Identificación, estimación y predicción robusta en modelos ARIMA

1988

Identificación, estimación y predicción robusta en modelos *ARIMA*. Informe correspondiente al convenio entre el Instituto Nacional de Estadística y Censos y la Facultad de Ciencias Exactas y Naturales, UBA

POR ELENA MARTINEZ Y VÍCTOR J. YOHAI*

1. INTRODUCCIÓN

La metodología propuesta por Box- Jenkins está dirigida a desarrollar modelos para series de tiempo, en particular series económicas, que permitan predecir los valores futuros de la serie en función de los valores pasados de la propia serie.

Contrariamente a lo que ocurre con los modelos econométricos,la elección de los modelos en esta metodología no presupone ninguna teoría económica, la elección del modelo es totalmente empírica, es decir determinada por los datos.

Básicamente, se trata de descubrir las regularidades estadísticas de la serie y modelarlas en el marco de una familia de modelos llamados autoregresivos-promedios móviles integrados. Generalmente esta familia de modelos es conocida como modelos *ARIMA*, *AR* del inglés AutoRegressive, *I* de Integrated y *MA* de moving average (promedio movil).

Box y Jenkins(1976) desarrollaron procedimientos que cubren todas las etapas del desarrollo de un modelo *ARIMA*. Sin embargo esos procedimientos son muy sensibles a la presencia de observaciones atípicas ("outliers") que se apartan de la regularidad estadística observada por la mayoría. Estos "outliers" pueden deberse a errores groseros en la medición de la observación o al hecho de que el período considerado es atípico debido a algún fenómeno de orden político o económico que influye en la variable desviándola de los límites esperados de acuerdo a la regularidad estadística de la serie.

El propósito de este trabajo es desarrollar una metodología que cubra todas las etapas necesarias para la elección de un modelo *ARIMA* para una serie de tiempo, de manera que esa elección no sea mayormente afectada por unos pocos "outliers".

Supongamos que se considere una serie $x_t, t = 1, ..., T$ observada a intervalos regulares de tiempo, por ejemplo mensualmente, trimestralmente o anualmente.

Para comenzar consideraremos series estacionarias. Básicamente una serie estacionaria es una serie cuya distribución no cambia en el tiempo y que por lo tanto su media, su varianza y su estructura de correlación en el tiempo se mantienen constantes en todo el período bajo estudio. En Economía existen muchas series que no son estacionarias, ya que suelen existir tendencias o comportamientos estacionales.

Sin embargo, en el método de Box y Jenkins, el estudio de las series estacionarias es básico. El caso de series no estacionarias se reduce al caso de estacionarias mediante transformaciones adecuadas. Por lo tanto comensaremos introduciendo modelos para series estacionarias.

En una serie estacionaria podemos introducir los siguientes conceptos:

*Se agradece la colaboración de la Licenciada Ana Bianco.

(i)media o esperanza de la serie a la cual denominaremos μ, que corresponde a la media de cada una de las variables que integran la serie, es decir podemos escribir $\mu = E(x_t)$, donde E indica esperanza

(ii)σ_x^2, la varianza de cada x_t, o σ_x, su desviación standard.

(iii)Las autocorrelaciones de la serie. La autocorrelación de orden i, que se indicará con ρ_i, se define como el coeficiente de correlación entre x_t y x_{t-i}. Como la serie es estacionaria este valor no dependerá de t. Esto es, cualquiera sean t y s el coeficiente de correlación entre x_t y x_{t-i} es el mismo que el coeficiente de correlación entre x_s y x_{s-i}.

La primer familia de modelos que consideraremos es la formada por los modelos autoregresivos. Un modelo autoregresivo de orden p, que se simboliza por $AR(p)$ es básicamente un modelo de regresión lineal donde las variables explicativas son los p valores anteriores de la variable. Más precisamente se dirá que la serie x_t es un $AR(p)$ si

$$x_t = \mu + \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \ldots + \phi_p(x_{t-p} - \mu) + u_t,$$

donde μ es la media de la serie, $\phi_1, \phi_2, \ldots \phi_p$ son los parámetros autoregresivos, y u_t representa el error y satisface los siguientes requerimientos:

(i) $u_1, u_2, \ldots, u_t$ son variables aleatorias independientes

(ii) $u_1, u_2, \ldots, u_t$ tienen la misma distribución. Se denominará σ_u^2 a la varianza común de los u_t.

(iii)La media de los u_t es 0

Cualquier serie u_t satisfaciendo las propiedades (i) (ii) y (iii) se denomina ruido blanco. u_t corresponde a aquella parte de x_t que no depende del pasado, y que por lo tanto no se puede predecir utilizando las observaciones anteriores. Por esta razón los errores u_t también son denominados innovaciones.

En los modelos autoregresivos el número de parámetros coincide con el número de observaciones anteriores de las cuales depende la serie. Si se dispone de un número relativamente pequeño de observaciones, la utilización de un modelo con muchos parámetros generalmente conduce a estimaciones ineficientes de los mismos. Para permitir la utilización de modelos que utilizan menos parámetros (más parsimoniosos) para representar la dependencia de cada valor de la serie con varios valores del pasado, se introducen los modelos promedios móviles. Una serie estacionaria será modelada por un promedio móvil de orden q, simbolizado por $MA(q)$, si se puede expresar como

$$x_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} \quad - \ldots \theta_q u_{t-q},$$

donde nuevamente las innovaciones u_t forman un ruido blanco y representa la parte de la serie x_t que no depende del pasado.

Para ejemplificar las posibilidades de estos modelos consideremos el modelo $MA(1)$

$$x_t = u_t - \theta u_{t-1}.$$

Se puede mostrar que este modelo es equivalente a un modelo $AR(\infty)$, donde $\phi_i = -\theta^i$, es decir

$$x_t = \mu - \theta(x_{t-1} - \mu) - \theta^2(x_{t-2} - \mu) - \ldots \theta^i(x_{t-i} - \mu) \ldots + u_t.$$

Por lo tanto hemos modelado con un solo parámetro una serie donde x_t depende de todos los valores anteriores.

Finalmente se pueden combinar los 2 modelos básicos $AR(p)$ y $MA(q)$, y obtener modelos mixtos autoregresivos- promedios móviles $ARMA(p,q)$. Combinando las ecuaciones de los modelos $AR(p)$ y $MA(q)$ obtenemos

$$x_t = \mu + \phi_1(x_{t-1}-\mu) + \phi_2(x_{t-2}-\mu) + \ldots + \phi_p(x_{t-p}-\mu) - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \ldots \theta_q u_{t-q} + u_t,$$

donde nuevamente las innovaciones u_t son un ruido blanco.

Para modelar series no estacionarias, Box y Jenkins usan el recurso de diferenciar la serie hasta que se transforme en estacionaria. Por ejemplo, supongamos una serie x_t que tenga una tendencia lineal

$$x_t = at + b + v_t,$$

donde v_t es una serie estacionaria. Si se considera la diferencia

$$w_t = x_t - x_{t-1} = at + b + v_t - a(t-1) - b - v_{t-1} = a + v_t - v_{t-1},$$

se observa que el término lineal en t desaparece y w_t resulta una serie estacionaria. De la misma manera, si en vez de una tendencia lineal se tuviese una tendencia cuadrática, bastaría con diferenciar 2 veces para eliminarla. En general si la tendencia estuviese dada por un polinomio de grado d, sería necesario diferenciar sucesivamente la serie d veces para eliminarla. Si la tendencia fuese exponencial, sería necesario tomar primero logaritmos para transformarla en lineal y luego diferenciar. Se llama modelo $ARIMA(p,d,q)$a un modelo que diferenciado d veces da un modelo estacionario $ARMA(p,q)$

Box y Jenkins desarrollaron una metodología para la elección del modelo $ARIMA$ que presenta las siguientes etapas

(i)Identificación del modelo. Esto corresponde a la elección de p, d y q. Es decir, en esta etapa se determina el número de parámetros autoregresivos, el número de parámetros promedios móviles y el número de diferencias a incluir en el modelo. La elección de p y q se harán utilizando el principio de parsimonia: cuantos menos parámetros tenga el modelo, más eficientemente podrán ser estimados. Por lo tanto se tratará de encontrar un modelo que ajuste a los datos con la menor cantidad posible de parámetros. Es importante señalar que la metodología de Box y Jenkins es iterativa, y que la elección del modelo en esta etapa está sujeta a un chequeo posterior (ver (iii)). Este chequeo puede determinar que el modelo elegido en esta etapa deba ser modificado.

(ii)Estimación de los parámetros. Una vez identificado el modelo en la etapa (i), se estiman los parámetros autoregresivos y promedios móviles suponiendo que las innovaciones u_t son normales.

(iii)Prueba de la bondad de ajuste de modelo. Una vez estimados los parámetros, se realiza un test de bondad de ajuste para determinar si el modelo elegido es adecuado. En caso de que el modelo no pase este test, se reidentificará el modelo, generalmente aumentando p o q. El análisis de los residuos correspondientes al modelo ajustado dará pistas sobre cual de estos órdenes aumentar y cuánto. Con el nuevo modelo se pasa nuevamente a la etapa (ii).

(iv)Predicción. Una vez que el modelo pasa el test de la etapa (iii) podrá ser utilizado para realizar predicciones de los valores futuros de la serie. Box y Jenkins derivan la forma de realizar las predicciones de los valores futuros utilizando el modelo elegido.

En la metodología de Box y Jenkins todas estas etapas son desarrolladas suponiendo que no hay "outliers".

El método de identificación del modelo $ARIMA$ está basado en los siguientes resultados

(a)En un modelo estacionario $ARMA(p,q)$,las autocorrelaciones ρ_k decrecen exponencialmente. En cambio en un modelo no estacionario $ARIMA(p,d,q)$, con $d > 0$, los ρ_k decrecen de manera parecida a la lineal.

(b) En un modelo estacionario $AR(p)$, si se considera un modelo de orden $k > p$, el correspondiente coeficiente $\phi_k = 0$.

