamente se ha probado que el fenómeno de correlación espuria se presenta en los modelos estudiados, aún en los que consideran la correlación entre observaciones.

La ocurrencia de este fenómeno depende principalmente de la no estacionaridad de los procesos analizados que intervienen en el modelo.

Las 49 combinaciones de procesos generadores de los datos consideradas en este trabajo son un grupo, si no extenso, medianamente amplio de situaciones en las que se puede incurrir en correlación espuria al tratar con observaciones de serie de tiempo o de mediciones repetidas de estudios longitudinales.

### 5.2 Estudio posterior

Los siguientes son algunos de los aspectos relacionados con los resultados obtenidos en el presente trabajo que merecen ser investigados.

- Observar el comportamiento de la devianza media como medida adicional para detectar la presencia o ausencia de correlación espuria.
- Determinar si la distribución en los estadísticos de prueba utilizados actualmente se mantiene cuando las variables que intervienen en el modelo

se generan por los procesos de serie de tiempo utilizados en el presente trabajo.

- Investigar el comportamiento del tamaño de la prueba en los modelos de respuesta binaria estudiados utilizando variables relacionadas generadas con los procesos de serie de tiempo consideradas en el presente trabajo. Adicionalmente observar la eficiencia en la predicción de los modelos obtenidos.
- Generar la metodología estadística adecuada en presencia de autocorrelación que resulte en inferencias válidas.

## LITERATURA CITADA

Albert A. y Anderson J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 71(1):1-10.

Collet, (1991). *Modeling Binary Data*. Chapman and Hall, London.

Dobson A. J. (1990). *An Introduction to Generalized Linear Models*. First edition. Chapman and Hall, London.

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

Hardin J. W. y Hilbe J. M. (2003). *Generalized Estimating Equations*. Chapman and Hall, London.

Harvey A. C. (1990). *The Econometric Analysis of Time Series*. Philip Allan. Great Britain.

Hendry D. F. (1980) Econometrics-Alchemy or science?. *Economica*. 47(188):387-406.

Liang K. Y. y Zeger S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*. 74(1):13-22.

Mood A. M., Graybill, F. A. y Boes D. C. (1974). *Introduction to the Theory of Statistics*. Third edition. McGraw-Hill. Singapur.

Murray M. P. (1994). A drunk and her dog: an illustration of cointegration and error correction. *American Statistical Association*. 48:37-39.

Myers R. H., Montgomery D. C. y Vining G. G. (2002). *Generalized Linear Models with applications in engineering and the sciences*. John Wiley and Sons, USA.

Noriega E. A. y Ventosa-Santaulária D. (2005). Spurious regression under deterministic and stochastic trends. Escuela de economía, Universidad de Guanajuato.

Noriega E. A. y Ventosa-Santaulária D. (2006?). Spurious regression and econometric trends. Dirección de Investigación Económica. Banco de México. Documento de Investigación No. 2006(05).

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

Ross S. M. (1996). *Stochastic Processes*. Second edition. John Wiley and Sons, USA.

Ross. S. M. (1999). *Simulación*. Segunda edición. Prentice Hall Hispanoamericana. México.

Stern R. D. y Coe R. (1984). A model fitting analysis of daily rainfall data. *Royal Statistical Society* 147(1):1-34.

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):1-63.

Zorn C. (2005). A solution to separation in binary response models. *Political Analysis*. 13(2):Forthcoming.



Anexo I. Tamaño de la prueba y valor máximo del estadístico t de algunos modelos

Cuadro A1. Tamaño de la prueba para  $H_0:\beta_1=0$ , modelo de regresión logístico.

Comb*	Tamaños de muestra, T									
	25	50	75	100	250	500	750	1000	2500	5000
11	0.033	0.042	0.042	0.047	0.053	0.046	0.031	0.04	0.049	0.054
12	0.024	0.041	0.047	0.047	0.045	0.05	0.058	0.051	0.047	0.062
13	0.026	0.036	0.068	0.037	0.049	0.058	0.052	0.05	0.061	0.055
14	0.023	0.035	0.04	0.054	0.047	0.048	0.047	0.048	0.056	0.047
15	0.026	0.042	0.044	0.046	0.041	0.061	0.047	0.05	0.053	0.051
16	0.043	0.038	0.041	0.046	0.058	0.045	0.048	0.054	0.043	0.043
17	0.027	0.041	0.046	0.039	0.07	0.048	0.045	0.043	0.058	0.034
21	0.025	0.051	0.036	0.038	0.044	0.038	0.056	0.044	0.058	0.041
22	0.031	0.047	0.058	0.067	0.122	0.182	0.208	0.271	0.422	0.568
23	0.041	0.038	0.055	0.042	0.088	0.311	0.541	0.688	0.877	0.896
24	0.033	0.043	0.062	0.083	0.215	0.437	0.541	0.576	0.724	0.812
25	0.037	0.084	0.081	0.106	0.211	0.353	0.45	0.51	0.649	0.725
26	0.031	0.081	0.106	0.13	0.307	0.505	0.654	0.732	0.844	0.897
27	0.044	0.08	0.082	0.126	0.291	0.484	0.606	0.686	0.849	0.893
31	0.03	0.046	0.05	0.048	0.054	0.052	0.054	0.067	0.043	0.053
32	0.029	0.041	0.042	0.047	0.079	0.326	0.638	0.784	0.895	0.924
33	0.038	0.033	0.05	0.048	0.08	0.613	0.998	1	1	1
34	0.035	0.04	0.044	0.053	0.173	0.723	0.896	0.932	0.985	0.984
35	0.038	0.036	0.048	0.047	0.128	0.624	0.822	0.874	0.944	0.955
36	0.03	0.046	0.03	0.053	0.205	0.879	0.991	0.998	1	1
37	0.024	0.056	0.042	0.045	0.171	0.801	0.965	0.981	0.999	1
41	0.023	0.038	0.041	0.045	0.059	0.04	0.045	0.043	0.048	0.053
42	0.022	0.065	0.081	0.122	0.394	0.59	0.694	0.747	0.794	0.805
43	0.036	0.039	0.039	0.057	0.379	0.884	0.956	0.971	0.991	0.998
44	0.029	0.057	0.098	0.185	0.647	0.892	0.935	0.941	0.969	0.972
45	0.039	0.072	0.119	0.233	0.575	0.803	0.861	0.885	0.884	0.902
46	0.033	0.06	0.13	0.279	0.672	0.884	0.95	0.979	0.994	0.998
47	0.037	0.074	0.129	0.263	0.661	0.87	0.919	0.938	0.984	0.996
51	0.026	0.04	0.044	0.053	0.056	0.043	0.039	0.06	0.054	0.043
52	0.038	0.108	0.221	0.255	0.456	0.587	0.64	0.667	0.772	0.838
53	0.013	0.039	0.058	0.078	0.451	0.808	0.896	0.921	0.966	0.979
54	0.029	0.136	0.247	0.374	0.731	0.857	0.873	0.913	0.926	0.958
55	0.088	0.269	0.398	0.465	0.699	0.787	0.833	0.829	0.885	0.905
56	0.086	0.299	0.407	0.488	0.749	0.864	0.905	0.927	0.956	0.967
57	0.102	0.285	0.400	0.501	0.733	0.827	0.836	0.907	0.942	0.972
61	0.022	0.032	0.042	0.057	0.054	0.052	0.046	0.04	0.047	0.055
62	0.045	0.175	0.253	0.314	0.512	0.593	0.62	0.658	0.661	0.675
63	0.025	0.034	0.063	0.081	0.527	0.9	0.962	0.983	0.998	0.99
64	0.032	0.135	0.304	0.448	0.8	0.915	0.925	0.936	0.939	0.927
65	0.089	0.316	0.438	0.523	0.708	0.796	0.797	0.817	0.802	0.817
66	0.117	0.343	0.492	0.578	0.802	0.924	0.944	0.958	0.96	0.957
67	0.11	0.331	0.462	0.553	0.786	0.867	0.894	0.91	0.942	0.963
71	0.021	0.036	0.040	0.045	0.038	0.040	0.048	0.036	0.054	0.040
72	0.041	0.135	0.247	0.300	0.521	0.622	0.704	0.727	0.744	0.775
73	0.027	0.051	0.058	0.062	0.532	0.903	0.965	0.976	0.994	0.997
74	0.029	0.126	0.303	0.441	0.791	0.896	0.929	0.937	0.970	0.966
75	0.104	0.302	0.427	0.516	0.708	0.812	0.842	0.841	0.873	0.901
76	0.107	0.328	0.478	0.556	0.817	0.918	0.946	0.964	0.983	0.983
77	0.113	0.316	0.464	0.575	0.77	0.904	0.929	0.936	0.981	0.987

