Pronóstico mediante Modelos Probabilísticos: Una Herramienta en la Toma de Decisiones

Ana María Yepes de Castaño
Martha Eugenia Alvarez Villa

APLICACIÓN ECONÓMICA: PRONÓSTICO DE LA INFLACIÓN

"La mejor forma de hacer pronósticos **depende de los intereses del usuario del pronóstico y de los recursos del pronosticador".**

Spyros Makridakis

Todo pronóstico es una información referida al futuro. Existen varias formas de realizarlo, sea de manera subjetiva mediante el aporte de expertos en el tema

Ana María Yepes de Castaño. Economista de la Universidad de Antioquia. Magister en Administración de Empresas (EAFIT). Profesora adscrita al Departamento de Economía Universidad EAFIT.

Martha Eugenia Alvarez Villa. Ingeniera Industrial de la Universidad Nacional. Especialista en Sistemas de Información (EAFIT). Profesora adscrita al Departamento de Informática y Sistemas Universidad EAFIT.

o de forma analítica a través del estudio de los datos históricos, por medio de técnicas de modelación determinísticas o probabilísticas.

En el proceso de la toma de decisiones que conlleva un análisis histórico del fenómeno objeto de estudio, es muy improbable que se repitan exactamente las mismas condiciones con las cuales se hizo el pronóstico. Es por ésto que "toda predicción [1] realista debe ser condicional con respecto a los posibles contextos alternativos que puedan prevalecer en el futuro, . . . Por tanto, si usted adopta la política A y si el escenario externo II se mantiene, entonces el resultado deberá ser X" (Pulido: 1989A).

Todo pronóstico es una información referida al futuro. Existen varias formas de realizarlo, sea de manera subjetiva mediante el aporte de expertos en el tema o de forma analítica a través del estudio de los datos históricos, por medio de técnicas de modelación determinísticas o probabilístico.

Las variables macroeconómicas determinan el entorno económico en el cual está inmersa la empresa. Es indispensable por lo tanto, conocer las políticas económicas que influyen en los futuros valores de los agregados más importantes y que a su vez afectan los indicadores de tipo empresarial. Las relaciones micro y macro económicas se pueden visualizar a través de predicciones realizadas con estas variables.

(1) Algunos autores usan indistintamente las palabras predicción y pronóstico. En el presente trabajo, se asume el concepto de pronóstico cuando se modela con el fin de encontrar valores futuros de la variable.

El indicador de mayor prioridad en la fijación de la política macroeconómica colombiana es la inflación o variación porcentual del índice de precios al consumidor. La Reforma Constitucional de 1991 determinó como principal función de la Junta Directiva del Banco de la República velar por el poder adquisitivo de la moneda, reflejando en esta forma la importancia del control de la inflación.

El empresario debe fijar sus políticas y estrategias basado entre otras cosas, en el conocimiento de la evolución de los precios en la economía, ya que ésto incide sobre su política salarial, sus presupuestos de ingresos y gastos, su competitividad externa, sus posibilidades de crédito etc. Es importante entonces hacer un seguimiento al pronóstico de la inflación para ver cómo se verá afectada la empresa.

Se dispone de diferentes formas de hacer un pronóstico, dependiendo sobre qué se fundamenta el estudio:

1. Con base en la apreciación subjetiva: realizada de acuerdo a opiniones de expertos sobre el futuro del fenómeno en cuestión.
2. Con base en el análisis aislado de la serie: es el estudio de la evolución de la variable, teniendo como único elemento fundamental, los valores de la serie en períodos anteriores. Entre ellas:

- Análisis de descomposición temporal de la serie en: tendencia, ciclo, estacionalidad e irregularidad.
- Análisis temporal de procesos estocásticos: permite hallar un modelo estadístico de comportamiento autorregresivo, influenciado por valores anteriores de la variable como el modelo AR (Auto Regressive); aquel basado en medias móviles: afectado por un componente aleatorio como el MA (Moving Average) y combinaciones de los anteriores como el ARMA (Auto Regressive - Moving Average) y el ARIMA (Auto Regressive Integrated Moving Average).

3. Con base en la determinación de la relación existente entre algunas variables estocásticas:

- Análisis de regresión y correlación , lineal o no lineal a través de la cuantificación del grado de dependencia
- Métodos de simulación: conocidas las distribuciones de probabilidad
- Modelos econométricos: análisis estructural de las relaciones causa - efecto entre fenómenos, con incorporación de factores aleatorios.
- Modelos probabilísticos tales como aquellos obtenidos mediante Cadenas de Markov, Programación Dinámica, Teoría de Colas, Manejo de Inventarios entre otros.

En este artículo se mostrará una metodología de pronóstico mediante los modelos de autorregresión (AR), medias móviles (MA) y combinación de ellos (ARMA y ARIMA) y se tomará la inflación (2) de la economía colombiana comprendida entre Enero de 1991 y Diciembre de 1995 como el agregado económico por pronosticar. Los datos de la serie analizada corresponden al acumulado a Diciembre de cada año y la fuente es el Departamento Administrativo Nacional de Estadísticas (DANE).

El empresario debe fijar sus políticas y estrategias basado entre otras cosas, en el conocimiento de la evolución de los precios en la economía, ya que ésto incide sobre su política salarial, sus presupuestos de ingresos y gastos, su competitividad externa, sus posibilidades de crédito etc. Es importante entonces hacer un seguimiento al pronóstico de la inflación para ver cómo se verá afectada la empresa.

