

Modelos ARFIMA: UNA APLICACIÓN AL ESTUDIO DEL DESPACHO TOTAL MENSUAL DE CEMENTO EN EL PERÚ

Domínguez C. Melina

Ysela Agüero Palacios

RESUMEN.- *Se estudia una clase de modelos ARIMA Fraccionalmente Integrados denominados modelos ARFIMA; los cuales tienen capacidad para mostrar dependencias significativas entre periodos de tiempo distantes (procesos de memoria larga). El parámetro de diferenciación fraccional d utilizado en estos modelos varía en el intervalo $(-0.5, 0.5)$.*

1. INTRODUCCIÓN

La modelización univariante de series cronológicas se ha convertido en la actualidad en una herramienta imprescindible para el análisis de datos, especialmente en el estudio de variables económicas. El objetivo de este tipo de análisis es capturar la dependencia y la regularidad que muestran los datos a lo largo del tiempo y su principal característica radica en que, para la construcción de un modelo, la única información muestral necesaria es la referida a la propia variable a estudiar.

Para quienes analizan series de tiempo económicas, es una práctica usual aplicar transformaciones a la serie de tiempo, con el objeto de volverla estacionaria, y así poder identificar y ajustar un modelo ARMA (p, q). La transformación usual es la diferenciación finita de orden d (d entero positivo mayor o igual a 1).

Los modelos ARMA (p, q) son conocidos como procesos de memoria corta, pues las autocorrelaciones decrecen exponencialmente a medida que crece el número de retardos. Desafortunadamente, esta clase de modelos no están bien adaptados para representar realizaciones de procesos con memoria larga los cuales ocurren con frecuencia en campos tan diversos como la hidrología, economía (Hipel & McLeod [8], Granger & Joyeux [7], Hosking [9]).

En el presente artículo se estudia una clase alternativa de procesos estocásticos denominada modelos ARFIMA (ARIMA fraccionalmente integrados), los cuales tienen una mayor performance cuando se analizan series de tiempo con comportamientos persistentes en periodos largos. En la sección II se presentan los conceptos básicos de los modelos ARFIMA y los pasos a seguir para el ajuste de modelos para series de memoria larga. En la sección III se realizó el análisis de la serie de datos referidos al *Despacho total mensual de cemento en el Perú*, comparando los pronósticos obtenidos mediante el ajuste de modelos ARIMA y ARFIMA.

Para estimar los parámetros y calcular los pronósticos se utilizó el módulo LONGMEM del programa computacional ITSM (Interactive Time Series Modelling ver 5.0).

2. MODELOS ARFIMA

En el enfoque del dominio del tiempo, se representa el comportamiento de un proceso estacionario mediante modelos lineales que incluyen términos del pasado de la misma variable y componentes puramente aleatorias. En este contexto la atención se centra en una clase de modelos lineales estacionarios denominados modelos autoregresivos de media móvil (ARMA).

En la práctica, la restricción de estacionariedad no necesariamente se verifica en la mayoría de las series de tiempo, especialmente en las referidas a variables económicas. Esto conduce a la necesidad de ampliar la clase de los modelos lineales incluyendo los procesos no estacionarios homogéneos; llamados así, porque pueden ser convertidos en estacionarios a partir de la aplicación de filtros lineales y transformaciones no lineales, dando origen a los modelos Autoregresivos Integrados Media Móvil o ARIMA (p, d, q).

En primer lugar, definimos a los modelos ARIMA como,

$$\phi(B)\Delta^d Y_t = \phi(B)(1-B)^d Y_t = \theta(B)a_t$$

Donde $\Delta^d Y_t = (1-B)^d Y_t$ es un proceso estacionario, d es un número entero positivo, B es el operador de retardo, $\{a_t\}$ es un proceso de ruido blanco con distribución $N(0, \sigma^2)$, y los términos $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, corresponden a la parte autoregresiva y media móvil, respectivamente.

