

## Observaciones atípicas en regresiones de corte transversal(\*)

por  
JESÚS MUR

Departamento de Análisis Económico  
Universidad de Zaragoza.

### RESUMEN

Existe una abundante literatura dedicada a la identificación y tratamiento de las observaciones atípicas, de los *outliers*, en un contexto de series temporales. Sin embargo, esta experiencia no puede extrapolarse directamente al caso espacial debido a las peculiaridades propias de la última dimensión. En el trabajo pretendemos avanzar en esa dirección, centrandó la discusión en las consecuencias de las observaciones atípicas sobre los contrastes de especificación más habituales en regresiones de corte transversal. Presentamos en primer lugar una serie de resultados analíticos cuya finalidad es intentar delimitar el impacto experimentado por los estadísticos de contraste. Completamos el análisis con un ejercicio de Monte Carlo diseñado para medir con mayor precisión los efectos de las anomalías existentes en la información muestral.

---

(\*) Este trabajo ha sido posible gracias al apoyo financiero brindado por el proyecto SEC 2002-02350 del Ministerio de Ciencia y Tecnología del Reino de España. El autor agradece igualmente la inestimable colaboración de Ana Angulo.

*Palabras clave:* atípicos, dependencia espacial, Monte Carlo

*Clasificación AMS:* 62P20, 90A19

## 1. INTRODUCCIÓN

Las series empleadas en trabajos aplicados son proclives a la generación de observaciones *extrañas*. Dixon (1950, p. 488) se refiere a observaciones que son '*dudosas a los ojos del analista*'; '*disidentes*', '*discordantes*' o '*sorprendentes*' son términos usados por Beckman y Cook (1983, p.120). Shekhar et al. (2002, p. 452) previenen sobre las consecuencias de '*observaciones que parecen ser inconsistentes con el resto*'. Hawkins (1980, p.1) es un poco más específico cuando alude a observaciones que se '*desvían tanto de las otras observaciones como para arrojar sospechas de que han venido generadas por un mecanismo diferente*'. Se trata, en definitiva, de las observaciones que Peña denomina *atípicas*(1) (Peña, 2001).

Ante su presencia, la reacción del analista suele combinar inquietud e incertidumbre. Inquietud porque esas anomalías revelan cierta debilidad en los datos la cual puede acabar erosionando la confianza en los resultados finales. También incertidumbre porque la causa que subyace a tales anomalías suele permanecer oculta. Anscombe (1960) cataloga esas observaciones en atribuibles a errores cometidos en la codificación de la información y en las asociadas a la variabilidad inherente al proceso estadístico que genera los datos. La medida que se impone en el primer caso es simple: revisar la información muestral para depurar los errores existentes. Si el atípico pertenece a la segunda categoría, la decisión ya no está tan clara (Evans, 1999): revisar el modelo, eliminar las observaciones afectadas, filtrar el conjunto de observaciones de acuerdo a algún criterio, transformar las variables, etc. son algunas de las sugeridas más a menudo.

Profundizando en la definición propuesta por Hawkins, deberíamos añadir un tercer caso concretado en la existencia de una ruptura estructural, de carácter parcial o local, que afecta al proceso generador de un número reducido de indivi-

---

(1) Durante el proceso de revisión he mantenido una cierta polémica con uno de los evaluadores del trabajo, más partidario del término '*outlier*'. El evaluador entendía que el término '*outlier*', al que da carta de naturaleza Anscombe (1960), es más general y se adecua mejor a casos, como el que se desarrolla en el trabajo, en los que no es posible discernir a priori la naturaleza o el motivo de la anomalía. En este sentido, el término '*atípico*' estaría más vinculado al hecho final de ser un punto '*extremo*' con respecto al resto. Sin embargo y aún reconociendo la sensatez de estas observaciones, en este trabajo el término '*atípico*' se emplea como sinónimo del término inglés '*outlier*'.

duos. Las observaciones producidas en tal caso emergerán como atípicas con respecto a las restantes aunque seguirán siendo legítimas, en el sentido de Barnett y Lewis (1994). Sin embargo, su mantenimiento en la muestra, si no se controlan de alguna forma, producirá efectos contaminantes muy perjudiciales para la estimación y el contraste del modelo.

La literatura especializada en el análisis de series temporales ha avanzado mucho en la discusión sobre el impacto y las consecuencias de los '*outliers*' con resultados muy interesantes. Los trabajos de Chang et al. (1988), Tsay (1988), Perron (1989) o Peña (1990) son sólo algunos de los pioneros en esta línea. Las referencias son numerosas para el caso multivariante más general; entre los más relevantes pueden citarse, por ejemplo, los trabajos de Rao (1973), Tukey (1977), Huber (1981), Sachs (1982), Rousseeuw y Leroy (1987) o Belsley (1991).

La situación existente en un ámbito espacial es peculiar. Por un lado, los datos de corte transversal son propensos a la generación de observaciones anómalas, de manera que términos como heterogeneidad, inestabilidad, irregularidad, etc. se emplean frecuentemente en este tipo de aplicaciones econométricas. Sin embargo, las referencias dedicadas al binomio '*Espacio-Observaciones Atípicas*' son escasas (la lista se reduce, básicamente, a los trabajos de Wartenberg, 1989, Martin, 1992, y Haining, 1994 y 1995; ver también Cressie, 1993). El presente trabajo quiere contribuir a paliar ese déficit, centrando la discusión en el caso específico de un modelo econométrico de corte transversal en el que alguna observación presenta rasgos peculiares. El enfoque que adoptamos coincide con el descrito por Hawkins (1981): las observaciones afectadas son legítimas pero no totalmente homologables con las restantes al existir una ruptura estructural local inadvertida por el analista.

Nuestra preocupación, en particular, se centra en la sensibilidad, en la robustez, de los contrastes de especificación habituales en este contexto, singularmente los dedicados a analizar la existencia de dependencia espacial. Para ello, en la sección segunda ofrecemos algunos resultados analíticos relativos a la batería de contrastes seleccionada. En la sección tercera se resuelve un ejercicio de simulación para intentar concretar mejor la sensibilidad de esos contrastes a la presencia de observaciones atípicas en la muestra. El artículo finaliza con una sección de conclusiones.

## **2. OBSERVACIONES ATÍPICAS EN MODELOS ECONOMÉTRICOS DE CORTE TRANSVERSAL**

Es evidente que la presencia de observaciones atípicas, de '*outliers*', en una muestra va a tener consecuencias indeseadas sobre los instrumentos de inferencia y contraste utilizados en análisis de tipo aplicado. Las referencias incluidas en la

sección de Introducción permiten sostener tal afirmación con rotundidad. Asumiendo ese efecto pernicioso, en la presente sección vamos a tratar de delimitar el impacto que ese tipo de observaciones tiene sobre los estadísticos de especificación empleados más habitualmente en análisis econométricos de corte transversal.

En concreto, nos referimos a los estadísticos I de Moran, LM-ERR, LM-EL y KR, relativos al análisis de dependencia espacial en la perturbación, junto al LM-LAG y al LM-LE cuyo objetivo se centra en analizar la estructura dinámica de la ecuación. A los anteriores añadimos el contraste SARMA cuya hipótesis nula es conjunta: estructura estática en la ecuación y un término de error ruido blanco. Es importante reseñar, con respecto al objetivo de este trabajo, que los siete contrastes se construyen en torno a los residuos de la estimación MCO (ver Apéndice 1). Dado que estos residuos responden a la presencia de atípicos, la misma sensibilidad tiene que trasladarse, en todo o en parte, a los estadísticos mencionados.

En el apartado 2.1 se examinan las distorsiones sufridas por los estimadores MCO obtenidos en presencia de observaciones atípicas y bajo diferentes regímenes de dependencia espacial. En el apartado 2.2 centramos la atención específicamente en la reacción de los contrastes de especificación comentados ante esas mismas anomalías.

## 2.1 Estimación MCO en modelos con dependencia espacial y observaciones atípicas

En términos generales podemos afirmar que el impacto de este tipo de puntos depende de las propias dimensiones de la anomalía a lo cual debe añadirse, en nuestro caso, la localización de la región contaminada y la estructura de dependencias existente en los datos. Para apreciarlo con más claridad, podemos utilizar el siguiente modelo, que incorpora un atípico en la observación  $r$ -ésima:

$$\left. \begin{aligned} y &= X\beta + \pi d + u \\ d_s &= \begin{cases} 0 & s \neq r \\ 1 & s = r \end{cases} = u \sim N(0; \sigma_u^2 I) \end{aligned} \right\} \quad [1]$$

Disponemos de  $R$  observaciones, en una muestra de corte transversal, tomada cada una de ellas en una región distinta del espacio. En consecuencia, el vector de observaciones de la variable endógena es de orden  $(R \times 1)$ , mientras que la matriz  $X$  de observaciones de las  $k$  variables explicativas es de orden  $(R \times k)$ ;  $\beta$  es un vector de parámetros de posición de orden  $(k \times 1)$  y  $u$  el vector  $(R \times 1)$  de perturbaciones aleatorias. Las  $(R-1)$  observaciones que podemos denominar '*típicas*' se obtienen como:  $y_s = x_s' \beta + u_s$ ;  $s \neq r$ , mientras que la procedente de la región  $r$ -ésima

sufre un desplazamiento por valor de  $\pi$  en la media:  $y_r = x_r'\beta + \pi + u_r$ . Supondremos que el vector de términos aleatorios,  $u$ , proviene de una distribución normal con vector de medias nulo y matriz de varianzas y covarianzas escalar. Inicialmente asumimos la hipótesis de independencia, que más adelante relajaremos.

Salvo en circunstancias especiales, el usuario no conocerá la peculiaridad asociada a la observación  $r$ -ésima por lo que desarrollará un modelo econométrico estándar,  $y=X\beta+v$ . La consecuencia será que la ecuación estará mal especificada, al haberse omitido una variable explicativa (la ficticia  $d$  asociada a la anomalía de la observación  $r$ -ésima). Este error se trasladará a la parte no controlada de la ecuación,  $v=u+\pi d$ , en cuya estructura intervendrá el término no incluido en la parte sistemática junto al componente aleatorio habitual,  $u$ . Las consecuencias de esta situación son bien conocidas (Aznar, 1988): los estimadores MCO son sesgados y la distribución de los residuos deja de estar centrada en el origen, aunque mantienen la matriz de covarianzas habitual. Análiticamente:

$$y = X\beta + \pi d + u = X\beta + v \quad \left\{ \begin{array}{l} E[v] = \pi d \\ V[v] = \sigma_u^2 I \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \hat{\beta} = [X'X]^{-1}X'y \rightarrow \left\{ \begin{array}{l} E[\hat{\beta}] = \beta + \pi[X'X]^{-1}X'd \\ V[\hat{\beta}] = \sigma_u^2[X'X]^{-1} \end{array} \right. \\ \hat{v} = y - X\hat{\beta} = Mv \rightarrow \left\{ \begin{array}{l} E[\hat{v}] = \pi Md = m_r \\ V[\hat{v}] = \sigma_u^2 M \end{array} \right. \end{array} \right. \quad [2]$$

siendo  $M$  la matriz  $[I - X(X'X)^{-1}X']$ . El vector  $m_r = Md$  se corresponde con la columna  $r$ -ésima de  $M$ .

Los resultados anteriores se han obtenido bajo el supuesto de que no existían mecanismos de interacción espacial entre las unidades de observación. A continuación vamos a introducir efectos espaciales en el modelo, manteniendo la anomalía en el proceso de generación de la observación  $r$ -ésima. En particular, vamos a contemplar dos posibles estructuras de dependencia transversal conocidas en la literatura como de *dependencia residual* (existen relaciones de tipo autoregresivo, SAR, o media móvil, SMA, pero solo en la perturbación) o *sustantiva* (los mecanismos de dependencia se producen en la ecuación principal del modelo).

En el caso de dependencia residual con estructura SAR, el modelo de referencia es:

$$\left. \begin{array}{l} y = X\beta + v \\ v = \rho Wv + u \\ u = \varepsilon + \pi d; \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} y = X\beta + v \\ v = [I - \rho W]^{-1}u = v^* + \pi d_{SAR}^* \\ v^* = [I - \rho W]^{-1}\varepsilon; d_{SAR}^* = [I - \rho W]^{-1}d \end{array} \right. \quad [3]$$

siendo  $\rho$  el parámetro del proceso SAR y  $W$  la denominada *matriz de contactos*, o *de pesos espaciales* que asumimos de tipo binario (Anselin, 1988). A diferencia de

lo que ocurre en el caso de [2], ahora el desplazamiento en la media afectará a todos los términos de error en una cuantía variable ( $\pi d_{SAR}^*$ , correspondiente a la  $r$ -ésima columna de la matriz  $[I - \rho W]^{-1}$  multiplicada por  $\pi$ ). El error de especificación, consecuencia de la anomalía existente en la observación  $r$ , sigue produciendo unos estimadores MCO sesgados:

$$\hat{\beta} = [X'X]^{-1}X'y \Rightarrow E[\hat{\beta}] = \beta + \pi[X'X]^{-1}X'd_{SAR}^* \quad [4]$$

Los residuos MCO tampoco están centrados en el origen. Además, la estructura de dependencia SAR en la perturbación afecta a la matriz de covarianzas del vector de residuos:

$$\hat{v} = y - X\hat{\beta} = Mv \rightarrow \begin{cases} E[\hat{v}] = \pi M d_{SAR}^* \\ V[\hat{v}] = \sigma_\varepsilon^2 M [I - \rho W]^{-2} M \end{cases} \quad [5]$$

Si en la perturbación domina una estructura media móvil (hacemos  $v = u - \theta Wu$  en la segunda ecuación de la llave izquierda de la expresión [3]), los estimadores MCO seguirán siendo sesgados y el vector de residuos también sufre un desplazamiento:

$$\hat{\beta} = [X'X]^{-1}X'y \Rightarrow E[\hat{\beta}] = \beta + \pi[X'X]^{-1}X'd_{SMA}^* \quad [6]$$

$$\hat{v} = y - X\hat{\beta} = Mv \rightarrow \begin{cases} E[\hat{v}] = \pi M d_{SMA}^* \\ V[\hat{v}] = \sigma_\varepsilon^2 M [I - \theta W]^2 M \end{cases} \quad [7]$$

Siendo  $d_{SMA}^* = [I - \theta W]d$ . La peculiaridad es que, ahora, la incidencia del atípico se concentra en los puntos localizados en las inmediaciones del contaminado. Por último, en el caso de autocorrelación espacial sustantiva el impacto del atípico alcanzará nuevamente a todas las observaciones:

$$\left. \begin{array}{l} y = \rho Wy + X\beta + v \\ v = \varepsilon + \pi d; \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \end{array} \right\} \Rightarrow \begin{cases} y = [I - \rho W]^{-1}[X\beta + \pi d] + [I - \rho W]^{-1}\varepsilon = \\ [I - \rho W]^{-1}X\beta + \pi d_{SAR}^* + [I - \rho W]^{-1}\varepsilon \end{cases} \quad [8]$$

El sesgo de los estimadores MCO es relevante y los ajustes en la distribución de los residuos también son de importancia:

$$E[\hat{\beta}] = [X'X]^{-1}X'[I - \rho W]^{-1}X\beta + \pi d_{SAR}^* \quad [9]$$

$$\hat{v} = y - X\hat{\beta} = My \rightarrow \begin{cases} E[\hat{v}] = M [I - \rho W]^{-1} X\beta + \pi M d_{SAR}^* \\ V[\hat{v}] = \sigma_\epsilon^2 M [I - \rho W]^2 M \end{cases} \quad [10]$$

## 2.2 Contrastes de dependencia espacial y observaciones atípicas

El contraste I de Moran es un contraste simple y muy utilizado en este tipo de aplicaciones donde ha adquirido gran protagonismo gracias a su fiabilidad (Anselin y Florax, 1995). Como indican Florax y de Graaff (2004, p. 35) se trata de un contraste '*contra una estructura de dependencia espacial lineal y aditiva*' en el término de perturbación. Sin embargo, la presencia de anomalías como las que estamos considerando provoca desajustes en el comportamiento del estadístico. Tal como se demuestra en el Apéndice 2, el valor esperado de la I de Moran, bajo la hipótesis nula de incorrelación en la perturbación e introduciendo una anomalía como la de [1], es(2):

$$E[I]_{H_0} = \left[ \frac{R}{S_0} \right] \left[ \frac{\sigma_u^2 \text{tr} MW + \pi^2 m_r' W m_r}{\sigma_u^2 (R - k) + \pi^2 (1 - p_{rr})} \right] \quad [11]$$

siendo  $S_0$  la suma de todos los elementos de la matriz de contactos  $W$ ,

$S_0 = \sum_{r=1}^R \sum_{s=1}^R w_{rs}$ , con  $w_{rs}$  el correspondiente elemento de  $W$ ;  $p_{rr}$  indica el elemento  $r$ -ésimo de la diagonal principal de la matriz  $P = X(X'X)^{-1}X'$ , el cual mide el grado de apalancamiento de la observación asociada. Este elemento se encuentra acotado,  $0 \leq p_{rr} \leq 1$ , por lo que  $\pi^2(1 - p_{rr}) \geq 0$ . Sin embargo, el impacto en el numerador es incierto ya que la forma cuadrática  $m_r' W m_r$  es indefinida. Las consecuencias sobre los momentos de segundo orden son difíciles de precisar en estos momentos.