1. MODELOS AUTORREGRESIVOS (AR) Y MODELOS DE MEDIAS MÓVILES (MA)

Un modelo autorregresivo es aquel que explica una variable por su valor en uno o varios períodos anteriores, más un término de error, estableciendo una relación funcional de tipo lineal. El modelo más sencillo es aquel que

(2) La inflación (o cambio porcentual del IPC) mide la evolución de los precios de un grupo de artículos que conforman la canasta familiar, resultado de la Encuesta de Ingresos y Gastos, realizada por el Dane entre 1984 y 1985 y cuya implementación empezó desde Enero de 1989. La información para calcular los nuevos índices se obtiene mes a mes y el Dane reporta la inflación del mes actual, de doce meses atrás y del año corrido (inflación acumulada desde el comienzo del año en estudio).

considera solamente un período de desfase [3]. La representación es:

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t$$

donde y_t es la variable por modelar, los coeficientes por estimar son ϕ_0 y ϕ_1 siendo a el término de error, que se caracteriza por comportarse como una variable aleatoria con media nula, varianza constante y ausencia de correlación entre los valores referidos a distintos períodos. Dicho error se denomina "ruido blanco".

Una variable puede ser explicada por su valor inmediatamente anterior (autorregresión) o en modelos con datos estacionales (S: Stationary), mensuales o trimestrales, por los valores correspondientes al mismo periodo del año o años precedentes, designándose como Stationary Auto Regressive: SAR (# de orden).

Para efectuar pronósticos también se pueden utilizar los valores de la serie corregidos por los errores observados en otros períodos debidamente ponderados, se habla entonces de modelos de medias móviles de orden q [4], Moving Average: MA(q) [5].

(3) Desfasar es expresar el valor de una variable en función de sus valores en períodos anteriores. Si el desfase es igual a 1: se establece la relación tomando en cuenta **solamente** el período inmediatamente anterior. Si el desfase es de 2: se toman los dos períodos anteriores.

(4) q : número de períodos.

(5) Al igual que en los modelos autorregresivos los desfases pueden ser estacionales.

Estos modelos quedan definidos como:

$$y_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q}$$

Entre los modelos autorregresivos y los de medias móviles existe una relación. De tal manera que un proceso autorregresivo de primer orden para valores del parámetro ϕ en el modelo AR comprendidos entre cero y uno, es equivalente a una media móvil de infinitos términos con una ponderación θ_i decreciente en forma exponencial. Lo anterior se debe a que las ponderaciones de los errores a partir de t-2 tienden a cero, en el modelo MA.

La combinación de modelos autorregresivos y de medias móviles da lugar al modelo Auto Regressive Moving Average: ARMA (p,q) [6] que en su forma más sencilla ARMA (1,1), se define como:

$$y_t = \mu + a_t + \phi_1 y_{t-1} + \theta_1 a_{t-1}$$

Los modelos ARMA sólo pueden ser aplicados a series que no muestren tendencia [7]; una de las formas de eliminar la tendencia (o estacionar la serie), es calculando diferencias sucesivas.

La representación de estas diferencias es:

$$\Delta y_t = y_t - y_{t-1}$$

(6) ARMA (p,q): Modelo autorregresivo de medias móviles, donde p designa el números de desfases de la variables en el AR y q el número de períodos ponderados para el término de error en el modelo MA.

(7) Tendencia: es un cambio constante de la variable que pudiera expresarse en términos de una función, sea lineal, cuadrática etc.

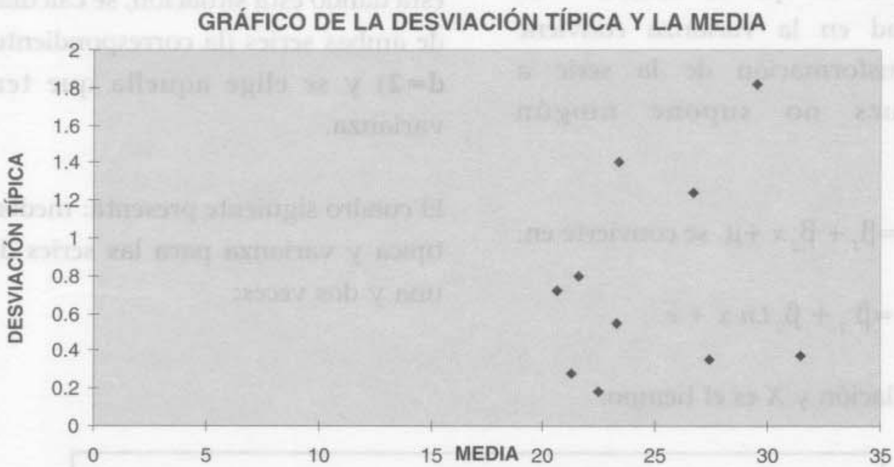
Antes de aplicar el modelo ARMA a una serie diferenciada hay que recalcularla deshaciendo la diferencia: *integrando* la serie. Surgen así los modelos Auto Regressive Integrated Moving Average: ARIMA (p,d, q), donde el primer número es el orden del proceso autorregresivo, el segundo el número de diferencias o de integraciones, y el tercero el orden del proceso de medias móviles.

2. TRANSFORMACIÓN PREVIA DE LA SERIE

La aplicación de modelos ARIMA debe realizarse con series estacionarias. Esto exige que la media y la varianza tengan un valor más o menos constante a lo largo del tiempo. Para ver si la varianza o la desviación típica es constante, se calcula su valor para distintos subperíodos, los cuales se eligieron semestralmente: de Enero a Junio y de Julio a Diciembre para cada año.