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro A2. Valor máximo de t para  $\beta_1$ , modelo de regresión logístico.

Comb*	Tamaños de muestra, T																
	25	50	75	100	200	300	400	500	600	700	800	900	1000	2000	3000	4000	5000
11	6.63	7.98	8.59	8.38	9.49	10.75	9.70	10.48	10.13	8.28	10.99	11.18	10.74	19.46	11.40	12.22	9.27
12	6.05	10.23	11.74	11.71	10.78	9.87	11.63	9.35	13.95	10.57	10.52	11.69	11.41	11.65	13.46	10.04	20.09
13	6.86	8.80	8.28	10.80	14.91	10.91	7.98	9.83	11.25	11.50	11.92	11.95	10.65	11.46	11.94	9.55	12.67
14	6.41	11.09	9.59	9.98	10.99	11.99	12.72	9.38	9.70	11.00	8.81	11.35	15.75	11.61	10.19	11.63	11.31
15	6.57	11.48	9.83	11.91	14.60	9.22	12.85	10.44	12.35	12.57	8.40	10.66	11.21	9.55	17.11	14.35	11.19
16	6.46	9.55	9.29	14.65	8.85	9.99	12.57	12.08	13.24	14.65	10.50	10.35	15.32	11.28	13.23	13.26	13.69
17	6.75	10.24	12.53	14.26	14.67	10.49	10.39	11.47	11.65	9.24	9.66	12.21	14.48	12.00	12.65	12.57	14.82
21	6.61	7.62	11.26	10.22	9.74	8.75	7.65	11.23	11.89	13.04	10.63	13.34	11.40	12.70	10.45	10.49	9.76
22	7.20	10.44	9.59	13.00	14.45	16.34	19.96	22.19	26.42	21.85	20.08	37.70	30.89	35.99	50.36	61.75	78.67
23	5.92	11.04	10.58	9.20	14.44	14.97	22.63	28.38	20.71	38.83	35.37	34.85	42.40	83.91	104.77	127.67	174.45
24	6.06	7.63	9.88	12.99	18.93	27.92	29.28	36.77	31.42	47.18	39.32	45.73	53.66	74.22	109.17	135.45	151.14
25	7.01	11.05	13.66	12.23	23.84	19.12	27.90	25.93	30.61	37.35	49.02	47.86	48.72	86.56	101.67	148.88	161.53
26	6.76	9.74	11.53	17.33	18.76	25.32	37.30	27.83	38.79	34.95	44.04	45.19	49.26	83.66	108.49	143.74	161.92
27	6.73	10.51	10.58	16.74	20.23	25.97	24.55	29.23	30.51	49.32	45.86	51.62	52.95	80.53	105.90	143.12	165.63
31	5.57	12.99	7.81	9.19	11.42	10.95	14.85	11.96	9.58	11.92	11.06	13.18	11.55	10.12	11.02	12.81	11.43
32	6.01	10.57	7.48	11.18	9.23	14.41	21.16	22.97	28.24	40.84	40.17	48.20	59.23	154.95	239.74	281.82	377.63
33	7.36	7.15	8.86	12.86	8.99	13.56	18.31	26.64	32.39	47.24	69.65	79.32	99.27	312.57	412.74	444.63	449.53
34	6.96	8.62	8.39	11.56	19.38	18.63	28.23	37.50	58.01	65.79	92.17	116.52	134.53	376.41	550.25	673.97	782.21
35	6.23	9.99	10.88	12.88	14.41	15.67	30.38	35.27	43.38	60.25	79.32	89.18	107.97	335.36	502.11	672.32	730.82
36	5.89	8.90	9.02	11.15	16.09	19.36	28.52	44.47	50.50	74.89	86.67	115.93	140.78	359.41	508.63	592.47	607.74
37	8.41	8.93	10.59	11.60	12.62	22.50	34.88	53.43	46.93	77.75	90.65	107.64	124.55	356.90	496.93	644.60	668.58
41	5.85	7.54	8.97	7.82	14.14	9.06	11.82	9.05	12.05	13.01	11.39	11.43	10.18	10.18	8.79	11.12	11.01
42	6.26	10.37	9.45	12.02	27.00	36.30	43.52	59.03	54.92	64.98	62.96	67.37	84.01	306.93	530.81	791.50	1042.31
43	6.40	9.44	9.49	10.52	16.87	29.29	39.38	51.12	62.98	74.16	99.88	132.18	171.76	557.53	857.08	1042.61	1179.31
44	6.17	10.24	10.45	19.58	45.64	59.77	75.01	94.03	94.56	111.83	147.83	174.69	201.27	615.23	931.37	1160.19	1379.54
45	6.08	10.53	15.36	20.31	44.76	61.79	74.47	91.53	87.99	113.65	125.23	179.48	197.76	614.63	1026.57	1315.55	1545.89
46	7.34	9.16	13.34	22.27	42.26	72.27	84.56	86.05	108.85	109.20	148.69	179.62	206.02	643.22	994.11	1193.39	1421.64
47	6.94	9.20	14.56	23.46	44.17	66.01	79.17	90.24	93.66	104.93	144.54	170.03	209.15	648.41	1013.54	1208.95	1332.32
51	5.69	9.65	7.32	8.22	9.60	11.61	9.59	10.24	15.39	11.42	11.48	11.67	15.27	13.27	10.13	12.55	9.52
52	6.19	12.01	16.42	19.11	46.72	63.44	88.78	107.79	132.66	151.98	180.00	184.28	226.05	434.43	603.13	920.71	1091.20
53	5.39	9.25	9.13	11.84	21.58	41.11	65.60	106.11	129.92	168.34	207.30	241.16	277.81	575.07	867.16	1170.31	1465.47

