Psicothema, 1994. Vol 6, n.º 2, pp. 503-524
ISSN 0214 - 9915 CODEN PSOTEG

EVALUACION DE LOS EFECTOS DE LA INTERVENCION EN DISEÑOS DE SERIES TEMPORALES EN PRESENCIA DE TENDENCIAS

Guillermo Vallejo Seco
Universidad de Oviedo

La solución comúnmente practicada para abordar el problema de la dependencia serial ha consistido en emplear el modelamiento ARIMA originalmente propuesto por Box y Tiao (1965). Con todo, el elevado número de puntuaciones exigido por el modelo para su correcta identificación, conlleva que el enfoque vea restringida su aplicación en muchos ámbitos socio-comportamentales. Para superar este hándicap diversos investigadores (Harrop y Velicer, 1985; 1990) han sugerido eliminar la problemática fase de identificación en aras de algún modelo asumido de antemano. Los resultados actuales parecen obstinarse en poner de manifiesto que las diferencias encontradas por nosotros (Vallejo y otros, 1992) en cuanto a la eficiencia de los estimadores de las distintas rutinas de análisis, no sólo no desaparecen a medida que complicamos los modelos desde los cuales las series son generadas, sino que se empeora.

Evaluation of the effects of intervention in designs of time series in the presence of trends. The most typical solution to the problem of serial dependence has been the use of the ARIMA models proposed by Box and Tiao (1965). But, for a reliable use of those models, too many observations are requiered, wich implies that they can't be properly used in many social and behavioral practical situations. To avoid this shortcoming, some investigators (Harrop and Velicer, 1985; 1990) have sugested to suppress the problematic phase of model identification in favor of using some model previously assumed. The current results coincide in conforming the differences found previosly (Vallejo et al., 1992) regarding the efficience of the estimators of the distintic procedures of analyses, not only does not disappear as the models are made more complex, but they deteriorate.

Los métodos de análisis de series temporales han ido ganando progresivamente aceptación entre los investigadores de las ciencias sociales y comportamentales. Tales análisis son de gran utilidad en una gran variedad de aplicaciones, como por ejemplo, establecimiento de los valores futuros de la variable cuyo comportamiento se quiere explicar, caracterización de la estructura estadística de un proceso y, sobre todo, para evaluar los efectos de las intervenciones efectuadas por los experimentadores. Como se han encargado de poner de manifiesto recientemente Huitema y McKean (1991) algunos investigadores en áreas tales como la psicología fisiológica tienen una larga historia en el uso de tales procedimientos, mientras que los investigadores pertenecientes a otras áreas de la psicología su experiencia es más reciente, aunque no por ello menos intensa, a juzgar por las publicaciones aparecidas durante los últimos quince años en algunas de las más importantes revistas pertenecientes al campo de la psicología escolar, social o clí-

Correspondencia a: Guillermo Vallejo Seco.
Facultad de Psicología.
Universidad de Oviedo
c/ Aniceto Sela, s/n. 33005 Oviedo

nica. A título de ejemplo, citaremos tan sólo el *Behavioral Assessment, Behavior Modification, Child Education and Treatment, Journal of Applied Behavior Analysis* o el *Social Science Review.*

Una característica común de los trabajos aparecidos en estas revistas, es que cuando se hace uso del experimento a fin de inferir nexos o relaciones causales, el diseño que más se utiliza es el de series temporales, o alguna variedad de éste. Típicamente, un diseño de series temporales interrumpidas implica efectuar una serie de observaciones repetidas a través del tiempo en una sóla unidad de análisis. En las ciencias comportamentales, la unidad suele ser el sujeto experimental y la interrupción se corresponde con la existencia de una intervención. El análisis compara la serie pre-intervención con la serie post-intervención en orden a evaluar si la introducción del tratamiento lleva asociado la producción de algún cambio. Por ejemplo, cambios que resultan muy familiares, son los cambios en el nivel y en la tendencia de las series.

Los procedimientos estadísticos tradicionales, tales como la prueba *t* de Student o el análisis de la varianza, han jugado un importante papel durante bastantes décadas a la hora de estimar y probar cambios entre las medias de diferentes grupos. Sin embargo, como nos recuerdan Box y Tiao (1975), estas pruebas solamente son válidas si las observaciones previas y posteriores al evento de interés varían en torno a μ_1 y μ_2 no sólo normalmente y con varianza constante, sino también independientemente. Ahora bien, existe una opinión bastante generalizada (Busk y Marascuilo, 1988; Sharpley y Alavosius, 1988; Suen, 1987, Suen y Ary, 1988; Greenwood y Matyas, 1990) según la cuál los datos registrados sucesivamente a lo largo del tiempo carecen de la gracia que generalmente confiere la aleatorización, y son usualmente dependientes y frecuentemente no estacionarios. Consecuentemente, todos aquellos procedimientos estadísticos, tanto paramétricos como noparamétricos, que requieren para su correcta aplicación el supuesto de independencia no deberían emplearse, pues la presencia de autocorrelación puede sesgar sustancialmente los resultados de las pruebas que no la tienen en cuenta.

Al hilo de lo dicho una pregunta que nos surge de inmediato se refiere a qué procedimiento analítico debemos utilizar con el fin de vencer la dependencia que usualmente presentan los datos obtenidos mediante el diseño de las series temporales interrumpidas. Pues bien, debemos manifestar que la solución más prometedora y también más practicada en el campo de las ciencias socio-comportamentales ha consistido en la adaptación efectuada por Glass, Wilson y Gottman (1975) de la técnica de modelamiento ARIMA, desarrollada inicialmente por Box y Tiao (1965) y Box y Jenkins (1970).

El procedimiento de Glass y otros (1975), tal y como es adaptado desde Box y Jenkins a fin de analizar series temporales consta de dos etapas. La primera de ellas, referida comúnmente como fase de identificación (a veces también se la denomina con los nombres de construcción del modelo, o formulación), implica tres pasos iterativos. En primer lugar, en base a información preliminar un modelo es tentativamente identificado. A continuación, basados en el modelo tentativo, los parámetros del modelo correspondiente son estimados. Y, por último, usando los coeficientes estimados, la bondad de ajuste del modelo es evaluada. En el caso supuesto de que el ajuste efectuado sea satisfactorio, el proceso de identificación del modelo es completo. Sin embargo, si el ajuste es insatisfactorio debemos retornar nuevamente al primer paso y tratar de identificar otro modelo tentativo. Comúnmente, el primer paso es referido como fase de especificación y en él la inspección de las funciones de la autocorrelación y de autocorrelación parcial propuestas en el trabajo original de

Box y Jenkins (1970) desempeñan un papel central, a pesar de los nuevos procedimientos de identificación existentes (Uriel, 1985). Al segundo paso se le suele describir con el nombre de fase de estimación de los parámetros del modelo. Si el modelo ha sido correctamente especificado las estimaciones de los parámetros correspondientes a los procesos autorregresivo y de promedio móvil, deben situarse dentro de los límites de estacionariedad e invertibilidad, de no ser así el modelo tentativamente especificado debe ser rechazado considerando de nuevo si las fuentes de no estacionariedad han sido extraidas adecuadamente. El último paso del ciclo de construcción del modelo se conoce con el nombre de fase de validación o fase de diagnóstico. Básicamente, en esta fase se trata de comprobar si los residuales del modelo especificado describen un proceso de ruido blanco. Para ello se comprueba si las funciones de autocorrelación correspondientes a los primeros retardos de los residuales difieren significativamente de cero. Además, de este análisis individual de los residuales del modelo tentativo, también se suele efectuar un análisis global de los coeficientes estimados, para ello se comprueba si el valor de las funciones de autocorrelación de los residuales son uniformemente cero.

Una vez superado el ciclo de construcción del modelo, en la segunda etapa del procedimiento de Glass y otros (1975), los datos son transformados en series independientes con la finalidad de que satisfagan los requerimientos del modelo lineal general.

Como hemos dicho con anterioridad el enfoque Box-Jenkins, adaptado por Glass y otros (1975) a las ciencias socio-comportamentales, ha generado un amplio interés entre los investigadores de estos campos. Las razones para ello son varias, sin embargo, a juicio de Velicer y McDonald (1984, pp. 33-34) las tres que siguen son las más importantes. En primer lugar, los análisis de series temporales son particularmente adecuados para estudiar problemas en ambientes aplicados donde los tradicionales diseños entre sujetos no reflejarían adecuadamente la situación implicada. En segundo lugar, los diseños de series temporales son especialmente apropiados para tratar con cuestiones de causalidad. Y, por último, este tipo de diseños poseen la ventaja adicional de permitirnos estudiar los efectos de diversos patrones de intervención. A nosotros nos gustaría añadir que el procedimiento en cuestión, además de contemplar la dependencia de los residuales, no requiere de una línea base estable y nos proporciona información separada de los cambios de nivel, de las tendencias e inclusive de los ciclos; esto es, de los patrones de cambio operados a lo largo del tiempo.