MEDIA (1991.01-1995.12)	DESVIACIÓN TÍPICA (1991.01-1995.12)
31.43333	0.37238
29.51667	1.820348
27.41667	0.354495
26.68333	1.236797
23.40000	1.401428
21.61667	0.793515
23.30000	0.547722
22.43333	0.18619
21.23333	0.280476
20.58333	0.716705

GRÁFICO 1



El que la serie sea estacionaria en la media implica que todos sus valores fluctúan en torno a un valor central que será la media de la serie; así también si la serie es estacionaria en varianza significa que la dispersión es constante durante todo el período muestral, no aumentando ésta con el paso del tiempo. Si la serie es estacionaria sólamente en la media, la grafica anterior se esperaría que diera como una línea recta vertical; si es estacionaria en varianza solamente, el diagrama de puntos mostraría una línea horizontal.

```
IDENT DLY
  Date: 8-12-1996 / Time: 9:20. SMPL range: 1991.02 - 1995.12. Number of observations: 59

    Autocorrelations      Partial Autocorrelations            ac        pac

        .    |****           .    |****            1     0.313     0.313
        .    |* .            .    | .              2     0.108     0.011
        .    |* .            .    |* .             3     0.113     0.084
      ***    | .          ****    | .              4    -0.194    -0.283
       .**   | .             .    | .              5    -0.150    -0.016
        .    | .             .    |* .             6    -0.019     0.052
        .*   | .             .    | .              7    -0.085    -0.035
        .*   | .            .*    | .              8    -0.079    -0.089
      ***    | .           ***    | .              9    -0.198    -0.241
      ***    | .            .*    | .             10    -0.221    -0.090
       .**   | .             .    | .             11    -0.118    -0.005
        .*   | .             .    | .             12    -0.092    -0.028
        .    | .             .    | .             13     0.034     0.015
        .    | .           .**    | .             14     0.011    -0.141
        .    | .            .*    | .             15    -0.020    -0.046
        .    |* .            .    |* .            16     0.068     0.049
        .*   | .            .*    | .             17    -0.041    -0.094
        .    |** .           .    |***            18     0.170     0.206
        .    |****           .    |* .            19     0.283     0.094
        .    |* .           .*    | .             20     0.097    -0.046
        .    |* .           .*    | .             21     0.097    -0.040
        .    | .             .    | .             22     0.036     0.004
      ***    | .           ***    | .             23    -0.269    -0.200
     ****    | .           ***    | .             24    -0.288    -0.221
        .*   | .             .    |* .            25    -0.101     0.066
       .**   | .            .*    | .             26    -0.142    -0.071
        .    | .             .    | .             27    -0.033     0.019
        .    | .             .    | .             28     0.021    -0.018
        .    | .             .    | .             29    -0.012     0.012
        .    | .             .    | .             30    -0.012     0.005
        .    | .            .*    | .             31    -0.028    -0.075
        .    |* .            .    | .             32     0.057     0.020
        .    |* .           .*    | .             33     0.078    -0.053
        .    |* .           .*    | .             34     0.050    -0.075
        .    |* .            .    | .             35     0.061    -0.024
        .*   | .           .**    | .             36    -0.042    -0.128

  Box-Pierce Q-Stat 38.29   Prob 0.3661   SE of Correlations 0.130
  Ljung-Box Q-Stat 54.72    Prob 0.0235
```

Se obtiene un peor resultado con la serie de dos diferencias, por ello se selecciona la serie DLY.

Se ha comprobado que la serie es estacionaria en media y en varianza en la parte regular, pero también hay que comprobarlo para la parte estacional. En el caso de la estacionariedad en varianza, se asume tanto para la parte regular como para la parte estacional. La estacionariedad en media, hay que comprobarla también sobre la parte estacional. Para ello se establece que si al evaluar las funciones de autocorrelación y autocorrelación parcial anuales (12, 24, 36, ...) se sale al menos un valor, de la banda de confiabilidad en el correlograma, se necesita una diferenciación en la parte estacional, como se verá en el numeral 4.

El dato 24 se sale de las bandas, en el correlograma, por tanto la serie necesita una diferenciación en la parte estacional, cada 12 períodos, la que se denominará DDLE: Doble Diferenciación del Logaritmo de *Y* en la parte Estacional.

El modelo presentará el valor D=1; aunque el nombre implica dos diferencias, la primera vez que se hizo la diferenciación, no fue sobre la parte estacionaria.

4. IDENTIFICACIÓN DEL MODELO

El instrumento técnico básico para identificar un modelo ARIMA es la denominada *función de autocorrelación* que mide el grado de correlación entre cada valor de la variable y los desfasados 1, 2,..., h períodos. Pero no basta con la función de autocorrelación, es preciso disponer además de la ***función de autocorrelación parcial*** (9).

Dado sin embargo, que de cada modelo se sabe qué gráfico de autocorrelación y autocorrelación parcial debe corresponder (teóricamente), el proceso de identificación o selección del modelo se simplifica al tener que elegir entre una serie de alternativas conocidas a priori.

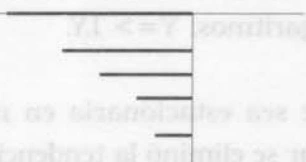
Función de autocorrelación	*Función de autocorrelación parcial*	*Modelo ARMA*
Decrece	1 ó 2 valores significativos	AR(1);AR(2)
1 ó 2 valores significativos	Decrece	MA(1);MA(2)
Decrece	Decrece	ARMA(1,1)
1 ó 2 valores significativos	1 ó 2 valores significativos	No existe

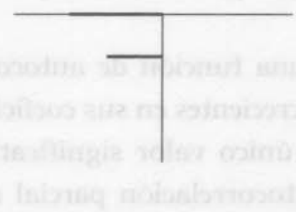
(9) La función de autocorrelación parcial esta constituida por los coeficientes de correlación parcial para diferentes retardos, los cuales miden la aportación que a las variaciones de una variable tiene otra en particular, aislados los efectos de las posibles restantes variables explicativas (céteris páribus).