Los contrastes basados en el principio del Multiplicador de Lagrange (Davidson, 1994) también sufren ajustes. Por ejemplo, la estimación máximo-verosímil (MV en adelante) del modelo de [3], introduciendo un error de falta de identificación de un punto atípico en la observación  $r$ -ésima, conduce al vector gradiente (3):

---

(2) Si el atípico no existe,  $\pi=0$ , el valor esperado se reduce a  $E[I]_{H_0} = \left( \frac{R}{S_0} \right) \left( \frac{\text{tr} MW}{R - k} \right)$ .

(3) Los detalles relativos a la estimación MV de los diferentes modelos contemplados en esta sección pueden consultarse en Anselin (1988).

$$g(\gamma) = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \rho} \\ \frac{\partial l}{\partial \sigma_\varepsilon^2} \end{bmatrix} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} X'B'(\pi d + \varepsilon) \\ (\pi d + \varepsilon)'B^{-1}W(\pi d + \varepsilon) - \text{tr}WB^{-1} \\ -\frac{R}{2} + \frac{(\pi d + \varepsilon)'(\pi d + \varepsilon)}{2\sigma_\varepsilon^2} \end{bmatrix} \quad [12]$$

siendo  $l$  el logaritmo de la función de verosimilitud del vector  $y$ ,  $\gamma$  el vector  $[\beta, \rho, \sigma^2]'$  y  $B$  la matriz  $[I - \rho W]$ . Su valor esperado, bajo la hipótesis nula de incorrelación ( $H_0: \rho=0$ ), no será cero:

$$E[g(\gamma)]_{H_0: \rho=0} = \frac{\pi}{\sigma_\varepsilon^2} \begin{bmatrix} X'd \\ 0 \\ \pi \\ 2\sigma_\varepsilon^2 \end{bmatrix} = \frac{\pi}{\sigma_\varepsilon^2} \begin{bmatrix} X'_r \\ 0 \\ \pi \\ 2\sigma_\varepsilon^2 \end{bmatrix} \quad [13]$$

donde  $x_r$  es el vector  $(1 \times k)$  de observaciones de las variables explicativas correspondientes al punto afectado. La matriz de información muestral, nuevamente bajo el supuesto de incorrelación, deja de ser diagonal por bloques:

$$H(\gamma) = -E \left[ \frac{\partial^2 l}{\partial \gamma \partial \gamma'} \right]_{H_0: \rho=0} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} (X'X) & 2\pi X'Wd & \frac{\pi}{\sigma_\varepsilon^2} X'_r \\ 2\pi d'W'X & 2\sigma_\varepsilon^2 S_0 + \pi^2 s_r & 0 \\ \frac{\pi}{\sigma_\varepsilon^2} X_r & 0 & \frac{R}{\sigma_\varepsilon^2} \left( 1 + \frac{2\pi^2}{R\sigma_\varepsilon^2} \right) \end{bmatrix} \quad [14]$$

siendo  $s_r$  el elemento  $r$ -ésimo de la diagonal principal de  $W'W$  (si esta matriz es de tipo binario,  $s_r$  se corresponde con el número de contratos de la región  $r$ -ésima; ver Cliff y Ord, 1981).

Sin embargo, el usuario, al desconocer la incidencia que afecta a la observación en cuestión, actuará de la forma acostumbrada. En concreto, para contrastar la hipótesis nula de que la perturbación es un ruido blanco ( $H_0: \rho=0$ ) frente a la alternativa de dependencia espacial ( $H_A: \rho \neq 0$ ), el contraste LM-ERR que se especificará será el habitual(4) (Anselin y Florax, 1995):

(4) El superíndice A indica que se trata de una aproximación al estadístico, o matriz, relevante en cada caso.

$$LM - ERR^A = [g(\hat{\gamma})]' [H^A(\hat{\gamma})]^{-1} [g(\hat{\gamma})] \quad [15]$$

El Multiplicador de Lagrange que debería haberse utilizado es el siguiente(5):

$$LM - ERR = [g(\hat{\gamma}) - Eg(\hat{\gamma})]' [H(\hat{\gamma})]^{-1} [g(\hat{\gamma}) - Eg(\hat{\gamma})]_{as} \sim \chi^2(1) \quad [16]$$

donde 'as' significa '*asintóticamente*' (esto es, que el estadístico de la parte izquierda alcanza la distribución que se especifica en la derecha sólo en un contexto asintótico). Los estadísticos de [15] y de [16] pueden relacionarse al descomponer la matriz de información de [14] como:

$$H(\gamma) = H^A(\gamma) + H^B(\gamma) = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} (X'X) & 0 & 0 \\ 0 & 2\sigma_\varepsilon^2 S_0 & 0 \\ 0 & 0 & \frac{R}{2\sigma_\varepsilon^2} \end{bmatrix} + \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} 0 & 2\pi X'Wd & \frac{\pi}{\sigma_\varepsilon^2} x_r' \\ 2\pi d'W'X & \pi^2 s_r & 0 \\ \frac{\pi}{\sigma_\varepsilon^2} x_r & 0 & \frac{\pi^2}{\sigma_\varepsilon^4} \end{bmatrix} \quad [17]$$

Introduciendo la última expresión en el Multiplicador de [16] se obtiene:

$$\begin{aligned} LM - ERR &= [g(\hat{\gamma}) - Eg(\hat{\gamma})]' [H(\hat{\gamma})]^{-1} [g(\hat{\gamma}) - Eg(\hat{\gamma})] = \\ &= [g(\hat{\gamma}) - Eg(\hat{\gamma})]' \left\{ [H^A(\hat{\gamma})]^{-1} - H^*(\hat{\gamma}) \right\} [g(\hat{\gamma}) - Eg(\hat{\gamma})] = \\ &= g(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} g(\hat{\gamma}) + Eg(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} Eg(\hat{\gamma}) - 2g(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} Eg(\hat{\gamma}) \\ &\quad - [g(\hat{\gamma}) - Eg(\hat{\gamma})]' H^*(\hat{\gamma}) [g(\hat{\gamma}) - Eg(\hat{\gamma})] = \\ &= g(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} g(\hat{\gamma}) + Imerr_1 - 2Imerr_2 - Imerr_3 \end{aligned} \quad [18]$$

Esto es:

$$LM - ERR^A = g(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} g(\hat{\gamma}) = LM - ERR - Imerr_1 + 2Imerr_2 + Imerr_3 \quad [19]$$

---

(5) Anselin et al. (1996) se enfrentan a una situación similar en la que, como en nuestro caso, es necesario centrar el vector gradiente para obtener un Multiplicador de Lagrange con buenas propiedades. La discusión en esa referencia conduce a los denominados Multiplicadores de Lagrange *robustos* a errores locales de especificación en la hipótesis nula denotados, en nuestro trabajo, como LM-EL y LM-LE.

La matriz  $H^*(\hat{\gamma})$  recoge los errores cometidos en la especificación de la matriz de información muestral,

$$H^*(\hat{\gamma}) = [H^A(\hat{\gamma})]^{-1} \left\{ [H^A(\hat{\gamma})]^{-1} + [H^B(\hat{\gamma})]^{-1} \right\}^{-1} [H^A(\hat{\gamma})]^{-1}.$$

Los términos  $Imerr_j$ , para  $j = 1, 2$  y  $3$ , son tres formas cuadráticas,

$$Imerr_1 = \text{Eg}(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} \text{Eg}(\hat{\gamma}), \quad Imerr_2 = \text{g}(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} \text{Eg}(\hat{\gamma}) \text{ y}$$

$Imerr_3 = [\text{g}(\hat{\gamma})' - \text{Eg}(\hat{\gamma})'] H^*(\hat{\gamma}) [\text{g}(\hat{\gamma}) - \text{Eg}(\hat{\gamma})]$ , cuyo comportamiento se puede caracterizar en cierta medida. El término  $Imerr_1$  converge a cero en probabilidad cuando el tamaño muestral tiende a infinito. Expresado de otra forma, si obtenemos el límite probabilístico (que denotaremos por  $p\lim$ ) de este término:

$$\begin{aligned} p\lim_{R \rightarrow \infty} Imerr_1 &= p\lim_{R \rightarrow \infty} \{ \text{Eg}(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} \text{Eg}(\hat{\gamma}) \} = \\ p\lim_{R \rightarrow \infty} \left\{ \left( \frac{\text{Eg}(\hat{\gamma})}{\sqrt{R}} \right)' \left[ \frac{H^A(\hat{\gamma})}{R} \right]^{-1} \left( \frac{\text{Eg}(\hat{\gamma})}{\sqrt{R}} \right) \right\} &= \quad [20] \\ p\lim_{R \rightarrow \infty} \left\{ \left( \frac{\text{Eg}(\hat{\gamma})}{\sqrt{R}} \right)' [I_{AS}^A(\gamma)]^{-1} \left( \frac{\text{Eg}(\hat{\gamma})}{\sqrt{R}} \right) \right\} &= 0 \end{aligned}$$

donde  $I_{AS}^A(\gamma)$  es la matriz de convergencia asintótica de  $H^A(\gamma)$ :

$$\begin{aligned} I_{AS}^A(\gamma) = p\lim_{R \rightarrow \infty} \frac{H^A(\gamma)}{R} &= \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} p\lim_{R \rightarrow \infty} \frac{(X'X)}{R} & 0 & 0 \\ 0 & 2\sigma_\varepsilon^2 p\lim_{R \rightarrow \infty} \frac{S_0}{R} & 0 \\ 0 & 0 & \frac{1}{2\sigma_\varepsilon^2} \end{bmatrix} = \\ \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} \sum_{xx} & 0 & 0 \\ 0 & 2\sigma_\varepsilon^2 \omega & 0 \\ 0 & 0 & \frac{1}{2\sigma_\varepsilon^2} \end{bmatrix} & \quad [21] \end{aligned}$$

En esta expresión,  $\omega$  debe interpretarse como el número promedio de contactos alcanzado en el sistema de regiones a medida que aumentamos el tamaño de la muestra (el número total de contactos se corresponde con  $S_0$ ). El término  $lmerr_2$  converge a una constante positiva:

$$\begin{aligned} \text{plim}_{R \rightarrow \infty} \text{limerr}_2 &= \text{plim}_{R \rightarrow \infty} \left\{ \mathbf{g}(\hat{\gamma})' \left[ \mathbf{H}^A(\hat{\gamma}) \right]^{-1} \mathbf{Eg}(\hat{\gamma}) \right\} = \\ \text{plim}_{R \rightarrow \infty} \left\{ \left( \frac{\mathbf{g}(\hat{\gamma})}{R} \right)' \left[ \frac{\mathbf{H}^A(\hat{\gamma})}{R} \right]^{-1} \mathbf{Eg}(\hat{\gamma}) \right\} &= \left\{ \mathbf{Eg}(\hat{\gamma})' \left[ \mathbf{I}_{AS}^A(\hat{\gamma}) \right]^{-1} \mathbf{Eg}(\hat{\gamma}) \right\} = \end{aligned} \quad [22]$$

$$\left( \frac{\pi}{\sigma_\varepsilon} \right)^2 \left[ \mathbf{x}_r \sum_{XX}^{-1} \mathbf{x}_r' + \frac{\pi^2}{2\sigma_\varepsilon^2} \right] = \left( \frac{\pi}{\sigma_\varepsilon} \right)^2 \left[ \mathbf{P}_{rr} + \frac{\pi^2}{2\sigma_\varepsilon^2} \right]$$

siendo  $\mathbf{P}_{rr} = \mathbf{x}_r \sum_{XX}^{-1} \mathbf{x}_r'$  una medida de apalancamiento de la observación  $r$ -ésima con respecto a la matriz de convergencia de los momentos muestrales correspondiente a los regresores del modelo,  $\Sigma_{XX}$ . Finalmente, el término  $lmerr_3$  es una forma cuadrática de un vector aleatorio,  $[\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]$ , sobre una matriz genérica,  $\mathbf{H}^*(\hat{\gamma})$ . Es importante indicar que ese vector cumple un Teorema Central del Límite bien definido:  $\sqrt{R} [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]_{as} \rightarrow N \left[ \mathbf{0}; \mathbf{I}_{AS}^A(\hat{\gamma})^{-1} \right]$ , siendo  $\mathbf{I}_{AS}^A(\hat{\gamma})$  la matriz de información asintótica del modelo de [3]. En consecuencia, la distribución de probabilidad de esa forma cuadrática se corresponde con la de una suma de variables *chi-cuadrado* mutuamente independientes, con pesos  $\lambda_j$  (Johnson y Kotz, capítulos 28 y 29, 1970):

$$\text{lmerr}_3 = [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]' \mathbf{H}^*(\hat{\gamma}) [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})] = \sum_{j=1}^R \lambda_j z_j^2 \quad [23]$$

Las variables aleatorias  $z_j$  se distribuyen de acuerdo a una ley normal  $N(0,1)$  y los elementos  $\{\lambda_j, j=1, 2, \dots, R\}$  son las raíces características de la matriz  $\mathbf{L}(\hat{\gamma})' \mathbf{H}^*(\hat{\gamma}) \mathbf{L}(\hat{\gamma})$ , siendo  $\mathbf{L}(\hat{\gamma})$  la matriz que factoriza la matriz de información de [14]:  $\mathbf{H}(\hat{\gamma}) = \mathbf{L}(\hat{\gamma}) \mathbf{L}(\hat{\gamma})'$  (Kendall y Stuart, 1977).