(c)En un modelo $MA(q)$, si $k > q$, la correspondiente autocorrelación $\rho_k = 0$.

De estos resultados se desprende que tanto las autocorrelaciones ρ_k como los coeficientes ϕ_k son muy informativos para la identificación del modelo. Como ninguno de estos coeficientes es conocido, se comienza por estimarlos.

La autocorrelación ρ_k se estima por la autocorrelación muestral, que se indica con r_k y se define por

$$r_k = \frac{\sum_{t=k+1}^{T} x_t x_{t-k}}{\sum_{t=1}^{T} x_t^2}.$$

Para estimar ϕ_k se considera un modelo autoregresivo de orden k

$$x_t = \phi_{k1} x_{t-1} + \ldots + \phi_{kk} x_{t-k} + u_t,$$

y se lo ajusta por el método de Yule-Walker. Este método está basado en los coeficientes r_k y se puede resolver en forma iterativa. En cada iteración se estiman los coeficientes del modelo de un orden dado en función de los estimadores correspondientes a un modelo de un grado menor, para mayores detalles ver Box y Jenkins (1976). El valor $\hat{\phi}_{kk}$ que resulta de estimar ϕ_{kk} se denomina autocorrelación parcial de orden k y estimará a ϕ_k. Siguiendo los resultados (a), (b), y (c) establecidos anteriormente, el método de Box y Jenkins procede para la identificación del modelo de la siguiente manera:

(a*)Se diferencia la serie hasta que las autocorrelaciones muestrales aparezcan con un decrecimiento tipo exponencial.

Una vez obtenida la serie estacionaria por diferenciación, se identifica un modelo $ARMA(p,q)$ para la misma de acuerdo a las siguientes reglas:

(b*)Si para $k > p$ todos los $\hat{\phi}_{kk}$ no son significativamente diferentes de 0, se puede identificar un modelo $AR(p)$.

(c*)Si para todo $k > q$, r_k no es significativamente diferente de 0, se puede identificar un modelo $MA(q)$.

Para identificar modelos mixtos $ARMA(p,q)$, el procedimiento es más complicado, pudiendose proceder por prueba y error.

Los métodos de estimación utilizados corresponden a distintas aproximaciones al estimador de máxima verosimilitud correspondiente a innovaciones normales. Una de las aproximaciones más utilizadas es el estimador de mínimos cuadrados. En el caso particular de un modelo $AR(p)$, este estimador se obtiene por la minimización de

$$S(\vec{\phi},\mu)=\sum_{p+1}^{T}\hat{u}_t^2(\vec{\phi},\mu), \tag{1-1}$$

donde $\hat{u}_t(\vec{\phi},\mu)=\hat{u}_t$ está definido por

$$\hat{u}_t = x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} \tag{1-2}$$

para $t=p+1,...,T$.

Para el caso de un modelo $ARMA(p,q)$ con $q>0$ la definición de los residuos $\hat{u}_t$ es mas compleja.

Suponiendo que los errores son normales, los estimadores de mínimos cuadrados (como todas las variantes de máxima verosimilitud) son asintóticamente eficientes.

Sin embargo tanto las autocorrelaciones muestrales(y por lo tanto las autocorrelaciones parciales), como los estimadores de mínimos cuadrados (o cualquiera de las variantes de máxima verosimilitud) son muy sensibles a

(i)la hipótesis de normalidad de los errores

(ii) la presencia de unas pocas observaciones anómalas ("outliers")

Es decir pequeñas desviaciones de la normalidad o unos pocos "outliers" pueden dar identificaciones incorrectas y producir grandes sesgos en los estimadores.

Sin embargo, es muy frecuente en las series económicas violaciones a la hipótesis de normalidad como así también la presencia de "outliers". Por lo tanto es necesario buscar procedimientos alternativos para la identificación del modelo, estimación de párametros y predicción de valores futuros de la serie. Estos procedimientos se denominan robustos y tienen la propiedad de ser poco sensibles a las desviaciones de normalidad como también a la presencia de "outliers".

Se han propuesto varias clases de estimadores como alternativas robustas a mínimos cuadrados, por ejemplo los M-estimadores, GM-estimadores y RA- estimadores. Una revisión de todos estos estimadores puede encontrarse en Martin y Yohai (1985). Sin embargo, a pesar de que todos estos estimadores son más robustos que mínimos cuadrados pueden ser muy sensibles a grandes "outliers", especialmente en modelos donde hay coeficientes de tipo promedio móvil, o también en autoregresivos de alto orden. Esto se basa en el hecho de que todos estos estimadores están basados en los residuos definidos en (1.2). Por ejemplo los M- estimadores para un modelo $AR(p)$ se definen por la minimización de

$$\sum_{t=p+1}^{T}\rho\left(\frac{\hat{u}_t(\vec{\phi},\mu)}{\hat{\sigma}_u}\right),$$

donde ρ es una función simétrica que crece menos rápidamente que la función cuadrática. Por ejemplo se puede elegir como $\rho(u)=|u|$, de esta manera se le da menos peso a los residuos grandes, los cuales pueden corresponder a "outliers".

Sin embargo los residuos $\hat{u}_t$ dados por (1.2) están definidos de tal manera de que un "outlier" afecta a varios $\hat{u}_t$. Por ejemplo en un modelo $AR(p)$ se tiene que

$$\hat{u}_t = x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p}, \tag{1-3}$$

y por lo tanto si x_t es un "outlier", afecta a $\hat{u}_t, \hat{u}_{t+1}, \ldots, \hat{u}_{t+p}$, es decir $p+1$ residuos. Esto significa por ejemplo que si hay un 10 % de "outliers", y $p = 5$, puede haber hasta un 60 % de residuos afectados. La situación es aun peor cuando hay operadores promedios móviles, ya que en este caso si x_t es un "outlier", afecta a todos los residuos $\hat{u}_s$ con $s \geq t$. Como ejemplo podemos citar el $MA(1)$ donde

$$\hat{u}_t = x_t + \theta x_{t-1} + \theta^2 x_{t-2} + \ldots + \theta^{t-1} x_1.$$

Luego en este caso si x_1 es un "outlier" afecta a todos los residuos. Por lo tanto un método de estimación robusta requiere que los residuos se calculen en forma diferente a la dada en (1.2).

Una propuesta para computar los residuos en forma más robusta fue dada por Martin, Samarov y Vandaele (1983). Para ésto se propone realizar un filtrado de la serie en forma recursiva de manera de reemplazar las observaciones que aparentan ser "outliers" por otros valores más acordes con el resto de la serie. Supongamos se considera que x_t sigue un modelo $AR(p)$, que ya se haya filtrado la serie hasta el periodo $t-1$, y que esta serie sea $y_1, y_2, \ldots, y_{t-1}$, de manera que y_j coincide con x_j cuando x_j no aparenta ser un "outlier" y es modificada cuando lo aparenta. Luego se puede definir un nuevo residuo $\tilde{u}_t$ por

$$\tilde{u}_t = x_t - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p}. \tag{1-4}$$

De esta manera contrariamente a lo que sucedía con la formula (1.2), los "outliers" anteriores a x_t no afectarán el cálculo correspondiente al residuo $\tilde{u}_t$. Por otro lado como se explicará en la Sección 3, en base al residuo $\tilde{u}_t$ se podrá modificar o no el valor x_t para obtener el valor filtrado y_t.

Sin embargo para poder aplicar el filtro robusto que nos permite obtener la serie y_t limpia de "outliers" se requiere un modelo inicial y estimadores de los parámetros correspondientes, aunque no es fundamental que el modelo o los estimadores sean los definitivos. Martin, Samarov y Vandaele (1983) propone utilizar modelos autoregresivos y utilizar GM-estimadores. Dado que los GM-estimadores no son totalmente robustos el filtro correspondiente puede fallar en la supresión de "outliers".

En este trabajo se presenta una nueva propuesta que combina las ideas de Martin, Samarov y Vandaele de utilizar un filtro que limpie la serie de "outliers" con una forma robusta de obtención de un estimador inicial para sucesivos modelos autoregresivos que se utilizarán para construir la serie filtrada. En este informe se describen procedimientos robustos basados en estas ideas para las distintas etapas del análisis de una serie de tiempo: identificación, estimación y predicción.

2. Esquema de la metodología propuesta

La metodología puede ser esquematizada en los siguientes pasos

1. Estimación de los parámetros de un modelo autoregresivo $AR(p)$ inicial. Este estimador será denominado $\vec{\phi}^{(0)} = (\phi_1^{(0)},...,\phi_p^{(0)})$. Esto se hará por un método paso a paso, donde iterativamente se estimarán modelos autoregresivos de orden $1,2,...,p$. En cada paso será necesario estimar un solo parámetro, lo cual se hará en forma robusta. En cada uno de estos pasos se realizará un filtrado de la serie usando el modelo autoregresivo estimado en la etapa anterior.

2.Filtrado inicial de la serie. Una vez obtenidos los estimadores del modelo autoregresivo inicial se los utilizará para filtrar la serie, obteniendo una nueva serie $y_t, t = 1,...T$ limpia de "outliers". Simultáneamente se obtendrá una serie σ_t que estima la dispersión de la predicción de x_t utilizando la serie filtrada $y_1,...,y_{t-1}$

3.Identificación del modelo. Para realizar la identificación del modelo $ARMA(p,q)$ que ajusta a la serie x_t, se calcularán las autocorrelaciones y autocorrelaciones parciales de la serie y_t.

4.Estimación del modelo. Una vez identificado el modelo $ARMA(p,q)$, los estimadores correspondientes se obtendrán por la minimización de

$$\sum_{t=p+1}^{T} \rho(\frac{x_t - \hat{x}_{t.t-1}(\vec{\phi},\vec{\theta},\mu)}{\sigma_t}),$$

donde $\hat{x}_{t.t-1}(\vec{\phi},\vec{\theta},\mu)$ es el valor predicho de x_t usando $y_1,...y_{t-1}$ cuando los parámetros del modelo $ARMA(p,q)$ son $\vec{\phi}$ y $\vec{\theta}$ y μ, y donde ρ es una función par y acotada.