Como se puede observar en el cuadro anterior:

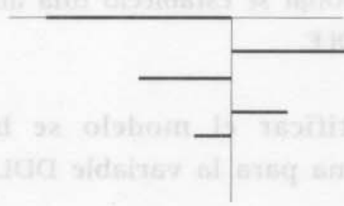
* En un modelo AR la función de autocorrelación decrece (en forma regular, alternando valores positivos y negativos o en forma de ondas sinuosoidales) y la de autocorrelación parcial presenta entonces un número de coeficientes igual al orden del proceso. Se puede apreciar lo anterior en los siguientes correlogramas:

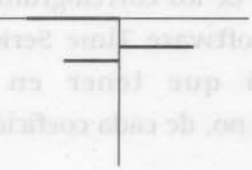
CORRELOGRAMA SIMPLE AR (1)	CORRELOGRAMA PARCIAL AR (1)





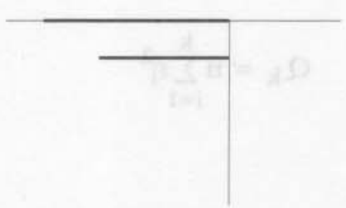
CORRELOGRAMA SIMPLE AR (2)	CORRELOGRAMA PARCIAL AR (2)

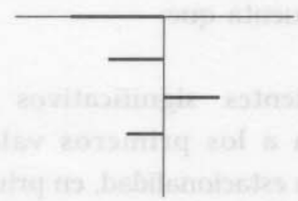




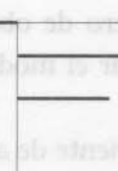
* En forma simétrica, un modelo MA(1) tiene sólo el primer coeficiente distinto de cero en la función de autocorrelación (y la de autocorrelación parcial decreciendo); un modelo MA(2) tendrá dos coeficientes no nulos.

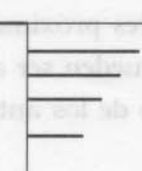
CORRELOGRAMA SIMPLE MA (1)	CORRELOGRAMA PARCIAL MA (1)





CORRELOGRAMA SIMPLE MA (2)	CORRELOGRAMA PARCIAL MA (2)





* Un modelo ARMA(1,1) presentará diversas combinaciones, pero siempre decreciendo ambas funciones, la de autocorrelación y la de autocorrelación parcial.

Para la identificación de la componente estacional, las reglas son similares, pero aplicadas a los coeficientes de autocorrelación y autocorrelación parcial cada S períodos.

Por ejemplo, una función de autocorrelación con valores decrecientes en sus coeficientes 12, 24, 36, y un único valor significativo en la función de autocorrelación parcial en 12, se identificaría como un SAR(12).

En las gráficas de los correlogramas obtenidas a través del software Time Series Processor (TSP), habrá que tener en cuenta la significación o no, de cada coeficiente.

Las gráficas presentan automáticamente unas bandas de significación a partir de las cuales se podrían considerar los coeficientes significativos con un nivel de confianza bastante elevado (95%). Conocidas las bandas, a la hora de la identificación del modelo también se tendrá en cuenta que:

- Los coeficientes significativos que no correspondan a los primeros valores o a múltiplos de la estacionalidad, en principio, no deberán considerarse, dado que no tienen tanta influencia como aquellos.

- Los coeficientes próximos a los retardos de estacionalidad pueden ser asimilados como un posible contagio de los anteriores.

- Los coeficientes teóricos relativamente pequeños no es fácil detectarlos como significativamente distintos de 0.

DDLE es la serie que se utiliza para identificar el modelo, ya que es estacionaria en media y en varianza:

- Para que sea estacionaria en varianza se ha aplicado logaritmos: Y=> LY.

- Para que sea estacionaria en media en la parte regular se eliminó la tendencia aplicando una diferencia: LY => DLY.

- Para que sea estacionaria en media en la parte estacional se estableció una diferencia: DLY => DDLE.

Para identificar el modelo se hace un correlograma para la variable DDLE en 36 período:

En el momento de definir el modelo adecuado en términos de la no correlación de los residuos, se deben emplear los estadísticos de Box-Pierce y el de Ljung-Box. Así el Box-Pierce se calcula:

$$Q_k = n\sum_{i=1}^{k} r_i^2$$

Donde:

k : número de elementos del correlograma residual

n : número de observaciones usadas para ajustar el modelo.

r_i^2 : coeficiente de autocorrelación simple de la serie de los residuos

IDENT DDLE

Date: 8-12-1996 / Time: 9:20. SMPL range: 1992.02 - 1995.12. Number of observations: 47

Autocorrelations	Partial Autocorrelations		ac	pac
. |******	. |******	1	0.433	0.433
. |****	. |* .	2	0.281	0.115
. |*** .	. |* .	3	0.230	0.090
. ** | .	*****| .	4	-0.142	-0.364
. * | .	. |* .	5	-0.060	0.097
. * | .	. * | .	6	-0.107	-0.067
.*** | .	. * | .	7	-0.212	-0.063
. * | .	. | .	8	-0.055	0.018
**** | .	**** | .	9	-0.271	-0.285
**** | .	. ** | .	10	-0.339	-0.180
. ** | .	. |* .	11	-0.167	0.040
**** | .	. ** | .	12	-0.302	-0.118
. ** | .	. | .	13	-0.122	0.028
. | .	. | .	14	0.018	-0.022
. | .	. | .	15	-0.028	-0.014
. |* .	. * | .	16	0.094	-0.083
. | .	. ** | .	17	-0.009	-0.136
. |* .	. |***.	18	0.095	0.204
. |** .	. | .	19	0.185	-0.033
. |* .	. * | .	20	0.045	-0.093
. |** .	. * | .	21	0.123	-0.054
. |** .	. |* .	22	0.151	0.053
. * | .	. ** | .	2	-0.080	-0.187
. * | .	. * | .	24	-0.071	-0.107
. * | .	. | .	25	-0.104	0.021
. ** | .	. * | .	26	-0.164	-0.079
. | .	. | .	27	-0.012	0.019
. | .	. |* .	28	-0.038	0.056
. | .	. | .	29	0.026	-0.000
. |** .	. |* .	30	0.136	0.055
. | .	. | .	31	0.010	0.002
. |* .	. | .	32	0.045	-0.018
. | .	. ** | .	33	0.022	-0.163
. * | .	. | .	34	-0.043	0.017
. | .	. | .	35	0.001	-0.032
. | .	. * | .	36	0.008	-0.045

Box-Pierce Q-Stat 41.92 Prob 0.2294 SE of Correlations 0.146

Ljung-Box Q-Stat 57.88 Prob 0.0118

Si el modelo adecuado para cierta serie es un ARIMA (p,d,q), Q_k sigue una distribución X^2 con (k - p - q) grados de libertad y dejará una probabilidad de cola derecha correspondiente en la distribución X^2, mayor que un nivel de significancia deseado (usualmente 0.05), para probar la hipótesis nula de la existencia de ruido blanco.