Para analizar las consecuencias sobre el estadístico LM-LAG es necesario utilizar como referencia el modelo de [8]. El vector gradiente correspondiente a la última especificación es:

$$g(\gamma) = \frac{\partial l}{\partial \gamma} \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \rho} \\ \frac{\partial l}{\partial \sigma_\varepsilon^2} \end{bmatrix} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} X'(\pi d + \varepsilon) \\ y' W'(\pi d + \varepsilon) - \text{tr}WB^{-1} \\ -\frac{R}{2} + \frac{(\pi d + \varepsilon)'(\pi d + \varepsilon)}{2\sigma_\varepsilon^2} \end{bmatrix} \quad [24]$$

Debe recordarse que ahora  $\rho$  es el parámetro que acompaña al retardo de la variable endógena en el modelo de [8] y  $B$  es la matriz  $[I - \rho W]$ . El valor esperado del gradiente de [24], bajo el supuesto de ausencia de efectos espaciales en la ecuación ( $H_0: \rho=0$ ), también es diferente de cero:

$$E[g(\gamma)]_{H_0: \rho=0} = \frac{\pi}{\sigma_\varepsilon^2} \begin{bmatrix} x_r' \\ \beta' X' W d \\ \frac{\pi}{2\sigma_\varepsilon^2} \end{bmatrix} \quad [25]$$

La matriz de información muestral, bajo la hipótesis nula, es de tipo general:

$$H(\gamma) = -E \left[ \frac{\partial^2 l}{\partial \gamma \partial \gamma'} \right]_{H_0: \rho=0} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} (X'X) & X'W(X\beta + \pi d) & \frac{\pi}{\sigma_\varepsilon^2} x_r' \\ (X\beta + \pi d)' W' X & \left\{ \beta' X' W^2 (X\beta + 2\pi d) \right. \\ & \left. + \pi^2 s_r + 2\sigma_\varepsilon^2 S_0 \right\} & \frac{\pi}{\sigma_\varepsilon^2} \beta' X' W d \\ \frac{\pi}{\sigma_\varepsilon^2} x_r' & \frac{\pi}{\sigma_\varepsilon^2} d' W' X \beta & \frac{R}{2\sigma_\varepsilon^2} \left( 1 + \frac{2\pi^2}{R\sigma_\varepsilon^2} \right) \end{bmatrix} \quad [26]$$

siendo  $W^2 = W'W$ . Esta matriz admite una descomposición similar a la obtenida en el caso del LM-ERR de modo que:

$$H(\gamma) = H^A(\gamma) + H^B(\gamma) =$$

$$\frac{1}{\sigma_\varepsilon^2} \left\{ \begin{bmatrix} (X'X) & X'WX\beta & 0 \\ \beta' X' W' X & \left\{ \beta' X' W^2 X\beta + \right. \\ & \left. 2\sigma_\varepsilon^2 S_0 \right\} & 0 \\ 0 & 0 & \frac{R}{2\sigma_\varepsilon^2} \end{bmatrix} + \begin{bmatrix} 0 & \pi X' W d & \frac{\pi}{\sigma_\varepsilon^2} x_r' \\ \pi d' W' X & \left\{ 2\pi \beta' X' W^2 d + \right. \\ & \left. \pi^2 s_r \right\} & \frac{\pi}{\sigma_\varepsilon^2} \beta' X' W d \\ \frac{\pi}{\sigma_\varepsilon^2} x_r' & \frac{\pi}{\sigma_\varepsilon^2} d' W' X \beta & \frac{\pi}{\sigma_\varepsilon^2} \end{bmatrix} \right\} \quad [27]$$

El último resultado permite desarrollar el estadístico LM-LAG en una suma de cuatro términos:

$$\begin{aligned}
 \text{LM-LAG} &= [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]' [\mathbf{H}(\hat{\gamma})]^{-1} [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})] = \\
 &\mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{g}(\hat{\gamma}) + \mathbf{Eg}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \\
 &\mathbf{Eg}(\hat{\gamma}) - 2\mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{Eg}(\hat{\gamma}) - [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]' \mathbf{H}^*(\hat{\gamma}) [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})] = \quad [28] \\
 &\mathbf{g}(\hat{\gamma})' (\mathbf{H}^A(\hat{\gamma}))^{-1} \mathbf{g}(\hat{\gamma}) + \text{lmlag}_1 - 2\text{lmlag}_2 - \text{lmlag}_3
 \end{aligned}$$

La matriz  $\mathbf{H}^*(\gamma)$  mantiene, formalmente, la misma estructura que en el caso anterior, basta con actualizar el contenido de las matrices de base  $[\mathbf{H}^A(\hat{\gamma})]$  y  $[\mathbf{H}^B(\hat{\gamma})]$ , por lo que podemos escribir:

$$\text{LM-LAG}^A = \mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{g}(\hat{\gamma}) = \text{LM-LAG} - \text{lmlag}_1 + 2\text{lmlag}_2 + \text{lmlag}_3 \quad [29]$$

El término  $\text{lmlag}_1$  converge, con  $R$ , a cero:

$$\begin{aligned}
 \text{p lim}_{R \rightarrow \infty} \text{lmlag}_1 &= \text{p lim}_{R \rightarrow \infty} \{ \mathbf{Eg}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{Eg}(\hat{\gamma}) \} = \\
 \text{p lim}_{R \rightarrow \infty} &\left\{ \left( \frac{\mathbf{Eg}(\hat{\gamma})}{\sqrt{R}} \right)' \left[ \frac{\mathbf{H}^A(\hat{\gamma})}{R} \right]^{-1} \left( \frac{\mathbf{Eg}(\hat{\gamma})}{\sqrt{R}} \right) \right\} = \quad [30] \\
 \text{p lim}_{R \rightarrow \infty} &\left\{ \left( \frac{\mathbf{Eg}(\hat{\gamma})}{\sqrt{R}} \right)' [\mathbf{I}_{AS}^A(\hat{\gamma})]^{-1} \left( \frac{\mathbf{Eg}(\hat{\gamma})}{\sqrt{R}} \right) \right\} = \mathbf{0}
 \end{aligned}$$

Nuevamente,  $\mathbf{I}_{AS}^A(\gamma)$  es la matriz de convergencia asintótica de la matriz  $\mathbf{H}^A(\gamma)$  de [27]:

$$I_{AS}^A(\gamma) = p \lim_{R \rightarrow \infty} \frac{H^A(\gamma)}{R} =$$

$$\frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} p \lim_{R \rightarrow \infty} \frac{(X'X)}{R} & p \lim_{R \rightarrow \infty} \frac{(X'WX)}{R} \beta & 0 \\ \beta' p \lim_{R \rightarrow \infty} \frac{(X'WX)}{R} & p \lim_{R \rightarrow \infty} \beta' \frac{(X'W^2X)}{R} \beta + 2\sigma_\varepsilon^2 p \lim_{R \rightarrow \infty} \frac{S_0}{R} & 0 \\ 0 & 0 & \frac{1}{2\sigma_\varepsilon^2} \end{bmatrix} = \quad [31]$$

$$\frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} \Sigma_{xx} & \Sigma_{xwx} \beta & 0 \\ \beta' \Sigma_{xwx} & \beta' \Sigma_{xw^2x} \beta + 2\sigma_\varepsilon^2 \omega & 0 \\ 0 & 0 & \frac{1}{2\sigma_\varepsilon^2} \end{bmatrix}$$

El término  $lmlag_2$  converge a una constante positiva:

$$p \lim_{R \rightarrow \infty} \{ g(\hat{\gamma})' [H^A(\hat{\gamma})]^{-1} Eg(\hat{\gamma}) \} =$$

$$p \lim_{R \rightarrow \infty} \left\{ \frac{\pi^2}{\sigma_\varepsilon^4} \left[ \sigma_\varepsilon^2 P_{rr} + V(\rho) [(d - p_r)' WX \beta]^2 + \frac{\pi^2}{2R} \right] \right\} = \quad [32]$$

$$\frac{\pi^2}{\sigma_\varepsilon^4} \left[ \sigma_\varepsilon^2 P_{rr} + V(\rho) [(d - p_r)' WX \beta]^2 \right]$$

siendo  $P_{rr}$  la misma medida de apalancamiento que interviene en [22],  $p_r$  es la columna  $r$ -ésima de la matriz  $P$  y  $V(\rho)$  es el límite de convergencia, bajo la hipótesis nula, de la varianza del estimador MV de  $\rho$ :

$$V(\rho) = p \lim_{R \rightarrow \infty} \left[ \left( \frac{\beta' X' W' M W X \beta}{\sigma_\varepsilon^2} \right) / R \right]^{-1} = \left[ \frac{1}{\sigma_\varepsilon^2} \beta' (\Sigma_{xw^2x} - \Sigma_{xwx} \Sigma_{xx}^{-1}) \beta \right]^{-1}$$

El comportamiento del término  $lmlag_3$  resulta más impreciso, aunque puede mantenerse (adaptado) el resultado de [23]:

$$lmlag_3 = [g(\hat{\gamma}) - Eg(\hat{\gamma})]' H^*(\hat{\gamma}) [g(\hat{\gamma}) - Eg(\hat{\gamma})] = \sum_{j=1}^R \lambda_j z_j^2 \quad [33]$$

Las variables  $z_j$  son variables normales unitarias  $N(0,1)$  y los pesos  $\{\lambda_j, j=1, 2, \dots, R\}$  son las raíces características de  $L(\gamma)'H^*(\gamma)L(\gamma)$ , siendo  $L(\gamma)$  la matriz que factoriza la matriz de información de la expresión de [26].

Para concluir esta discusión dedicada a los Multiplicadores de Lagrange resta por considerar el caso del contraste SARMA. La hipótesis nula es conjunta (la ecuación principal es estática y la perturbación es un ruido blanco), por lo que debe ampliarse el modelo de referencia:

$$\left. \begin{aligned} y &= \rho W y + X\beta + v \\ v &= \theta W v + u \\ u &= \varepsilon + \pi d; \varepsilon \sim \text{iid}N(0, \sigma_\varepsilon^2) \end{aligned} \right\} \quad [34]$$

Los resultados relativos al gradiente se obtienen de la forma usual:

$$\mathbf{g}(\gamma) = \frac{\partial l}{\partial \gamma} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \rho} \\ \frac{\partial l}{\partial \theta} \\ \frac{\partial l}{\partial \sigma^2} \end{bmatrix} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} X'D'(\pi d + \varepsilon) \\ y'WD'(\pi d + \varepsilon) - \text{tr}WB^{-1} \\ (\pi d + \varepsilon)'D^{-1}W(\pi d + \varepsilon) - \text{tr}WD^{-1} \\ -\frac{R}{2} + \frac{(\pi d + \varepsilon)'(\pi d + \varepsilon)}{2\sigma_\varepsilon^2} \end{bmatrix} \Rightarrow \quad [35]$$

$$E[\mathbf{g}(\gamma)]_{H_0: \rho=\theta=0} = \frac{\pi}{\sigma_\varepsilon^2} \begin{bmatrix} x_r' \\ \beta'X'Wd \\ 0 \\ \frac{\pi}{2\sigma_\varepsilon^2} \end{bmatrix}$$

siendo  $B=[I-\rho W]$  y  $D=[I-\theta W]$ . La presencia del atípico no tiene incidencia en el término del gradiente asociado al estimador MV de  $\theta$ , cuyo valor esperado es cero, aunque sí afecta a los restantes. La estructura de la matriz de información muestral es compleja, incluso bajo la hipótesis de no interacción espacial:

$$H(\gamma) = -E \left[ \frac{\partial^2 l}{\partial \gamma \partial \gamma'} \right]_{H_0: \rho=0} =$$

$$\frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} (X'X) & X'W(X\beta + \pi d) & 2\pi X'Wd & \frac{\pi}{\sigma_\varepsilon^2} X'_r \\ (X\beta + \pi d)'W'X & \left\{ \beta'X'W^2(X\beta + 2\pi d) + \pi^2 s_r + \sigma_\varepsilon^2 S_0 \right\} & \left\{ 2(\pi\beta'X'W^2d - \pi^2 s_r + \sigma_\varepsilon^2 S_0) \right\} & \frac{2\pi\beta'X'Wd}{\sigma_\varepsilon^2} \\ 2\pi d'W'X & \left\{ 2(\pi d'W^2X\beta - \pi^2 s_r + \sigma_\varepsilon^2 S_0) \right\} & 2\sigma_\varepsilon^2 S_0 + \pi^2 s_r & 0 \\ \frac{\pi}{\sigma_\varepsilon^2} X'_r & \frac{2\pi d'W'X\beta}{\sigma_\varepsilon^2} & 0 & \frac{R}{2\sigma_\varepsilon^2} \left( 1 + \frac{2\pi^2}{R\sigma_\varepsilon^2} \right) \end{bmatrix} \quad [36]$$

Esta matriz puede descomponerse en la suma de otras dos matrices, tal y como se ha hecho con anterioridad:

$$H(\gamma) = H^A(\gamma) + H^B(\gamma) =$$

$$\frac{1}{\sigma_\varepsilon^2} \left\{ \begin{bmatrix} X'X & X'WX\beta & 0 & 0 \\ B'X'W'X & \left\{ \beta'X'W^2X\beta + 2\sigma_\varepsilon^2 S_0 \right\} & 2\sigma_\varepsilon^2 S_0 & 0 \\ 0 & 2\sigma_\varepsilon^2 S_0 & 2\sigma_\varepsilon^2 S_0 & 0 \\ 0 & 0 & 0 & \frac{R}{2\sigma_\varepsilon^2} \end{bmatrix} + \begin{bmatrix} 0 & \pi X'Wd & 2\pi X'Wd & \frac{\pi}{\sigma_\varepsilon^2} X'_r \\ \pi d'W'X & \left\{ 2\pi\beta'X'W^2d + \pi^2 s_r \right\} & \left\{ 2\pi\beta'X'W^2d - \pi^2 s_r \right\} & \frac{2\pi\beta'X'Wd}{\sigma_\varepsilon^2} \\ 2\pi d'W'X & \left\{ 2\pi dW^2X\beta - 2\pi^2 s_r \right\} & \pi^2 s_r & 0 \\ \frac{\pi}{\sigma_\varepsilon^2} X'_r & \frac{2\pi d'W'X\beta}{\sigma_\varepsilon^2} & 0 & \frac{\pi^2}{\sigma_\varepsilon^4} \end{bmatrix} \right\} \quad [37]$$

Finalmente, el estadístico SARMA se resume en:

$$\begin{aligned}
\text{SARMA} &= [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]' [\mathbf{H}(\hat{\gamma})]^{-1} [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})] = \\
&\mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{g}(\hat{\gamma}) + \mathbf{Eg}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{Eg}(\hat{\gamma}) - 2\mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{Eg}(\hat{\gamma}) \\
&\quad - [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]' \mathbf{H}^*(\hat{\gamma}) [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})] = \\
&\mathbf{g}(\hat{\gamma})' (\mathbf{H}^A(\hat{\gamma}))^{-1} \mathbf{g}(\hat{\gamma}) + \text{sarma}_1 - 2\text{sarma}_2 - \text{sarma}_3
\end{aligned} \tag{38}$$

La estructura de la matriz  $\mathbf{H}^*(\hat{\gamma})$ , y la de las matrices constitutivas  $\mathbf{H}^A(\hat{\gamma})$  y  $\mathbf{H}^B(\hat{\gamma})$ , replica resultados conocidos. El estadístico SARMA erróneamente especificado al no haberse introducido la anomalía de la observación  $r$ -ésima puede expresarse como:

$$\text{SARMA}^A = \mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{g}(\hat{\gamma}) = \text{SARMA} - \text{sarma}_1 + 2\text{sarma}_2 + \text{sarma}_3 \tag{39}$$

El término  $\text{sarma}_1$  converge a cero a medida que el tamaño muestral crece mientras que el término  $\text{sarma}_2$  tiende al mismo límite de convergencia que el obtenido para  $\text{lmlag}_2$ :

$$\begin{aligned}
\text{p lim}_{R \rightarrow \infty} \text{sarma}_2 &= \text{p lim}_{R \rightarrow \infty} \text{lmlag}_2 = \text{p lim}_{R \rightarrow \infty} \left\{ \mathbf{g}(\hat{\gamma})' [\mathbf{H}^A(\hat{\gamma})]^{-1} \mathbf{Eg}(\hat{\gamma}) \right\} = \\
&\text{p lim}_{R \rightarrow \infty} \left\{ \frac{\pi^2}{\sigma_\varepsilon^4} \left[ \sigma_\varepsilon^2 \mathbf{P}_{rr} + \mathbf{V}(\rho) [(d - p_r)' \mathbf{W}\mathbf{X}\beta]^2 + \frac{\pi^2}{2R} \right] \right\} = \\
&\quad \frac{\pi^2}{\sigma_\varepsilon^4} \left[ \sigma_\varepsilon^2 \mathbf{P}_{rr} + \mathbf{V}(\rho) [(d - p_r)' \mathbf{W}\mathbf{X}\beta]^2 \right]
\end{aligned} \tag{40}$$

El término  $\text{sarma}_3$  vuelve a ser una forma cuadrática de un vector de normales estandarizadas sobre una matriz diagonal con elementos  $\{\lambda_j, j=1, 2, \dots, R\}$ , de modo que:

$$\text{sarma}_3 = [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})]' \mathbf{H}^*(\hat{\gamma}) [\mathbf{g}(\hat{\gamma}) - \mathbf{Eg}(\hat{\gamma})] = \sum_{j=1}^R \lambda_j z_j^2 \tag{41}$$

Estos pesos  $\{\lambda_j, j=1, 2, \dots, R\}$  coinciden con las raíces características de la matriz  $\mathbf{L}(\gamma)' \mathbf{H}^*(\gamma) \mathbf{L}(\gamma)$ , , siendo  $\mathbf{L}(\gamma)$  la matriz que factoriza la matriz de información de [36].