Sin embargo, el empleo de este enfoque también presenta una serie de desventajas como son la falta de generalización y, sobre todo, el largo número de observaciones que es requerido para identificar adecuadamente el modelo. En relación con el primero de los inconvenientes señalados, debemos decir que el problema nos parece menor, pues además de ya haber dejado constancia en otro lugar (Vallejo, 1991) de nuestro particular punto de vista en los aspectos relacionados con puntos como el que estamos tocando, en la actualidad existen algunas publicaciones de diseños de series de tiempo de carácter trans-seccional que se han ocupado de evaluar la generalización de los efectos de la intervención a través de diversos sujetos, lo cual contribuye sin duda a paliar el defecto olvidado (véase, por ejemplo, Velicer y McDonald, 1991). Obviamente, este tipo de trabajos al comparar múltiples series de tiempo registradas desde una misma unidad experimental o desde unidades diferentes, representa una solución potencial al problema de la falta de generalidad. Por lo que respecta al segundo problema las cosas no están tan claras, o al menos nosotros no somos conscientes de que lo estén, ya que de por sí es bas-

tante difícil precisar cuál es el número mínimo de puntuaciones necesarias para que no aparezcan problemas de fiabilidad y exactitud en la tarea de identificación; sin embargo, son muchos los autores (Box y Jenkins, 1976, p. 33; McCleary y Hay, 1980, p. 20; Glass, Wilson y Gottman, 1975, p. 115) que han sugerido que son requeridas al menos 50 observaciones y preferiblemente 100 por fase experimental para llevar a cabo la especificación del modelo con ciertas garantías de éxito.

Así pues, para que el enfoque ARIMA describa correctamente el componente estocástico de una serie de carácter cronológico es necesario ajustar adecuadamente el modelo mediante la correcta identificación de la parte sistemática de dicho componente. Esta parte es la responsable de la dependencia serial y refleja la acción de los errores aleatorios previos sobre las observaciones actuales, y dado que, la especificación de esta estructura está influenciada por las fluctuaciones aleatorias de las puntuaciones, el modelo descrito por los errores será mejor identificado cuanto mayor sea el número de datos con los que contemos.

Con el fin de evitar los problemas surgidos en la etapa de especificación del modelo, dado que como acabamos de exponer el número mínimo de puntos exigidos para identificar con ciertas garantías el modelo puede resultar prohibitivo, sobre todo, en diferentes ámbitos aplicados, e inclusive disponiendo de él la correcta identificación puede aún ser problemática, en los últimos años han sido realizadas tres propuestas que prescinden completamente de esta discutible fase. Entre los procedimientos alternativos que evitan la problemática fase de identificación destacan: La asunción de un simple modelo AR(1) en todas las series (Simonton, 1977), la estimación de la matriz de transformación desde los datos empíricos empleando un analísis de perfiles (Algina y Swaminathan, 1977, 1979; Algina y Olejnik, 1982; Swaminathan y Algina, 1977) y el ajuste de los modelos de series temporales asumiendo un modelo autorregresivo de quinto orden, AR(5).

Con la finalidad de recabar evidencia empírica del comportamiento de los modelos asumidos de antemano [ARIMA (1,0,0) y ARIMA (5,0,0)] en relación con el modelo correctamente identificado (MCI), Harrop y Velicer (1985) llevaron a cabo una investigación en la que utilizaron múltiples series con observaciones simuladas mediante el conocido procedimiento de Monte Carlo, excepción hecha del enfoque de la Transformación General que fue aproximado por un modelo ARIMA (3,0,0). Las series generadas eran representativas de 26 modelos ARIMA [(1,0,0) (2,0,0) (0,0,1) (0,1,0) y (0,1,1)] en los cuales, además de manipular el grado de dependencia serial, también manejaron el tamaño muestral mediante la utilización de series de 40 y 100 observaciones. Para cada modelo y tamaño muestral simularon 10 series con un efecto de la intervención abrupto y constante de valor uno y una varianza del término aleatorio ruido blanco igual a la unidad. Los descubrimientos de estos autores, en base a comparar los tres modelos en relación con la varianza del error de los residuales, nivel previo a la intervención y cambio de nivel posterior a la intervención, parecen sugerir que la problemática fase de identificación puede ser eliminada del ciclo de construcción del modelo y reemplazada completamente por cualquiera de los modelos especificados. De este modo, según estos autores, al reducirse tanto el grado de sofisticación estadística como la cantidad de observaciones requeridas, se incrementa el número potencial de aplicaciones de la técnica ARIMA dentro de las Ciencias Sociales, Comportamentales y de la Salud.

Con todo, existen una serie de cuestiones que como mínimo, nos obligan a ser cautelosos con los resultados obtenidos por Harrop y Velicer (1985). En primer lugar, en la fase de identificación puede ocurrir, y de hecho ocurre en este caso, que se especifiquen

varios modelos alternativos y que una vez estimados cada uno de ellos satisfaga los criterios examinados. En segundo lugar, si bien es verdad que en el trabajo de estos autores se incluyen los modelos más frecuentemente encontrados en las Ciencias Sociales, excepción del ARIMA (1,1,0), ARIMA (0,0,2) y del ARIMA (1,0,1), no podemos decir que ocurra lo mismo con la representación de las regiones del área de estacionariedad-invertibilidad. En tercer lugar, el hecho de que no estimasen ni los errores estándar de los parámetros ni las probabilidades de cometer errores Tipo I o errores Tipo II bajo los diferentes enfoques, conduce a que sus conclusiones sean vistas, esencialmente, como incompletas. Por último, como los propios autores reconocen, los resultados de esta investigación también están limitados por el hecho de que el programa de ordenador no fuese programable para obtener procesos autorregresivos de orden más alto y, en menor medida, por el propio algoritmo de cálculo que el programa tenía implementado; ya que para los procesos complejos, trabajaba con valores de los parámetros que variaban con incrementos de 10^{-1} en vez de los valores 10^{-3} ó 10^{-4}, lo cual a decir de Harrop y Velicer le convertía en insuficiente.

A raíz de lo dicho, resulta obvio, que se necesitaban nuevas pruebas con el objeto de poder determinar si las sugerencias de Harrop y Velicer iban más allá de lo meramente tentativo y podemos aceptar la recomendación de eliminar definitivamente la fase de identificación en la construcción empírica del modelo. Pues, como se deduce de lo dicho en el apartado anterior, los autores argumentan convincentemente acerca de las ventajas de suprimir la fase de especificación en la formulación del modelo; sin embargo, las ausencias referidas con anterioridad debilitan su postura. Por este motivo, nosotros (Vallejo y otros, 1992) llevamos a cabo una investigación que se centró principalmente en el estudio de los problemas que restringían el alcance de los resultados descubiertos por Harrop y Velicer (1985). Concretamente en nuestro trabajo, además de centrarnos en la exploración de la sesgadez que mostraban los parámetros estimados bajo los diferentes enfoques, nos concentramos principalmente en el estudio de las tasas de error Tipo I, en la potencia de las pruebas bajo los tres enfoques y en la precisión de las estimaciones mediante la comparación de los errores estándar resultantes.

En orden a evaluar los objetivos expuestos llevamos a cabo el experimento de simulación de Monte Carlo, en el cual, al igual que en el estudio de Harrop y Velicer, manipulamos el tamaño de muestra, el tipo de modelos seleccionados y el valor de los parámetros. Sin embargo, existieron ciertas diferencias en cuanto a los valores de las variables manipuladas y en cuanto a la forma de identificar el modelo. Por lo que respecta a la primera cuestión, en nuestra investigación los tamaños muestrales manipulados fueron N=50 y N=100, los modelos seleccionados fueron los mismos que en el estudio de Harrop y Velicer, excepción hecha de los modelos ARIMA (0,0,0) y ARIMA (0,1,0) que fueron sustituidos por los modelos ARIMA (0,0,2) y ARIMA (1,0,1). A su vez, los valores de los parámetros incluidos en nuestra investigación no representaban solamente las áreas 1, 3, 4 y 6 de las regiones de estacionariedad, como ocurría con el estudio de Harrop y Velicer, sino que eran representativos de las áreas 1, 2, 3, 4, 5 y 6 de las regiones de estacionariedad e invertibilidad delimitadas por las gráficas elaboradas por Stralkowski, Wu y Devor (1970, 1974).

Por lo que respecta a la segunda cuestión, esto es, en cuanto a la forma de especificar el modelo, los autores referidos utilizaron el enfoque de Glass, Wilson y Gottman (1975) descrito con anterioridad; dado que este es el procedimiento que tiene incorporado el programa TSX (desarrollado por Glass, Bower y Padia, 1974) empleado en su investigación. Nosotros por nuestra parte, una vez generadas las series, llevamos a cabo un análi-

regiones de estacionariedad e invertibilidad, en lo tocante a la potencia y a los errores estándar de los estimadores.

Objetivos Generales

A raíz de lo dicho dos dudas nos han incomodado. La primera tiene que ver con la confianza que a nosotros como investigadores nos merecían los descubrimientos obtenidos, pues, si bien es cierto que nuestros resultados fueron en parte coincidentes con los hallados previamente por Harrop y Velicer, lo cual hasta cierto punto era esperable, no lo es menos que también fueron parcialmente divergentes. Las discrepancias aludidas pudieron deberse a otras causas diferentes a las esgrimidas en nuestros trabajos. Así, por ejemplo, podíamos dudar de nuestros resultados por el hecho de que nuestras series de tiempo no estuvieran correctamente generadas e inclusive por el hecho de que nuestro procedimiento de simulación no fuese el más adecuado. El segundo problema es una extensión de la hipótesis adelantada en el apartado anterior; esto es, ¿hasta qué punto es cierto que las diferencias entre los tres enfoques, dejan de ser tales a medida que las realizaciones encontradas dependen de modelos ARIMA más complejos?