Los modelos ARFIMA se escriben de la misma forma que los ARIMA, con la diferencia que en los primeros, el orden de diferenciación d varía en el intervalo $(-0.5, 0.5)$, esto es;

$$\phi(B)(1-B)^d Y_t = \phi(B)\Delta^d T_t = \theta(B)a_t. \quad (1)$$

Donde el operador de diferenciación es definido por la serie binomial;

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB + \frac{d(d-1)B^2}{2!} - \frac{d(d-1)(d-2)B^3}{3!} + \dots$$

2.1 Análisis de series de Memoria Larga

En el trabajo con la clase de los modelos de Box y Jenkins, la identificación consiste en determinar cuál es el modelo ARMA que generó la serie de tiempo en estudio. Dado que la función de autocorrelación caracteriza estos modelos, utilizamos la función de autocorrelación estimada para identificar el proceso generador de los datos.

Puesto que, los modelos ARMA (p, q) han sido desarrollados para representar procesos estocásticos estacionarios, el primer requisito que debe tenerse en cuenta en la etapa de identificación es verificar si la serie de tiempo objeto de estudio, es estacionaria. Con este fin, realizamos el análisis exploratorio de la serie de tiempo. Así por ejemplo, el gráfico de la serie mostrará la presencia de tendencias, cambios en el nivel, heterocedasticidad, etc. El correlograma muestral ayudará a determinar si la serie tiene memoria larga (decrecimiento lento). Un espectro estimado con valores muy elevados en frecuencias próximas de cero también da indicios de persistencia en periodos largos.

Otra herramienta que podemos incorporar para explorar si la serie es de memoria larga es el gráfico de dispersión relacionando el logaritmo del periodograma versus el logaritmo de $\{4 \operatorname{sen}^2(w_j/2)\}$, donde w_j es la frecuencia del Periodograma. Una relación aproximadamente lineal indicará persistencia en periodos largos. También se pueden utilizar métodos formales como los contrastes de raíces unitarias.

Para observar la persistencia en periodos largos se requiere calcular un número razonablemente grande de autocorrelaciones estimadas. Se sugiere calcular por lo menos 50 autocorrelaciones estimadas, lo cual significa que necesitamos series de 150 a 200 datos para poder realizar un análisis adecuado.

Una vez que se establece si la serie requiere diferenciación (entera o fraccional), el siguiente paso es determinar el orden de diferenciación. Si este es entero, el asunto se reduce a determinar el valor adecuado para d (1, 2 o 3 como máximo). Por otro lado, si existen indicios de que la serie tiene memoria larga recurriremos a las diferencias de orden fraccional. Nuevamente, el valor de d se determinará primero mediante el análisis exploratorio y luego a partir de métodos más formales, como veremos más adelante.

Finalmente, se analiza la función de autocorrelación simple de la serie diferenciada para identificar el modelo ARMA más adecuado. Si el orden de diferenciación fue entero el modelo identificado será un ARIMA (p, d, q) , si fue fraccional el modelo será un ARFIMA (p, d, q) .

El conjunto de parámetros $\beta = (\phi, \theta, d, \sigma^2)$ del modelo ARFIMA (p, d, q) , puede ser estimado por el método de mínimos cuadrados partiendo de la función de densidad espectral o por el método de máxima verosimilitud (Brockwell - Davis [2], Geweke - Porter - Hudack [5], Diebold - Rudebrush [3]).

Para la evaluación de la bondad del ajuste y la adecuación del modelo ajustado se utilizan las mismas técnicas utilizadas en esta etapa en el caso de los modelos ARIMA esto es el contrastes de hipótesis y análisis de residuos.

Granger - Joyeux [7] y Geweke - Porter - Hudak [5] proponen realizar las predicciones a partir del modelo ARFIMA ajustado utilizando la representación autoregresiva de orden infinito dada por,

$$y_t = \sum_{j=0}^{\infty} \pi_j B^j y_t + a_t \quad (2)$$

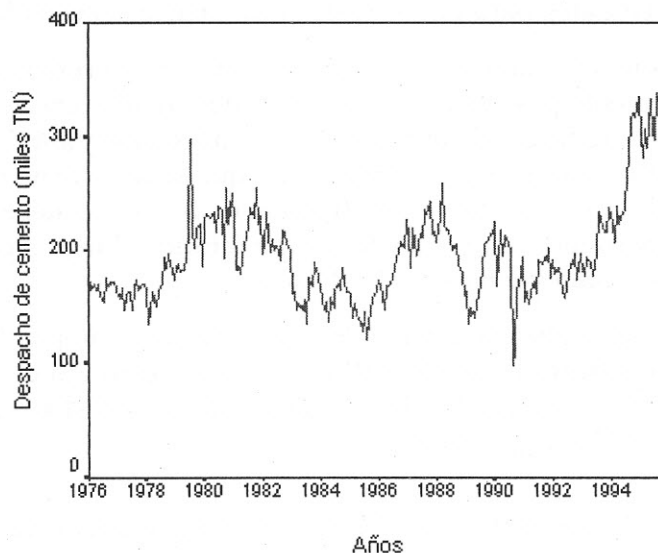
Las predicciones se basan en una versión truncada de la ecuación (2) después de k retardos. Pieris [12] y Pieris - Perera [13] tratan algunas de las fórmulas para calcular predicciones de una representación autoregresiva.