### 3. CONTRASTES DE DEPENDENCIA ESPACIAL Y OBSERVACIONES ATÍPICAS. UN EJERCICIO DE MONTE CARLO PARA MUESTRAS FINITAS

La literatura especializada ha demostrado que la presencia de puntos atípicos en una muestra crea distorsiones en el funcionamiento de los instrumentos de inferencia. Esta observación es válida con carácter general y es de aplicación al caso espacial al que se ha dedicado la sección anterior. Los resultados obtenidos muestran la sensibilidad de los estadísticos analizados a este tipo de puntos, cuya incidencia se extenderá también a la función de potencia de los contrastes. La solución analítica completa del problema no es factible en estos momentos, por lo que un ejercicio de Monte Carlo emerge como una opción recomendable. El primer apartado de la sección lo dedicamos a discutir las características del ejercicio resuelto, para centrar el contenido del segundo en los resultados obtenidos.

#### 3.1 Características del ejercicio de Monte Carlo

El problema de las observaciones atípicas en modelos econométricos de corte transversal tiene muchas vertientes (número de atípicos, posición geográfica de los mismos, dispersión, estructura de dependencias, etc.), por lo que sería necesario desarrollar un ejercicio de gran envergadura para acotar todas sus ramificaciones. En este sentido, los resultados que vamos a presentar son limitados puesto que cubren solo unas pocas variantes.

De acuerdo con la discusión planteada en la sección anterior, en el ejercicio se han simulado dos tipos de modelos: estático con estructura de dependencia residual, y dinámico con perturbación ruido blanco. El modelo simulado en el caso estático tiene una estructura simple:

$$y_r = \beta_0 + \beta_1 x_r + v_r; \quad r = 1, 2, \dots, R \quad [42]$$

El término de perturbación  $v_r$  responde a los supuestos planteados, SAR ó SMA:

$$\begin{array}{l} \text{ESTRUCTURA SAR} \\ \left\{ \begin{array}{l} y = X\beta + v \\ v = \rho Wv + u \\ u = \varepsilon + \pi d; \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \end{array} \right. \end{array} \quad \begin{array}{l} \text{ESTRUCTURA SMA} \\ \left\{ \begin{array}{l} y = X\beta + v \\ v = u - \theta Wu \\ u = \varepsilon + \pi d; \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \end{array} \right. \end{array} \quad [43]$$

La ecuación simulada en el caso dinámico se ajusta al planteamiento habitual:

## ESTRUCTURA DINÁMICA

$$\begin{cases} \mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \\ \mathbf{v} = \boldsymbol{\varepsilon} + \pi \mathbf{d}; \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{1}) \end{cases} \quad [44]$$

El término  $\mathbf{d}$  se corresponde con la variable ficticia usada en [1] y  $\pi$  es un parámetro que mide el tamaño del atípico. En [43] y [44],  $\mathbf{X}$  es una matriz de orden  $(R \times 2)$  cuya primera columna, de acuerdo a [42], es de unos mientras que la segunda recoge las observaciones de la variable  $x_r$ . Estas últimas se han obtenido de una distribución uniforme  $U(0,1)$ . En todos los casos, los parámetros  $\beta_j$  ( $j=0,1$ ) se han mantenido igual a 2 y  $\sigma_{\boldsymbol{\varepsilon}}^2$  se ha hecho igual a 1, lo cual asegura un coeficiente de determinación medio en la regresión, sin efectos espaciales, de 0.80 aproximadamente.

Se han definido tres sistemas de regiones utilizando cuadrículas regulares de  $5 \times 5$ ,  $10 \times 10$  y  $15 \times 15$  lo que resulta en tamaños muestrales de 25, 100 y 225 observaciones. La matriz  $\mathbf{W}$  se ha especificado de tipo binario normalizado. Únicamente se han simulado valores positivos en los parámetros  $\rho$  y  $\theta$ , comprendidos entre 0 y 0.99. Además, se han ensayado dos posibles localizaciones para el atípico, periférica o central. La primera supone intervenir una región situada, geográficamente, en la parte externa del sistema mientras que en la segunda se ha actuado sobre una región situada en el núcleo central del mismo. Por último, el tamaño del atípico se ha hecho depender de la dispersión del vector de perturbaciones  $\boldsymbol{\varepsilon}$ . En concreto, si por  $\hat{\sigma}_{\boldsymbol{\varepsilon}}$  denotamos la desviación típica observada en este vector, el parámetro  $\pi$  se ha hecho igual a 0 (no hay anomalía),  $2.5 \hat{\sigma}_{\boldsymbol{\varepsilon}}$  (la anomalía es pequeña),  $5 \hat{\sigma}_{\boldsymbol{\varepsilon}}$  (la anomalía es relevante) ó  $7.5 \hat{\sigma}_{\boldsymbol{\varepsilon}}$  (la anomalía es muy grande). Para simplificar la exposición, sólo vamos a presentar los resultados correspondientes al caso de un atípico grande ( $\pi$  es igual  $7.5 \hat{\sigma}_{\boldsymbol{\varepsilon}}$ ) con una posición central.

### 3.2 Resultados de la simulación

Los resultados obtenidos se resumen en la Tabla 1, dedicada al nivel de significación empírico de los contrastes, y en las Figuras 1 a 7 donde se presenta la potencia estimada.

No hay sorpresas relevantes en los resultados de la Tabla 1. La presencia de una observación atípica en la muestra tiene efectos perniciosos sobre el tamaño de los contrastes, especialmente cuando la muestra es pequeña y la anomalía es de envergadura. La influencia del atípico es evidente en muestras de tamaño 25. Los recortes observados en el nivel de significación empírico son apreciables, de forma que este último cae con frecuencia por debajo del 1.0% en respuesta a la anomalía (el teórico es el usual del 5.0%). El contraste KR parece ser el más robusto a los

atípicos mientras que los distintos Multiplicadores de Lagrange (singularmente los corregidos, LM-EL y LM-LE) reaccionan vivamente. La situación se repite cuando se utilizan muestras de tamaño 100. El impacto sigue siendo apreciable por cuanto la distorsión creada por el atípico reduce a la mitad, aproximadamente, el tamaño de los contrastes. Finalmente, la existencia de la anomalía tiende a pasar desapercibida en una muestra con 225 observaciones.

**Tabla 1**

TAMAÑO EMPÍRICO PARA UN NIVEL DE SIGNIFICACIÓN DEL 5%.(\*)

	<i>R</i> = 25		<i>R</i> = 100		<i>R</i> = 225	
	<i>Sin atípicos</i>	<i>1 atípico</i>	<i>Sin atípicos</i>	<i>1 atípico</i>	<i>Sin atípicos</i>	<i>1 atípico</i>
I de Moran	0.034	0.009	0.043	0.029	0.050	0.047
LM-ERR	0.044	0.014	0.049	0.030	0.048	0.049
LM-EL	0.044	0.027	0.054	0.024	0.052	0.050
KR	0.061	0.033	0.050	0.035	0.050	0.051
LM-LAG	0.068	0.033	0.047	0.025	0.061	0.061
LM-LE	0.070	0.017	0.054	0.024	0.063	0.065
SARMA	0.053	0.007	0.048	0.021	0.052	0.057

(\*) El intervalo de confianza teórico para  $p$  (probabilidad de rechazar la hipótesis nula), en un ejercicio de 1000 réplicas, es  $0.036 < p < 0.064$ .

Los resultados obtenidos para la función de potencia estimada se presentan en las Figuras 1 a 7 (los detalles pueden obtenerse directamente del autor). Algunos aspectos de estos gráficos ya eran conocidos con anterioridad. La debilidad de los contrastes de dependencia espacial en un contexto de muestras pequeñas es uno de ellos. Otro es el peor funcionamiento de todos los contrastes cuando en la hipótesis alternativa se simula un proceso media móvil. Tampoco es una novedad la aparente superioridad del contraste I de Moran para detectar patrones de dependencia espacial en la perturbación, ni tampoco las carencias del contraste KR, sobre todo en estructuras SMA (Florax y de Graaff, 2004).

Las figuras ponen de manifiesto, igualmente, el comportamiento poco selectivo de los estadísticos tradicionales, sensibles a todo tipo de procesos de dependencia espacial (esto es, que afecten sólo a la perturbación o al retardo de la endógena en la ecuación principal). Esta falta de criterio es la razón de ser de los Multiplicadores robustos (LM-EL y LM-LE), los cuales tienden a discriminar, efectivamente, atendiendo a la naturaleza del proceso de dependencias (residual o sustantiva). El coste es una cierta pérdida de potencia y un comportamiento más irregular en su funcionamiento. Por ejemplo, la simulación de valores próximos a la unidad en el parámetro de autocorrelación espacial provoca distorsiones, como las de la Figura 4, en el caso del contraste LM-EL bajo dependencia sustantiva y tamaño muestral

225. El repunte final observado en la función de potencia es inesperado al igual que la caída producida, en esa función, para el mismo caso (dependencia sustantiva) y una muestra de 25 observaciones. Las anomalías en el comportamiento del contraste LM-LE no son tan manifiestas pero también se dan, como se aprecia en la Figura 6. La extraña relación existente entre las funciones de potencia estimada, con y sin atípicos, para el caso dependencia sustantiva y 25 observaciones creemos que es atribuible al pequeño tamaño de la muestra.

Descontando esas circunstancias puntuales, podemos afirmar que la existencia de observaciones atípicas se traduce en un peor funcionamiento para todos los contrastes de especificación examinados. Las caídas son más significativas cuando la muestra utilizada es pequeña y, especialmente, cuando el proceso simulado en la hipótesis alternativa es de tipo SMA. En tales circunstancias (ver Figuras 3 y 4), los contrastes KR y LM-EL son poco fiables, y su comportamiento no mejora ni siquiera utilizando coeficientes de autocorrelación con valores superiores a 0.90. Sin embargo, el contraste KR parece ser el más robusto a la presencia de atípicos en la muestra cuando en la perturbación domina una estructura SAR.

Las distorsiones creadas por las anomalías en los datos no son tan apreciables cuando se difunden mediante un proceso de dependencia sustantiva. En este caso, tal como se aprecia en las Figuras 5, 6 y 7, la pérdida de potencia sufrida por los diferentes contrastes es menor. El estadístico LM-LAG es el que funciona mejor en esta situación, aunque su predominio se diluye conforme aumenta el tamaño muestral. De esta forma, cuando se utilizan 225 observaciones, las diferencias entre el estadístico LM-LAG y el LM-LE (robusto a errores locales de especificación de la hipótesis nula) son inapreciables tanto en potencia como con respecto a la forma de la propia función de potencia.

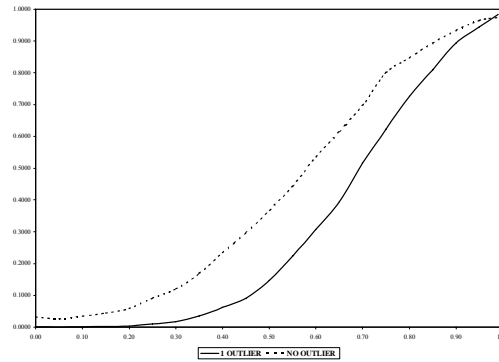
Para finalizar, otro aspecto que conviene subrayar es que, si bien es evidente que la incidencia de las observaciones anómalas se hace menor conforme el tamaño muestral aumenta, todavía se perciben ciertas disfunciones en el comportamiento de los estadísticos utilizando una muestra de tamaño 225. En este último caso, debe tenerse en cuenta que sólo se ha intervenido un punto a pesar de aumentar el tamaño muestral, lo cual no parece muy razonable en circunstancias reales. Por el contrario, a medida que aumenta el tamaño de la muestra, la probabilidad de contar con más observaciones atípicas también crece. La observación es interesante aunque supera los objetivos del presente trabajo.

Figura 1 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE I DE MORAN EN PRESENCIA DE ATÍPICOS