Un estadístico alternativo del anterior pero más exigente es el Ljung-Box que se calcula de la siguiente manera:

$$Q_k = T(T+2)\sum_{i=1}^{k} r_i^2 / (T-1)$$

Donde T es la longitud de la serie. El criterio de aceptación o de rechazo, es similar al anterior estadístico.

En el correlograma se puede ver que el valor de Q-Statistic de Box - Pierce es igual a 41,92 y que deja una probabilidad de cola derecha en la Chi-cuadrado de 0.2294 mayor que un nivel de significancia de 0.05, por lo tanto no hay evidencia para rechazar la hipótesis de no correlación de los residuos o la existencia de ruido blanco y además se puede pensar en la viabilidad de modelar por medio de ARIMA y pronosticar los valores de la variable.

No resulta fácil seleccionar el modelo ARIMA a partir de la función de autocorrelación (total y parcial) estimada a través de los datos. Se ha seleccionado tres modelos alternativos y posteriormente se comprobará cuál de ellos resulta finalmente más idóneo. No se debe considerar la etapa de identificación del modelo como una decisión irreversible, sino como el inicio de un proceso de selección, el cual se revisará en etapas posteriores.

Según el correlograma se puede identificar en la parte regular un AR(1). En la parte estacional se presentan dudas para identificar el proceso, teniendo como alternativas un SARMA(1,1), un SAR(1) y un SMA(1), por tanto se probarán para ver cuál presenta un mejor resultado.

5. ESTIMACIÓN DE COEFICIENTES

En TSP se calculan los coeficientes de la regresión múltiple con AR(1), SAR(12) y SMA(12) como variables predictoras y DDLE como variable dependiente:

*** Prueba 1: AR(1) SAR(12) SMA(12)**

LS // Dependent Variable is DDLE
Date: 8-12-1996 / Time: 9:21.
SMPL range: 1993.03 - 1995.12.
Number of observations: 34.
Convergence achieved after 5 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
SMA(12)	0.7861407	0.0756888	10.386491	0.0000
AR(1)	0.4940823	0.1266606	3.9008360	0.0005
SAR(12)	-0.7963099	0.1647372	-4.8338202	0.0000

R-squared	0.525483	Mean of dependent var	-0.001288
Adjusted R-squared	0.494869	S.D. of dependent var	0.037593
S.E. of regression	0.026718	Sum of squared resid	0.022130
Log likelihood	76.48851	F-statistic	17.16481
Durbin-Watson stat	1.858888	Prob(F-statistic)	0.000010

*** Prueba 2: AR(1) SAR(12)**

LS // Dependent Variable is DDLE
Date: 8-12-1996 / Time: 9:22.
SMPL range: 1993.03 - 1995.12
Number of observations: 34.
Convergence achieved after 3 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
AR(1)	0.5016061	0.1593329	3.1481628	0.0035
SAR(12)	-0.3484466	0.1823762	-1.9105929	0.0651

R-squared	0.335871	Mean of dependent var	-0.001288
Adjusted R-squared	0.315117	S.D. of dependent var	0.037593
S.E. of regression	0.031111	Sum of squared resid	0.030972
Log likelihood	70.77345	F-statistic	16.18338
Durbin-Watson stat	1.989502	Prob(F-statistic)	0.000328

* **Prueba 3: AR(1) SMA(12)**

LS // Dependent Variable is DDLE
Date: 8-12-1996 / Time: 9:22
SMPL range: 1992.03 - 1995.12
Number of observations: 46
Convergence achieved after 9 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
SMA(12)	-0.8757845	0.0355321	-24.647698	0.0000
AR(1)	0.4846431	0.0994675	4.8723767	0.0000

R-squared	0.600144	Mean of dependent var	-3.83E-05
Adjusted R-squared	0.591056	S.D. of dependent var	0.033751
S.E. of regression	0.021583	Sum of squared resid	0.020497
Log likelihood	112.1998	F-statistic	66.03950
Durbin-Watson stat	1.923417	Prob(F-statistic)	0.000000

La salida del computador indica: el total de observaciones, la variable dependiente (DDLE) y el número de iteraciones que se han necesitado para alcanzar una suficiente convergencia en los valores de los coeficientes estimados.

En segundo lugar se recogen para cada componente AR o MA el valor del coeficiente, su desviación típica, el estadístico t y la probabilidad de que, dados esos resultados, el parámetro teórico sea nulo, con un nivel de confianza del 95%.

6. CONTRASTE DE LA VALIDEZ DEL MODELO

Después de determinados los modelos factibles, es necesario realizar algunas pruebas o contrastes, con los cuales se pueden tener criterios más objetivos y complementarios, para determinar aquel que pueda ayudar a hacer un mejor pronóstico. Es deseable un modelo que tenga mínima varianza, mayor porcentaje de explicación de la variable y en general lograr la mayor verosimilitud.

* ***Suma de los residuos***

En las tres pruebas la suma de los cuadrados de los residuos (SSR) es reducida, por lo que los tres modelos se podrían aceptar como buenos. Pero se tienen que analizar más aspectos.