5.Estimación de la matriz de covarianza. Se estiman las varianzas y covarianzas de los estimadores calculados en el paso 4.

6.Filtrado final de la serie. Se filtra nuevamente la serie x_t usando el modelo $ARMA(p,q)$ y los estimadores del paso 4. Esto da lugar a una nueva serie y_t.

7.Predicción. Se desarrollará un procedimiento para, a partir del modelo elegido, realizar predicciones robustas. Para predecir x_{t+h} utilizando $x_1, x_2,...x_t$, se utilizará el procedimiento usual de predicción pero remplazando $x_1, x_2,...x_t$ por $y_1,...,y_t$. Por ejemplo en el caso de un modelo $AR(p)$, si $\hat{\phi}_1,..,\hat{\phi}_p$ son los estimadores de los parámetros autoregresivos y $\hat{\mu}$ es el estimador de la media, la predicción de x_{t+1} vendrá dada por

$$\hat{\phi}_1(y_t - \hat{\mu}) + \hat{\phi}_2(y_{t-1} - \hat{\mu}) + ... + \hat{\phi}_p(y_{t-p+1} - \hat{\mu}).$$

3. Filtro para eliminar "outliers"

Supongamos para simplificar que una serie se rige por un modelo $AR(p)$ y $\mu = 0$. El caso más general de un $ARMA(p,q)$ y $\mu \neq 0$ se trata similarmente. Simbolicemos por $\mathbf{x}_t = (x_t, x_{t-1},..., x_{t-p+1})$, y llamemos

$$h(\mathbf{x}_{t-1}, \vec{\phi}) = \phi_1 x_{t-1} + ... + \phi_p x_{t-p}. \tag{3-1}$$

Este valor será la predicción de x_t, basado en los valores pasados de la serie en caso de que $x_{t-1}, x_{t-2},...x_{t-p}$ no fuesen "outliers". Supongamos además que conozcamos la dispersión σ_u de las innovaciones. Luego recursivamente podemos definir 2 series

(i) La serie filtrada y_t . Esta serie coincidirá con la x_t para casi todas las observaciones.Cuando x_t aparente ser un "outlier", el y_t se obtendrá modificando x_t de manera que sea consistente con los valores anteriores .

(ii) Una serie σ_t que mide la dispersión en la predicción de x_t usando $y_1,...,y_{t-1}$. La precisión de esta predicción va a ser variable ya que va a depender de si entre los valores anteriores y próximos al periodo t hay "outliers". Si entre los valores más próximos a x_t, por ejemplo x_{t-1} o x_{t-2} hay "outliers" la dispersión de la predicción σ_t deberá ser más grande que la dispersión cuando no hay "outliers", la cual es σ_u.

A continuación describiremos cómo se realiza iterativamente el filtrado. Supongamos que se hayan computado $y_1,...,y_{t-1}$, $\sigma_1,...,\sigma_t$ luego se define

(3-2)
$$y_t = h(\mathbf{y}_{t-1},\vec{\phi}) + \sigma_t\psi(\frac{x_t - h(\mathbf{y}_{t-1},\vec{\phi})}{\sigma_t}),$$

donde $\psi(u)$ es una función impar y acotada que coincide con la identidad para $|u|$ pequeño. Observese que $h(\mathbf{y}_{t-1},\vec{\phi})$ corresponde al valor de x_t que se prediciría cuando los valores anteriores son $y_{t-1},...y_{t-p}$. Por otra parte $x_t - h(\mathbf{y}_{t-1},\vec{\phi})/\sigma_t$ corresponde al residuo correspondiente standarizado. Cuando esta cantidad no sea muy grande, de acuerdo a lo expresado, la función ψ coincidirá con la identidad, y por lo tanto de acuerdo con la fórmula (3.2), y_t coincidirá con x_t. A continuación damos algunas posibles funciones ψ que aclararán el mecanismo del filtro.

(i)**rechazo total de "outliers"**

$$\psi(u) = \begin{cases} u & \text{si } |u| \leq k \\ 0 & \text{si } |u| > k. \end{cases}$$

$$y_t = \begin{cases} x_t & \text{si } |x_t - h(\mathbf{y}_{t-1}\vec{\phi})| \leq k\sigma_t \\ h(\mathbf{y}_{t-1},\vec{\phi}) & \text{si } |x_t - h(\mathbf{y}_{t-1},\vec{\phi})| > k\sigma_t. \end{cases}$$

En este caso, si la observación x_t es sospechosa de ser un "outlier", lo cual ocurre cuando el residuo standarizado tiene módulo mayor que k, el valor de y_t corresponde con la predicción utilizando la serie filtrada.

(ii)**Función ψ de Huber o winsorización**

$$\psi(u) = \begin{cases} u & \text{si } |u| \leq k \\ k\text{sign}(u) & \text{si } |u| > k. \end{cases}$$

$$y_t = \begin{cases} x_t & \text{si } |x_t - h(\mathbf{y}_{t-1},\vec{\phi})| \leq k\sigma_t \\ h(\mathbf{y}_{t-1},\vec{\phi}) + k\sigma_t & \text{si } x_t - h(\mathbf{y}_{t-1},\vec{\phi}) > k\sigma_t \\ h(\mathbf{y}_{t-1},\vec{\phi}) - k\sigma_t & \text{si } x_t - h(\mathbf{y}_{t-1},\vec{\phi}) < -k\sigma_t. \end{cases}$$

En este caso, si x_t es sospechosa de ser un "outlier" debido a un residuo standarizado con valor absoluto mayor que k, se elige y_t de manera de que el residuo sea en valor absoluto igual a k y del mismo signo que el correspondiente a x_t.

(iii)**Función de tipo Hampel.**Como esta función es impar la describiremos únicamente para $u \geq 0$.

$$\psi(u) = \begin{cases} u & \text{si } 0 \leq u \leq k_1 \\ k\text{sign}(u) & \text{si } k_1 < u \leq k_2 \\ (k_1 k_3 - u k_1)/(k_3 - k_2) & \text{si } k_2 < u \leq k_3 \\ 0 & \text{si } u > k_3. \end{cases}$$

La expresión para σ_t es mas compleja y puede encontrarse en Martin y Yohai (1985). En el caso de la función de rechazo total la fórmula de σ_t corresponde a la dispersión de la predicción de x_t cuando las observaciones anteriores rechazadas son consideradas como faltantes.

4. ESTIMACIÓN DEL MODELO AUTOREGRESIVO INICIAL

Para estimar los coeficientes del modelo autoregresivo inicial se irán estimando paso a paso modelos autoregresivos de orden creciente. Se comenzará por un modelo de orden 1 y en cada paso se incrementará el orden del modelo en 1.

Para eliminar el problema de la elección simultánea de μ, en el caso estacionario se centra la serie restándole la mediana. Para no introducir nueva notación, la serie centrada será también llamada x_t. Se elige la mediana para centrar la serie debido a que este estadístico es robusto, es decir la mediana no es mayormente afectada por unos pocos "outliers".

En el primer paso se debe estimar un modelo autoregresivo de orden 1

$$x_t = \phi_1 x_{t-1} + u_t.$$

El único parámetro a estimar en este caso es ϕ_1 y se lo estima por

$$\hat{\phi}_1 = \text{mediana}\frac{x_t}{x_{t-1}}.$$

A continuación describiremos la iteración para pasar de un modelo de orden k al siguiente de orden $k+1$. Supongamos que se tiene estimado un modelo autoregresivo de orden k, cuyos parámetros son $\phi_1, \phi_2, ..., \phi_k$. Sea σ_u el estimador de la dispersión de los residuos. Para determinar el modelo de orden $k+1$ se deben realizar los siguientes pasos.

(i) Calculamos la serie filtrada y_t usando el modelo de orden k

(ii) Calculamos los residuos $\tilde{u}_t$

$$\tilde{u}_t = x_t - \phi_1 y_{t-1} - ... - \phi_k y_{t-k}, \qquad t = k+1, ...T. \tag{4-1}$$

La ventaja de usar estos residuos consiste en que al reemplazar un "outlier" x_{t-i}, $i = 1,..k$ por el valor filtrado y_{t-i}, se atempera el efecto de posibles "outliers".

(iii) Seguidamente se calcula un nuevo σ_u por

(4- 2)
$$\sigma_u = \frac{\text{mediana}|\tilde{u}_t|}{.6745}.$$

(iv) Con el nuevo σ_u se filtra nuevamente la serie x_t y se obtiene una nueva serie y_t. Nuevamente se calcula la serie de residuos $\tilde{u}_t$ usando (4.1).

(v) Calculamos los residuos hacia atrás $\breve{u}_t$ por

(4-3)
$$\breve{u}_t = x_t - \phi_1 y_{t+1} - ...\phi_k y_{t+k}.$$

La razón para definir $\breve{u}_t$ es la siguiente.En el paso $k+1$ se deberá utilizar x_{t-k-1} para explicar x_t, o lo que queda sin explicar de x_t que es $\tilde{u}_t$. En vez de utilizar x_{t-k-1}, se utilizará una variable equivalente a aquella y que sea ortogonal a todas las variables utilizadas anteriormente: $x_{t-1}, ...x_{t-k}$. Esta variable que será equivalente a x_{t-k-1}, pero ortogonal a todas las anteriores será $\breve{u}_{t-k-1}$.

(vi)Se calcula una estimación robusta de la pendiente de $\tilde{u}_t$ usando $\breve{u}_{t-k-1}$. Se utiliza el estimador γ dado por

(4-4)
$$\gamma = \text{mediana}\frac{\tilde{u}_t}{\breve{u}_{t-k-1}}.$$

(vii) Se calculan los nuevos coeficientes autoregresivos. Como podemos escribir que "aproximadamente"

$$x_t - \phi_1 y_{t-1} - ... - \phi_k y_{t-k} = \gamma(x_{t-k-1} - \phi_1 y_{t-k} - ... - \phi_k y_{t-1}).$$

Igualando coeficientes tendremos los nuevos ϕ_i, los cuales estarán dados por

$$\phi_i = \begin{cases} \phi_i - \gamma\phi_{k-i+1} & \text{si } 1 \leq i \leq m-1 \\ \gamma & \text{si } i = k+1. \end{cases}$$

(viii) Calculamos la serie filtrada y_t usando los nuevos ϕ_i y el σ_u calculado en (iv).