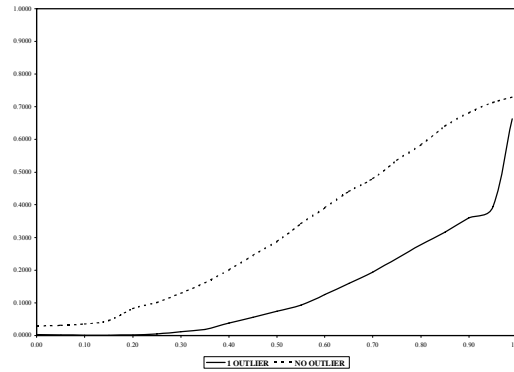
## Dependencia residual. Proceso SAR

R=25



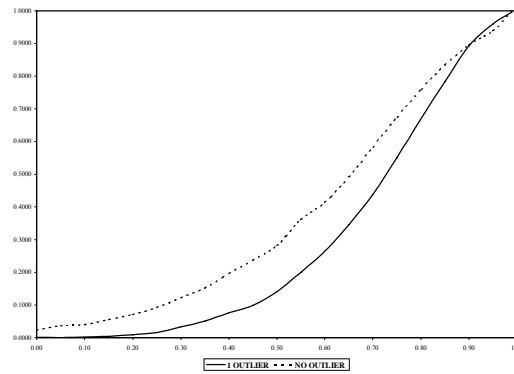
## Dependencia residual .Proceso SMA

R=25



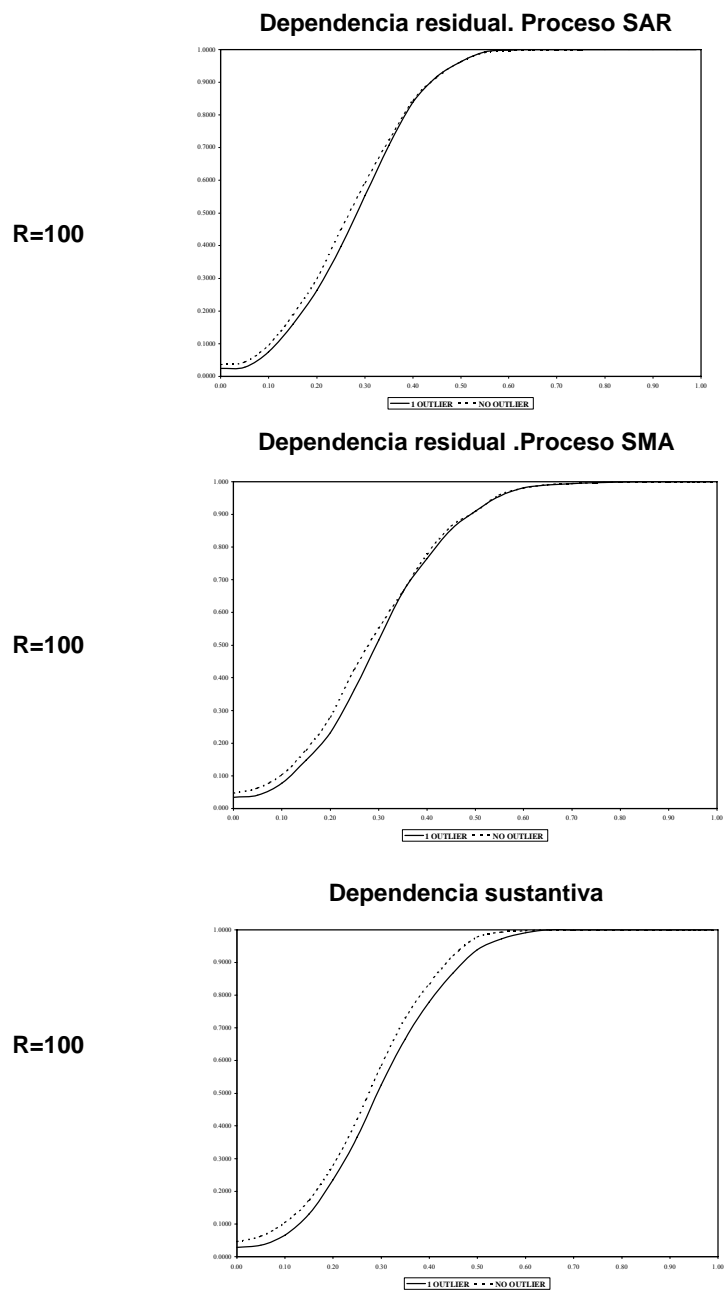
## Dependencia sustantiva

R=25



**Figura 1 (Continuación)**

TAMAÑO Y POTENCIA DEL CONTRASTE I DE MORAN EN PRESENCIA DE ATÍPICOS



**Figura 1 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE I DE MORAN EN PRESENCIA DE ATÍPICOS

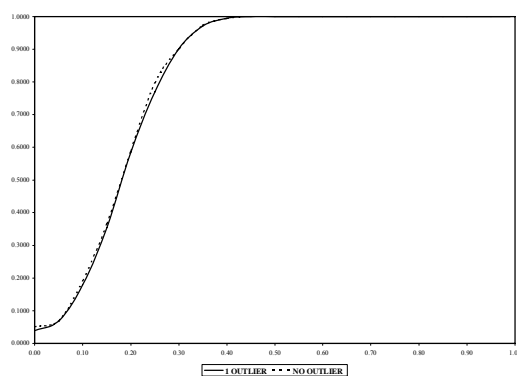
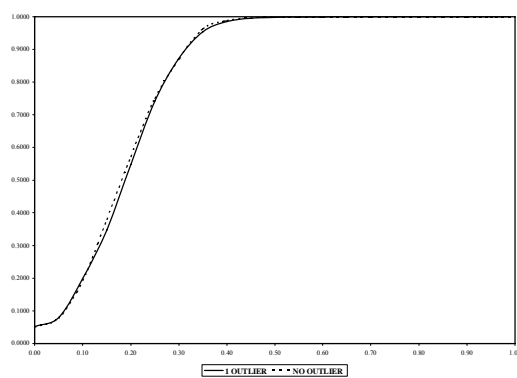
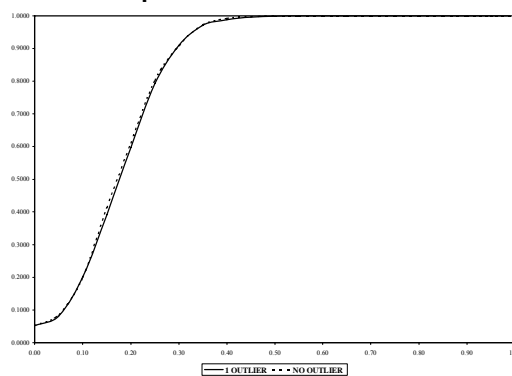
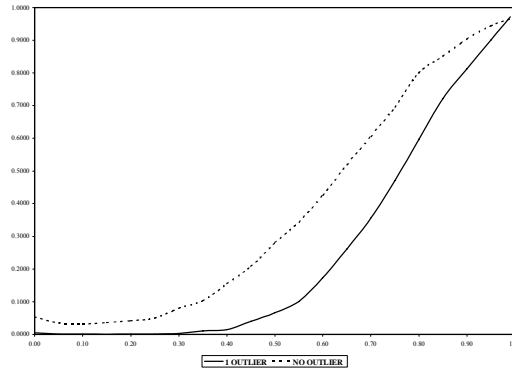
**Dependencia residual. Proceso SAR****R=225****Dependencia residual .Proceso SMA****R=225****Dependencia sustantiva****R=225**

Figura 2 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE LM-ERR EN PRESENCIA DE ATÍPICOS

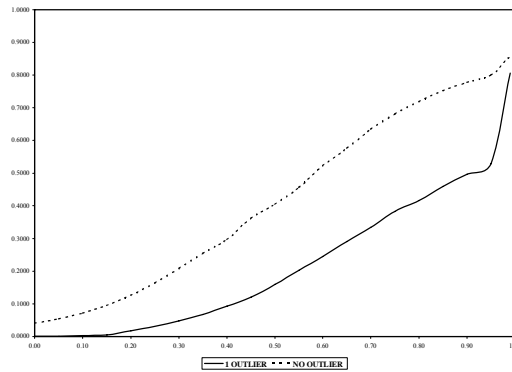
## Dependencia residual. Proceso SAR

R=25



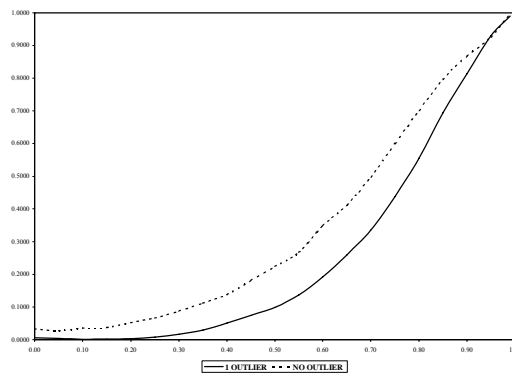
## Dependencia residual .Proceso SMA

R=25



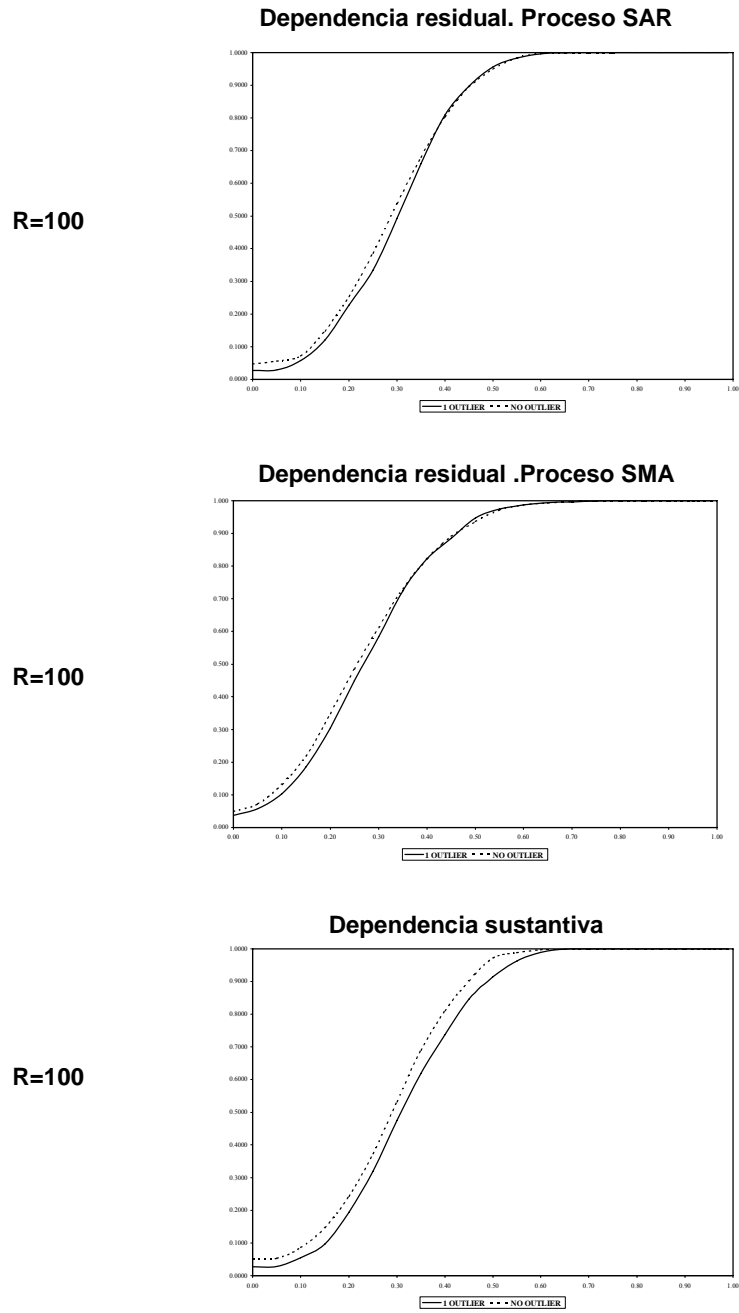
## Dependencia sustantiva

R=25



**Figura 2 (Continuación)**

TAMAÑO Y POTENCIA DEL CONTRASTE LM-ERR EN PRESENCIA DE ATÍPICOS



**Figura 2 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE LM-ERR EN PRESENCIA DE ATÍPICOS

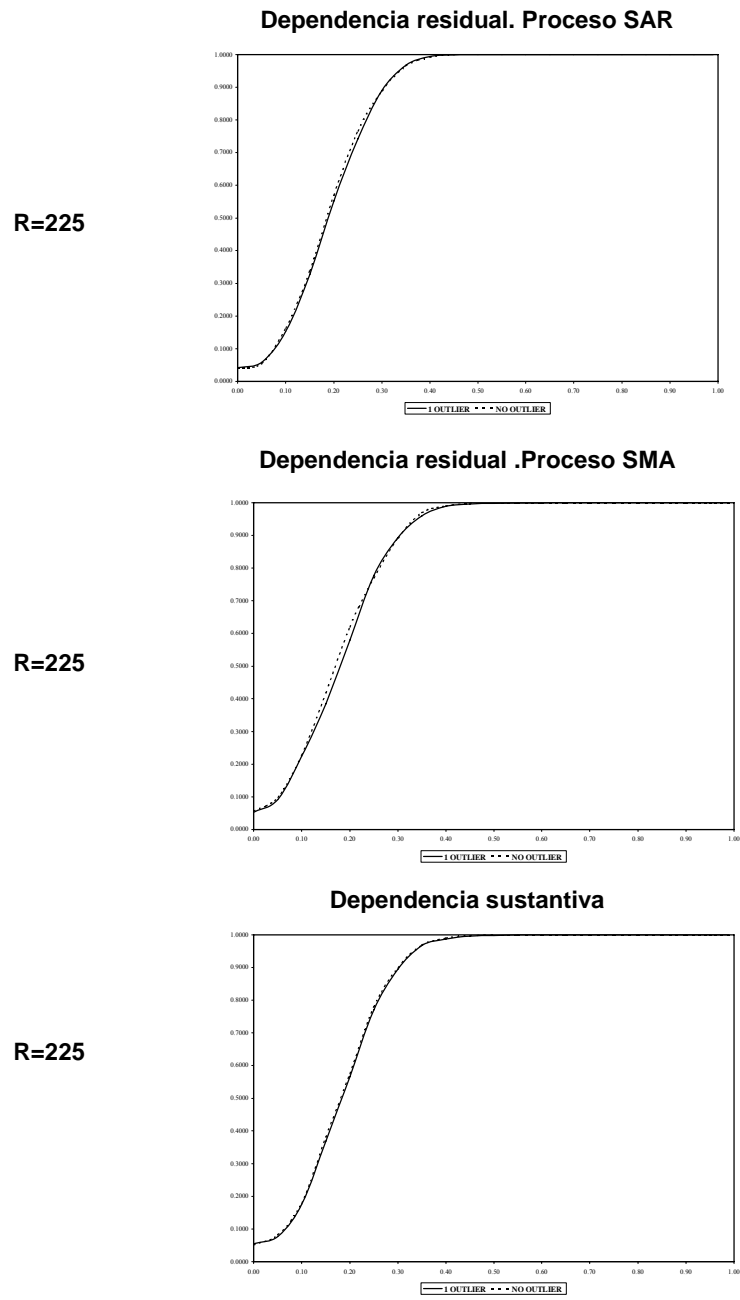
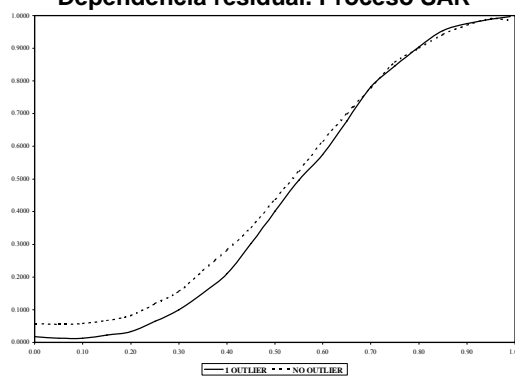


Figura 3 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE KR EN PRESENCIA DE ATÍPICOS

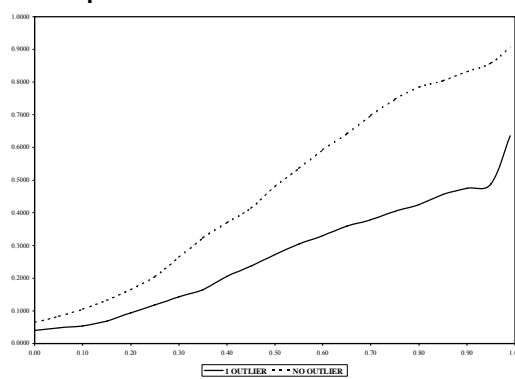
## Dependencia residual. Proceso SAR

R=25



## Dependencia residual. Proceso SMA

R=25



## Dependencia sustantiva

R=25

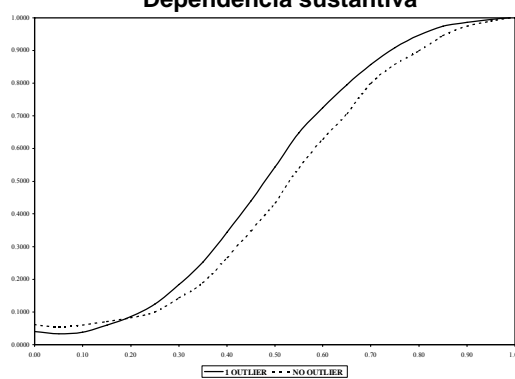
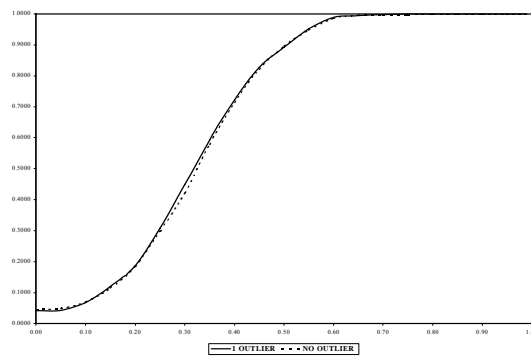


Figura 3 (Continuación)