* ***Desviación típica de la regresión***

También en las tres pruebas la desviación típica (SD) es reducida, por lo que los tres modelos se podrían aceptar como igualmente idóneos. Aunque se deben analizar otros criterios.

*** Coeficiente de determinación**

Este coeficiente nos da el grado de explicación del modelo a través de las variables exógenas. Valores elevados del mismo mostrarán la mayor bondad del modelo, considerándose bastante apropiados resultados próximos a 0,9.

En cualquier aplicación es importante aclarar si el coeficiente de correlación está referido a la variable originaria o a la transformada. En el caso del TSP, R^2 se calcula sobre la variable transformada y eso supone trabajar con unos valores sensiblemente más bajos que los de otros programas que calculan R^2 en términos de la variable inicial. Como punto de referencia, coeficientes de correlación de 0,5 a 0,7 en una variable en diferencias pueden ser equivalentes a coeficientes frecuentemente superiores a 0,9 en la variable original.

Después de determinados los modelos factibles, es necesario realizar algunas pruebas o contrastes, con los cuales se pueden tener criterios más objetivos y complementarios, para determinar aquel que pueda ayudar a hacer un mejor pronóstico. Es deseable un modelo que tenga mínima varianza, mayor porcentaje de explicación de la variable y en general lograr la mayor verosimilitud.

En el presente caso el mejor modelo será el tercero -que corresponde a AR(1) y SMA(2)-, al tener un R^2 de 0,6 frente al 0,52 y 0,33 de la prueba 1 y 2 respectivamente.

*** Contraste T de significación de parámetros**

Un parámetro podrá considerarse como estadísticamente significativo (a niveles de confianza próximos al 95%) si el estimador tiene un valor que supere dos veces su desviación típica o, lo que es lo mismo, si el estadístico T es superior a dos (en valor absoluto).

- En la prueba 1: todos los parámetros son estadísticamente significativos.
- En la prueba 2: SAR(12) tiene un estadístico T de -1,9 con lo que no es estadísticamente significativo. Por lo cual se rechazaría este modelo.
- En la prueba 3: todos los parámetros son estadísticamente significativos.

*** Contraste F**

Es el contraste que muestra la significación conjunta del modelo. Para comprobarlo, se compara la probabilidad que deja el estadístico F (de cola derecha) con el nivel de significancia α deseado. Si la probabilidad es menor que tal nivel se acepta que el modelo es apropiado significativamente. En este caso, en los tres modelos las probabilidades están muy cercanas a cero.

*** Logaritmo de verosimilitud**

Será preferible aquel modelo que presente un valor de L (Log likelihood) más bajo. En el caso estudiado, este contraste presenta unos valores de 76,48851, 70,77345 y 112,1998 respectivamente.

*** *Contraste de Durbin-Watson***

El contraste establecido por Durbin y Watson permite verificar la hipótesis alternativa de un proceso autorregresivo de primer orden en los residuos ($a_t = \rho a_{t-1}+e_t$) respecto a la hipótesis nula Ho: $\rho = 0$. El cálculo del estadístico d se realiza según la siguiente fórmula, a partir de los residuos del modelo, y guarda una estrecha relación con el propio coeficiente de autocorrelación estimado:

$$d = \frac{\sum_{t=2}^{n} (at - a_{(t-1)})^2}{\sum_{t=1}^{n} at^2} \approx 2\,(1-\rho)$$

y puesto que ρ varía entre -1 y +1, d oscilará entre 0 y 4, en forma tal que

$\rho = 1$(autocorrelación positiva perfecta) $d = 0$

$\rho = -1$(autocorrelación negativa perfecta) $d = 4$

$\rho = 0$ (ausencia de autocorrelación) $d = 2$

Dado que al elaborar un modelo ARMA se quiere obtener unos residuos que se comporten como ruido blanco y que, por tanto, no presenten autocorrelación de ningún orden, el valor ideal de d sería de 2. En la práctica, se consideran como aceptables valores relativamente cercanos a 2, podríamos decir que entre 1,5 y 2,5.

Aparte de este contraste de validez sólo parcial para una autocorrelación de primer orden entre los residuos, es habitual utilizar el contraste Q, aplicado directamente a la serie de residuos del modelo.

CRITERIOS DE VALIDEZ OBTENIDOS PARA LOS TRES MODELOS

	Modelo 1	Modelo 2	Modelo 3
Suma de cuadrados de los residuos	0,022130	0,030972	0,020497*
Desviación típica	0,026718	0,031111	0,021583*
Coeficiente de determinación	0,525483	0,335871	0,600144*
Coeficiente de determinación corregido	0,494869	0,315117	0,591056*
Logaritmo de verosimilitud	76,48851	70,77345	112,1998
Estadístico de Durbin-Watson	1,858888	1,989502	1,923417
Contraste F	17,16481	16,18338	66,03950*

Con todos los resultados obtenidos tras las regresiones, se considera más eficiente la **Prueba 3** por los mejores contrastes alcanzados por ésta (que se puede observar en aquellos señalados con asterisco). Por tanto nuestro modelo final es un AR(1) SMA(12).