En vista de lo dicho, decidimos que era necesario replicar crucialmente nuestro trabajo anterior en base a la manipulación de otros tipos de modelos más complejos, protegiéndonos a la vez de la posible presencia de propiedades extrañas tanto en las series de tiempo generadas como en el procedimiento. Para ello hemos planificado una investigación dividida en tres fases. La primera tiene como objetivo desechar que las series utilizadas por nosotros estuvieran viciadas por la presencia de propiedades extrañas. La segunda, disipar la duda de si nuestro procedimiento de identificar el modelo desde los residuales proporciona evidencia sustancialmente distinta de la disponible hasta la fecha en base a las puntuaciones originales. La tercera, ampliar el estudio de los problemas que restringen el alcance de los descubrimientos anteriores complementando el rango de modelos manipulados y los efectos de la intervención.

Experimento 1

Método

En orden a evaluar el primer objetivo que nos hemos marcado en el apartado anterior hemos diseñado un experimento de simulación de Monte Carlo en base a dos áreas de interés: Tipo de modelo seleccionado y valor de los parámetros manejados.

Con respecto al primer criterio, se han seleccionado los modelos ARIMA (0,0,0), ARIMA (1,0,0), ARIMA (2,0,0), ARIMA (0,0,1) y ARIMA (0,0,2). Esta selección se debe a que estos procesos son, en opinión de varios autores (Gottman y Glass, 1978; Judd y Kenny, 1981; Marsh y Shibano, 1984; Glass, Wilson y Gottman, 1975) representativos de las series que uno puede encontrarse con mayor frecuencia en las ciencias sociocomportamentales.

En lo que se refiere al segundo criterio, los valores de los parámetrosyse han elegido de manera que sean representativos del rango de valores aceptables para satisfacer la condición de estacionariedad ($-1 < \phi_1 < 1$) y de invertibilidad ($-1 < \theta_1 < 1$). Igualmente, tanto los valores de los parámetros de los procesos autorregresivos como de promedio móvil de segundo orden se seleccionaron de forma que quedasen representadas las seis áreas de las regiones de estacionariedad e invertibilidad de acuerdo con las delimitaciones gráficas elaboradas por Stralkowski, Wu y Devor (1970, 1974).

Los modelos seleccionados para este experimento así como los valores elegidos como representativos de los parámetros son los mismos que los empleados por Vallejo y otros en la investigación anterior, con la sola excepción del modelo ARIMA (0,0,0) que fue utilizado en la presente investigación por

primera vez. Los valores de los parámetros correspondientes a los modelos seleccionados son mostrados en la tabla 1.

En orden a asegurar que únicamente muestras con propiedades estocásticas estables eran incluidas en nuestro experimento para cada modelo, en la construcción de las series, cien puntos fueron generados, pero solamente los cincuenta últimos fueron analizados. Para cada una de las 16.500 series (33 modelos × 500 repeticiones) el nivel y cambio del nivel previo y posterior introducción del tratamiento se igualaron a cero y el componente aleatorio ruido blanco lo seleccionamos de forma que estuviera normal e independientemente distribuido con $\mu=0$ y $\sigma=1$. Los datos normales aleatorios fueron generados desde la subrutina GGNML del programa IMSL (1989) versión 9.2. Finalmente, desarrollamos un programa de ordenador para producir series de tiempo simuladas a partir de los modelos ARIMA seleccionados.

Una vez generadas las series temporales llevamos a cabo un análisis con cada una de ellas mediante la prueba *t* de Student para grupos independientes, con la intención de evaluar si se habían producido cambios de nivel estadísticamente significativos. Como hemos indicado más arriba, dicha prueba fue realizada con el propósito de verificar si los datos utilizados en el procedimiento de generación arrojan unas tasas de error Tipo I acorde con lo que cabría esperar, tanto cuando las series constituyen solamente realizaciones de ruido blanco [ARIMA (0,0,0)], como cuando constituyen realizaciones con distintos tipos de dependencia serial [ARIMA (1,0,0), ARIMA (2,0,0), ARIMA (0,0,1) y ARIMA (0,0,2)].

Resultados y Discusión

Las medidas empíricas relativas a la probabilidad de cometer errores Tipo I aparecen en las tablas 1 y 2. Dichas medidas se han obtenido tabulando el porcentaje con el que la hipótesis nula fue indebidamente rechazada a los niveles de significación 0.01, 0.05 y 0.10, tras aplicarse la prueba *t* de Student para grupos independientes. Como se puede apreciar, los resultados mostrados en la tabla 1 para el modelo ARIMA (1,0,0) replican esencialmente los trabajos de Padia (1974, citados por Gottman y Glass, 1978 y Glass y otros, 1975) y de Greenwood y Matyas (1990). Como era de esperar, en ausencia de autocorrelación, es decir bajo el modelo ARIMA (0,0,0), los valores empíricos α de se aproximan muy estrechamente a sus correspondientes valores teóricos ($\alpha = 0.01$, $\hat{\alpha} = 0.012$, $\alpha = 0.05$, $\hat{\alpha} = 0.054$, $\alpha = 0.10$, $\hat{\alpha} = 0.098$). Sin embargo, cuando está presente la dependencia serial las tasas de error Tipo I se desvían notablemente de los valores nominales. Así por ejemplo, y continuando con el modelo ARIMA (1,0,0), cuando los valores del parámetro ϕ_1 son positivos ($\phi_1 = 0.8$, $\phi_1 = 0.4$) el valor empírico de α sobreestima de una manera exagerada el valor teórico. A su vez, cuando los valores del parámetro ϕ_1 son negativos ($\phi_1 = -0.8$, $\phi_1 = -0.4$) el valor empírico de α infraestima ampliamente el valor teórico.

Por lo que se refiere al otro modelo de un solo parámetro, esto es, el modelo ARIMA (0,0,1), si bien, como se muestra en la tabla 2, no hemos encontrado el efecto simétrico que acabamos de referir, si que hemos obtenido que la presencia de dependencia serial provoca tasas muy elevadas de error Tipo I, siendo éstas más exageradas si cabe en aquellos casos en los cuales los valores del parámetro θ_1 eran negativos.

Continuando con el comentario de los resultados obtenidos para los modelos de dos parámetros, podemos comprobar cómo el modelo ARIMA (2,0,0) también arroja tasas de error Tipo I que se desvían marcadamente de los valores nominales. En concreto, observando la tabla 1 se desprende, que para los seis primeros valores de ϕ_1 y ϕ_2 los valores empíricos de α para los parámetros de las seis últimas filas de la tabla 1 infraestiman

Tabla 1
Tasas de error Tipo I derivadas empíricamente usando una prueba t de grupos independientes ($n_1 = n_2 = 25$).

Coeficientes autorregresivos	Nominal		
	$\alpha = 0.010$	$\alpha = 0.050$	$\alpha = 0.100$
ARIMA (1,0,0)			
0.8	0.210	0.428	0.466
0.4	0.092	0.222	0.318
0.0	0.012	0.054	0.098
-0.4	0.000	0.022	0.052
-0.8	0.000	0.000	0.002
ARIMA (2,0,0)			
0.7/0.2	0.478	0.676	0.690
0.4/0.4	0.290	0.482	0.548
1.2/-0.3	0.400	0.564	0.620
1.5/-0.5	0.672	0.708	0.708
0.8/-0.4	0.024	0.094	0.190
1.2/-0.4	0.210	0.366	0.432
-0.7/0.2	0.000	0.000	0.006
-0.4/0.4	0.002	0.014	0.030
-1.2/-0.3	0.000	0.000	0.000
-1.5/-0.5	0.000	0.000	0.000
-0.8/-0.4	0.000	0.002	0.022
-1.2/-0.4	0.000	0.000	0.000

Tabla 2
Tasas de error Tipo I derivadas empíricamente usando una prueba t de grupos independientes ($n_1 = n_2 = 25$).

Coeficientes de media móvil	Nominal		
	$\alpha = 0.010$	$\alpha = 0.050$	$\alpha = 0.100$
ARIMA (0,0,1)			
0.8	0.318	0.526	0.620
0.4	0.248	0.452	0.526
0.0	0.012	0.054	0.098
-0.4	0.248	0.756	0.928
-0.8	0.634	0.910	0.952
ARIMA (0,0,2)			
0.7/0.2	0.062	0.134	0.228
0.4/0.4	0.060	0.152	0.230
1.2/-0.3	0.004	0.054	0.126
1.5/-0.5	0.002	0.028	0.092
0.8/-0.4	0.006	0.032	0.050
1.2/-0.4	0.012	0.030	0.098
-0.7/0.2	0.000	0.006	0.014
-0.4/0.4	0.000	0.010	0.036
-1.2/-0.3	0.000	0.000	0.002
-1.5/-0.5	0.000	0.000	0.000
-0.8/-0.4	0.000	0.000	0.000
-1.2/-0.4	0.000	0.000	0.000

muy notablemente sus correspondientes valores teóricos (repárese que esta asimetría se corresponde con la elección de unas u otras áreas de estacionariedad, a la vez que depende de los valores de los parámetros elegidos dentro de las distintas áreas). Por su parte, en el modelo ARIMA (0,0,2) la mayoría de los valores de los parámetros mostrados en la tabla 2 tienden a infraestimar los valores nominales de α. En los casos donde esto no ocurre, la sobreestimación es mucho menos acentuada que en el modelo ARIMA (2,0,0).