II. APLICACIÓN DE LOS MODELOS ARFIMA AL ESTUDIO DE LA SERIE DESPACHO TOTAL MENSUAL DE CEMENTO EN EL PERÚ.

El volumen de cemento comercializado (despacho) por las empresas de explotación de canteras está fuertemente relacionado con los cambios en la economía y la política del país, pues la demanda es afectada por la inversión privada, y el crecimiento a largo plazo de la industria, además el mayor volumen de este producto es demandado por los organismos del estado que se encargan de las obras públicas directas; las cuales se ejecutan en mayor número en periodos de elecciones municipales o presidenciales.

La serie de datos de despacho total mensual de cemento que analizaremos corresponde al periodo de enero de 1976 hasta julio de 1996, publicada por el Instituto Nacional de Estadística (INEI).

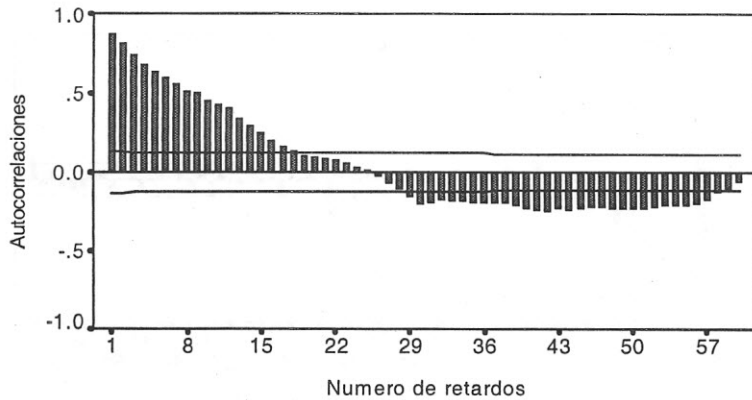
*Figura N° 1. Despacho mensual de cemento en el Perú.
Período 1976 - 1996*



La serie de datos tiene un comportamiento no estacionario con cambios frecuentes en el nivel y una tendencia creciente a partir del año 1995 (Fig. 1).

Esta tendencia a crecer se explica por la campaña de reelección presidencial en ese año. El gobierno de turno invirtió fuertes sumas de dinero en la construcción de escuelas (FONCODES), además de otras obras públicas como carreteras, puentes, etc., en el interior del país.

Fig. N° 2. Función de autocorrelación estimada de la serie



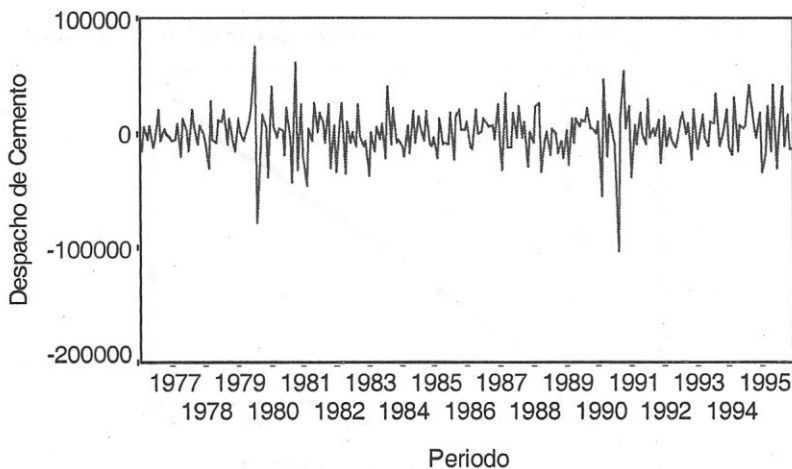
La función de autocorrelación simple estimada de esta serie de datos (Fig. 3) presenta un decrecimiento lento; lo cual, es un claro indicio de que la serie es no estacionaria y posiblemente tiene memoria larga.