(ix) Se calculan nuevamente los residuos por (4.1).

(x) Se calcula nuevamente σ_u por (4.2).

(xi) Se compara el nuevo σ_u con el calculado en (iv). Si el nuevo σ_u es mas chico, los ϕ calculados en (vii) seran los que correspondan al modelo de orden $k+1$. En caso contrario, se deja como modelo de orden $k+1$ el de orden k, es decir, para $i \leq k$ se dejan los ϕ_i del modelo de orden k y $\phi_{k+1} = 0$.

Esto se continua hasta que k alcance un valor fijado a priori. Actualmente calculamos modelos hasta de orden 10. Finalmente, para elegir el modelo inicial se utiliza un criterio de Akaike robustificado. Este criterio elige el modelo de orden p que minimiza

$$n\ln\sigma_u^2 + 2p.$$

5. Extensiones a modelos no estacionarios y estacionales

Caso de modelos con operadores estacionales estacionarios. En este caso se introducen tambien paso a paso para explicar x_t las variables $x_{t-s}, x_{t-2s}, ...$, donde s es la frecuencia de estacionalidad de la serie, usando un procedimiento similar al descripto anteriormente.

Casos de modelos ARIMA(p,d,q) no estacionarios. En este caso los coeficientes ϕ del modelo diferenciado estacionario se calculan en la forma descripta. Sin embargo no conviene diferenciar la serie original ya que ésto podría duplicar los "outliers". Supongamos que la serie original es x_t, y w_t es la serie diferenciada. Luego $w_t = x_t - x_{t-1}$. Por lo tanto si x_t es un "outlier", esto afectará a w_t y a w_{t-1}. Para evitar esta duplicación de "outliers", la diferenciación va a realizarse conjuntamente con el filtrado de la serie. La operación de diferencia se puede expresar en un modelo autoregresivo de un orden mayor. Por ejemplo, supongamos que w_t sea autoregresivo de orden 1. Luego se tiene $w_t = \phi w_{t-1} + u_t$, o sea $x_t - x_{t-1} = \phi(x_{t-1} - x_{t-2}) + u_t$. Despejando se obtiene que esto es equivalente $x_t = (1+\phi)x_{t-1} - \phi x_{t-2} + u_t$. Esto corresponde a que x_t sigue un modelo autoregresivo de orden 2 no estacionario con coeficientes $\phi_1 = (1+\phi)$ y $\phi_2 = -\phi$.

Por lo tanto después de la estimación de cada modelo se aplicará el filtro correspondiente al modelo autoregresivo no estacionario que resulta de componer los coeficientes del modelo estacionario con las diferencias involucradas. Queda sin resolver cómo se eligirá un modelo no estacionario y cuántas diferencias se incluirán en el mismo. Esto se puede hacer usando un correlograma inicial no robusto, tal como está descripto en Box y Jenkins (1976). Salvo en casos extremos, la presencia de "outliers" no afecta la forma global del correlograma que lleva a interpretar una serie como estacionaria o no estacionaria. Otra posibilidad consiste en realizar un correlograma inicial robusto estimando la correlación de orden k por

$$r_k = \text{mediana}\{\frac{x_t}{x_{t-k}}, t = k+1, ..., T\}.$$

El análisis de este correlograma permitirá decidir si hay o no necesidad de incluir operadores no estacionarios en el modelo. En caso de encontrar no estacionaridad, se diferenciará la serie y se procederá a realizar el correspondiente correlograma robusto de la serie diferenciada.

Otra posibilidad consiste en incorporar un operador no estacionario si el ϕ_1 calculado en Sección 4 está muy próximo a 1. La ventaja de esta solución consiste en que se podría implementar automaticamente.

6. Ejemplos de aplicación

En esta sección ilustraremos el procedimiento robusto propuesto para modelar series de tiempo mediante 3 ejemplos, uno real y 2 simulados.

Ejemplo 1. Consideremos la serie de tasas de inflación, definida como el incremento porcentual de los precios al consumidor, para el período junio de 1973 a diciembre de 1974. La serie está graficada en la Figura 1. Aparecen claramente 3 "outliers", diciembre de 1973, enero del 1974 y diciembre de 1974. El correlograma y correlograma parcial basado en las autocorrelaciones muestrales aparece en la Tabla 1. Como se ve, de acuerdo a esta tabla ninguna de estas autocorrelaciones es significativamente diferente de 0, y por lo tanto la serie puede ser considerada un ruido blanco. Esto significa que de acuerdo a este correlograma no habría manera de utilizar el pasado para realizar predicciones. Si realizamos predicciones con un modelo autoregresivo de orden 1 usando el método de mínimos cuadrados, obtenemos el siguiente modelo

$$x_t = -.11x_{t-1} + 2.1 + u_t,$$

el cual tiene pendiente negativa. aunque no significativamente distinta de 0.

Veamos ahora los resultados usando el procedimiento robusto. Los valores de la serie modificados por el filtro aparecen marcados con un triángulo en Figura 1, y corresponden a los 3 "outliers" citados anteriormente. Por otro lado si estimamos un modelo autoregresivo por el procedimiento robusto se obtiene el siguiente modelo $AR(1)$

$$x_t = .48x_{t-1} + .98 + u_t.$$

Las predicciones utilizando ambos modelos aparecen en la Figura 2. Como se ve las predicciones usando el modelo obtenido por el procedimiento robusto son muy superiores a las obtenidas con el modelo estimado por el método de mínimos cuadrados.

Ejemplo 2. En este ejemplo simulamos 200 observaciones del siguiente modelo $AR(1)$

$$x_t = .9x_{t-1} + u_t,$$

donde u_t tiene distribución con media 0 y varianza 1. Esta serie puede observarse en la Figura 3. Los valores filtrados, cuando son diferentes de los valores reales, se indican en el mismo gráfico, observándose que en este caso la serie filtrada coincide con la real. El modelo estimado por mínimos cuadrados para esta serie es

$$x_t = .79x_{t-1} + .046x_{t-2} + u_t,$$

y el modelo estimado por el procedimiento robusto es

$$x_t = .89x_{t-1} - .08x_{t-2} + u_t.$$

En la Figura 4 se presentan las primeras 50 observaciones de la serie con los respectivos valores predichos por ambos modelos. Se observa que ambos modelos dan resultados similares, lo cual es lógico ya que la serie no contiene "outliers".

A continuación 20 observaciones elegidas al azar de la serie fueron reemplazadas por "outliers". Estos "outliers" son del tipo aditivo, es decir se obtienen sumando a ciertas observaciones un ruido. En este caso las 20 observaciones que se modifican son obtenidas reemplazando el valor x_t por un nuevo valor $x_t + 5$. Llamaremos x_t^* a la serie modificada por estos "outliers", y está representada en la Figura 5. En el mismo gráfico se representan los valores filtrados. En la Figura 6 se representan los 50 primeros valores de la serie x_t^* y los correspondientes valores filtrados.Como se ve, el filtro reemplaza los "outliers" por valores acordes con los valores pasados de la serie, y de esta manera se logra un suavizado de la curva.

El modelo obtenido en este caso por mínimos cuadrados es

$$x_t^* = .409x_{t-1}^* + .184x_{t-2}^* + .131x_{t-3}^* + u_t.$$

El modelo estimado con el método robusto es

$$x_t^* = .91x_{t-1}^* - .08x_{t-2}^* - .03x_{t-3}^* + u_t,$$

el cual es muy próximo al modelo con el cual fueron generados las observaciones antes de cambiar las correspondientes a los "outliers". En la Figura 7 se pueden ver las predicciones obtenidas con ambos modelos para las primeras 50 observaciones, y puede observarse que las obtenidas con el método robusto ajustan mejor, especialmente en los períodos que siguen a la aparición de un "outlier".

Ejemplo 3. Este ejemplo consiste en 200 observaciones generadas a partir de un modelo no estacionario $ARI(1,1)$

$$x_t - x_{t-1} = .5(x_{t-1} - x_{t-2}) + u_t$$

En la Figura 8 se grafica esta serie, como asimismo los valores filtrados, que en este caso son 4 , aunque las correcciones son muy pequeñas. El modelo obtenido por mínimos cuadrados es

$$x_t - x_{t-1} = .414(x_{t-1} - x_{t-2}) + .08(x_{t-2} - x_{t-3}) + u_t,$$

y el modelo obtenido por el procedimiento robusto es

$$x_t - x_{t-1} = .46(x_{t-1} - x_{t-2}) + .08(x_{t-2} - x_{t-3}) + u_t.$$

En la Figura 9 se grafican los primeros 50 valores de la serie junto con las predicciones usando ambos modelos. Nuevamente, debido a que no hay "outliers", ambos modelos dan resultados muy similares. A continuación generamos 20 "outliers" sumando el valor 5 a 20 observaciones elegidas al azar. La serie modificada será llamada nuevamente x_t^* y se grafica en la Figura 10 juntamente con los valores filtrados. En la Figura 11 se grafican las primeras 50 observaciones. Usando el método usual de diferenciar la serie y luego aplicar mínimos cuadrados, obtenemos el modelo

$$x_t^* - x_{t-1}^* = -.379(x_{t-1}^* - x_{t-2}^*) + u_t.$$

El modelo obtenido por el procedimiento robusto es el siguiente

$$x_t^* - x_{t-1}^* = .23(x_{t-1}^* - x_{t-2}^*) + .19(x_{t-2}^* - x_{t-3}^*) + u_t.$$

En la Figura 12 se pueden apreciar las primeras 50 observaciones de la serie x_t^* y las predicciones obtenidas con ambos métodos. Como se ve en este gráfico, los resultados obtenidos con el procedimiento robusto son mucho más satisfactorios.

REFERENCIAS

Box,G.E.P y Jenkins,G.M. (1976). *Time Series forecasting and control.* Holden Day, San Francisco.