Los resultados obtenidos permiten efectuar muchos comentarios, sin embargo, a nuestro juicio los dos que siguen son los de mayor importancia. El primero se refiere a la presencia de correlación serial positiva, en concreto, a la presencia de correlación serial positiva en los modelos autorregresivos en los diseños de series temporales interrumpidas. Para nosotros este fenómeno tiene importantes consecuencias a nivel práctico, pues como acabamos de ver la correlación de este signo conlleva a un sustancial rechazo de la hipótesis nula cuando de hecho es verdadera, y son precisamente este tipo de autocorrelaciones las más frecuentes en las ciencias socio-comportamentales (Maddala, 1977; Liu, 1989); de ahí que su no detectación y posterior corrección se traduzca en una aceptación muy elevada de intervenciones significativas cuando de hecho no son tales. El segundo es consustancial al problema que venimos discutiendo. Y se refiere al hecho de si las series utilizadas en nuestra in-

vestigación gozan de las propiedades deseables, o si por el contrario no han sido generadas correctamente. Pues bien, a tenor de los resultados que acabamos de mostrar en las tablas 1 y 2, pensamos que nada se opone a que dudemos de la honradez de nuestras series temporales. Para reforzar esta idea queremos recordar al lector que cuando las series carecían de dependencia serial los valores estimados de α tendían a aproximarse estrechamente a sus correspondientes valores esperados. Además, también existen otros dos aspectos que refuerzan nuestra conclusión. Por un lado, merece la pena destacar que cuando las series fueron generadas a partir del modelo ARIMA (1,0,0) nuestros resultados replican adecuadamente los resultados encontrados previamente por Padia (1974) y por Greenwood y Matyas (1990) y, por otro, que también son plenamente coincidentes las expectaciones teóricas concernientes a la ejecución de pruebas estadísticas clásicas cuando las asunciones del modelos son incumplidas (p.e. Scheffé, 1959, pp. 338-339).

Experimento 2

Método

En orden a evaluar si el procedimiento seguido en la identificación de las series a partir del análisis de los residuales como ha sido sugerido por algunos investigadores (Huitema, 1985; Chatfield, 1989; Gorsuch, 1983; Judd y Kenny, 1981), en vez de hacerlo desde las puntuaciones previas a la intervención o algún otro de los métodos descritos con anterioridad, como es lo corriente, pudo haber sido el responsable de las diferencias encontradas entre los trabajos de Harrop y Velicer (1985) y el de Vallejo y otros (1992), hemos analizado las funciones de autocorrelación y autocorrelación parcial de parte de las salidas proporcionadas por el programa SPSS-X (1988) y utilizadas en el trabajo de Vallejo y otros para identificar las series cuyo valor de los parámetros no se desviaban más de doscientas centesimas de los valores de los parámetros utilizados en la generación de las series. Como ya fue dicho con anterioridad, en nuestro trabajo precedente no reparábamos en el análisis de las funciones citadas, tan sólo utilizábamos las salidas para decidir qué series iban a ser utilizadas en la transformación posterior.

Así pues, al habernos comportado de la manera aludida no podemos saber si nuestro procedimiento de identificar las series a partir de los residuales pudo influir en los resultados obtenidos, lo que si sabemos es que encontrar las series para alguno de los modelos utilizados fue muy laborioso, por ejemplo, las del modelo ARIMA (1,0,1) para los parámetros de las áreas de estacionariedad e invertibilidad 3 y 6. Por este motivo, con el fin de comprobar hasta que punto el procedimiento seguido pudo condicionar los resultados alcanzados, hemos llevado a cabo un examen de las funciones de autocorrelación y autocorrelación parcial de 100 de las series generadas para los modelos ARIMA (1,0,0) ARIMA (0,0,1) y ARIMA (0,1,1), para N=50 y N=100 y para cuatro valores de dependencia a los que nos hemos venido refiriendo ($\pm$ 0.4 y $\pm$ 0.8). Dicho examen, además de permitirnos verificar si nuestras expectaciones teóricas [p.e., la identificación del modelo generador será tanto más fácil cuanto mayor sea el valor de N, las series generadas bajo el modelo ARIMA (1,0,0) serán más fáciles de identificar que las series generadas bajo los otros dos modelos, etc] coinciden con los resultados obtenidos, también nos va a permitir comprobar el grado de coincidencia con los resultados reportados por otros investigadores (p.e., Uriel, 1985; Matyas y Greenwood, 1991; Velicer y Harrop, 1983) empleando modelos y/o procedimientos de análisis similares.

Bajo el modelo ARIMA (1,0,0) el número de series que al final presentaba una identificación clara lo obtuvimos verificando las dos propiedades siguientes de las funciones de autocorrelación (FAC) y de autocorrela-

ción parcial estimadas (FACP): En primer lugar, si los signos de los cinco primeros valores de las FAC estimadas coincidían con los cinco primeros valores de las FAC derivadas teóricamente. En segundo lugar, verificando si únicamente el primer valor de las FACP estimadas excedía los límites o bandas de confianza ($\pm 2SE$; $SE_{kk} = \sqrt{1/N}$), esto es, si era significativamente distinto de cero al aplicar el testa de Quenouille (1949). Solamente en aquellos casos en los cuales los dos criterios se mostraron coincidentes aceptamos que las series generadas desde el modelo ARIMA (1,0,0) estaban identificadas correctamente.

De igual modo, el número de series que bajo el modelo ARIMA (0,0,1) podía considerarse con identificación clara también lo hallamos examinando las dos propiedades anunciadas más arriba de las FAC y FACP estimadas. En concreto, en este caso lo que hicimos fue verificar, primeramente, si las distintas series generadas bajo el modelo MA(1) poseían una FAC estimada con $\hat{\rho}_1$ como único coeficiente significativamente distinto de cero. A tal efecto comprobamos si era el único coeficiente que excedía los límites o bandas de confianza

$$[\pm 2SE;\ SE_k = \sqrt{1/N\ (1+2\ \sum_{i=1}^{k} \hat{\rho}_1^2)}\ ].$$

Seguidamente, verificamos si los signos de los cinco primeros valores de las FACP estimadas coincidían con los signos de los cinco primeros valores de las FACP derivadas teóricamente. Al igual que en el caso anterior tan sólo aquellas series que poseían simultáneamente las dos propiedades relatadas en sus FAC y FACP estimadas eran retenidas. Finalmente, por lo que respecta al modelo ARIMA (0,1,1), las series retenidas por presentar una identificación clara fueron obtenidas siguiendo el procedimiento descrito para la identificación de las series del modelo ARIMA (0,0,1), obviamente tras la previa diferenciación de las series (d=1).

Resultados y Discusión

En la tabla 3 presentamos el porcentaje de identificaciones correctas para los tres modelos analizados y para los tamaños muestrales utilizados, promediados a través de las cien replicaciones empleadas y de los diferentes valores de los parámetros de dependencia.

Tabla 3
Porcentaje de identificaciones claras bajo los modelos AR(1), MA(1) e IMA(1,1)

	Número de observaciones		
Tipo de modelo	N=50	N=100	Total
ARIMA (1,0,0)	39.00 (400)*	44.25 (400)	**41.625** (800)
ARIMA (0,0,1)	32.25 (400)	40.75 (400)	**36.625** (800)
ARIMA (0,1,1)	17.50 (400)	21.75 (400)	**19.625** (800)
TOTAL	**29.66** (1200)	**35.58** (1200)	**32.625** (2400)

*Los valores que van entre paréntesis representan el número total de identificaciones posibles.

Como se desprende de la tabla el porcentaje global de identificaciones correctas fue aproximadamente del 33% (783 de 2400). Observando los porcentajes totales representados en las columnas podemos comprobar cómo las dificultades de identificación disminuyen conforme se incrementa el número de observaciones de que se dispone. Similarmente, desde los porcentajes totales correspondientes a las filas se desprende cómo las mayores dificultades de identificación acaecen bajo el modelo ARIMA (0,1,1), y las menores con el modelo ARIMA (1,0,0); por su parte, el modelo ARIMA (0,0,1) ocupa una dificultad de identificación intermedia.

A continuación, en la tabla 4 presentamos

Tabla 4
Porcentaje de identificaciones claras clasificadas por grado de dependencia y tipo de modelo

Tipo de modelo	Grado de dependencia	Número de observaciones N = 50	N = 100	Totales
ARIMA (1,0,0)	0.80	51	60	55.5
	0.40	14	22	18.0
	-0.40	26	28	27.0
	-0.80	65	67	66.0
ARIMA (1,0,1)	0.80	58	64	61.0
	0.40	17	28	22.5
	-0.40	12	19	15.5
	-0.80	43	52	47.5
ARIMA (1,1,1)	0.80	32	35	33.5
	0.40	11	14	12.5
	-0.40	4	9	6.5
	-0.80	23	29	26.0
TOTAL		**29.66** (1200)	**35.58** (1200)	**32.625** (2400)

los porcentajes de identificaciones correctas para los tres modelos ARIMA, para los tamaños muestrales utilizados y para los diferentes valores de los parámetros empleados.