Para proceder a identificar el modelo determinaremos el orden de diferenciación. En primer lugar se identificará un modelo ARIMA, luego un ARFIMA y finalmente compararemos los pronósticos obtenidos por ambos métodos.

3.1 Identificación y ajuste de un modelo ARIMA

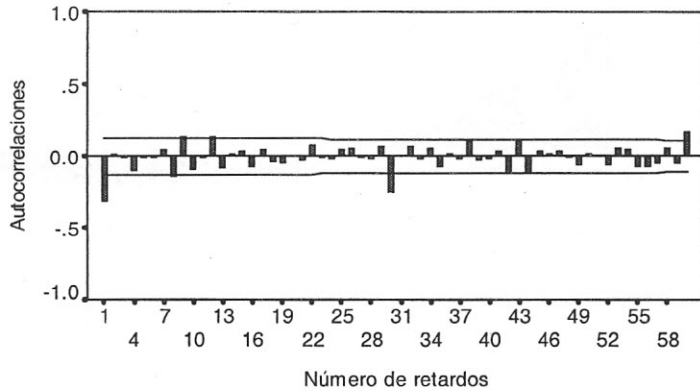
Para determinar el número de diferencias finitas necesario calculamos la varianza de la serie original ($\sigma^2 = 19659E5$), de la primera diferencia ($\sigma^2 = 43625E4$) y la segunda diferencia ($\sigma^2 = 606E6$) y observamos que la varianza disminuye al diferenciar la serie por primera vez pero vuelve a, crecer en la segunda diferencia.

Figura N° 3. Serie con una diferencia finita



Siguiendo este criterio decidimos analizar la serie con una diferencia finita. La serie transformada usando el filtro de diferencias finitas $w_t = (1 - B) y_t$, se exhibe en la Fig. 3. Esta nueva serie muestra un comportamiento estacionario.

Figura N° 4. Función de autocorrelación de la Serie con una diferencia finita ($d = 1$)

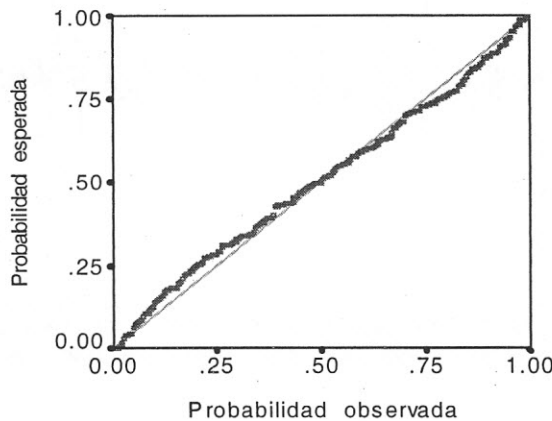


En la Fig. 4 se observa que la autocorrelación simple para un retardo de orden 1 es grande comparado con las autocorrelaciones calculadas para los siguientes retardos 2, 3, 4, etc. Este comportamiento sugiere un modelo media móvil de orden 1 (MA(1)). Luego, el modelo identificado es un ARIMA (0, 1, 1), cuya ecuación es