Martin,R.D, Samarov,A. y Vandaele,W. (1981). Robust methods in *ARMA* models. En *Applied Time Series Analysis of Economic Series*, editor A.Zellner. Economic Research Report ER5-, Bureau of the Census, Washington, DC.

Martin,R.D. y Yohai,V.J. (1985)Robustness in time series and estimating *ARMA* models. En *Handbook of statistics, vol 5*, editores E.J.Hannan, P.R.Krishnaiah y M.M.Rao. Elsevier Science, New York.

Autocorrelaciones	Autocorrelac. parciales		ac	d.s.	pac
. .	. .	1	-.0169	.23	-.0169
. .**	. .**	2	.1675	.23	.1673
. .	. .	3	.0284	.24	.0346
. .	. .	4	.0299	.24	.0030
. .	. .	5	.0230	.24	.0135
. .*	. .*	6	.0669	.24	.0630
. .**	. .**	7	.2082	.24	.2107
. .	. .	8	-.0208	.25	-.0347
. *.	. **.	9	-.0699	.25	-.1560
. *.	. *.	10	-.0859	.25	-.1070
. ***.	. ***.	11	-.3246	.25	-.3303
. .**	. .**	12	.1806	.27	.2252

Tabla 1: Tasas de inflación
Correlograma y correlograma parcial

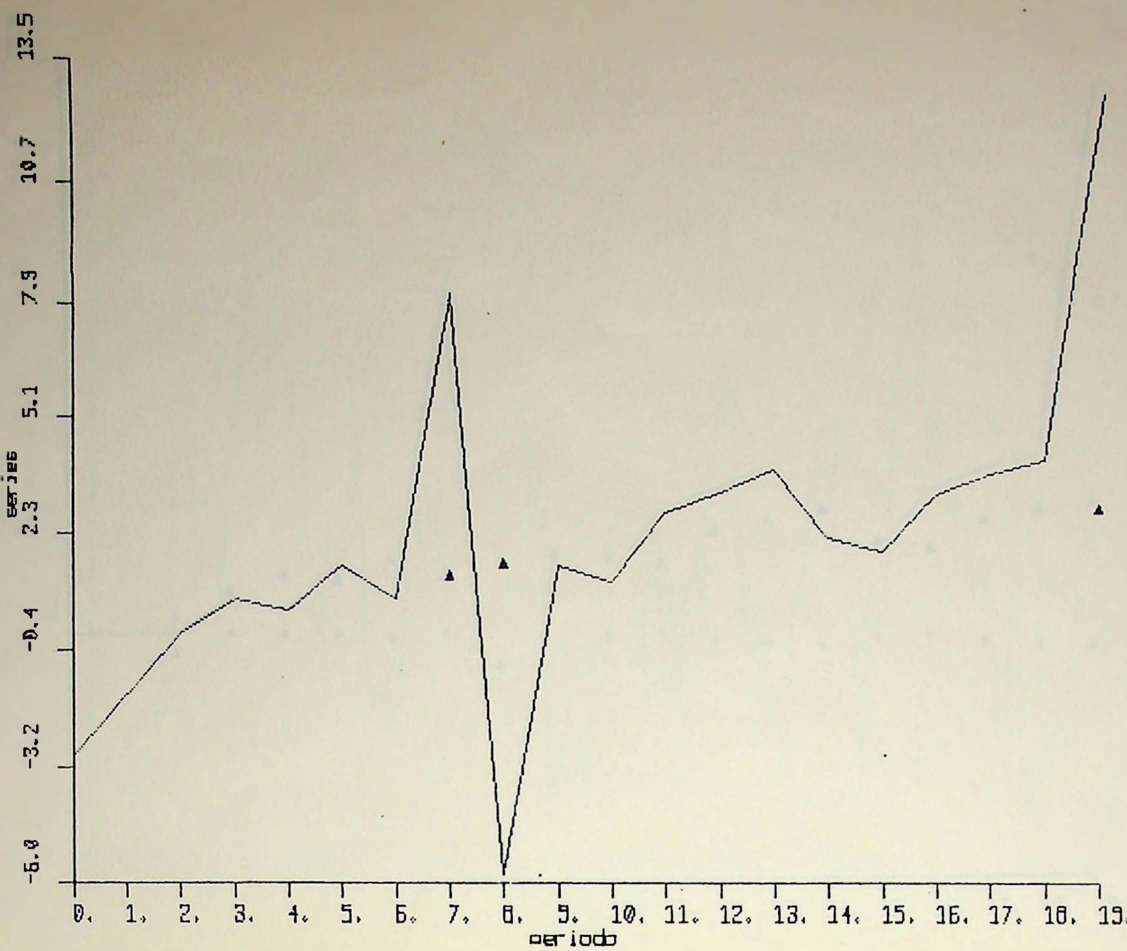

Figura 1: Tasas de inflación (Junio 1973 a Diciembre 1974)

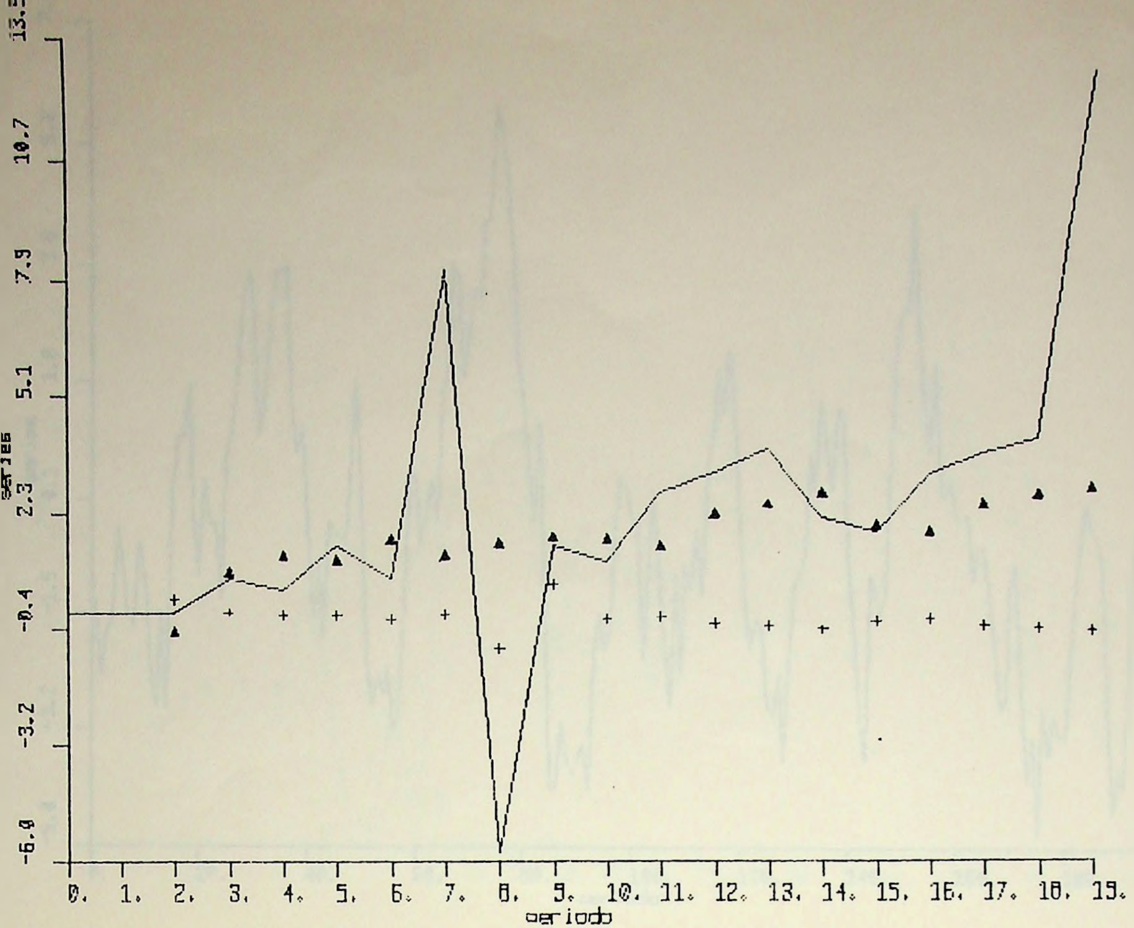

Figura 2: Tasas de inflación
+ predicción por mínimos cuadrados
▲ predicción robusta