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_i$ , el segundo el proceso que generó a  $x_i$ .

Cuadro A2. Valor máximo de t para  $\beta_1$ , modelo de regresión logístico. Continuación.

Comb*	Tamaños de muestra, T																
	25	50	75	100	200	300	400	500	600	700	800	900	1000	2000	3000	4000	5000
54	6.40	11.63	17.03	25.38	56.64	91.46	124.09	149.77	181.54	219.35	244.96	274.47	312.00	625.59	939.98	1248.31	1501.84
55	8.28	14.89	22.78	32.36	65.84	96.06	119.90	155.96	202.70	226.79	260.40	277.17	325.02	674.73	871.67	1270.02	1624.90
56	7.39	16.61	24.58	27.88	70.46	103.93	139.71	170.06	214.04	242.13	272.38	320.90	317.30	630.31	976.22	1280.82	1605.08
57	8.59	17.06	22.69	29.08	67.07	100.81	145.18	165.45	214.62	238.41	278.71	301.38	345.03	657.60	994.79	1215.27	1630.30
61	5.85	8.32	7.51	10.02	11.51	10.66	9.43	11.50	11.38	11.89	11.77	11.58	11.82	10.90	14.32	11.13	12.22
62	7.51	13.27	16.86	19.62	44.08	59.02	86.91	100.11	126.02	152.18	170.09	187.32	193.37	356.02	637.14	507.35	709.79
63	5.68	7.63	10.34	11.42	21.12	41.70	68.82	108.81	139.58	172.37	205.41	238.01	269.30	530.73	606.17	579.28	682.07
64	5.87	10.70	16.80	27.83	57.96	90.14	119.99	150.36	184.00	216.34	241.88	275.54	293.08	569.40	878.53	792.37	749.08
65	8.16	16.86	21.93	32.34	62.62	104.60	116.41	169.31	176.48	195.47	249.90	285.99	305.01	553.27	755.26	721.97	793.66
66	8.31	15.65	27.64	31.55	66.04	102.74	137.42	172.18	189.50	224.57	238.76	277.50	308.34	539.73	617.41	640.65	588.71
67	7.87	14.63	25.28	34.80	64.34	95.66	128.58	155.85	189.74	220.45	256.12	274.35	250.73	474.81	539.33	578.19	574.93
71	6.71	7.89	10.94	9.33	11.96	9.22	9.21	10.38	9.23	12.40	10.79	12.84	9.45	16.28	10.06	10.08	9.09
72	6.82	10.76	14.99	20.79	39.43	64.28	95.16	110.76	136.47	155.57	172.21	197.92	214.06	399.22	671.79	872.14	1009.87
73	5.70	7.48	12.20	8.16	21.44	43.22	73.90	109.37	136.99	173.32	208.62	245.57	270.37	565.42	818.30	1116.84	1296.32
74	5.58	10.68	19.27	25.62	59.00	90.91	121.51	153.71	187.54	209.84	247.00	270.54	307.92	607.30	924.69	1132.87	1451.37
75	7.99	14.74	24.41	32.77	62.25	96.31	125.84	167.20	205.28	268.74	254.66	286.25	318.43	559.16	1101.42	1078.41	1635.43
76	7.34	15.62	25.60	32.79	69.98	97.50	139.68	180.99	211.67	224.01	272.83	309.20	328.93	637.46	929.65	1199.27	1315.58
77	7.35	14.81	22.44	30.77	65.80	101.51	128.14	164.33	185.46	230.46	259.97	309.66	310.36	578.02	868.85	1101.70	1594.48

\*Combinación de procesos de la expresión (3.1), el primer dígito indica el proceso que generó a y, el segundo el proceso que generó a x.

Cuadro A3. Tamaño de la prueba para  $H_{00}: \beta_{01}=0, \eta_t = \beta_0 + \beta_1 x_t$ . Modelo de cadena de Markov.