Ciñiéndonos al modelo ARIMA (1,0,0) los resultados de la tabla 4 ponen de relieve tres aspectos importantes. En primer lugar, que los modelos ARIMA (1,0,0) presentan mayores dificultades de identificación cuando los valores del parámetro ϕ_1 eran positivos ($\phi_1 = 0.8$, $\phi_1 = 0.4$) que cuando los valores del parámetro ϕ_1 eran negativos ($\phi_1 = 0.8$, $\phi_1 = -0.4$). En concreto, cuando los valores del parámetro ϕ_1 eran positivos el porcentaje total de series correctamente identificadas resultó ser del 36.75%, mientras que cuando los valores del parámetro ϕ_1 eran negativos el porcentaje de series correctamente identificadas fue del 46.50%. En segundo lugar, reseñar que el incremento del tamaño de muestra, además de llevar parejo una disminución de las dificultades de identificación de las series (29.66% frente a 35.58%), también tiende a reducir las diferencias entre el grado de dificultad que presenta la identificación de modelos autorregresivos con $\phi_1 > 0$ y $\phi_1 < 0$ Por ejemplo, cuando el número de observaciones era de 50 la diferencia entre los porcentajes de series correctamente identificadas bajo $\phi_1 > 0$ y $\phi_1 < 0$ y fue 8.5; mientras que cuando el número de observaciones era de 100, la diferencia entre los porcentajes de series correctamente identificadas bajo $\phi_1 > 0$ y $\phi_1 < 0$ fue 6.5. Y, por último, que el grado de dificultad que presenta la identificación del modelo ARIMA (1,0,0) disminuye a medida que el parámetro autorregresivo se aproxima al límite de la región de estacionariedad.

Por lo que respecta al modelo ARIMA (0,0,1), los resultados esquematizados en la tabla 4 ponen de relieve que, al igual que sucedía con el modelo AR(1), las dificultades de identificación se reducían a medida que incrementábamos el número de datos disponibles y conforme el valor absoluto del

parámetro de promedio móvil θ_1 se aproximaba al límite de la región de invertibilidad. Además, y al revés de lo que ocurría con el modelo anterior, con el mismo valor absoluto del parámetro θ_1, las dificultades de identificación fueron mayores cuando los valores del parámetro eran negativos ($\theta_1 = -0.8$, $\theta_1 = -0.4$) que cuando los valores del parámetro eran positivos ($\theta_1 = 0.8$, $\theta_1 = 0.4$). En concreto cuando los valores deeran negativos el porcentaje de series correctamente identificadas fue tan sólo del 31.5%, mientras que cuando los valores del parámetro θ_1 eran positivos, el porcentaje de series con identificación clara ascendió al 41.75%. Resultados similares a los comentados en este apartado fueron observados para el modelo ARIMA (0,1,1), nos referimos claro está desde el punto de vista cualitativo, pues desde el punto de vista cuantitativo resulta obvio que las dificultades de identificación se incrementaron sustancialmente; de todos modos, este hecho era esperable y concuerda estrechamente con los resultados obtenidos por Velicer y Harrop (1983) utilizando los mismos valores del parámetro ($\pm$ 0.4 y $\pm$ 0.8), pero distinta metodología.

En el trabajo de estos autores se utilizaron otros tres modelos, además de los utilizados por nosotros, los tamaños muestrales fueron N=40 y N=100 y la fase de identificación se basaba en las decisiones de dos grupos de 6 sujetos cada uno, previamente entrenados en el enfoque de Glass y otros (1975). Si bien Velicer y Harrop (1983, p. 554) nos dicen que el procedimiento de identificación fue llevado a cabo examinando las FAC y FACP estimadas haciendo uso del programa CORREL (Glass y Bower, 1974), no nos especifican qué propiedades de estas funciones fueron exactamente examinadas. Sea como fuere sus resultados y los nuestros son bastante coincidentes. Tan sólo hemos encontrado una discrepancia notable y que, por supuesto, nos ha llamado la atención. Concretamente, ellos encontraron que el porcentaje de identificaciones correctas bajo el modelo ARIMA (1,0,0) fue del 32% y nosotros del 41%, mientras que para el modelo ARIMA (0,0,1) el porcentaje de identificaciones correctas encontradas en el trabajo de Velicer y Harrop es del 56%, mientras que en nuestro caso fue tan sólo del 37%. Esto no quiere decir que dudemos de nuestros resultados, sino antes bien que nos sorprende el que estos autores obtuvieran un mayor porcentaje de identificaciones claras para el modelo MA(1), que para el modelo AR(1), cuando en pura teoría les debería haber ocurrido precisamente lo contrario. Pues, como se puede observar desde las derivaciones teóricas en una estructura de promedios móviles los coeficientes teóricos que configuran las FAC y FACP tienen cotas inferiores a las que configuran las FAC y FACP teóricas de las estructuras autorregresivas (Vallejo, en prensa). En concreto, desde el trabajo citado, se puede comprobar cómo en el modelo MA(1) el valor absoluto de ρ es ≤ 0.5.

Por último, apuntar que un aspecto que sin duda puede llamar la atención del atento lector, es la dualidad existente entre los procesos autorregresivos y de promedio móvil; lo cual como hemos anunciado, conduce a que en los procesos AR(1) cuando la autocorrelación es positiva el número de identificaciones sea menor que cuando la autocorrelación es negativa; mientras que en el modelo MA(1) el comportamiento es el inverso del señalado para el modelo AR(1). Nosotros en el presente trabajo no nos hemos detenido a investigar el porqué de este hecho, no obstante, si nos gustaría adelantar que tal vez esta cuestión pueda ser adecuadamente manejada apelando al posible sesgo de los coeficientes estimados.

Experimento 3

Método

Para evaluar el tercer objetivo hemos diseñado un nuevo experimento de simulación de Monte Carlo en base a tres áreas de interés: Tamaño del efecto de la intervención,

tipo de modelos seleccionados y valor de los parámetros manejados.

En relación con el primer criterio, todas las intervenciones ocurrían en el punto medio de las series, esto es, en el punto 26. En todos los casos el cambio de nivel fue abrupto y permanente, mientras que el tamaño de la intervención utilizado en unos casos fue de una desviación estándar y en otros de una desviación estándar y media (I=1, I=1.5). Intervenciones de esta clase son las más comúnmente descubiertas en las ciencias socio-comportamentales, o al menos las que son con mayor frecuencia hipotetizadas y probadas. Cada uno de estos efectos que podemos considerar de suficiente magnitud como para ser detectado analíticamente, pero lo bastante pequeños como para poderse detectar gráficamente de una manera fiable, fue combinado con un cambio de tendencia pre y post-intervención (CT). Además, son idénticos a los que emplearon Harrop y Velicer (1990) en una evaluación cuantitativa de los programas TSX, GENTS, BMDP-2T, SAS-ETS e ITSE en la que utilizaron los mismos modelos y valores de dependencia que en su anterior estudio de 1985. Los valores seleccionados de 15° para *T* y de –15° para *CT* nos parecen adecuados, pues al igual que sucedía con el tamaño de los efectos de la intervención, estas elevaciones y descensos representan valores que son lo suficientemente pequeños como para ser discriminados a ojo, pero de suficiente magnitud como para ser detectados analíticamente.

Con respecto al segundo criterio, hemos seleccionado los modelos ARIMA (1,1,0) y ARIMA (0,1,1). La razón principal para incluir modelos más complejos, reside en nuestra inquietud por saber si las diferencias encontradas anteriormente entre los tres enfoques para los modelos AR(1), AR(2), MA(1) y MA(2) se mantienen, o si por el contrario tienden a disminuir como sucedía previamente con los modelos IMA (0,1,1) y ARMA (1,0,1). Antes de proseguir advertir de tres cuestiones. En primer lugar, que si eliminamos la tendencia de los modelos actuales, los órdenes de los parámetros autorregresivos y de promedio móvil son los mismos que los de nuestro trabajo anterior. Como fue dicho, el motivo que determinó nuestra elección se debió a que los trabajos diseñados para tal fin han puesto de relieve que existen muchas posibilidades de que estos modelos sean representativos de una buena parte de las series encontradas en las ciencias socio-comportamentales (veánse, por ejemplo, Marsh y Shibano, 1984; Glass y otros, 1975; Revenstorf y otros, 1980). En segundo lugar, el modelo manipulado actualmente, esto es, el ARIMA (1,1,0) lleva incorporado una tendencia determinística; mientras que el modelo ARIMA (0,1,1) lleva incorporado tanto una tendencia determinística, como estocástica. Y, por último, que los modelos manipulados incorporan cuatro parámetros, no tan sólo dos como ocurría con la investigación precedente. En concreto, estos modelos contemplan el nivel previo a la introducción del tratamiento, la tendencia, el cambio de nivel debido a la intervención y el cambio de tendencia de la fase post-tratamiento. En la tabla 5 aparecen las ecuaciones correspondientes a los modelos seleccionados en el presente experimento.

Por lo que respecta al último criterio, como en las investigaciones precedentes, los valores de los parámetros ϕ_1 y θ_1 se han elegido de manera que fuesen representativos del rango de valores aceptables para satisfacer la condición de estacionariedad ($-1<\phi_1<1$) y de invertibilidad ($-1<\theta_1<1$).