$$W_t = (1 - B) Y_t = a_t - \hat{\theta}_1 a_{t-1}$$

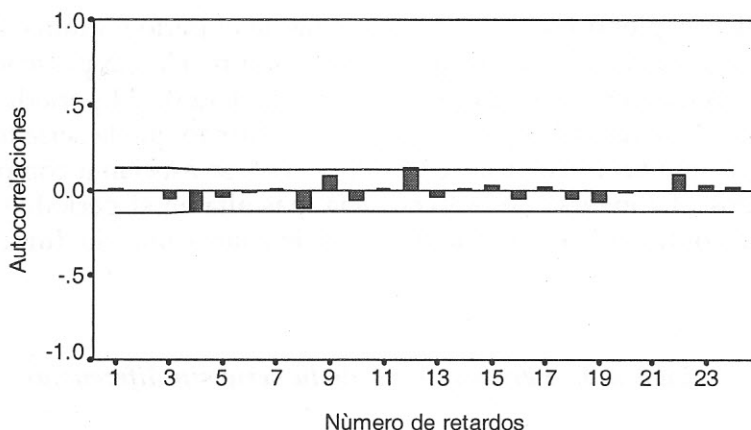
Las estimaciones máximo verosímiles de los parámetros del modelo identificado fueron: $\theta_1 = -0.35$ ($p \leq 0.05$) y la varianza estimada del ruido blanco fue $\hat{\sigma}_a^2 = 39359$. Observar que los residuos del modelo ajustado tienen un comportamiento aproximadamente normal (Fig. 5). Utilizamos la prueba de Kolmogorov - Smirnov para contrastar la hipótesis de normalidad ($K = 0.82$, $p \leq 0.52$), concluyendo que no hay evidencia estadísticamente significativa para rechazar la hipótesis de normalidad. Además, los residuos son incorrelacionados (Fig. 6).

Fig. N° 5 Normalidad de residuos del modelo ajustado



El contraste de la hipótesis nula que los errores del ajuste corresponden a una realización de un proceso de ruido blanco no puede ser rechazado ($\chi^2 = 16.98$, $P \geq 0.52$). Este es un indicador de que el modelo ajustado es adecuado.

Fig. N° 6 Autocorrelaciones de los residuos del ajuste



Observemos ahora el efecto de la diferenciación en el periodograma (fig. 7), y en la densidad espectral (Fig. 8). Después de aplicar el filtro de diferencias finitas a la serie; notamos claramente como el periodograma y la densidad espectral cambian, eliminando el comportamiento a largo plazo de la serie (ciclos), el cual esta asociado a las frecuencias bajas y aumentando la importancia de las frecuencias altas asociadas a componentes de periodos corto.

Fig. N° 7. Periodograma estimado de la serie con una diferencia de orden uno

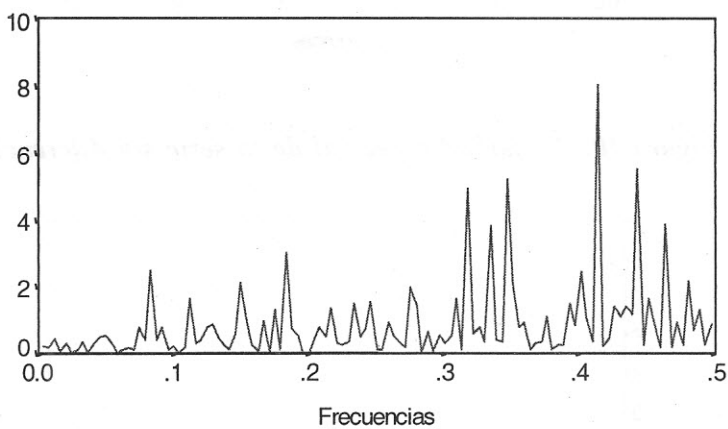
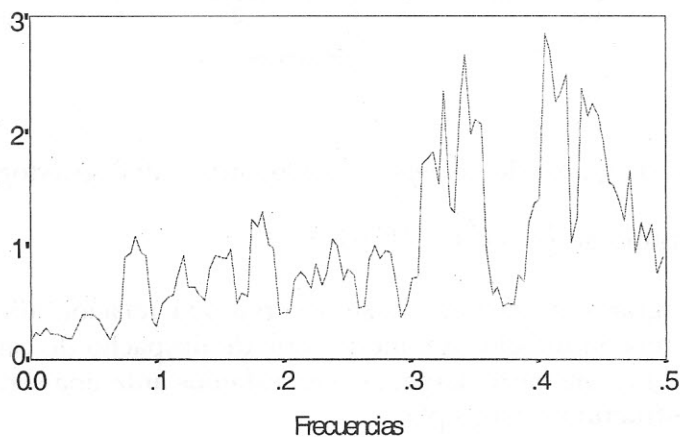


Fig. N° 8. Densidad espectral: Serie con una diferencia de orden uno



3.2 Ajuste de la serie mediante modelos ARFIMA

El análisis exploratorio de la serie, mediante el periodograma y la función de densidad espectral de la serie sin diferenciación revelan la presencia de comportamientos persistentes a largo plazo, (memoria larga). El periodograma (Fig. 9), y la función de densidad espectral (Fig. 10) confirman que la serie tiene memoria larga, notar que las frecuencias cercanas a cero, se asocian a componentes de periodos largos. La inversa de la frecuencia más alta en el periodograma es, 50 meses, lo cual confirma la existencia de ciclos de cuatro años de duración aproximadamente.