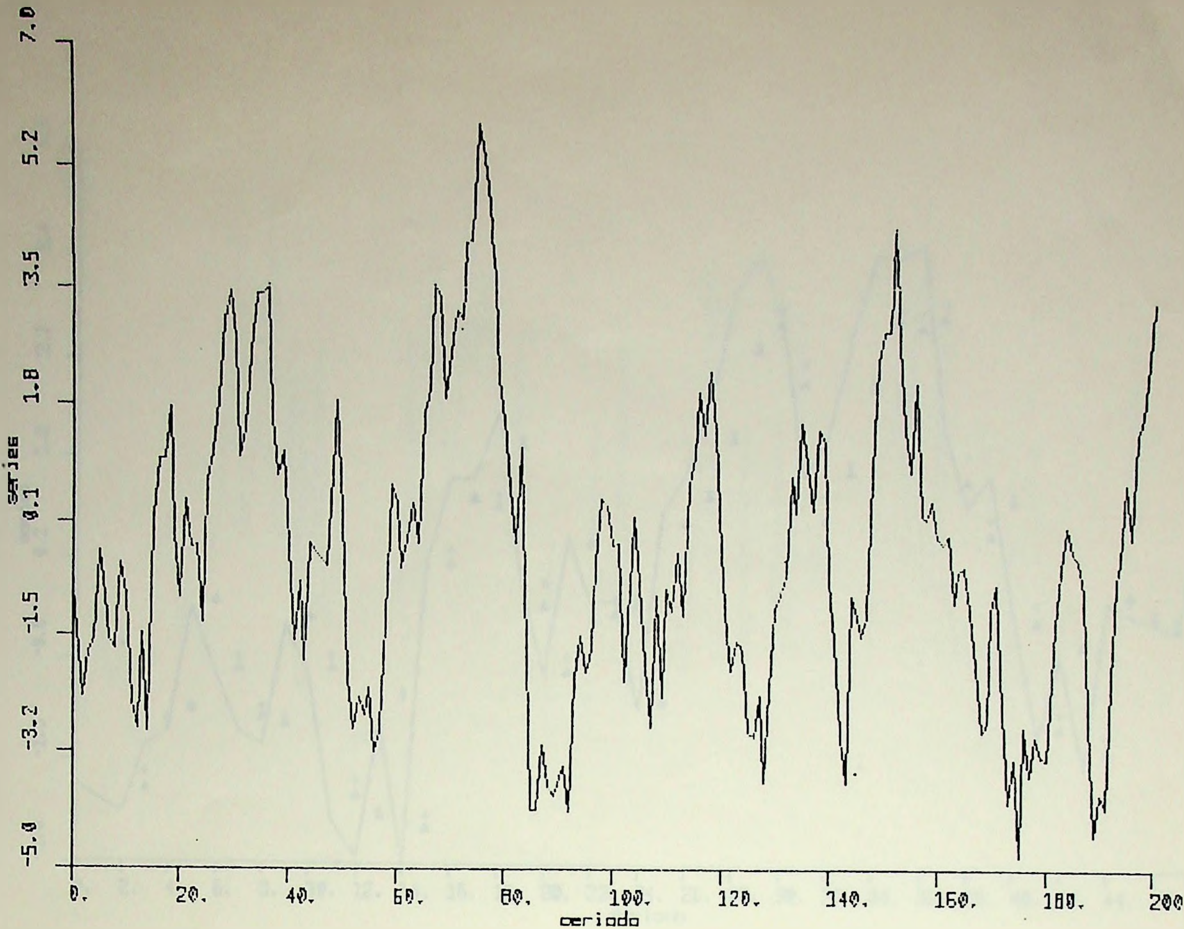

Figura 3: Proceso AR(1)