TAMAÑO Y POTENCIA DEL CONTRASTE KR EN PRESENCIA DE ATÍPICOS

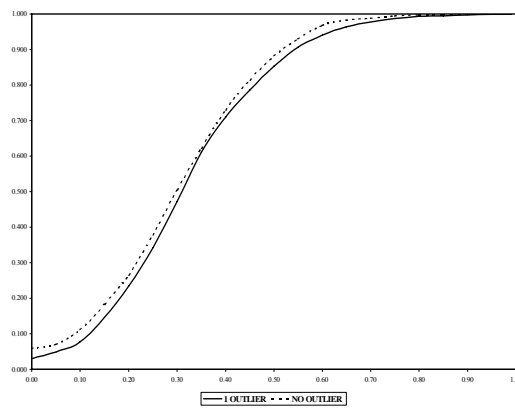
Dependencia residual. Proceso SAR

R=100



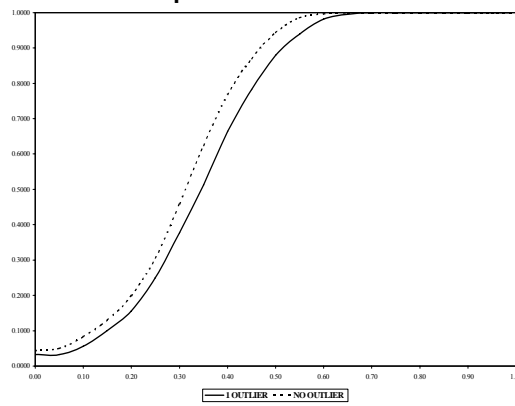
Dependencia residual. Proceso SMA

R=100



Dependencia sustantiva

R=100



**Figura 3 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE KR EN PRESENCIA DE ATÍPICOS

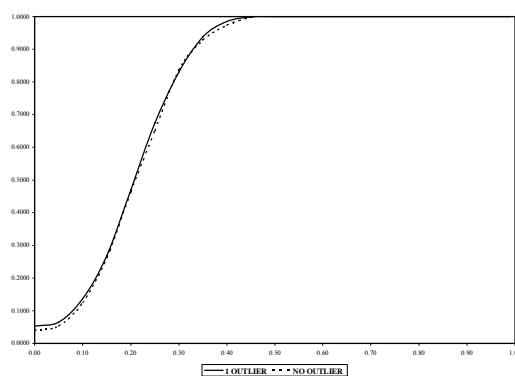
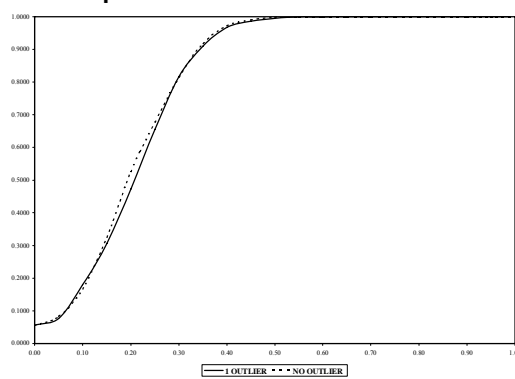
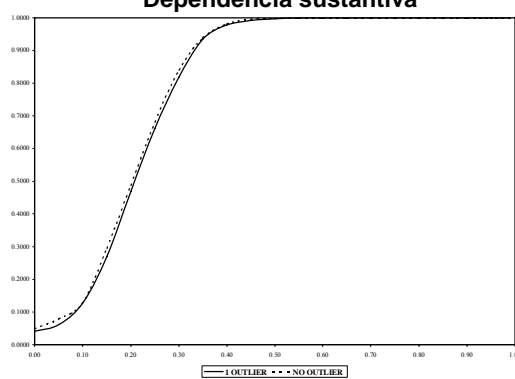
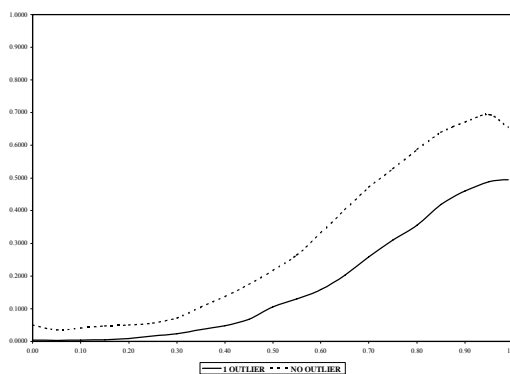
**Dependencia residual. Proceso SAR****R=225****Dependencia residual. Proceso SMA****R=225****Dependencia sustantiva****R=225**

Figura 4 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE LM-EL EN PRESENCIA DE ATÍPICOS

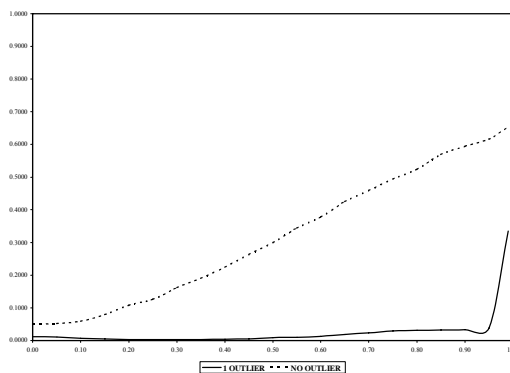
## Dependencia residual. Proceso SAR

R=25



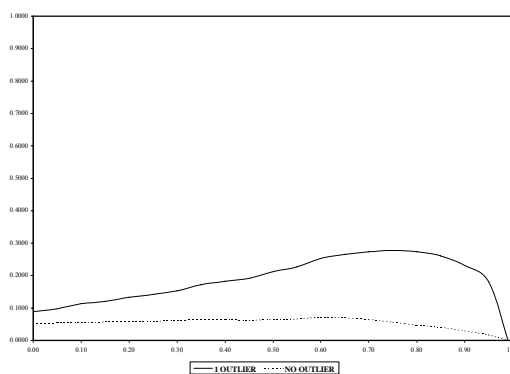
## Dependencia residual. Proceso SMA

R=25



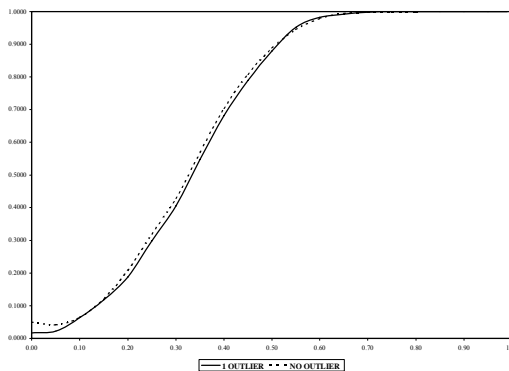
## Dependencia sustantiva

R=25

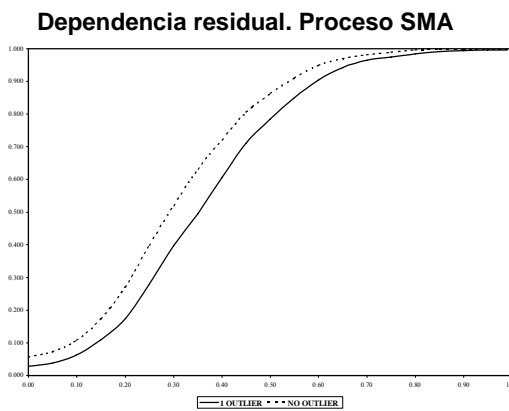


**Figura 4 (Continuación)**  
**TAMAÑO Y POTENCIA DEL CONTRASTE LM-EL EN PRESENCIA DE ATÍPICOS**  
**Dependencia residual. Proceso SAR**

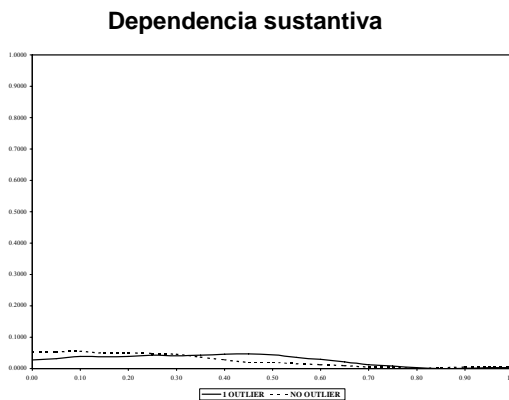
**R=100**



**R=100**



**R=100**



**Figura 4 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE LM-EL EN PRESENCIA DE ATÍPICOS

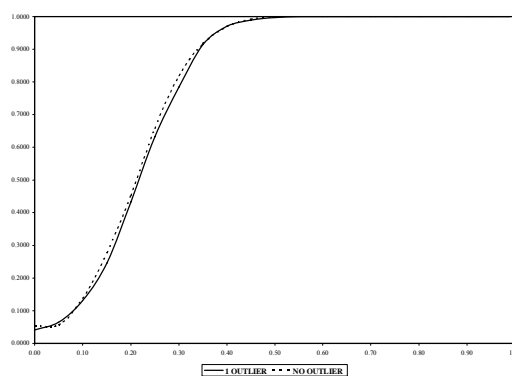
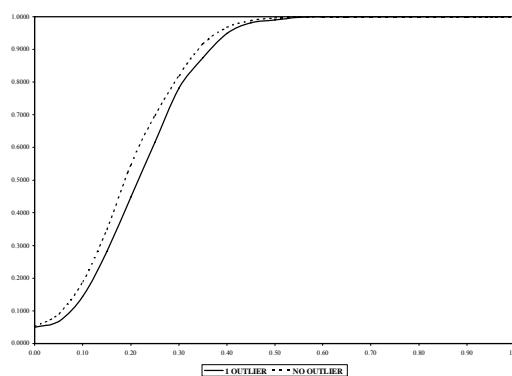
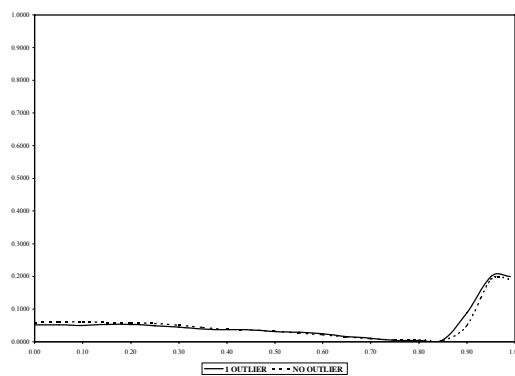
**Dependencia residual. Proceso SAR****R=225****Dependencia residual. Proceso SMA****R=225****Dependencia sustantiva****R=225**

Figura 5 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE LM-LAG EN PRESENCIA DE ATÍPICOS

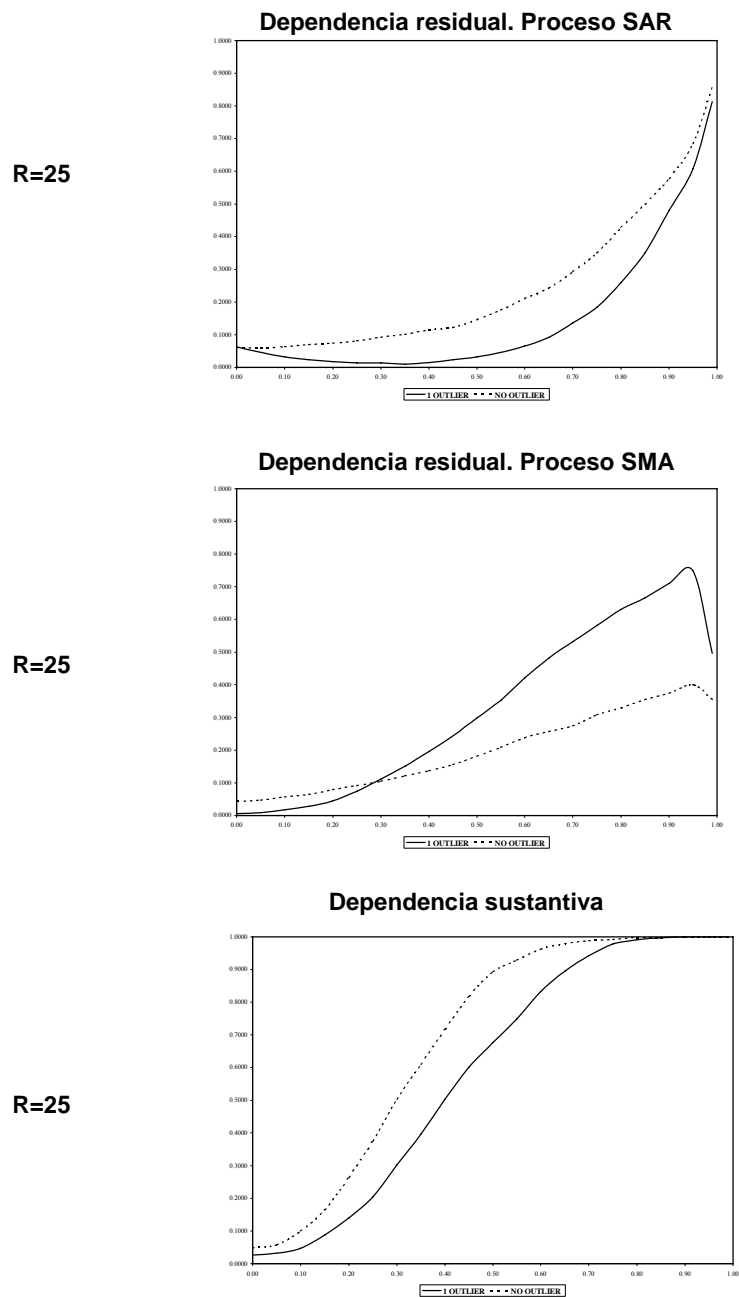
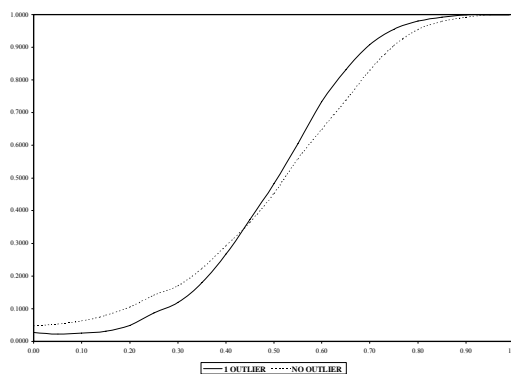


Figura 5 (Continuación)