Comb*	Tamaños de muestra, T.											
	25	50	100	200	300	400	500	600	700	800	900	1000
11	0.03	0.05	0.05	0.05	0.06	0.04	0.06	0.06	0.04	0.05	0.05	0.05
12	0.01	0.03	0.02	0.03	0.03	0.04	0.03	0.05	0.03	0.04	0.02	0.03
13	0.02	0.04	0.04	0.05	0.05	0.03	0.03	0.02	0.02	0.02	0.01	0.01
14	0.03	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.02	0.04	0.03	0.05	0.05	0.05	0.04	0.05	0.05	0.06	0.06	0.04
22	0.02	0.04	0.03	0.03	0.04	0.04	0.04	0.03	0.04	0.05	0.05	0.04
23	0.02	0.04	0.03	0.05	0.05	0.03	0.03	0.04	0.03	0.04	0.03	0.03
24	0.02	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.02	0.01
25	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
26	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02
27	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.03
31	0.03	0.05	0.05	0.05	0.06	0.05	0.04	0.06	0.06	0.04	0.06	0.05
32	0.02	0.02	0.03	0.03	0.04	0.03	0.04	0.05	0.05	0.07	0.15	0.22
33	0.02	0.04	0.05	0.04	0.03	0.04	0.02	0.02	0.08	0.19	0.47	0.79
34	0.02	0.04	0.02	0.01	0.01	0.01	0.01	0.01	0.06	0.15	0.44	0.66
35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.14	0.33	0.53
36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.07	0.30	0.68	0.90
37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.10	0.31	0.66	0.91
41	0.03	0.04	0.04	0.06	0.04	0.05	0.04	0.06	0.07	0.05	0.05	0.05
42	0.02	0.02	0.03	0.13	0.29	0.37	0.42	0.46	0.50	0.53	0.56	0.60
43	0.02	0.04	0.04	0.07	0.29	0.53	0.68	0.75	0.80	0.87	0.89	0.91
44	0.02	0.03	0.03	0.21	0.51	0.60	0.67	0.77	0.82	0.87	0.87	0.88
45	0.00	0.00	0.00	0.22	0.43	0.56	0.63	0.63	0.70	0.73	0.78	0.76
46	0.00	0.00	0.01	0.28	0.52	0.65	0.71	0.80	0.83	0.87	0.90	0.93
47	0.00	0.00	0.01	0.27	0.48	0.61	0.68	0.76	0.83	0.85	0.86	0.90
51	0.01	0.03	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.06	0.04	0.04
52	0.02	0.04	0.08	0.15	0.20	0.24	0.28	0.32	0.36	0.34	0.40	0.39
53	0.01	0.03	0.04	0.10	0.21	0.35	0.50	0.60	0.66	0.71	0.74	0.73
54	0.01	0.05	0.12	0.36	0.49	0.57	0.64	0.65	0.68	0.70	0.71	0.73
55	0.01	0.02	0.15	0.33	0.42	0.50	0.50	0.56	0.60	0.60	0.62	0.64
56	0.00	0.03	0.14	0.38	0.50	0.59	0.65	0.68	0.72	0.74	0.74	0.76
57	0.00	0.03	0.15	0.40	0.50	0.53	0.63	0.66	0.69	0.69	0.72	0.73
61	0.01	0.03	0.04	0.04	0.04	0.04	0.06	0.04	0.04	0.04	0.05	0.05
62	0.01	0.04	0.09	0.22	0.30	0.33	0.39	0.42	0.45	0.47	0.48	0.45
63	0.01	0.04	0.05	0.13	0.28	0.51	0.72	0.80	0.87	0.91	0.93	0.94
64	0.01	0.04	0.18	0.49	0.66	0.76	0.79	0.82	0.86	0.85	0.89	0.85
65	0.00	0.03	0.20	0.43	0.52	0.57	0.59	0.65	0.66	0.65	0.70	0.69
66	0.00	0.04	0.21	0.51	0.67	0.75	0.79	0.83	0.83	0.86	0.89	0.88
67	0.01	0.03	0.21	0.47	0.61	0.69	0.72	0.77	0.77	0.80	0.80	0.80
71	0.02	0.03	0.05	0.05	0.05	0.04	0.06	0.05	0.04	0.05	0.04	0.05
72	0.02	0.04	0.08	0.21	0.27	0.33	0.36	0.42	0.44	0.46	0.50	0.50
73	0.02	0.04	0.06	0.11	0.27	0.47	0.67	0.77	0.83	0.87	0.92	0.93
74	0.02	0.04	0.16	0.46	0.60	0.68	0.76	0.79	0.81	0.82	0.86	0.85
75	0.00	0.03	0.18	0.40	0.50	0.58	0.61	0.63	0.67	0.68	0.71	0.72
76	0.00	0.03	0.21	0.49	0.63	0.73	0.78	0.83	0.84	0.87	0.87	0.91
77	0.00	0.04	0.17	0.45	0.61	0.67	0.73	0.77	0.78	0.82	0.84	0.85

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro A4. Valor máximo de t para el coeficiente  $\beta_{01}$ ,  $\eta_t = \beta_0 + \beta_1 x_t$ . Modelo de cadena de Markov.