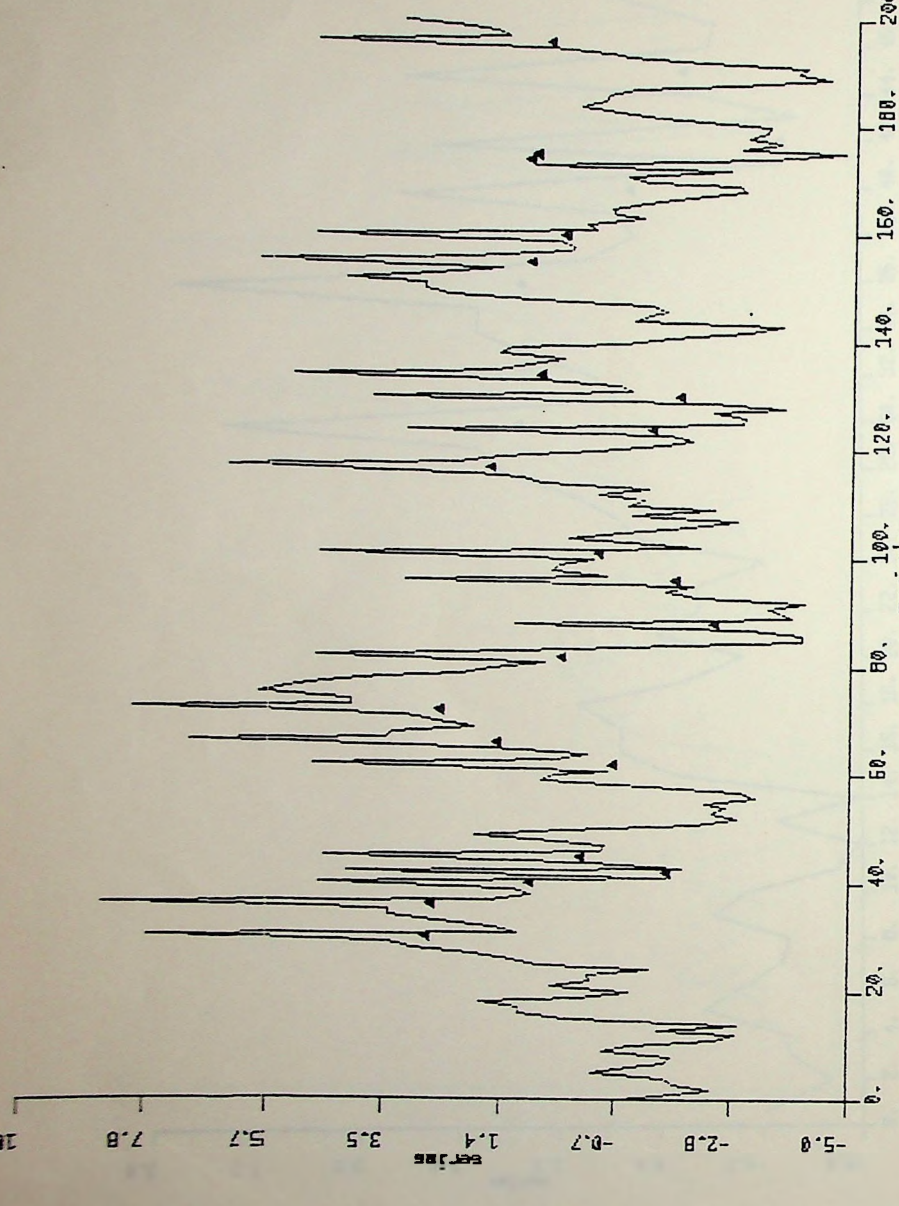

Figura 5: Proceso AR(1) contaminado

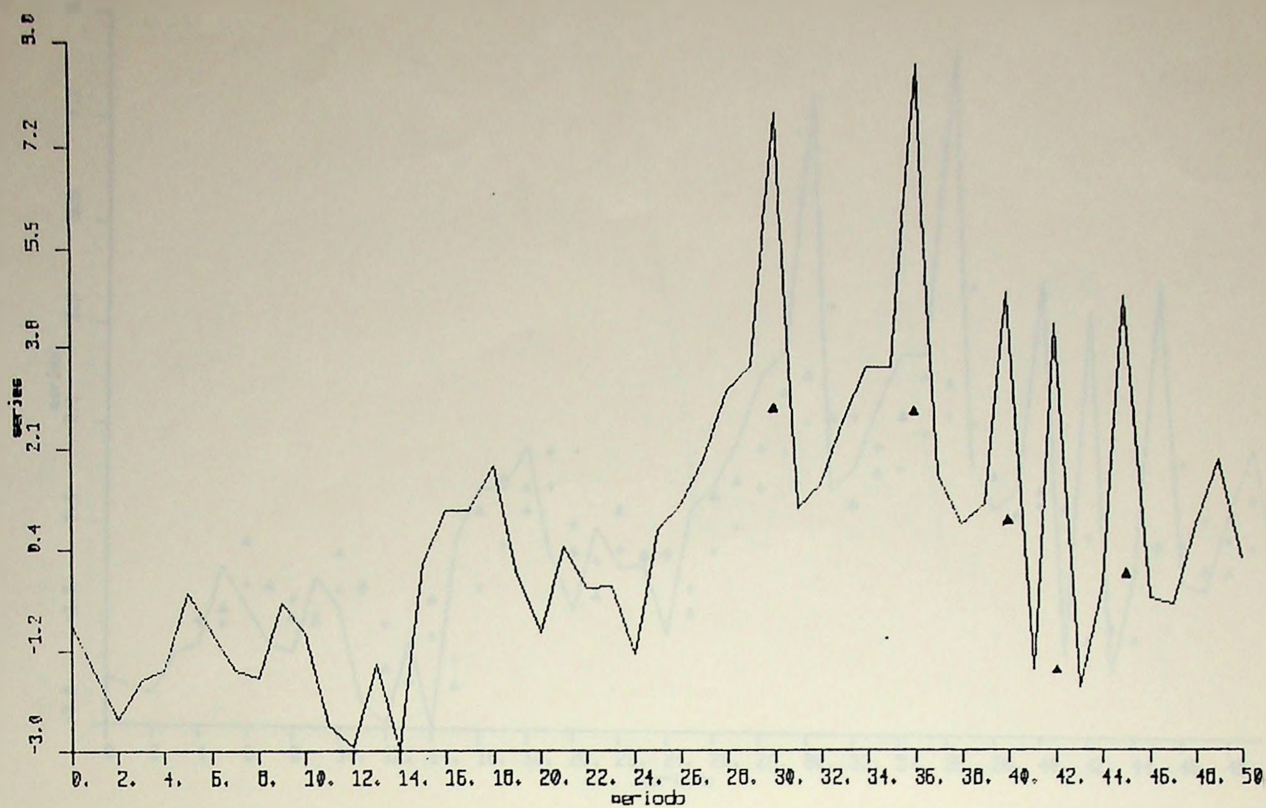

Figura 6: Proceso AR(1) contaminado
▲ valores filtrados

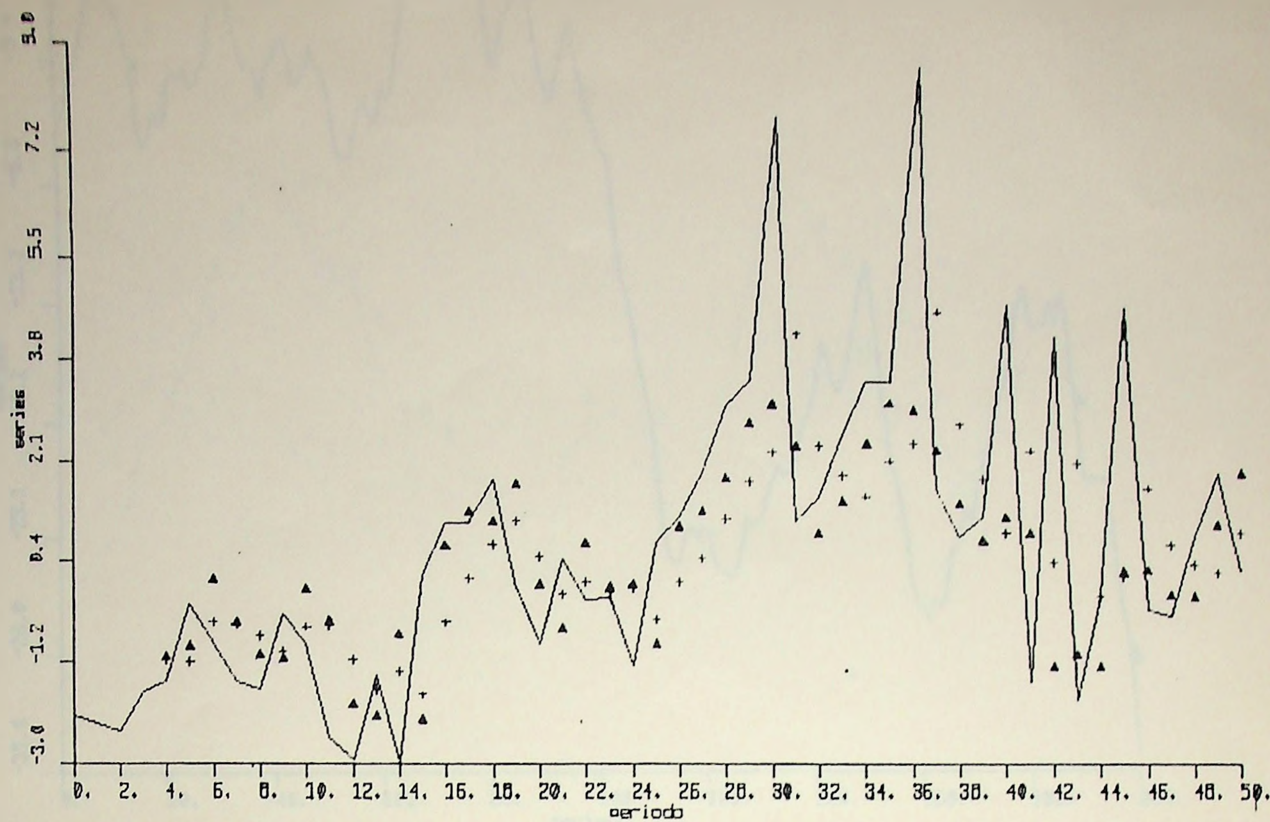

Figura 7: Proceso AR(1) contaminado
+ predicción por mínimos cuadrados
▲ predicción robusta

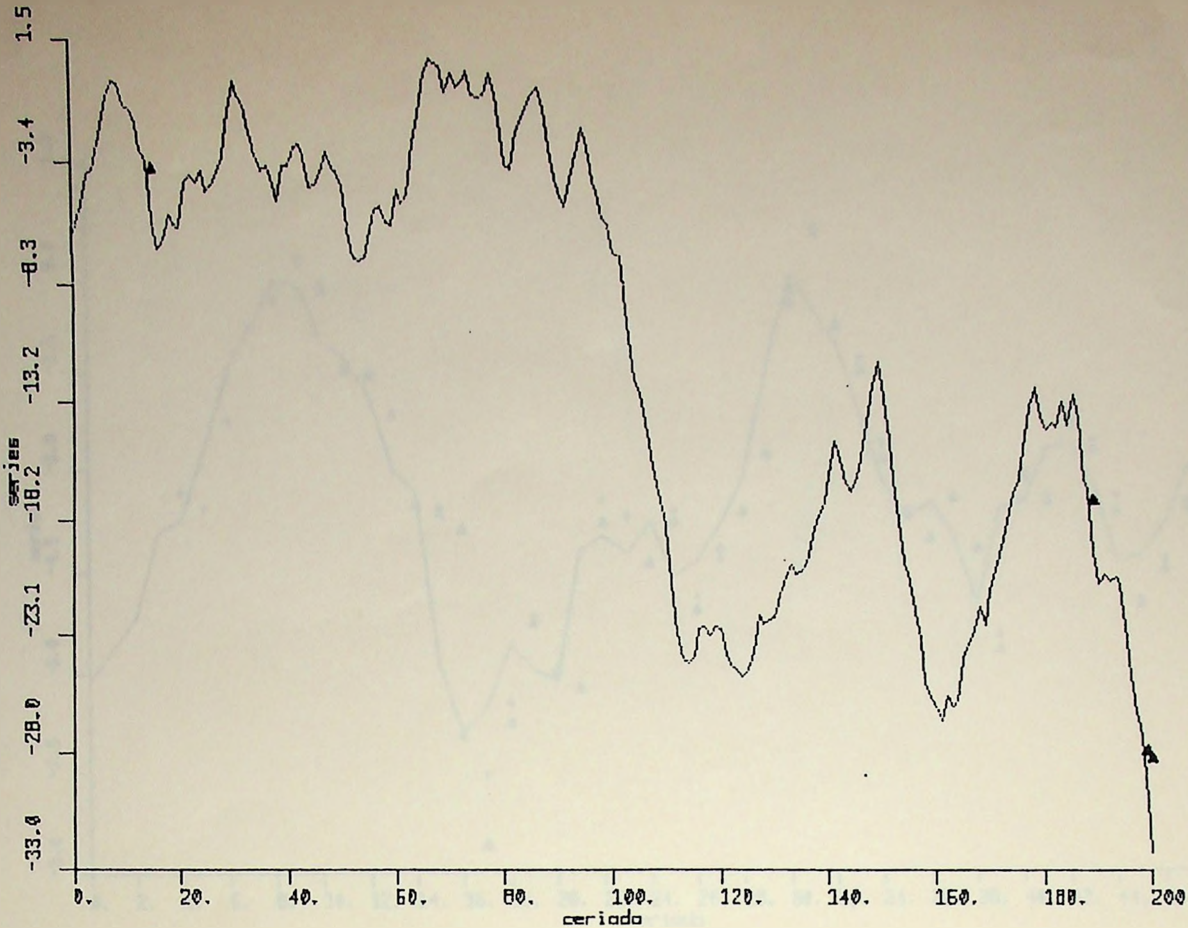

Figura 8: Proceso ARI(1,1)
▲ valores filtrados

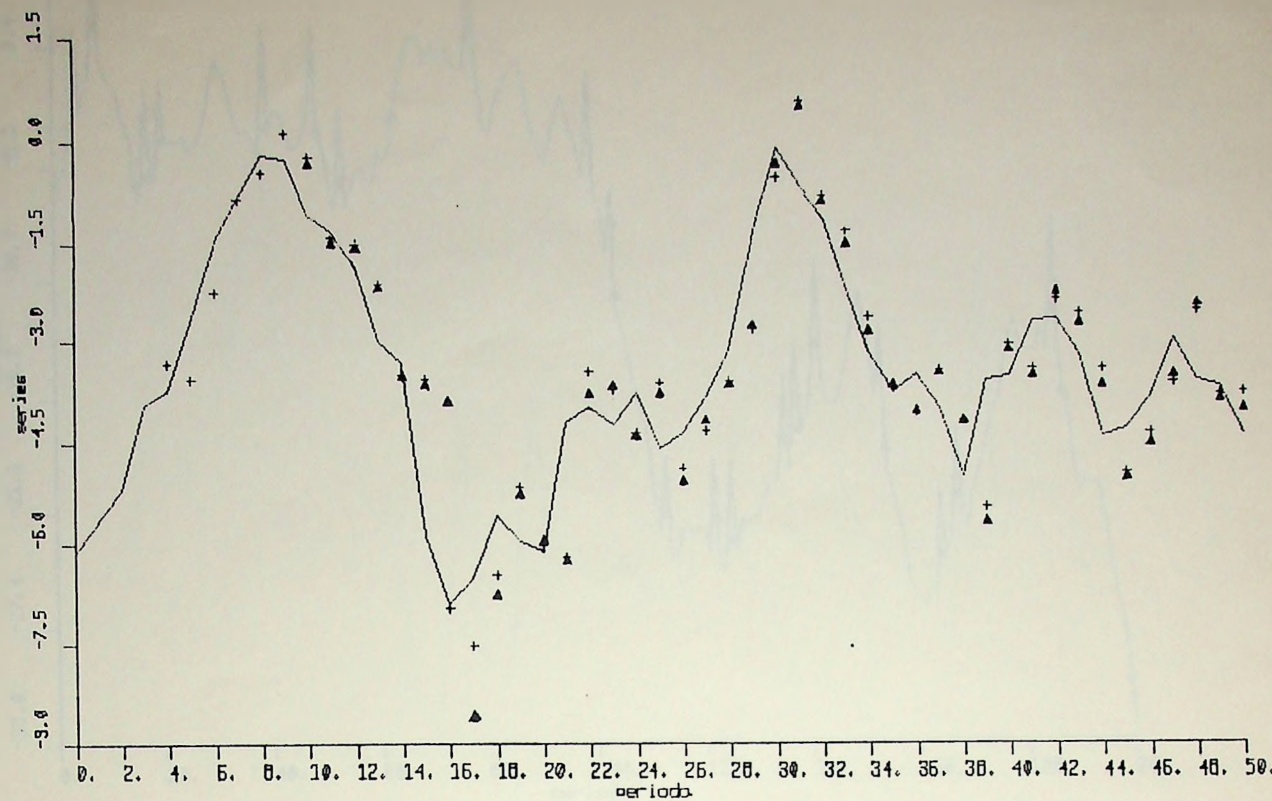

Figura 9: Proceso ARI(1,1)
+ predicción por mínimos cuadrados
▲ predicción robusta