Figura 9. Periodograma de la serie sin diferenciar

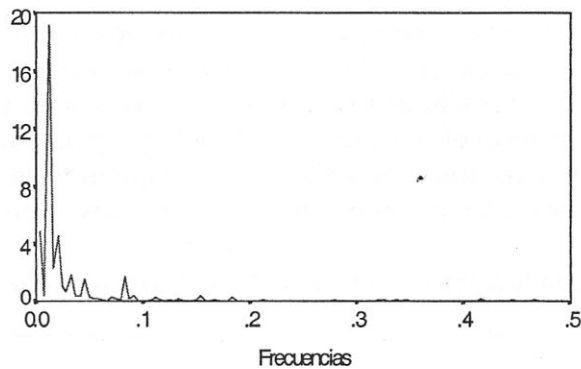
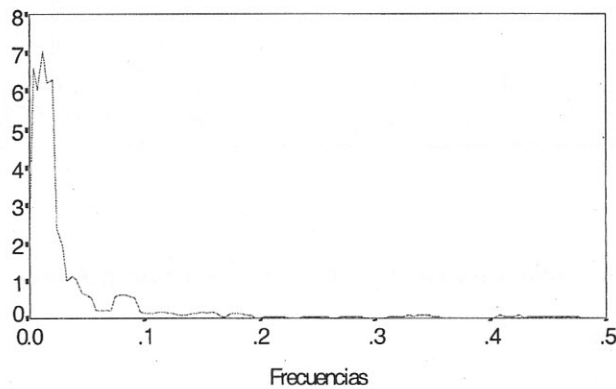


Figura 10. Densidad espectral de la serie sin diferenciar

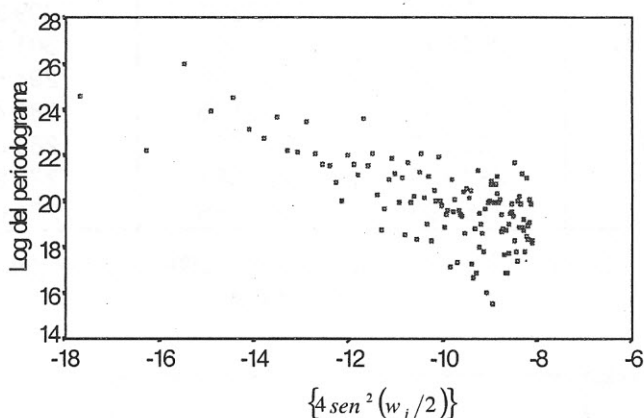


Construimos el diagrama de dispersión del logaritmo del periodograma $(I_n(w_j))$ versus el logaritmo de $\{4 \text{ sen}^2(w_j/2)\}$ (Fig. 11).

En el diagrama de dispersión se observa una relación aproximadamente lineal, lo cual nos da indicios de que la serie de despacho de cemento efectivamente puede ser de memoria larga, esto es, estamos ante una serie no estacionaria, con una estructura a largo plazo.

Fig. 11. Diagrama de dispersión del logaritmo del Periodograma

$$(I_n(w_j)) \text{ y el logaritmo de } \{4 \text{ sen}^2(w_j/2)\}$$



La metodología de estimación de parámetros en los modelos ARFIMA, se basa en la densidad espectral de la serie, y la máxima verosimilitud. Primero estimamos el valor para d , ajustando un modelo de regresión a \log del periodograma y $\log \{4 \text{ sen}^2(w_j/2)\}$, el cual resulta en un valor de $d = 0.3$ (pendiente de la recta). Exploramos con varios modelos siendo el mejor un ARMA(1,1). La estimación final de parámetros fue obtenida por el método de máxima verosimilitud.