Comb*	Tamaños de muestra, T.											
	25	50	100	200	300	400	500	600	700	800	900	1000
11	5.8	8.4	12.1	11.0	13.7	8.4	12.6	10.2	9.5	11.8	9.9	14.9
12	5.3	5.8	8.8	8.6	9.7	11.0	8.8	10.7	9.1	12.6	7.8	9.0
13	4.8	8.5	10.0	8.3	8.2	11.7	9.7	8.0	7.8	5.7	10.7	5.9
14	5.2	8.0	10.5	13.1	5.9	9.1	4.9	3.4	5.4	5.6	5.8	5.4
15	3.9	4.6	5.1	5.7	3.3	7.0	3.2	2.9	4.3	3.7	6.1	5.8
16	3.7	4.6	4.5	4.0	3.8	4.2	3.2	3.5	4.1	5.2	3.8	3.7
17	4.1	4.7	7.6	4.6	6.7	4.0	4.3	4.2	4.2	5.2	3.3	6.2
21	5.1	9.7	8.9	12.9	12.1	9.6	10.4	12.5	9.9	13.5	11.6	11.2
22	5.2	7.8	7.5	10.8	9.4	9.1	10.9	9.2	9.1	10.3	12.6	9.8
23	5.1	9.6	9.0	12.7	11.1	7.6	7.5	10.8	8.4	9.0	7.0	9.0
24	6.4	6.3	13.1	10.8	7.1	9.4	5.4	7.5	7.0	6.3	8.9	7.1
25	3.8	5.3	5.8	4.0	5.0	5.2	5.9	4.5	6.3	7.9	11.6	9.0
26	3.6	5.4	5.1	5.5	8.1	6.4	10.1	5.5	7.2	8.6	7.4	9.1
27	4.1	9.3	5.0	7.3	6.6	4.8	5.3	7.2	6.1	7.5	8.1	7.8
31	6.3	9.2	8.8	9.5	9.3	10.3	15.2	10.9	11.9	11.0	12.7	12.5
32	5.8	8.3	7.6	8.6	8.5	8.6	9.1	9.3	9.1	11.0	14.9	17.2
33	4.8	8.5	8.3	12.1	12.6	11.0	9.6	9.5	10.8	14.4	21.4	23.5
34	5.5	7.8	8.4	10.6	7.6	5.6	6.1	8.2	13.0	13.9	22.1	23.9
35	4.2	5.8	5.2	4.2	4.8	3.4	4.6	5.2	9.0	12.7	17.2	24.7
36	4.6	4.3	5.3	4.0	3.6	5.5	5.7	6.1	10.4	14.3	21.1	28.1
37	4.3	4.6	3.6	4.7	3.4	4.9	4.2	9.6	9.7	12.8	17.8	25.6
41	5.2	7.3	7.9	9.7	12.2	11.9	7.9	10.6	13.0	8.5	9.6	15.2
42	5.3	7.7	11.5	20.7	21.3	27.5	38.1	33.6	42.5	50.7	45.3	42.8
43	5.4	9.1	9.1	11.4	21.6	30.8	31.7	38.7	42.5	43.5	62.2	56.2
44	5.6	8.7	7.4	16.9	30.5	40.3	50.2	55.6	66.6	65.6	70.4	76.0
45	4.6	4.1	6.8	20.7	30.4	40.2	53.6	59.1	58.4	60.9	68.5	67.4
46	3.6	4.7	6.4	19.1	29.5	40.9	50.4	57.1	58.6	60.6	67.9	72.6
47	4.3	4.7	5.5	19.2	29.2	39.8	47.1	54.3	54.3	64.3	62.3	68.3
51	6.0	7.9	11.5	11.7	13.4	14.3	10.9	9.8	10.9	14.9	10.7	11.9
52	5.6	7.9	11.3	21.0	21.5	27.5	23.6	35.3	38.2	35.2	39.6	39.4
53	4.9	7.4	11.7	10.8	18.4	24.9	31.5	36.8	41.8	44.6	60.7	56.8
54	5.1	8.0	10.2	19.4	28.7	32.3	47.4	48.7	56.3	68.9	70.7	76.0
55	4.9	6.7	12.2	18.0	29.5	32.1	44.8	45.5	47.8	59.2	58.8	71.5
56	4.3	6.3	12.8	18.6	26.6	33.7	42.2	45.3	50.7	59.0	61.2	64.4
57	4.4	6.0	10.3	20.9	23.9	32.6	42.7	47.4	59.8	53.3	57.7	73.0
61	6.3	8.3	9.3	10.5	8.3	15.9	10.6	8.8	9.6	9.5	9.0	9.6
62	5.6	6.7	10.5	16.5	20.3	25.6	28.1	33.5	31.4	42.0	38.8	56.9
63	5.2	7.3	10.1	13.5	15.8	22.3	31.8	46.8	41.6	48.7	56.6	58.5
64	4.9	7.3	12.0	22.4	27.4	33.8	43.0	50.1	60.6	61.7	70.8	73.8
65	4.2	7.2	12.4	25.7	32.3	30.9	47.9	48.4	47.7	56.7	57.8	62.7
66	4.1	7.5	9.6	18.5	26.4	38.0	41.7	47.9	51.9	60.3	58.8	72.9
67	4.2	7.1	9.9	21.0	29.8	34.6	37.0	48.6	48.0	54.5	71.7	69.3
71	5.4	7.5	10.1	8.9	8.4	12.6	10.9	13.7	9.3	9.8	14.4	8.3
72	5.3	8.3	10.3	13.3	23.1	24.4	26.3	35.1	35.9	34.4	35.7	40.9
73	5.1	8.4	8.1	13.1	16.2	20.5	37.2	33.2	44.9	49.2	54.8	61.7
74	5.4	7.3	9.9	21.7	27.3	37.5	40.6	53.4	55.6	64.2	68.8	84.1
75	4.0	5.7	13.5	21.3	25.1	41.4	39.6	49.0	57.4	57.7	61.6	70.0
76	4.4	6.3	11.4	20.6	26.8	34.9	40.9	44.8	59.4	57.4	64.3	64.8
77	4.3	7.7	10.3	19.8	25.3	35.4	40.0	45.8	49.6	57.4	77.9	70.0

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro A5. Tamaño de la prueba para la hipótesis  $H_{10}: \beta_{11}=0, \eta_t = \beta_0 + \beta_1 x_t$ .  
Modelo de cadena de Markov.