Al igual que en los experimentos anteriores, en orden a asegurar que únicamente muestras con propiedades estocásticas estables eran incluidas en nuestro experimento para cada modelo, en la fase de construcción de las series, cien puntos fueron generados, pero tan sólo los cincuenta últimos han sido empleados. Para cada una de las 2.400 series (8 modelos × 100 repeticiones × 3 tamaños del efecto) el nivel previo a la presentación del tratamiento se igualó a cero por simpli-

Tabla 5.
Ecuaciones modelo generadoras de las series temporales utilizadas en el experimento

Modelo ARIMA (1,1,0)

$$Pre\text{-}I:\ Y_{n_1} = L + \phi_1 (Y_{n-1} - L) + \lambda + \varepsilon_{n_1}$$

$$Post\text{-}I:\ Y_N = L + \phi_1 (Y_{N-1} - L - I) + \lambda + I + \Delta + \varepsilon_N$$

Modelo ARIMA (0,1,1)

$$Pre\text{-}I:\ L + (n_1 - 1)\,\lambda\,(1 - \theta_1) + \lambda + (1 - \theta_1)\,(\varepsilon_1 + \ldots + \varepsilon_{n_1-1}) + \varepsilon_{n_1}$$

$$Post\text{-}I:\ L + (N - 1)\,\lambda\,(1 - \theta_1) + \lambda + I + \Delta + (n_2 - 1)\,\Delta\,(1 - \theta_1) + (1 - \theta_1) \sum_{i=1}^{N-1} \varepsilon_i + \varepsilon_N$$

cidad y el cambio de nivel posterior a la intervención fue abrupto y permanente y en 800 series lo fijamos en 1, en otras 800 lo fijamos en 1.5 y en las restantes en cero con el fin de calcular la probabilidad de cometer errores Tipo I. A su vez, el componente aleatorio ruido blanco lo seleccionamos de forma que estuviera normal e independientemente distribuido con $\mu = 0$ y $\sigma^2 = 1$. Los datos normales aleatorios fueron generados con la subrutina GGNML del programa IMSL (1989), versión 9.2. Posteriormente, mediante un programa de ordenador en el cual estaban programadas las secuencias de la tabla 5 producimos las series de tiempo utilizadas en nuestra investigación.

Resultados y Discusión

Estimadores empíricos del nivel previo a la introducción del tratamiento (L), de la tendencia pre-intervención (λ), del cambio de nivel posterior a la presentación del tratamiento (I) y del cambio de tendencia post-intervención (Δ) para cada uno de los modelos ARIMA [(1,1,0) y (0,1,1)] y bajo los dos tamaños de I (1,1.5) utilizados son presentados en las tablas 6 y 7.

El valor criterio para el nivel previo L, se corresponde con el valor utilizado en la generación de las series; es decir, cero. Lo mismo sucede con los valores de los parámetros restantes. Esto es, el valor criterio de λ se corresponde con 0.27, el valor criterio de I con 1 y 1.5 desviaciones estándar y el valor criterio de Δ con –0.27.

En general, los datos ponen de relieve que los tres procedimientos utilizados en nuestros analísis [MCI, AR(1) y AR(5)] proporcionan resultados muy similares para los efectos de *L*, λ, *I* y Δ, tanto para los promedios parciales dentro de cada uno de los diferentes modelos ARIMA, como para el promedio global efectuado a través de ellos. Sin embargo, las desviaciones típicas (S) son sistemáticamente más elevadas en los enfoques AR(1) y AR(5), lo cual no hace más que corroborar los resultados de nuestro trabajo anterior.

Por lo que al sesgo se refiere, si comparamos los estimadores empíricos con los valores de los parámetros conocidos de la población (*L=0*, λ=0.27, *I*=1, *I*=1.5 *y* Δ=0.27), observamos que los valores promediados de los estimadores $\hat{\lambda}$ *y* $\hat{\Delta}$ se aproximan estrechamente al de los parámetros λ *y* Δ, esto es, $E(\hat{\lambda}) \approx \lambda$ y $E(\hat{\Delta}) \approx \Delta$. En concreto, el valor estimado de λ presenta un sesgo positivo de dos centésimas, mientras que el valor estimado de Δ manifiesta un sesgo negativo de tres centésimas. Por su parte, los valores promediados de los estimadores $\hat{L}$ y $\hat{I}$ difieren

Tabla 6.
Estimación del nivel preintervención, tendencia preintervención, cambio de nivel y cambio de tendencia promediados a través de los modelos (tamaño del efecto de la intervención = 1)

Modelo	MCI Idéntico		AR(1) ARIMA(1,0,0)		AR(5) ARIMA(5,0,0)	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
Nivel pre-intervención						
(1,1,0)	-0.600	0.829	-0.393	1.117	-0.419	1.081
(0,1,1)	-0.130	0.384	-0.032	1.146	0.021	0.712
	-0.365	**0.606**	**-0.212**	**1.131**	**0.199**	**0.896**
Tendencia pre-intervención						
(1,1,0)	0.296	0.181	0.284	0.195	0.328	0.262
(0,1,1)	0.305	0.223	0.273	0.449	0.294	0.317
	0.301	**0.202**	**0.278**	**0.322**	**0.311**	**0.289**
Cambio de nivel post-intervención						
(1,1,0)	0.930	0.528	1.070	1.100	1.126	0.847
(0,1,1)	1.461	0.495	1.301	0.713	1.359	0.699
	1.195	**0.515**	**1.185**	**0.906**	**1.242**	**0.773**
Cambio de tendencia post-intervención						
(1,1,0)	-0.320	0.317	-0.289	0.342	-0.337	0.412
(0,1,1)	-0.332	0.215	-0.279	0.431	-0.314	0.394
	-0.326	**0.176**	**-0.284**	**0.386**	**-0.325**	**0.403**

ligeramente del valor de los parámetros *L* e *I*, así mientras el valor estimado de *L* presenta un sesgo negativo de casi dos centésimas (–0.1846), el valor estimado de *I* presenta un sesgo positivo de algo más de una décima para el promedio de las dos intervenciones (0.095 y 0.1383). Estos resultados replican razonablemente los descubrimientos de los trabajos previos (Harrop y Velicer, 1985; Vallejo y otros, 1992).

Como en nuestro trabajo anterior, las pruebas referidas a la propiedad de insesgadez de los estimadores no nos son de gran utilidad a la hora de discriminar entre las tres rutinas de análisis, ya que como hemos anunciado las diferencias más acentuadas que presentan los datos analizados se refieren a las mayores desviaciones típicas encontradas bajo los enfoques AR(1) y AR(5) para las estimaciones de $\hat{L}$, $\hat{\lambda}$, $\hat{I}$, y $\hat{\Delta}$. Por este motivo vamos a pasar a examinar la precisión de los tres procedimientos de estimación.

Conclusiones concernientes a la relativa eficiencia de los enfoques MCI, AR(1) y AR(5) son rápidamente extraidas comparando los respectivos errores estándar (SE) presentados en la tabla 8 para cada uno de los diferentes valores del parámetro autorregresivo ϕ y del parámetro de promedios móviles θ a través de los modelos ARIMA (1,1,0) y ARIMA (0,1,1). Como puede obaservarse desde la tabla 8, los errores estándar calculados en función de los dos tamaños de la intervención (I=1 e I=1.5) para cada uno de los efectos estimados de los tratamientos arrojan un alto grado de coincidencia bajo los tres enfoques (aunque no se ilustra lo mismo sucedía para los efectos estimados de *L*, λ y Δ.

Tabla 7

Estimación del nivel preintervención, tendencia preintervención, cambio de nivel y cambio de tendencia promediados a través de los modelos (tamaño del efecto de la intervención = 1.5)

Modelo	MCI Idéntico $\bar{X}$	s	AR(1) ARIMA(1,0,0) $\bar{X}$	s	AR(5) ARIMA(5,0,0) $\bar{X}$	s
Nivel pre-intervención						
(1,1,0)	-0.512	0.713	-0.386	0.805	-0.357	0.791
(0,1,1)	-0.217	0.253	-0.110	0.986	0.121	0.675
	-0.364	**0.483**	**-0.248**	**0.895**	**-0.118**	**0.733**
Tendencia pre-intervención						
(1,1,0)	0.282	0.127	0.273	0.341	0.279	0.337
(0,1,1)	0.294	0.216	0.294	0.312	0.306	0.324
	0.288	**0.171**	**0.283**	**0.326**	**0.292**	**0.331**
Cambio de nivel post-intervención						
(1,1,0)	-1.372	0.453	-1.439	1.310	-1.481	1.592
(0,1,1)	-2.010	0.519	-1.637	1.235	-1.893	1.093
	-1.690	**0.486**	**-1.538**	**1.272**	**-1.687**	**1.342**
Cambio de tendencia post-intervención						
(1,1,0)	-0.302	0.140	-0.309	0.227	-0.296	0.215
(0,1,1)	-0.316	0.243	-0.291	0.348	-0.282	0.317
	-0.309	**0.191**	**-0.300**	**0.287**	**-0.289**	**0.266**

Tabla 8

Estimación del efecto del tratamiento y de su correspondiente error estándar para cada uno de los diferentes valores de ϕ *y* θ de los dos procesos ARIMA a través de los tres procedimientos analíticos y para I=1 e I=1.5

Modelo	Valor de los parámetros	Idéntico α	Idéntico $\hat{\alpha}$	Idéntico SE	ARIMA(1,0,0) α	ARIMA(1,0,0) $\hat{\alpha}$	ARIMA(1,0,0) SE	ARIMA(5,0,0) α	ARIMA(5,0,0) $\hat{\alpha}$	ARIMA(5,0,0) SE
Tamaño del efecto = 1										
(1,1,0)	0.8	1	0.97	0.75	1	1.08	0.81	1	1.23	0.79
(1,1,0)	0.4	1	1.09	0.91	1	1.12	0.93	1	1.16	0.96
(1,1,0)	-0.4	1	0.86	1.04	1	1.14	1.31	1	1.12	1.19
(1,1,0)	-0.8	1	0.78	1.02	1	0.94	1.29	1	0.99	1.34
Promedio global		**1**	**0.93**	**0.93**	**1**	**1.07**	**1.09**	**1**	**1.12**	**1.07**
(0,1,1)	0.8	1	1.15	0.46	1	1.08	0.72	1	1.19	0.94
(0,1,1)	0.4	1	1.05	1.20	1	1.59	1.31	1	1.64	1.25
(0,1,1)	-0.4	1	2.14	1.03	1	1.88	1.43	1	1.71	1.38
(0,1,1)	-0.8	1	1.49	0.38	1	0.69	2.14	1	0.86	1.59
Promedio global		**1**	**1.46**	**0.76**	**1**	**1.31**	**1.40**	**1**	**1.35**	**1.29**