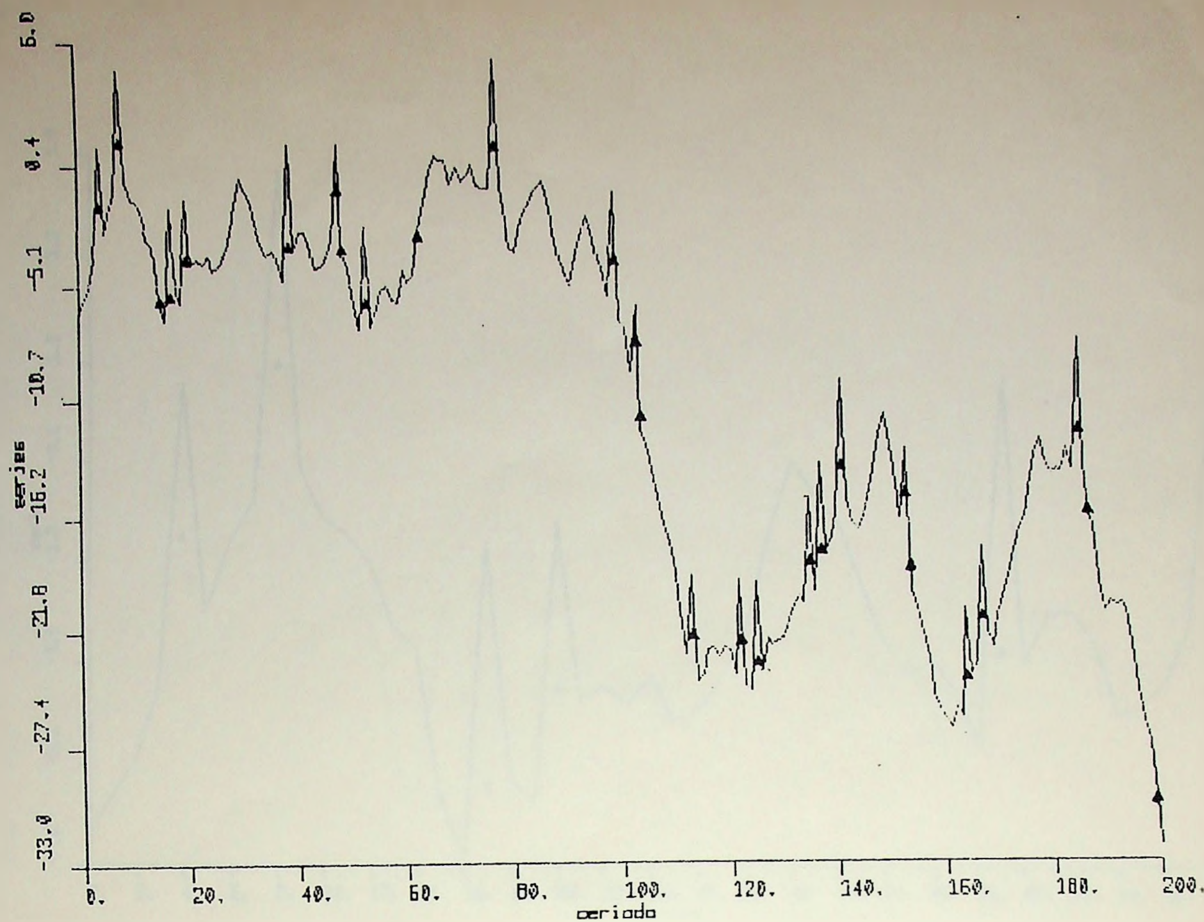

Figura 10: Proceso ARI(1,1) contaminado
▲ valores filtrados

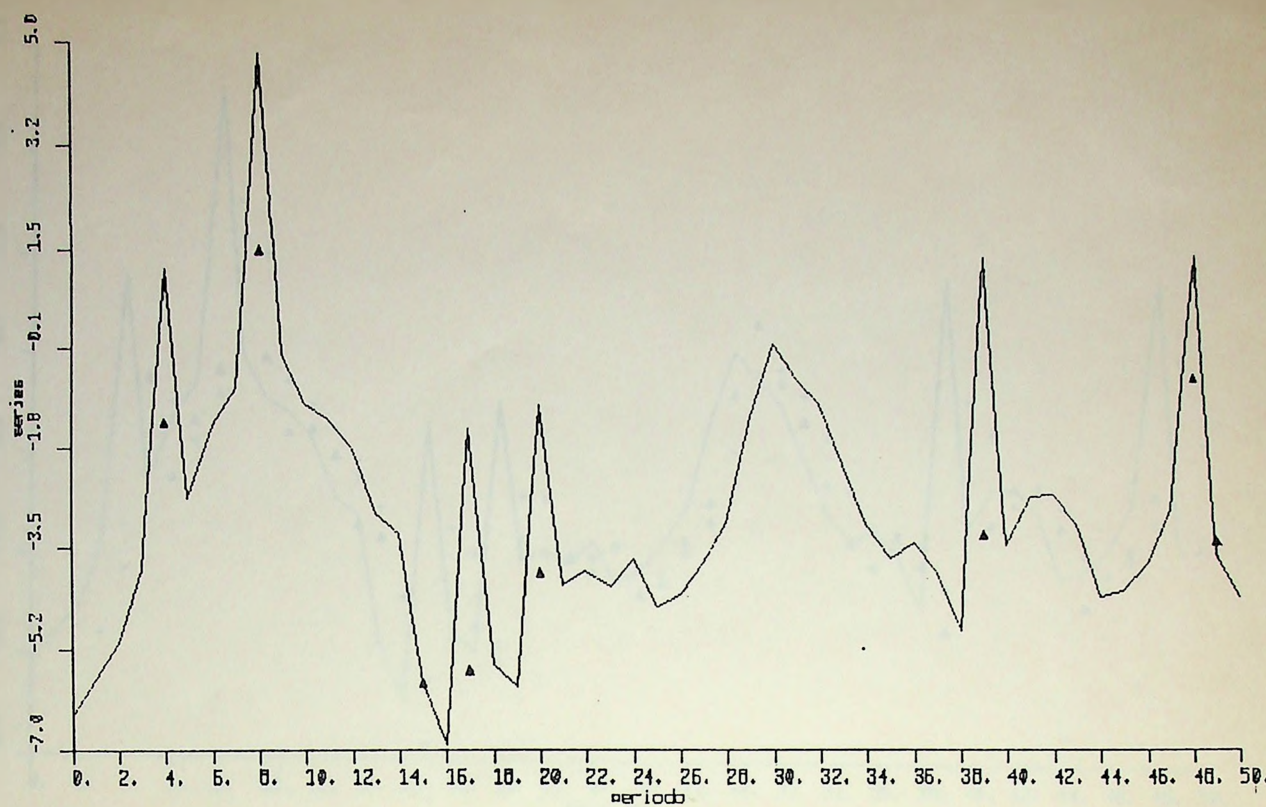

Figura 11: Proceso ARI(1,1) contaminado (50 primeras observaciones)
▲ valores filtrados

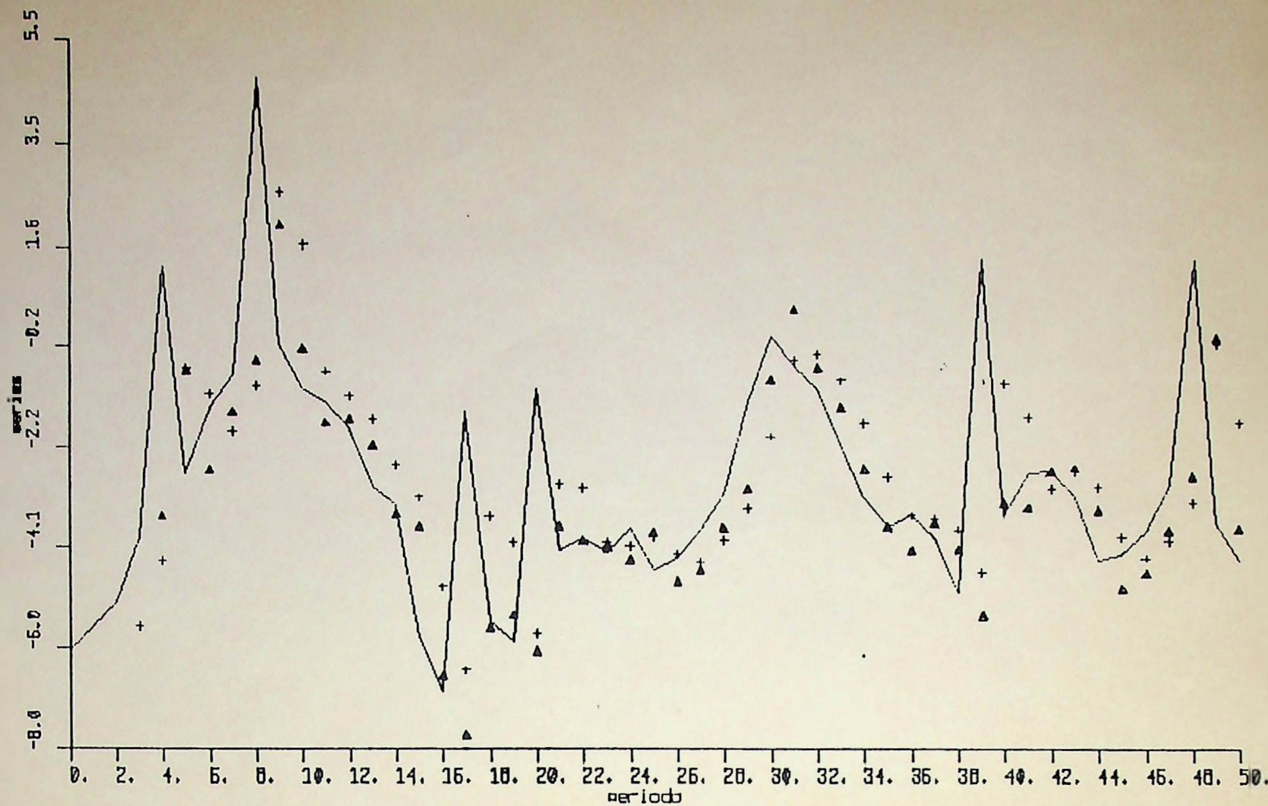

Figura 12: Proceso ARI(1,1)
+ predicción por mínimos cuadrados
▲ predicción robusta