Comb*	Tamaños de muestra, T											
	25	50	100	200	300	400	500	600	700	800	900	1000
11	0.01	0.03	0.05	0.04	0.04	0.04	0.06	0.05	0.06	0.05	0.06	0.04
12	0.01	0.05	0.05	0.06	0.07	0.07	0.08	0.07	0.06	0.05	0.06	0.07
13	0.01	0.03	0.05	0.05	0.07	0.05	0.07	0.07	0.07	0.09	0.07	0.11
14	0.01	0.04	0.06	0.08	0.11	0.10	0.11	0.10	0.12	0.12	0.12	0.10
15	0.03	0.08	0.09	0.11	0.11	0.12	0.11	0.12	0.11	0.11	0.12	0.11
16	0.03	0.07	0.08	0.11	0.11	0.11	0.12	0.10	0.10	0.12	0.11	0.12
17	0.03	0.07	0.09	0.10	0.12	0.11	0.10	0.10	0.13	0.12	0.12	0.11
21	0.02	0.04	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.06	0.06
22	0.03	0.06	0.08	0.12	0.12	0.19	0.20	0.22	0.26	0.24	0.27	0.30
23	0.01	0.04	0.05	0.07	0.14	0.25	0.35	0.44	0.56	0.61	0.66	0.70
24	0.02	0.07	0.11	0.22	0.30	0.42	0.49	0.53	0.56	0.59	0.61	0.64
25	0.04	0.10	0.18	0.27	0.30	0.38	0.43	0.46	0.49	0.50	0.54	0.54
26	0.04	0.11	0.20	0.33	0.40	0.48	0.57	0.62	0.67	0.69	0.73	0.75
27	0.05	0.13	0.20	0.33	0.38	0.50	0.53	0.57	0.62	0.64	0.68	0.71
31	0.01	0.04	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.05	0.04	0.05
32	0.02	0.05	0.07	0.09	0.16	0.25	0.40	0.55	0.67	0.75	0.81	0.84
33	0.01	0.03	0.05	0.07	0.14	0.38	0.71	0.95	1.00	1.00	1.00	1.00
34	0.02	0.05	0.07	0.17	0.31	0.57	0.76	0.86	0.89	0.92	0.94	0.95
35	0.03	0.07	0.10	0.17	0.30	0.49	0.66	0.76	0.83	0.84	0.89	0.91
36	0.03	0.08	0.11	0.19	0.44	0.70	0.90	0.97	0.99	1.00	0.99	1.00
37	0.03	0.09	0.10	0.19	0.39	0.68	0.85	0.94	0.96	0.97	0.98	0.99
41	0.01	0.03	0.05	0.05	0.04	0.05	0.04	0.04	0.05	0.04	0.05	0.06
42	0.01	0.05	0.13	0.40	0.53	0.60	0.65	0.71	0.74	0.75	0.74	0.76
43	0.01	0.05	0.06	0.29	0.64	0.77	0.88	0.91	0.95	0.97	0.97	0.98
44	0.01	0.05	0.23	0.59	0.75	0.86	0.88	0.93	0.94	0.95	0.95	0.96
45	0.03	0.11	0.31	0.54	0.70	0.79	0.82	0.84	0.87	0.87	0.88	0.89
46	0.03	0.12	0.33	0.62	0.76	0.87	0.90	0.95	0.96	0.96	0.97	0.98
47	0.03	0.11	0.30	0.59	0.72	0.84	0.87	0.91	0.93	0.95	0.95	0.96
51	0.02	0.04	0.04	0.05	0.06	0.04	0.05	0.04	0.05	0.05	0.05	0.05
52	0.06	0.12	0.26	0.41	0.51	0.55	0.59	0.58	0.61	0.64	0.68	0.65
53	0.02	0.03	0.07	0.28	0.56	0.71	0.77	0.82	0.85	0.86	0.91	0.90
54	0.02	0.11	0.33	0.62	0.73	0.77	0.84	0.83	0.87	0.87	0.84	0.86
55	0.11	0.28	0.46	0.59	0.70	0.72	0.75	0.79	0.78	0.79	0.81	0.80
56	0.11	0.26	0.46	0.68	0.76	0.80	0.82	0.84	0.88	0.89	0.90	0.89
57	0.12	0.26	0.44	0.67	0.72	0.78	0.79	0.82	0.85	0.85	0.88	0.88
61	0.02	0.03	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.07
62	0.08	0.18	0.37	0.58	0.61	0.66	0.68	0.68	0.73	0.71	0.73	0.71
63	0.01	0.04	0.09	0.44	0.74	0.91	0.94	0.97	0.99	0.99	0.99	0.99
64	0.03	0.18	0.52	0.79	0.89	0.92	0.91	0.95	0.95	0.95	0.96	0.95
65	0.16	0.38	0.59	0.75	0.79	0.81	0.81	0.84	0.84	0.83	0.86	0.86
66	0.19	0.38	0.63	0.81	0.89	0.91	0.95	0.96	0.96	0.98	0.98	0.98
67	0.18	0.38	0.64	0.78	0.86	0.88	0.91	0.93	0.93	0.94	0.93	0.94
71	0.03	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.05	0.06	0.05	0.05
72	0.07	0.17	0.38	0.53	0.63	0.68	0.70	0.71	0.74	0.74	0.78	0.75
73	0.02	0.04	0.08	0.45	0.74	0.90	0.94	0.98	0.97	0.98	0.98	0.99
74	0.03	0.18	0.48	0.77	0.84	0.88	0.92	0.94	0.95	0.94	0.96	0.95
75	0.15	0.37	0.58	0.75	0.78	0.82	0.85	0.85	0.86	0.88	0.88	0.86
76	0.14	0.40	0.62	0.82	0.87	0.93	0.94	0.96	0.96	0.97	0.97	0.99
77	0.18	0.37	0.61	0.78	0.86	0.88	0.90	0.93	0.93	0.94	0.95	0.95

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .

Cuadro A6. Valor máximo de t para el coeficiente  $\beta_{11}$ ,  $\eta_t = \beta_0 + \beta_1 x_t$ . Modelo de cadena de Markov.