Tabla 8
(Continuación)

Modelo	Valor de los parámetros	Idéntico			ARIMA(1,0,0)			ARIMA(5,0,0)		
		α	$\hat{\alpha}$	SE	α	$\hat{\alpha}$	SE	α	$\hat{\alpha}$	SE
Tamaño del efecto = 1.5										
(1,1,0)	0.8	1.5	1.53	0.75	1.5	1.61	0.99	1.5	1.67	0.90
(1,1,0)	0.4	1.5	1.42	0.91	1.5	1.54	1.14	1.5	1.54	1.14
(1,1,0)	-0.4	1.5	1.37	0.99	1.5	1.32	1.07	1.5	1.46	1.18
(1,1,0)	-0.8	1.5	1.17	0.85	1.5	1.29	1.25	1.5	1.25	1.14
Promedio global		**1.5**	**1.37**	**0.87**	**1.5**	**1.44**	**1.11**	**1.5**	**1.48**	**1.09**
(0,1,1)	0.8	1.5	1.86	0.54	1.5	1.63	1.12	1.5	1.74	1.02
(0,1,1)	0.4	1.5	1.55	1.17	1.5	1.71	1.17	1.5	1.89	1.31
(0,1,1)	-0.4	1.5	2.64	0.98	1.5	2.11	1.34	1.5	2.40	1.52
(0,1,1)	-0.8	1.5	1.99	0.49	1.5	1.07	.93	1.5	1.53	1.35
Promedio global		**1.5**	**2.01**	**0.80**	**1.5**	**1.63**	**1.39**	**1.5**	**1.89**	**1.30**

Sin embargo, cuando comparamos los errores estándar de los tres procedimientos de estimación dentro de cada una de las condiciones manipuladas del tamaño del efecto vemos que los SE son diferentes. En todos los casos, los SE ligados al enfoque MCI fueron más pequeños que los SE vinculados a los enfoques asumidos de antemano, procesos AR(1) y AR(5). Así pues, si conjugamos este hecho con el descubrimiento señalado anteriormente de que las estimaciones de los efectos de los tratamientos eran esencialmente las mismas para los tres enfoques, podemos concluir que el procedimiento MCI una vez más nos proporciona estimaciones relativamente más eficaces que las obtenidas bajo los dos enfoques asumidos de antemano. Obviamente, estos resultados no confirman nuestras expectativas en el sentido de que las diferencias entre los tres enfoques se desvanecerían a medida que los modelos utilizados en la generación de las series se volvieran más complejas; al revés, los datos comentados se obstinan en hacer patentes las diferencias puestas de relieve en nuestro trabajo anterior con una representación más amplia de modelos.

Por último, las estimaciones empíricas concernientes a la probabilidad de cometer errores Tipo I y a la potencia de prueba de cada uno de los tres enfoques para los diferentes modelos ARIMA utilizados en la generación de las series se muestran en la tabla 9.

Como se puede apreciar observando la tabla 9, los datos indican que con respecto a la tasa de error Tipo I las tres rutinas de análisis arrojan resultados bastante similares. Reseñar que en todos los casos analizados el estimador empírico α sobrepasa el valor teórico especificado, siendo este hecho más destacable cuando el nivel de α era el 0.01. En concreto, como se puede comprobar echando una ojeada a la tabla 9, cuando las series generadas a partir de los modelos ARIMA (1,1,0) y ARIMA (0,1,1) eran analizadas por medio de los enfoques MCI y AR(1), respectivamente, el valor empírico de α resultó ser más de dos veces el valor teórico ($\hat{\alpha} = 0.025$, $\hat{\alpha} = 0.022$). En ninguno de los casos restantes ocurrió un fenómeno similar, esto es, ni cuando el nivel de significación al $\alpha = 0.01$, ni cuando lo fijamos al $\alpha = 0.05$.

Tabla 9
Estimación empírica de la tasa de error Tipo I y de la potencia

				Potencia de prueba			
Modelos ARIMA	Procedimientos analíticos	P (Error Tipo I)		I = 1		I = 1.5	
		α = **0.01**	α = **0.05**	α = **0.01**	α = **0.05**	α = **0.01**	α = **0.05**
(1,1,0)	MCI	0.025(10)*	0.057(23)	0.625(250)	0.690(276)	0.657(263)	0.757(303)
	AR(1)	0.012(5)	0.067(27)	0.570(228)	0.627(251)	0.607(243)	0.672(269)
	AR(5)	0.015(6)	0.070(28)	0.595(238)	0.665(266)	0.612(245)	0.720(288)
(0,1,1)	MCI	0.017(7)	0.075(30)	0.592(237)	0.635(254)	0.650(260)	0.702(281)
	AR(1)	0.022(9)	0.077(31)	0.520(208)	0.550(220)	0.575(230)	0.622(249)
	AR(5)	0.015(6)	0.090(36)	0.560(224)	0.607(243)	0.612(245)	0.637(255)

* Los números que van entre paréntesis indican el número de veces que la hipótesis alternativa es aceptada incorrectamente en la parte izquierda y correctamente en la parte derecha de la tabla.

Con todo, nos gustaría poner de relieve, que si bien es verdad que las estimaciones empíricas de α excedieron ligeramente los valores teóricos, implicando de esta manera que los tres procedimientos de análisis eran levemente liberales, no es menos cierto que tanto para el umbral del $\alpha = 0.01$, como para el umbral del $\alpha = 0.05$, las estimaciones empíricas de α se encontraban en todos los casos dentro de los límites de confianza de los dos errores estándar de los valores de α teóricos. Es decir, que los $\hat{\alpha}$ se encontraban dentro de $\pm 2SE$ ($SE = \sqrt{pq/N}$). Consecuentemente, este resultado sugiere que la tendencia liberal de las estimaciones aludidas pudo ser debida tan solamente a variación al azar.

Para ir terminando, por lo que a la potencia de prueba ser refiere, los datos de la tabla 9 ponen de relieve que cuando los efectos de los tratamientos están presentes (I=1 e I=1.5) el procedimiento MCI es siempre más poderoso a la hora de detectar dichos efectos que cualesquiera de los otros dos procedimientos. Esta tendencia se mantiene inalterable a lo largo de los dos modelos empleados en la generación de las series. Adicionalmente, desde la tabla 9 también se desprende que el enfoque de la Transformación General se muestra sistemáticamente más eficaz que el enfoque AR(1), tanto cuando la magnitud del efecto de la intervención era de una sigma, como cuando era de una sigma y media. Además, la potencia de prueba también se vio favorecida por el incremento de la magnitud del efecto.

Todos estos resultados están en la línea de nuestros descubrimientos anteriores y, por tanto, vienen a cuestionar nuestras expectativas de que las diferencias entre los tres procedimientos de análisis se irían estrechando a medida que los modelos se vuelven más complicados. En efecto, no es este hecho el que nosotros hemos constatado (dado que las diferencias se siguen manteniendo, al menos para los dos modelos que hemos analizado), sino otro que nos hace ser bastante pesimistas en relación con la solución al problema que venimos discutiendo. Con la cautela requerida, creemos estar en condiciones de afirmar que una vez que la inestabilidad es introducida en la serie, va a ser muy difícil, sino imposible removerla. Para apoyar esta idea no tenemos más que comparar la potencia descubierta en la presente investigación con la encontrada en nuestro trabajo anterior para el mismo tamaño de muestra y para la misma magnitud del efecto. Por ejemplo, con el nivel de significación fijado al 0.01 la potencia promediada a lo largo de los seis modelos empleados en nuestro trabajo previo fue de 0.715, mientras que ahora resultó ser de tan sólo 0.575. Un resultado similar se

observa cuando comparamos ambas potencias al nivel de 0.05, en concreto, la potencia de nuestro trabajo precedente fue de 0.765, mientras que la encontrada en nuestra investigación actual resultó ser de 0.625.

Así pues, vamos a concluir señalando que los datos parecen obstinarse en poner de manifiesto, por un lado, que las diferencias entre las distintas rutinas de análisis no desaparecen con la complicación de los modelos desde los cuales las series son generadas y, por otro, que la eficacia de los diferentes procedi nientos utilizados va empeorándose conforme la complejidad de los modelos se va incrementando.

Conclusiones finales

Los datos indican que con respecto a las tasas de error Tipo I y estimaciones de los efectos del nivel, tendencia, cambio de nivel y cambio de tendencia, los tres procedimientos de análisis producen resultados que podemos tildar básicamente de similares. La tendencia liberal para el nivel α aparecida en todos los casos analizados, no fue estadísticamente significativa. A nuestro modo de ver este hecho sugiere que no debemos preocuparnos en exceso por las posibles inexactitudes en relación con las probabilidades de cometer más errores Tipo I de los estipulados. Asimismo, las tres rutinas de análisis nos han proporcionado estimaciones de los parámetros en las que tan sólo han aparecido ligeros sesgos en lo referido a la estimación del nivel previo a la intervención, en los demás casos, nuestros datos han puesto una vez más de relieve el carácter insesgado de las estimaciones efectuadas bajo el modelo lineal.