IDENT RESID

Date: 8-08-1996 / Time: 12:38. SMPL range: 1992.02 - 1995.12. Number of observations: 47

```
Autocorrelations   Partial Autocorrelations        ac       pac

     .    |.          .    |. .         1    0.011    0.011
     .    |* .        .    |* .         2    0.048    0.048
     .    |* .        .    |* .         3    0.109    0.108
     .*** |.          .*** |.           4   -0.231   -0.238
     . ** |.          . ** |.           5   -0.140   -0.152
     . ** |.          . ** |.           6   -0.181   -0.181
     .    |* .        .    |** .        7    0.082    0.167
     .    |* .        .    |* .         8    0.042    0.053
     . ** |.          .*** |.           9   -0.183   -0.252
     . ** |.          *****|.          10   -0.148   -0.369
     .    |* .        .    |** .       11    0.106    0.154
     . ** |.          .    |.          12   -0.151    0.023
     .    |** .       .    |** .       13    0.118    0.170
     .    |* .        .*** |.          14    0.099   -0.207
     .    |.          .*** |.          15   -0.001   -0.204
     .    |** .       .    |** .       16    0.161    0.151
     **** |.          . ** |.          17   -0.315   -0.129
     .    |* .        .    |.          18    0.062    0.006
     .    |* .        .    |* .        19    0.100    0.045
     .  * |.          .  * |.          20   -0.100   -0.040
     .    |* .        .  * |.          21    0.080   -0.081
     .    |* .        .    |.          22    0.063   -0.008
     .  * |.          .    |.          23   -0.042   -0.029
     .*** |.          .*** |.          24   -0.197   -0.197
     .  * |.          .    |.          25   -0.041    0.031
     .    |.          .  * |.          26   -0.018   -0.098
     .    |.          .  * |.          27    0.029   -0.069
     .    |.          .    |.          28   -0.028    0.034
     .    |** .       .  * |.          29    0.127   -0.061
     .    |* .        .  * |.          30    0.040   -0.048
     .  * |.          .    |.          31   -0.082   -0.011
     .    |* .        .    |.          32    0.096   -0.002
     .  * |.          .  * |.          33   -0.055   -0.069
     .    |* .        . ** |.          34    0.046   -0.129
     .    |.          .  * |.          35   -0.015   -0.069
     .    |.          .  * |.          36    0.009    0.082
```

Box-Pierce Q-Stat 22.74   Prob 0.9581   SE of Correlations 0.146

Ljung-Box Q-Stat 36.71   Prob 0.4358

7. ANÁLISIS DETALLADO DE LOS ERRORES

Más información para determinar la validez del modelo la proporciona el gráfico de los residuos el cual permite una visión de conjunto de la magnitud de los errores, sesgos sistemáticos y puntos de errores excepcionales. Las salidas del TSP señalan las bandas más-menos una desviación típica (68% de los casos en una distribución normal).

Si el modelo es correcto los residuos han de comportarse como ruido «blanco», es decir, tienen media nula, varianza constante y ausencia de autocorrelación.

En la salida del TSP, se puede observar que los puntos están dentro de las bandas, no existe ningún valor significativo, y Q-Statistic(36) es igual a 22,74 dejando una probabilidad de 0.9581, por tanto el modelo está bien estimado, los residuos se comportan como «ruido blanco».

8. SELECCIÓN DEL MODELO

A través del proceso seguido se puede determinar el modelo

ARIMA (p,d,q) * SARIMA (P,D,Q)

es igual a:

ARIMA (1,1,0) * SARIMA (0,1,12)

- p es la componente autorregresivo (AR), su valor puede ser:

  0: la endógena desplazada no interviene, es decir: no depende del valor anterior.

  1: la variable depende del valor anterior, así:

  y_t depende de y_{t-1}
  y_{t-1} depende de y_{t-2}
  .
  .
  .

  2: y_t depende de y_{t-1} y de y_{t-2} con un peso para y_{t-1} y otro para y_{t-2}

- q es la componente de medias móviles (donde se incorporan los errores cometidos), su valor puede ser:
  0: no interviene ningún error, sólo el del período.
  1: influye el error anterior.
  2: intervienen hasta dos errores.

- la I de ARIMA indica el número de veces que hay que diferenciar. El valor de d puede ser:
  0: es válida la serie original.
  1: se calcula una diferencia.
  2: se calculan dos diferencias (diferencia de diferencia)

P,D,Q es lo mismo que p,d,q pero en la parte estacional.

9. PREDICCIÓN

Una vez seleccionado el modelo definitivo que cumple satisfactoriamente los criterios de evaluación establecidos, se pasa a la etapa de predicción.

En términos de TSP los pasos a seguir son los siguientes:

- Definir los períodos de predicción
- Hacer el pronóstico de la variable DDLE

Luego se recalculan las variables originales, y se recomponen las transformaciones haciendo el procedimiento inverso al establecido cuando se sacó logaritmo y se hicieron las diferenciaciones:

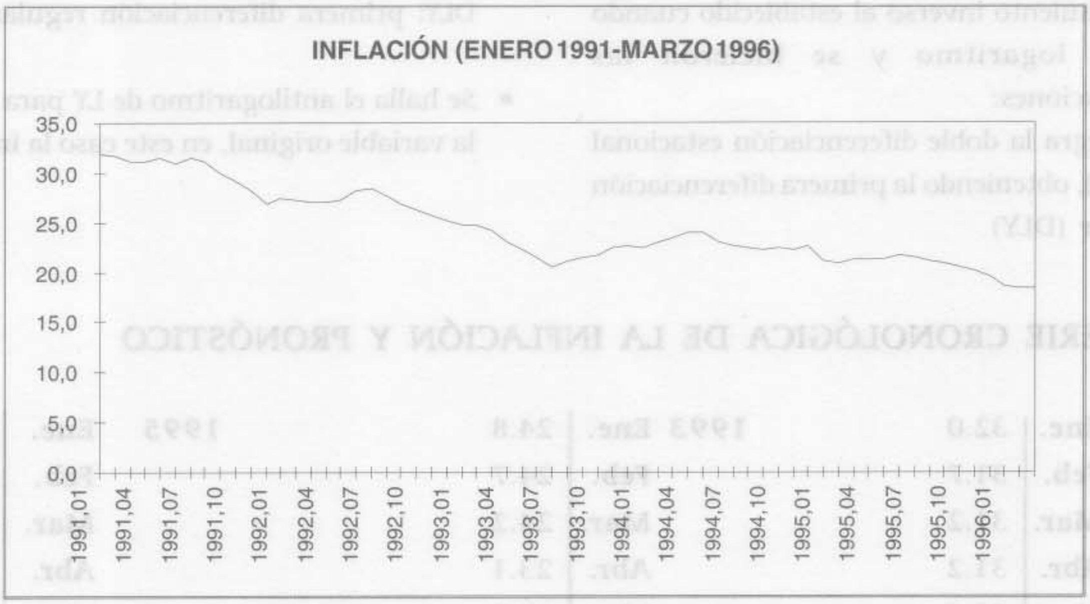
- Se integra la doble diferenciación estacional (DDLE), obteniendo la primera diferenciación regular (DLY)
- Se obtiene el logaritmo de la variable original (LY), mediante la integración de la variable DLY: primera diferenciación regular
- Se halla el antilogaritmo de LY para obtener la variable original, en este caso la Inflación.