Comb*	Tamaño de muestra, T											
	25	50	100	200	300	400	500	600	700	800	900	1000
11	5.21	7.29	8.13	11.91	7.96	9.68	11.44	10.53	10.42	10.89	12.41	9.24
12	5.15	7.45	10.41	12.76	17.07	18.16	9.88	11.85	9.96	11.26	12.63	9.85
13	5.24	7.97	7.38	11.29	11.42	10.94	13.67	13.73	12.31	14.37	14.69	15.74
14	4.82	8.67	9.56	11.96	12.52	22.05	17.05	16.62	20.18	15.58	17.95	19.19
15	5.95	9.60	12.68	17.61	16.09	13.41	20.72	16.93	22.68	16.57	13.78	18.55
16	6.69	10.21	13.06	19.56	18.12	16.35	16.34	15.14	18.06	16.28	16.42	14.89
17	6.46	9.74	12.66	15.51	16.57	16.48	19.56	21.13	15.56	22.86	24.02	16.49
21	5.23	6.37	7.24	9.16	8.55	9.67	14.12	11.52	8.87	11.87	9.38	11.07
22	5.51	9.84	15.53	15.82	15.18	20.99	25.67	21.71	32.41	22.83	45.06	28.23
23	5.30	6.25	8.71	11.74	14.18	24.74	28.62	31.62	41.87	40.38	61.08	51.48
24	5.45	9.63	12.51	17.62	32.30	31.40	38.05	45.75	39.06	55.40	55.57	59.49
25	8.13	13.73	18.18	25.08	22.22	36.10	45.37	40.97	45.53	49.76	48.68	77.61
26	8.53	12.27	19.78	24.12	37.35	31.95	47.11	41.51	44.53	50.84	57.85	74.51
27	7.14	10.63	18.86	23.12	29.31	36.58	49.37	41.64	52.86	46.89	56.29	58.97
31	5.26	7.15	11.30	10.86	12.69	11.14	10.79	10.72	12.39	9.61	9.22	9.96
32	6.75	7.89	11.41	10.42	19.75	23.53	29.22	33.12	39.29	50.47	72.26	75.44
33	5.78	7.97	9.90	10.13	15.75	29.38	31.91	46.22	65.13	92.06	115.16	147.28
34	5.01	7.98	13.64	14.65	22.65	42.22	55.88	76.22	83.72	121.75	158.08	186.74
35	7.26	9.99	14.51	18.91	26.77	48.55	49.41	69.13	85.74	112.78	139.47	166.22
36	5.46	11.23	13.01	19.15	28.23	40.52	54.75	66.45	107.01	121.60	161.27	190.95
37	6.69	9.92	13.23	21.85	25.94	44.36	59.37	75.14	98.77	122.34	155.32	190.10
41	5.34	6.69	10.55	10.16	18.08	10.11	10.95	12.22	11.73	9.10	13.08	8.41
42	6.13	7.71	21.40	38.20	50.55	61.86	68.16	81.44	88.83	89.62	93.15	108.56
43	5.52	7.69	10.40	23.98	41.77	52.03	72.69	80.36	94.21	101.73	124.37	155.96
44	5.36	8.96	26.01	56.15	83.22	100.53	120.37	131.83	144.07	159.77	181.63	204.10
45	5.87	12.53	32.96	56.95	90.86	108.06	120.53	135.45	143.18	158.32	159.92	215.82
46	7.73	13.26	31.93	65.44	86.51	113.99	119.51	134.62	138.64	138.32	175.28	215.67
47	6.32	12.38	28.33	60.93	92.27	104.23	121.61	131.20	137.41	148.54	175.53	228.12
51	6.31	7.08	7.00	11.79	10.96	8.90	10.86	12.14	14.27	10.29	9.93	9.43
52	7.68	12.75	27.03	42.72	65.73	95.63	113.73	139.59	143.55	181.45	200.30	213.20
53	6.26	7.57	10.46	23.48	38.75	68.86	114.93	141.87	177.82	205.99	244.02	275.04
54	6.01	12.49	26.38	58.89	89.13	123.45	150.77	182.27	218.76	248.09	274.98	303.99
55	7.60	17.13	34.64	63.67	102.93	128.35	159.13	192.81	235.74	255.78	318.28	318.08
56	7.78	15.00	34.26	65.27	93.51	133.15	168.46	204.40	228.44	268.56	303.08	336.14
57	7.54	15.80	32.07	65.87	100.12	127.77	171.68	212.75	243.31	277.35	277.22	335.98
61	5.65	6.31	10.48	10.39	15.04	10.55	14.21	10.23	17.23	12.64	17.42	14.09
62	7.30	12.48	25.35	45.07	66.74	91.53	110.58	126.84	168.88	167.53	190.96	216.77
63	6.47	10.18	12.73	31.61	55.08	75.28	106.78	139.77	176.99	211.98	239.40	268.27
64	5.98	11.92	28.18	57.95	88.47	121.36	152.43	186.11	214.30	245.26	278.16	287.69
65	8.79	15.96	35.32	70.59	104.77	126.53	174.74	187.77	228.74	235.12	297.35	267.02
66	8.48	16.09	32.40	69.21	98.25	132.52	161.52	198.27	230.39	261.65	259.64	330.10
67	8.34	16.96	34.59	63.28	101.97	131.79	163.12	201.60	216.86	234.83	271.37	357.65
71	6.01	9.57	10.10	8.85	12.88	12.32	10.64	12.95	15.16	11.30	14.07	11.96
72	7.10	14.35	25.72	44.49	68.22	94.36	112.73	135.14	161.22	209.25	199.73	241.53
73	5.83	10.93	11.62	27.99	43.97	75.10	106.21	146.85	183.88	214.42	242.17	275.84
74	5.87	12.54	28.20	59.41	91.64	121.48	149.27	182.94	211.75	249.40	280.20	310.35
75	8.86	17.89	32.49	62.53	106.19	131.98	154.85	194.86	229.36	260.31	288.06	311.10
76	8.29	16.80	37.78	68.66	97.66	147.24	173.59	212.59	230.21	272.69	317.14	338.27
77	8.91	18.21	35.96	70.07	99.67	142.55	165.53	213.93	222.80	260.11	309.67	329.84

\*Combinación de procesos de serie de tiempo de la expresión (3.1), el primer dígito indica el proceso que generó a  $y_t$ , el segundo el proceso que generó a  $x_t$ .



## Anexo II. Diagramas de flujo utilizados en cada modelo de regresión

La Figura A1 presenta el diagrama de flujo seguido para la obtención del tamaño de la prueba y el valor máximo del estadístico  $t$  en el modelo de regresión logístico. A continuación se describe lo indicado por cada una de las literales que intervienen en el diagrama.

R. Matriz de resultados que contiene el tamaño de la prueba y el valor máximo de  $t$  en cada tamaño de muestra.

T. Vector de tamaño  $n$  cuyos elementos son los diferentes tamaños de muestra.

r. Número de repeticiones del experimento, para cada tamaño de muestra se realizaron 1000 repeticiones.

nr. Número de rechazos de la hipótesis  $H_0: \beta_1=0$ .

i. Elemento contador sobre los elementos del vector T.

x. Vector de observaciones de la variable explicatoria.

y. Vector de observaciones de la variable de respuesta.

E(). Operador esperanza.

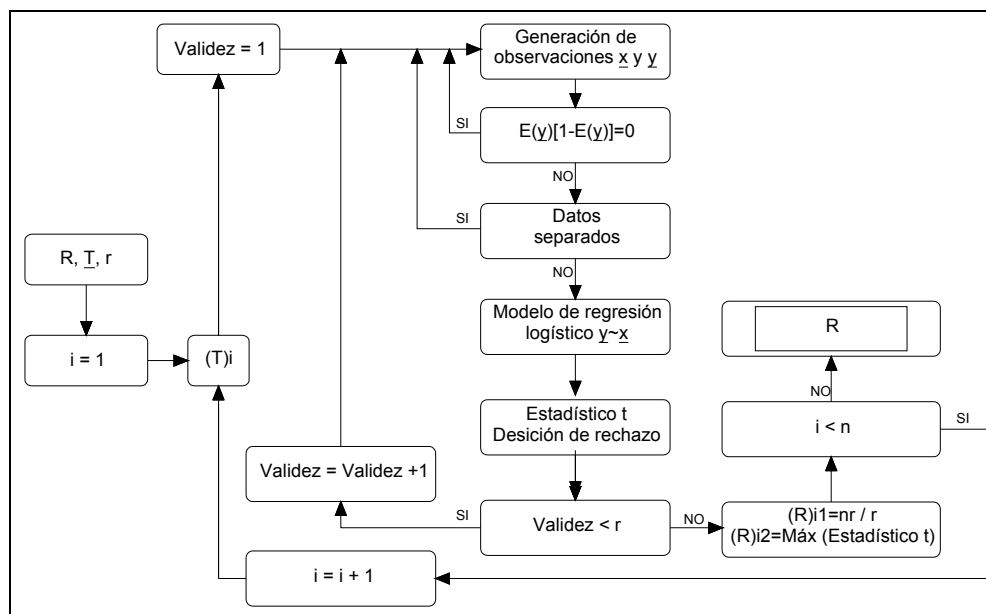


Figura A1. Diagrama de flujo, modelo de regresión logístico

La Figura A2 muestra el diagrama de flujo seguido para la obtención del tamaño de la prueba y el valor máximo del estadístico t en el modelo de ecuaciones de estimación generalizadas.