Sin embargo, los datos de nuestro estudio sí que nos proporcionan diferencias en lo referido a la potencia y a los errores estándar de los estimadores, tanto en el caso en que la magnitud del efecto de la intervención manipulada era una vez la desviación estándar, como cuando la magnitud del efecto manipualda era 1.5 veces la desviación estándar. El enfoque MCI se ha manifestado sistemáticamente como el más poderoso y también el que vierte estimaciones relativamente más eficientes, mientras que el procedimiento AR(1) se ha situado siempre en el extremo opuesto, ocupando, por tanto, el enfoque de la Transformación General el lugar intermedio.

En consecuencia, la eliminación de la problemática fase de identificación del ciclo de construcción del modelo en beneficio de algún modelo asumido de antemano va a disminuir nuestra capacidad (ya de por sí bastante mermada a tenor del reducido tamaño de muestra con el que usualmente suele contar el analista y, como acabamos de poner de relieve, por la creciente complejidad del proceso generador de los datos) para detectar cambios útiles y efectivos cuando en realidad existen. Con todo, es el investigador informado de esta cuestión y otras, por supuesto, quien tiene que decidir si está dispuesto a sacrificar potencia (no acabamos de tener claro si el enfoque sobreajustado también comete más errores Tipo I más allá del valor nominal estipulado, pues en los dos estudios presenta tasas de error ligeramente mayores que los otros dos procedimientos, pero sin superar las bandas o límites de $\pm 2SE$), en aras de una mayor flexibilidad, rango potencial de aplicaciones y simplicidad. Pues, si bien es verdad que las dificultades a vencer con los enfoques que sugieren pasar por alto la etapa de identificación en favor de algún modelo preconcebido son de peso, no es menos cierto que las oportunidades que se le ofrecen al analista de datos conductuales son escasas; a no ser que el propio diseño le permita llevar a cabo las permutaciones suficientes para proporcionarnos suficiente potencia y aplicar entonces pruebas de aleatorización (Edgington, 1984; Onghena, 1992; Wampold y Worsham, 1986), o bien opte por algún procedimiento abreviado como, por ejemplo, el basado en el estadístico C desarrollado por Young (1941), implementado por Tryon (1982) y fuertemente contestado por Blumberg (1984) y Crosbie (1989).

Referencias

Algina, J. y Olejnik, S. F. (1982). Multiple group time-series design: An analysis of data. *Evaluation Review, 6,* 203-232.

Algina, J. y Swaminathan, H. A. (1977). A procedure for the analysis of time series designs. *Journal of Experimental Education, 45,* 916-926.

Algina, J. y Swaminathan, H. A. (1979). Alternatives to Simonton's analysis of the interrupted and multiple-group time series designs. *Psychological Bulletin, 86,* 916-926.

Blumberg, C. J. (1984). Comments on «A simplified time-series analysis for evaluating treatment interventions». *Journal of Applied Behavior Analysis, 17,* 539-542.

Box, G. E. P. y Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control.* 2nd Edition. San Francisco, CA: Holden-Day.

Box, G. E. P. y Tiao, G. C. (1965). A change in level of a nonstationary time series. *Biometrika, 52,* 181-192.

Box, G. E. P. y Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association, 70,* 70-79.

Busk, P. L. y Marascuilo, L. A. (1988). Autocorrelation in single-subject research: A counterer argument to the myth of no autocorrelation. *Behavioral Assessment, 10,* 229-242.

Chatfield, C. (1989). *The Analysis of Time Series: An Introduction.* 4th Edition. London: Chapman and Hall.

Crosbie, V. (1989). The inappropiateness of the C statistic for assessing stability or treatment effects with single-subject data. *Behavioral Assessment, 11,* 315-325.

Edgington, E. S. (1984). Statistics and single case analysis. *Progress in Behavior Modification, 16,* 83-119.

Glass, G. V.; Willson, V. L. y Gottman, J. M. (1975). *Design and Analysis of Time-Series Experiments.* Boulder CO: Colorado Associated University Press.

Glass, G. V.; Bower, C. y Padia, W. L. (1974). *Computer Program.* Boulder, CO: University of Colorado Press.

Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N=1) data. *Behavioral Assessment, 5,* 141-154.

Gottman, J. M. y Glass, G. V. (1978). Analysis of interrupted time-series experiments. En J. R. Kratochwill (Ed.): *Single subject research: Strategies for Evaluation Change.* New York, NY: Academic Press.

Greenwood, K. M. y Matyas, T. A. (1990). Problems with the application of interrupted time-series analysis for brief single-subject data. *Behavioral Assessment, 12,* 355-370.

Harrop, J. W. y Velicer, W. F. (1985). A comparison of alternative approaches to the analysis of interrupted time-series. *Multivariate Behavioral Research, 20,* 27-44.

Harrop, J. W. y Velicer, W. F. (1990). Computer program for interrupted time series analysis: II. A quantitative evaluation. *Multivariate Behavioral Research, 25,* 233-248.

Huitema, B. E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment, 7,* 107-118.

Huitema, B. E. y McKean, J. W. (1991). Autocorrelation estimation and inference with small samples. *Psychological Bulletin, 110,* 291-304.

Judd, C. M. y Kenny, D. (1981). *Estimating the Effects of Social Interventions.* New York, NY: Cambridge University Press.

Liu, Lon-Mu (1989). Identification of seasonal ARIMA models using a filtering method. *Communications in Statistics. Theory and Methods, 18,* 2279-2288.

Maddala, G. S. (1977). *Econometrics.* New York, NY: McGraw-Hill (trad., 1985, McGraw-Hill).

Marsh, J. C. y Shibano, M. (1984). Issues in the statistical analysis of clinical time-series data. *Social Work Research and Abstracts, 20,* 7-12.

Matyas, T. A. y Greenwood, K. M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effect. *Journal of Applied Behavioral Analysis, 23,* 341-351.

McCleary, R. y Hay, R. A. Jr. (1980). *Applied Time Series Analysis for the Social Sciences.* Beverly Hills, CA:SAGE Publications.

Onghena, P. (1992). Randomization test for extensions and variations of ABAB single-case experimental designs. *Behavioral Assessment, 14,* 153-171.

Quenouille, M. H. (1949). Approximate test of correlation in time series. *Journal of the Royal Statistical Society, B, 11,* 68-84.

Scheffé, H. (1959). *The Analysis of Variance.* New York, NY: John Wiley.

Sharpley, C. F. y Alavosius, M. P. (1988). Autocorrelation in Behavioral data: An alternative perspective. *Behavioral Assessment, 10,* 243-251.

Simonton, D. K. (1977). Cross-sectional time-series experiments: Some suggested estatistical analysis. *Psychological Bulletin, 84,* 489-502.

Suen, H. K. (1987). On the epistemology of autocorrelation in applied behavior analysis. *Behavioral Assessment, 9,* 113-124.

Suen, H. K. y Ary, D. (1987). Autocorrelation in applied behavior analysis: Myth or reality? *Behavioral Assessment, 9,* 125-130.

Swaminathan, H. y Algina, J. (1977). Analysis of quasi-experimental time-series designs. *Multivariate Behavioral Research, 12,* 111-131.

Stralkowski, C. M.; Wu, S. M. y Devor, R. E. (1970). Charts for the interpretation and estimation of the second order autoregressive model. *Technometrics, 12,* 669-685.

Stralkowski, C. M.; Wu, S. M. y Devor, R. E. (1974). Charts for the interpretation and estimation of the second order moving average and mixed first order autoregressive-moving average models. *Technometrics, 16,* 275-285.

Tryon, E. (1982). A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis, 15,* 423-429.

Uriel, E. (1985). *Análisis de Series Temporales: Modelos ARIMA.* Madrid: Paraninfo.

Vallejo, G. (1991). La validez de las investigaciones en el ámbito experimental. En J. Pascual, M.ª T. Anguera, G. Vallejo y F. Salvador (Eds.): *Psicología Experimental,* pp. 41-75. Valencia: Nau Llibres.

Vallejo, G. (En prensa). *Diseños de Series Temporales Interrumpidas.* Madrid: Síntesis.

Vallejo, G; Herrero, J. y Cuesta, M. (1992). Comparación de la potencia y eficacia de diversos modelos ARIMA en series temporales interrumpidas: Un estudio de simulación. *Revista de Investigaciones Psicológicas, 10,* 174-214.

Velicer, F. W. y Harrop, J. (1983). The reliability and accuracy of time series model identification. *Evaluation Review, 7,* 551-560.

Velicer, F. W. y McDonald, R. P. (1984). Time series analysis without model identification. *Multivariate Behavioral Research, 19,* 33-47.

Velicer, F. W. y McDonald, R. P. (1991). Cross-sectional time series designs: A General Transformation Approach. *Multivariate Behavioral Research, 26,* 247-254.

Wampold, B. E. y Worsham, N. L. (1986). Randomization test for multiple-baseline designs. *Behavioral Assessment, 8,* 135-143.

Young, L. C. (1941). On randomness in ordered sequences. *Annals of Mathematical Statistics, 12,* 293-300.

Aceptado el 8 de marzo de 1994