TAMAÑO Y POTENCIA DEL CONTRASTE LM-LAG EN PRESENCIA DE ATÍPICOS

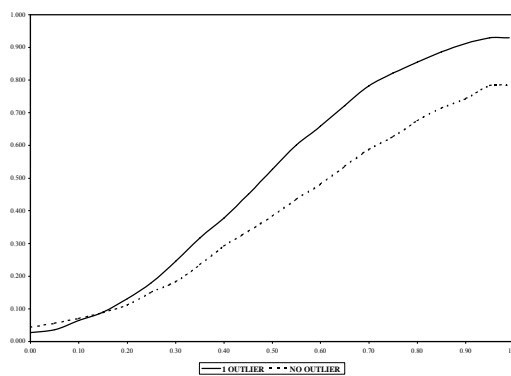
Dependencia residual. Proceso SAR

R=100



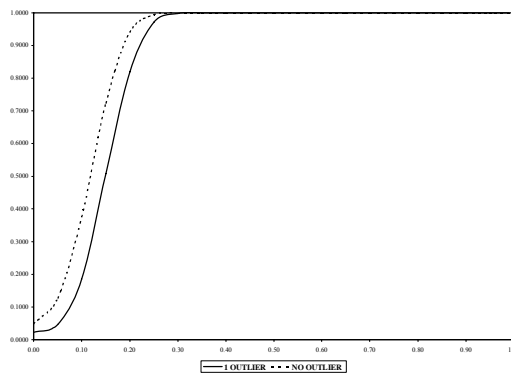
Dependencia residual. Proceso SMA

R=100



Dependencia sustantiva

R=100



**Figura 5 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE LM-LAG EN PRESENCIA DE ATÍPICOS

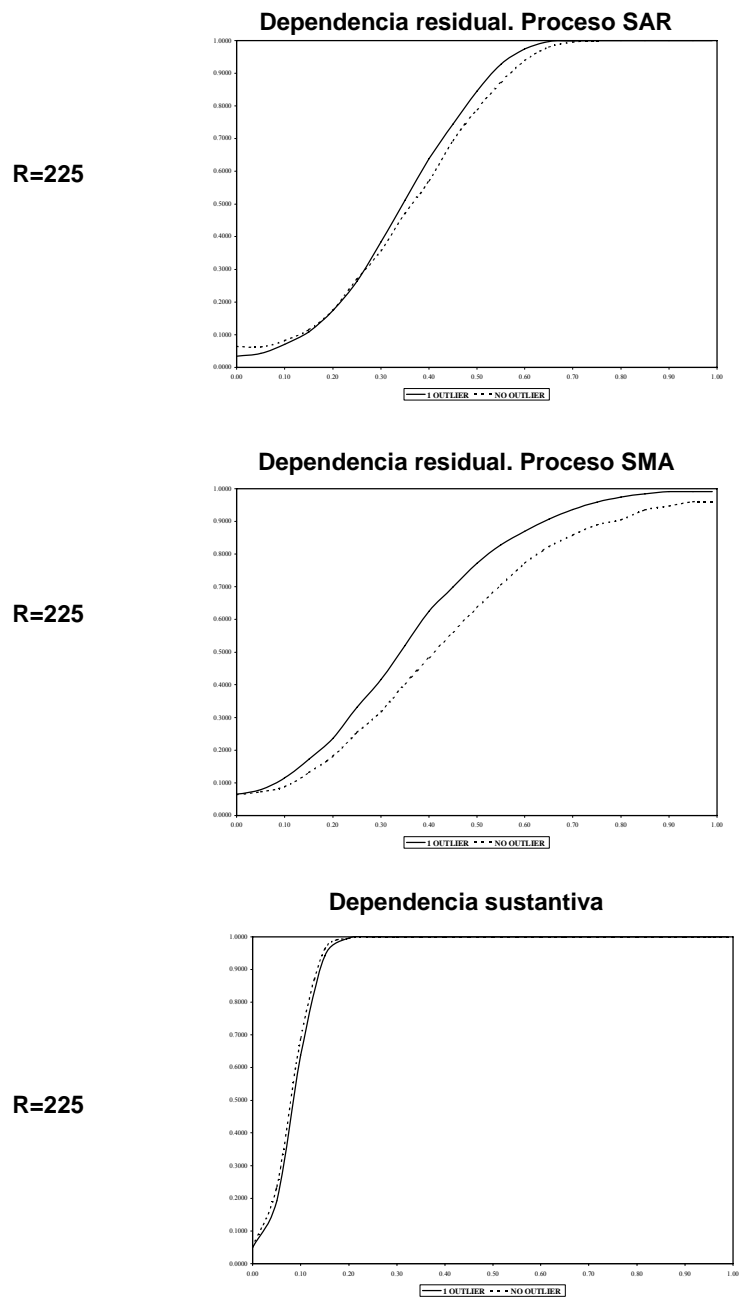
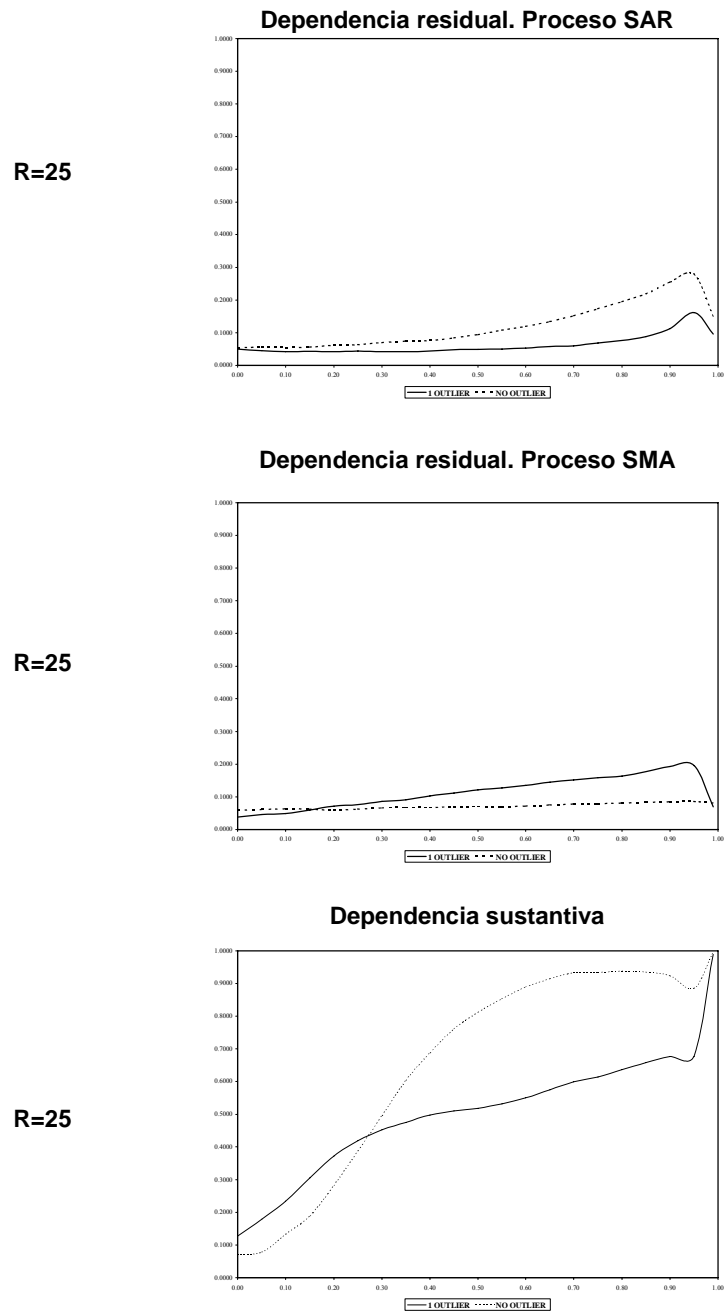


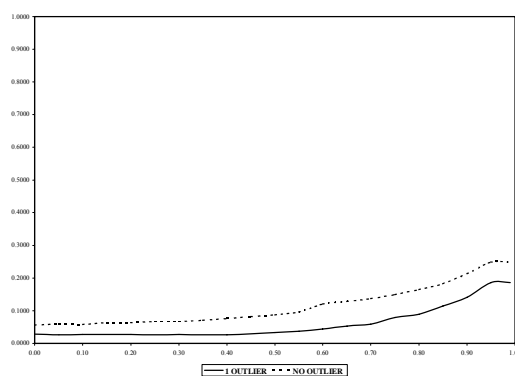
Figura 6 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE LM-LE EN PRESENCIA DE ATÍPICOS



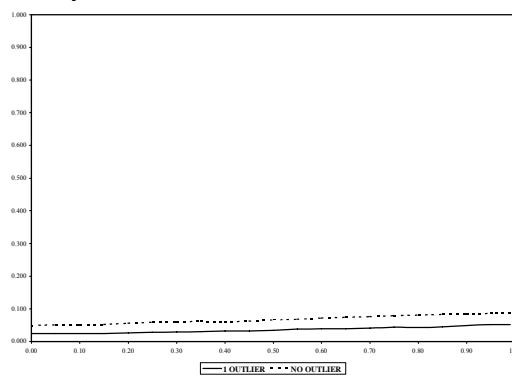
**Figura 6 (Continuación)**  
**TAMAÑO Y POTENCIA DEL CONTRASTE LM-LE EN PRESENCIA DE ATÍPICOS**  
**Dependencia residual. Proceso SAR**

**R=100**



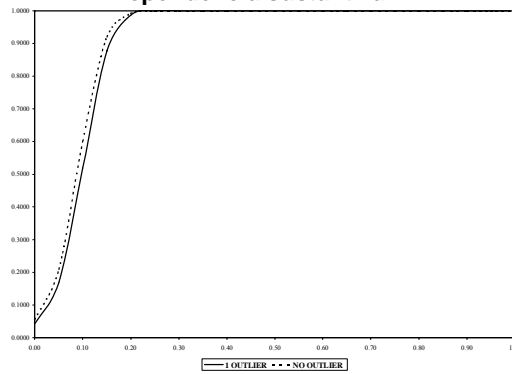
**Dependencia residual. Proceso SMA**

**R=100**



**Dependencia sustantiva**

**R=100**



**Figura 6 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE LM-LE EN PRESENCIA DE ATÍPICOS

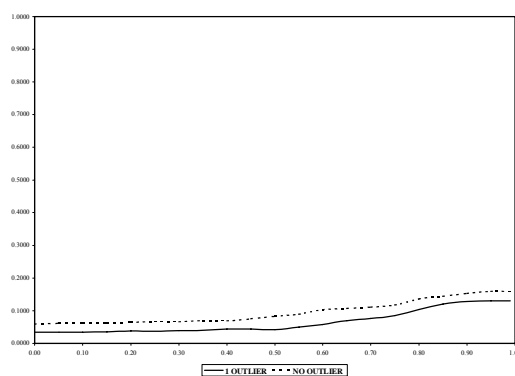
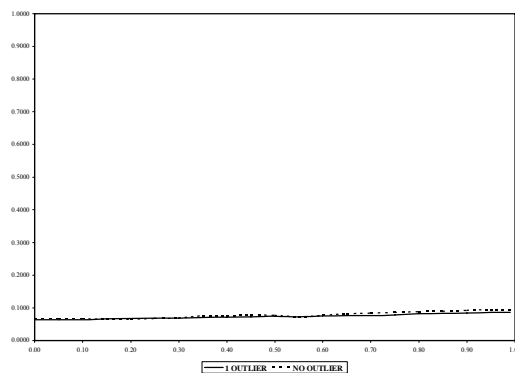
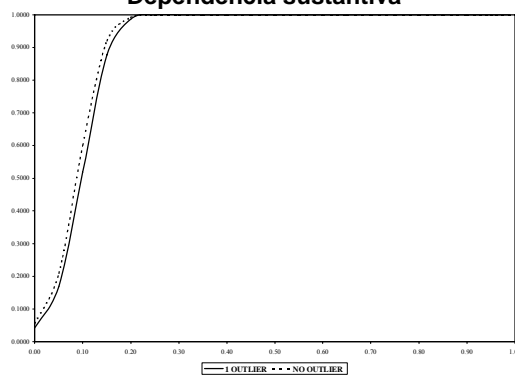
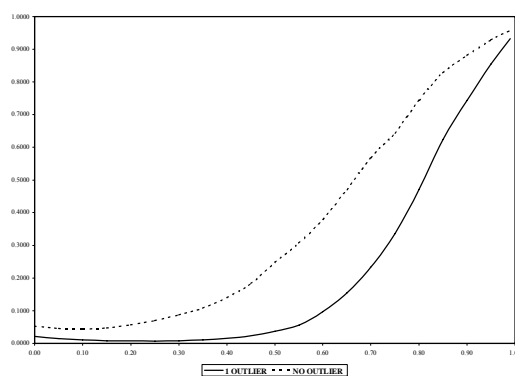
**Dependencia residual. Proceso SAR****R=225****Dependencia residual. Proceso SMA****R=225****Dependencia sustantiva****R=225**

Figura 7 (Continúa)

TAMAÑO Y POTENCIA DEL CONTRASTE SARMA EN PRESENCIA DE ATÍPICOS

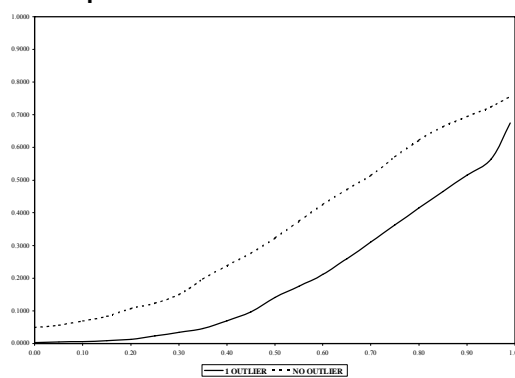
## Dependencia residual. Proceso SAR

R=25



## Dependencia residual. Proceso SMA

R=25



## Dependencia sustantiva

R=25

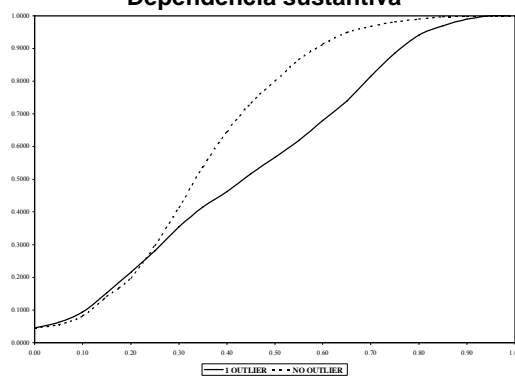
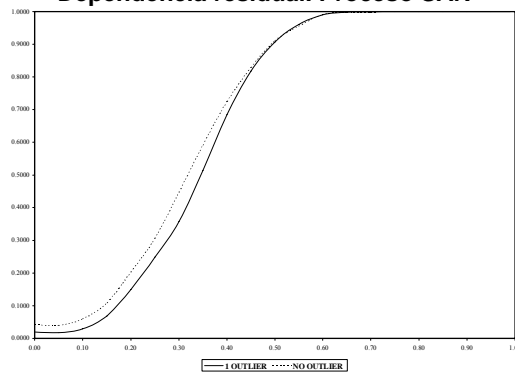


Figura 7 (Continuación)

TAMAÑO Y POTENCIA DEL CONTRASTE SARMA EN PRESENCIA DE ATÍPICOS

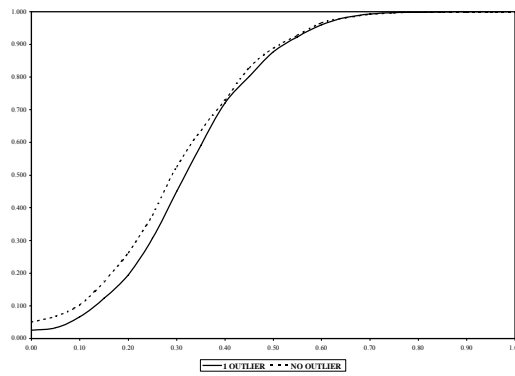
## Dependencia residual. Proceso SAR

R=100



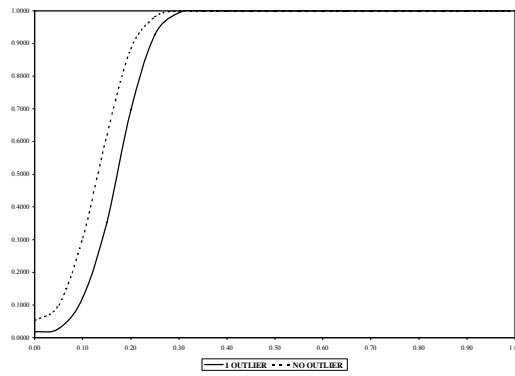
## Dependencia residual. Proceso SMA

R=100



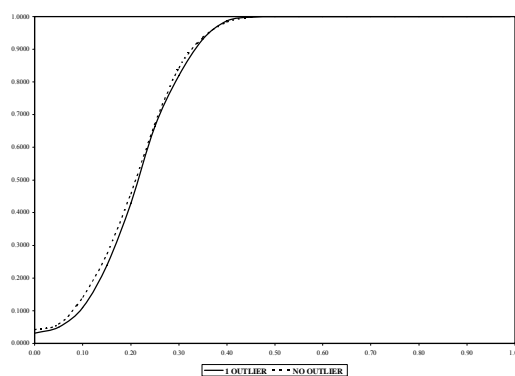
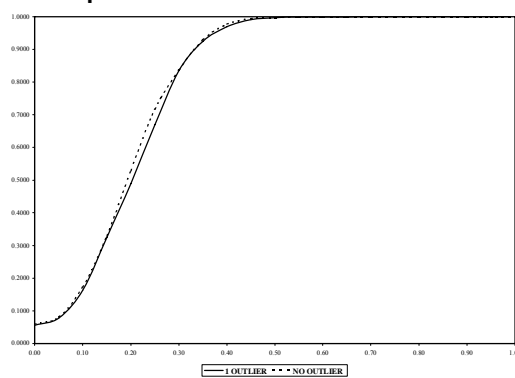
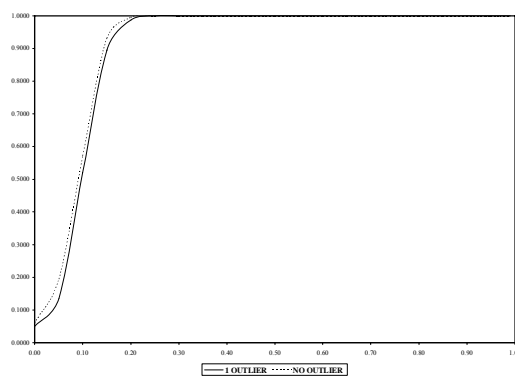
## Dependencia sustantiva

R=100



**Figura 7 (Conclusión)**

TAMAÑO Y POTENCIA DEL CONTRASTE SARMA EN PRESENCIA DE ATÍPICOS

**Dependencia residual. Proceso SAR****R=225****Dependencia residual. Proceso SMA****R=225****Dependencia sustantiva****R=225**

#### 4. CONCLUSIONES Y REFLEXIONES FINALES

El objetivo de este trabajo era tratar de precisar mejor la incidencia de las observaciones atípicas sobre los contrastes de especificación empleados más a menudo en un contexto espacial. La cuestión se ha discutido con detalle en el análisis de series temporales, donde ocupa un lugar prominente. Sin embargo, la evidencia aportada para el caso espacial es todavía muy limitada.