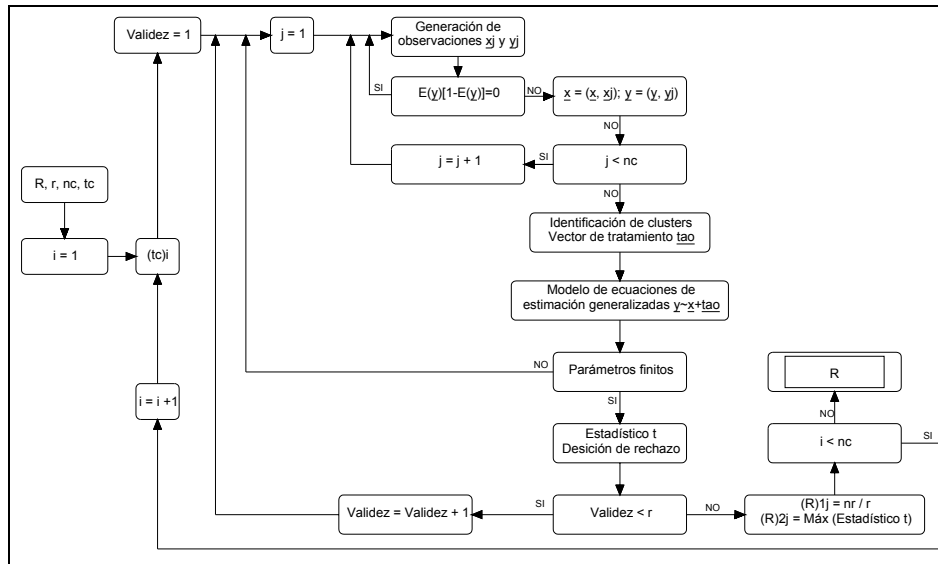


Figura A2. Diagrama de flujo, ecuaciones de estimación generalizadas

A continuación se describe lo que indica cada literal en la Figura 2.

R. Matriz de resultados que contiene el tamaño de la prueba y el valor máximo de t en cada tamaño de muestra.

tc. Vector de tamaño nc cuyos elementos son números de observaciones por individuo.

r. Número de repeticiones del experimento, para cada tamaño de muestra se realizaron 1000 repeticiones.

nc. Número de clusters o de individuos.

nr. Número de rechazos de la hipótesis  $H_0: \beta_1=0$ .

i. Elemento contador sobre los elementos del vector tc.

x. Vector de observaciones de la variable explicatoria.

y. Vector de observaciones de la variable de respuesta.

E(). Operador esperanza.

j. Elemento contador sobre el número de cluster en cada tamaño de muestra.

tao. Vector de tamaño  $(n_c \times t_c)$  cuyos elementos son 0 y 1 y describen la ausencia o presencia de un tratamiento.

La Figura A3 muestra el diagrama de flujo seguido para la obtención del tamaño de la prueba y el valor máximo del estadístico t en el modelo de cadena de Markov de primer orden. A continuación se indica el significado de cada una de las literales que intervienen en el diagrama.

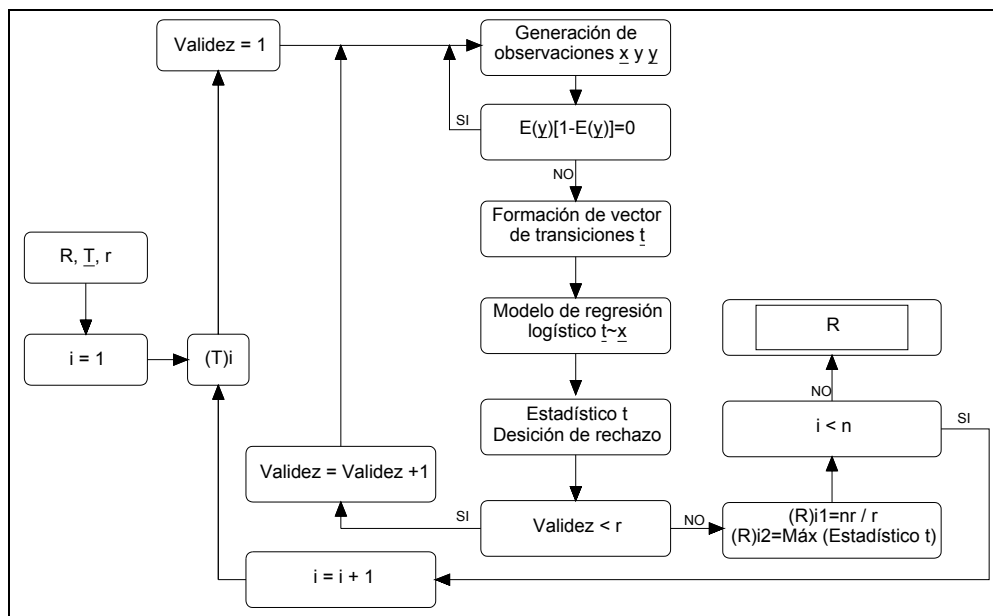


Figura A3. Diagrama de flujo, modelo de cadena de Markov

A continuación se describe lo que indica cada literal en la figura anterior.

R. Matriz de resultados que contiene el tamaño de la prueba y el valor máximo de t en cada tamaño de muestra.

T. Vector de tamaño n cuyos elementos son los diferentes tamaños de muestra.

r. Número de repeticiones del experimento, para cada tamaño de muestra se realizaron 1000 repeticiones.

x. Vector de observaciones de la variable explicatoria.

y. Vector de observaciones de la variable de respuesta.

$E()$ . Operador esperanza.

nr. Número de rechazos de la hipótesis  $H_0: \beta_1=0$ .

i. Elemento contador sobre los elementos del vector T.

t. Vector de elementos 0 o 1 que indica las transiciones de un estado de la cadena a otro.