10. SERIE CRONOLÓGICA DE LA INFLACIÓN Y PRONÓSTICO

1991	**Ene.**	32.0	**1993**	**Ene.**	24.8	**1995**	**Ene.**	21.0
	Feb.	31.7		**Feb.**	24.7		**Feb.**	20.9
	Mar.	31.2		**Mar.**	24.2		**Mar.**	21.3
	Abr.	31.2		**Abr.**	23.1		**Abr.**	21.2
	May.	31.5		**May.**	22.2		**May.**	21.3
	Jun.	31.0		**Jun.**	21.4		**Jun.**	21.7
	Jul.	31.6		**Jul.**	20.5		**Jul.**	21.5
	Ago.	31.2		**Ago.**	21.1		**Ago.**	21.1
	Sep.	30.0		**Sep.**	21.4		**Sep.**	20.8
	Oct.	29.3		**Oct.**	21.7		**Oct.**	20.5
	Nov.	28.2		**Nov.**	22.4		**Nov.**	20.1
	Dic.	26.8		**Dic.**	22.6		**Dic.**	19.5
1992	**Ene.**	27.4	**1994**	**Ene.**	22.5			
	Feb.	27.4		**Feb.**	23.0	**1996**	**Ene.**	18.5
	Mar.	27.1		**Mar.**	23.4		**Feb.**	18.4
	Abr.	27.2		**Abr.**	23.9		**Mar.**	18.4
	May.	27.3		**May.**	23.9			
	Jun.	28.1		**Jun.**	23.1			
	Jul.	28.4		**Jul.**	22.7			
	Ago.	27.7		**Ago.**	22.4			
	Sep.	26.9		**Sep.**	22.3			
	Oct.	26.3		**Oct.**	22.4			
	Nov.	25.7		**Nov.**	22.2			
	Dic.	25.1		**Dic.**	22.6			

Fuente: Guía Empresarial ANIF. 1991-1996.

GRÁFICO 2



11. CONCLUSIONES

Existen diversas formas de realizar pronósticos para una serie, alternativas al ARIMA. Entre ellas se cuenta con el método de descomposición de la serie en los factores que pueden afectarla: tendencial, cíclico y estacional. La tendencia es la proyección a largo plazo de la variable y muestra la dirección hacia donde apunta la serie; el factor cíclico determina las ondulaciones a lo largo de la tendencia y el estacional muestra movimientos a corto plazo dentro de un período (máximo un año). Tales factores se pueden relacionar de manera aditiva o multiplicativa. Estos modelos se clasifican dentro de los métodos determinísticos de pronósticos.

El pronóstico mediante ARIMA para los meses de Enero, Febrero y Marzo de 1996 (18.5, 18.4 y 18.4% respectivamente) se puede observar en el recuadro de la serie original presentado anteriormente.

Existen diversas formas de realizar pronósticos para una serie alternativas al ARIMA. Entre ellas se cuenta con el método de descomposición de la serie en los factores que pueden afectarla: tendencial, cíclico y estacional. La tendencia es la proyección a largo plazo de la variable y muestra la dirección hacia donde apunta la serie; el factor cíclico determina las ondulaciones a lo largo de la tendencia y el estacional muestra movimientos a corto plazo dentro de un período (máximo un año).

Utilizando la descomposición clásica a través del modelo multiplicativo de la serie se hallaron las predicciones para tales períodos, con los mismos datos. Sus resultados fueron: 19.3, 19.2 y 18.9 respectivamente.

Una comparación entre el ARIMA, el método clásico y los valores reales se muestran a continuación:

Meses de 1996 1996	Arima	Clásico por Descomposición	Valor Real
Enero	18.5	19.3	20.2
Febrero	18.4	19.2	20.8
Marzo	18.4	18.9	20.2

Como se puede observar, los valores pronosticados no corresponden exactamente al valor real debido a que consideran sólo el comportamiento histórico de la variable como única fuente para establecer el pronóstico y dejan de lado la influencia de tipo coyuntural de otras variables que pudieran afectar la evolución de la serie y lo que haría factible el uso de modelos multivariantes.

En el caso particular de la Inflación, aunque ella conlleva una influencia grande de tipo inercial (comportamiento autorregresivo), ya que los agentes económicos tratan de recuperar la inflación pasada con el fín de no perder poder adquisitivo (lo que repercute en el valor de la Inflación futura), existen otras variables que ejercen mayor o menor influencia tales como el aumento del gasto público, el incremento del circulante, la devaluación de la moneda, las expectativas políticas y económicas etc.

BIBLIOGRAFÍA

Carrión García, Andrés. Técnicas de Previsión. Valencia: Universidad Politécnica de Valencia. Departamento de Estadística e Investigación Operativa.

Gujarati, Damodar N. Econometría 2a. de. Bogotá: Mc Graw Hill, 1990. 2a. edición.

Pulido San Román, Antonio. Modelos Econométricos. 3a. de. Madrid: Ed. Pirámide, 1989.

Pulido San Román, Antonio. Predicción Económica y Empresarial. Madrid: Ed. Pirámide, 1989.