La aproximación analítica resuelta en la sección segunda corrobora resultados conocidos en el caso general: la presencia de anomalías en la muestra afecta a la fiabilidad de los instrumentos de inferencia. Las distorsiones sufridas por los estadísticos examinados dependen, entre otras cosas, del tamaño de la anomalía y del grado de apalancamiento correspondiente a las observaciones de las variables explicativas del punto contaminado. La peor situación combina un gran tamaño en la anomalía con un fuerte apalancamiento en las exógenas; es decir, se trata de una observación que es atípica tanto en el espacio  $X$  como en el espacio  $[X | y]$ . Debe recordarse que un apalancamiento elevado implica que la observación tiende a apoderarse del plano de regresión muestral forzándolo, en el límite, a que pase por sus inmediaciones. En consecuencia, el residuo asociado será pequeño (primer error: el residuo debería ser grande para alertar sobre la anomalía) a costa de alterar el contenido informativo suministrado por las restantes observaciones (segundo error: la malla de dependencias espaciales se distorsiona).

La simulación llevada a cabo en la sección tercera nos ha permitido confirmar algunos de estos resultados. Por ejemplo, el impacto de las observaciones atípicas guarda una relación inversa con el tamaño de la muestra pero es directa con el tamaño de la anomalía. Además, las distorsiones creadas por estos puntos se debilitan a medida que aumenta la intensidad de la dependencia espacial existente en los datos. Estos resultados eran esperados a priori. Sin embargo, también se han producido otros más sorprendentes. Por ejemplo, el hecho de que las anomalías en los datos tengan menor incidencia cuando se propagan mediante procesos de dependencia espacial sustantiva es uno de ellos. La distinta sensibilidad mostrada por los diferentes estadísticos es otro aspecto a significar, incluido el comportamiento aparentemente más robusto del contraste KR en el caso SAR.

Para finalizar, la principal conclusión que podemos avanzar es la importancia de analizar con detenimiento la información muestral, antes de acometer el trabajo de modelización econométrica, prestando especial atención al problema de los puntos extremos, atípicos u *outliers*. Cualquiera que sea la decisión tomada con respecto a esos puntos, la sugerencia de Chatterjee y Hadi (1988, p.80) nos parece singularmente acertada: '*No se deben rechazar mecánicamente los atípicos ('outliers' en el*

original) y *proceder con el análisis*. Si los atípicos son observaciones genuinas ('bona fide' en su terminología), pueden estar indicando la no adecuación del modelo bajo ciertas condiciones específicas. A menudo proporcionan claves valiosas para que el analista pueda construir un modelo mejor'. La evidencia aportada en este trabajo, aunque no sea más que una primera aproximación al problema, nos permite ratificar la pertinencia de esta observación.

## APÉNDICE 1. ESTADÍSTICOS DE ESPECIFICACIÓN ANALIZADOS

Los contrastes analizados se encuentran referidos siempre al modelo de la hipótesis nula; esto es  $y=X\beta+u$ . Este modelo se ha estimado por MCO, donde  $\hat{\sigma}^2$  y  $\hat{\beta}$  son los correspondientes estimadores MCO y  $\hat{u}$  las series de residuos. Los contrastes obtenidos son los siguientes (ver Anselin y Florax, 1995, para los detalles):

$$I \text{ de Moran: } I = \frac{R}{S_0} \frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}}; \quad S_0 = \sum_{r=1}^R \sum_{s=1}^R w_{rs} \quad [A1.1]$$

$$\text{Test LM-ERR: } LM-ERR = \left( \frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2} \right)^2 \frac{1}{T_1}; T_1 = \text{tr}[W'W + WW] \quad [A1.2]$$

$$\text{Test LM-EL: } LM-EL = \frac{\left( \frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2} - \frac{T_1}{R\hat{j}_{p-\beta}} \frac{\hat{u}'W\hat{y}}{\hat{\sigma}^2} \right)^2}{T_1 - \frac{T_1^2}{R\hat{j}_{p-\beta}}} \quad [A1.3]$$

$$\text{Test KR: } KR = h_R \frac{\hat{\phi}'Z'Z\hat{\phi}}{\hat{e}'\hat{e}} \quad [A1.4]$$

$$\text{Test LM-LAG: } LM-LAG = \frac{1}{R\hat{j}_{p-\beta}} \left( \frac{\hat{u}'W\hat{y}}{\hat{\sigma}^2} \right)^2 \quad [A1.5]$$

$$\text{Test LM-LE: } LM-LE = \frac{\left( \frac{\hat{u}'W\hat{y}}{\hat{\sigma}^2} - \frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2} \right)^2}{R\hat{j}_{p-\beta} - T_1} \quad [A1.6]$$

$$\text{Test SARMA: } SARMA = \frac{\left( \frac{\hat{u}'W\hat{y}}{\hat{\sigma}^2} - \frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2} \right)^2}{R\hat{j}_{p-\beta} - T_1} + \frac{1}{T_1} \left( \frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2} \right)^2 \quad [A1.7]$$

Además,  $R_{\hat{\beta}-\beta} = T_1 + (\hat{\beta}' X' WMWX \hat{\beta}) / \hat{\sigma}^2$  y  $M = [I - X(X'X)^{-1}X']$ . Por otra parte,  $\hat{e}$  es el vector de residuos de la regresión auxiliar del contraste de Kelejian-Robinson (KR), de orden  $(h_R \times 1)$ ,  $Z$  la matriz de variables explicativas incluidas en esa última regresión y  $\hat{\phi}$  la estimación del vector de parámetros asociado.

Como es conocido, la distribución asintótica del estadístico de Moran estandarizado es una  $N(0,1)$ ; los dos multiplicadores de Lagrange que continúan, LM-ERR y LM-EL, tienen una distribución asintótica  $\chi^2(1)$ , y la distribución del estadístico KR es una  $\chi^2(m)$ , siendo  $m$  el número de regresores que intervienen en la regresión auxiliar. La distribución de los tres últimos contrastes también es una *chi-cuadrado*, con un grado de libertad los dos primeros y con dos en el caso del contraste SARMA.

## APÉNDICE 2. VALOR ESPERADO DEL ESTADÍSTICO DE MORAN BAJO LA HIPÓTESIS NULA DE INCORRELACIÓN Y ANOMALÍAS EN LA GENERACIÓN DE LA OBSERVACIÓN $r$ -ÉSIMA

En este apéndice vamos a obtener el valor esperado del estadístico de Moran cuando se utiliza sobre los residuos MCO de un modelo como el de [2] en el que la perturbación se halla incorrelada espacialmente pero se ha cometido un error de especificación, al haberse omitido la variable ficticia  $d$  vinculada a la anomalía. Como se indica en [2], el resultado será un término de error compuesto,  $v = \pi d + u$ , unos estimadores MCO sesgados y unos residuos MCO,  $\hat{v} = y - X\hat{\beta} = M(u + \pi d)$ , no centrados en el origen:  $E[\hat{v}] = \pi M d = m_r$ , siendo  $M = [I - X(X'X)^{-1}X']$  y  $m_r = M d$ .

En cualquier caso, el estadístico de Moran se obtendrá de la forma usual:

$$I = \frac{R}{S_0} \frac{\hat{v}' W \hat{v}}{\hat{v}' \hat{v}} \quad [A2.1]$$

Con  $S_0 = \sum_{r=1}^R \sum_{s=1}^R w_{rs}$ . La resolución de los momentos del estadístico  $I$ , bajo la

hipótesis nula de independencia, es simple cuando utilizamos el argumento de Cliff y Ord (1981, p. 201) '*Podemos evaluar los momentos del estadístico  $I$  usando el teorema de Pitman-Koopmans. La aplicación directa de este teorema implica que:*

$$E[I^p] = \left( \frac{R}{S_0} \right)^p \frac{E[(\hat{v}' W \hat{v})^p]}{E[(\hat{v}' \hat{v})^p]}, p = 1, 2, \dots \quad [A2.2]$$

*dado que  $I$  y  $\hat{v}' \hat{v}$  son independientes*'. En nuestro caso, debe tenerse en cuenta la estructura del vector de residuos MCO:  $\hat{v} = M(u + \pi d) = Mv$ . Haciendo  $p=1$ , el valor esperado de la suma residual del denominador es:

$$\begin{aligned} E[\hat{v}' \hat{v}] &= E[(u + \pi d)' M(u + \pi d)] = E[u' M u + \pi^2 d' M d + 2\pi u' M d] = \\ &= \sigma^2(R-k) + \pi^2 d' [I - X(X'X)^{-1}X'] d = \\ &= \sigma^2(R-k) + \pi^2 [d' d - d' X(X'X)^{-1}X' d] = \\ &= \sigma^2(R-k) + \pi^2(1 - \rho_r) \end{aligned} \quad [A2.3]$$

Actuando sobre la forma cuadrática del numerador:

$$\begin{aligned}
 E[\hat{v}'W\hat{v}] &= E[(u + \pi d)'MWM(u + \pi d)] = \\
 E[u'MWMu + \pi^2 d'MWMd + 2\pi u'MWMd] &= \quad [A2.4] \\
 \sigma^2 \text{tr}MW + \pi^2 d'MWMd &= \sigma^2 \text{tr}MW + \pi^2 m_r' W m_r
 \end{aligned}$$

Por último, introduciendo [A2.3] y [A2.4] en [A2.2] obtenemos el resultado final correspondiente a la ecuación [11]. Los momentos de orden superior pueden derivarse replicando el mismo procedimiento. En Sen (1970) se detallan las condiciones, relativamente débiles, bajo las cuales el estadístico I de Moran alcanza la normalidad en un contexto asintótico.

## REFERENCIAS

- ANSCOMBE, F. (1960) «Rejection of Outliers». *Technometrics*, **2**, 123-147.
- ANSELIN, L. (1988) «Spatial Econometrics. Methods and Models». Dordrecht: Kluwer
- ANSELIN, L. y FLORAX, R. (1995) «Small Sample Properties of Tests for Spatial Dependence in Regression Models». En L. Anselin y R. Florax (eds.). *New Directions in Spatial Econometrics* (pp. 21-74). Berlin Springer.
- ANSELIN, L., BERA, A, FLORAX, R. y YOON, M. (1996) «Simple Diagnostic Tests for Spatial Dependence». *Regional Science and Urban Economics*, **26**, 77-104.
- AZNAR, A. (1989) «Econometric Model Selection. A New Approach». Dordrecht: Kluwer.
- BARNETT, V y LEWIS, T. (1994) «Outliers in Statistical Data», 3ª edición. Wiley. New York:
- BECKMAN, R. y COOK, R. (1983) «Outlier.....s», *Technometrics*, **25**, 119-149.
- BELSLEY, D. (1991) «Conditioning Diagnostics». Wiley. New York
- CHANG, I., TIAO, G. y CHEN, C. (1988) «Estimation of Time Series Parameters in the Presence of Outliers». *Technometrics*, **3**, 193-204.
- CHATTERJEE, S. y HADI, A. (1988) «Sensitivity Analysis in Linear Regression». Wiley. New York:

- CLIFF, A. y ORD, J. (1981) «Spatial Processes. Models & Applications». Pion. Londres
- CRESSIE, N. (1993) «Statistics for Spatial Data». Wiley. New York
- DAVIDSON, J. (1994) «Stochastic Limit Theory». Oxford University Press. Oxford
- DIXON, W. (1950) «Analysis of Extreme Values». *Annals of Mathematical Statistics*, **21**, 488-506.
- EVANS, V. (1999) «Strategies for Detecting Outliers in Regression Analysis. An Introductory Premier». En B. Thompson (ed.). *Advances in Social Science Methodology* (Vol. 5, pp. 213-233). Stamford JAI Press.
- FLORAX, R. y GRAAFF, T. de (2004) «The Performance of Diagnostics Tests for Spatial Dependence in Linear Regression Models. A Meta-Analysis of Simulation Studies». En L. Anselin, R. Florax y S. Rey (eds.) *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin Springer
- HAINING, R. (1994) «Diagnostics for Econometric Modeling in Spatial Econometrics». *Journal of Regional Science*, **34**, 325-341.
- HAINING, R. (1995). «Data Problems in Spatial Econometric Modeling». En L. Anselin y R. Florax (eds). *New Directions in Spatial Econometrics* (pp. 156-171). Berlin. Springer.
- HAWKINS, D. (1980) «Identification of Outliers». Chapman and Hall. Londres
- HUBER, P. (1981) «Robust Statistics». John Wiley & Sons. Nueva York
- JOHNSON, M. y KOTZ, S. (1970) «Distributions in Statistics». *Continuous Univariate Distributions*. John Wiley & Sons. Nueva York
- KENDALL, M. y STUART, A. (1977) «The Advanced Theory of Statistics», 4ª edición.: Charles Griffin & Company Limited. Londres
- MARTIN, R. (1992) «Leverage, Influence and Residuals in Regression Models when Observations are Correlated». *Communications in Statistics. Theory & Methods*, **21**, 1183-1212.
- PEÑA, D. (1990) «Influential Observations in Time Series». *Journal of Business & Economic Statistics*, **8**, 235-241.
- PEÑA, D. (2001) «Análisis de datos multivariantes. McGraw Hill. Madrid
- PERRON, P. (1989) «The Great Crash, the Oil Price Shock and the Unit Root Hypothesis». *Econometrica*, **57**, 1361-1401.
- RAO, R. (1973) «Linear Statistical Inference and Its Applications». John Wiley & Sons. Nueva York

- ROUSSEEUW, P. y LEROY, A. (1987) «Robust Regresión and Outlier Detection». John Wiley & Sons. Nueva York
- SACHS, L. (1982) «Applied Statistics. A Handbook of Techniques». Springer-Verlag. Nueva York:
- SEN, A. (1976) «Large Sample-Size Distribution of Statistics Used in Testing for Spatial Correlation». *Geographical Analysis*, **9**, 175-184.
- SHEKHAR, S., LU, C. y ZHANG, P. (2002) «Detecting graph-based spatial outliers». *Intelligent Data Analysis*, **6**, 451-468.
- TSAY, R. (1988) «Outliers, Level Shifts and Variance Changes in Time Series». *Journal of Forecasting*, **7**, 1-20.
- TUKEY, J. (1977) «Exploratory Data Analysis». Reading: Addison-Wesley.
- WARTENBERG, D. (1989) «Exploratory Spatial Analysis: Outliers, Leverage Points and Influence Functions». En D. Griffith (ed.). *Spatial Statistics: Past, Present and Future*. Monograph n° 12 (pp. 133-156). Institute of Mathematical Geography. Syracuse University. Syracuse

## OUTLIERS IN CROSS-SECTIONAL REGRESSIONS

### ABSTRACT

There is an abundant literature dedicated to the identification and treatment of *outliers* in the time series context. However, that cannot be extrapolated directly to the spatial case because of its particular properties. In the present paper our intention is to advance in this direction focussing the discussion on the consequences of outliers on the most habitual specification tests in cross sectional regressions. We present, firstly, a series of analytical results that try to delimit the impact suffered by the most relevant specification statistics. We complete the analysis with a Monte Carlo exercise designed to evaluate, with greater precision, the effect of outliers in the sample.

*Keywords:* outliers, spatial dependence, Monte Carlo

AMS Classification: 62P20, 90A19