La estimación máxima verosímil, requiere valores iniciales para estimar los parámetros del modelo, por lo tanto usaremos como valor inicial $d = 0.3$, mientras que para ϕ y θ , usaremos valores iniciales iguales a cero, (el máximo global para ϕ y θ , es aproximadamente el mismo usando otros valores iniciales). El modelo estimado

$$(1 - 0.75B)\nabla^{0.4} \hat{Y}_t = (1 - 0.4B) a_t$$

Expresándolo como un modelo ARMA tenemos;

$$(1 - 0.75B)\hat{U}_t = (1 - 0.4B) a_t$$

donde $\nabla^{0.4} \hat{Y}$ y $\sigma_a^2 = 471546500$.

3.3 Comparación de pronósticos

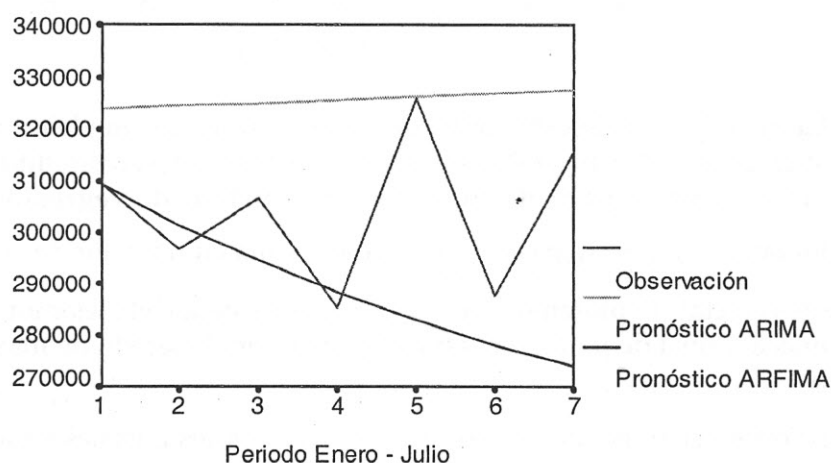
En la tabla 1, mostramos los pronósticos para los dos modelos con los correspondientes valores observados, aún cuando la varianza del ruido blanco para el modelo ARMA es menor, los cuatro primeros pronósticos con el modelo ARFIMA resultan muy próximos a los datos observados.

La Fig. 12 muestra los pronósticos obtenidos a partir de los dos modelos ajustados.

Tabla N° 1. Pronósticos con modelos ARIMA y ARFIMA para el periodo Enero - Julio 1996

Mes	Observación	ARIMA (0, 1, 1)	ARFIMA (1, 0, 3, 1)
Enero	309 249	323 821	309 239
Febrero	296 846	324 428	301 498
Marzo	306 701	325 036	294 577
Abril	285 280	325 643	288 443
Mayo	325 898	326 250	283 027
Junio	287 705	326 857	278 241
Julio	315 595	327 465	274 014

Fig. N° 12. Pronósticos: Enero - Julio 1996 con modelos ARIMA y ARFIMA



Observamos que los pronósticos para el ajuste ARMA tienden a crecer, esto se debe a que están basados en el análisis de la información proporcionada en un pasado más o menos inmediato, es decir los pronósticos son calculados tomando en cuenta los últimos datos conocidos de la serie. Por otro lado los pronósticos ARFIMA tienden a decrecer siguiendo más de cerca a la serie original.

CONCLUSIONES

La modelización de series de tiempo univariantes mediante los modelos ARIMA requiere que la serie estudiada sea estacionaria. Esto conduce generalmente a calcular diferencias finitas de la serie con el fin de convertirlas en estacionarias cuando no lo son. Posteriormente, la modelización se realiza sobre la serie diferenciada.

La aplicación de diferencias enteras a las series para convertirlas en estacionarias, tiene como consecuencia la eliminación de componentes a largo plazo. Esto se observa muy claramente al calcular el espectro de la serie original y diferenciada.

La diferenciación fraccional tiene la ventaja de mantener la información a largo plazo contenida en las series de memoria larga.

Los modelos ARFIMA producen pronósticos a corto y largo plazo que pueden estar más próximos de la realidad que los modelos ARIMA.

BIBLIOGRAFÍA

- [1] Box G.E.P.; Jenkins, G.M. *Time Series Analysis, Forecasting and Control*. Prentice Hall (1976).
- [2] Brockwell, Peter J.; Richard A. Davis *Time Series: Theory and Methods*. New York, Springer - Verlag, 2da ed (1991).
- [3] Diebold; Rudebusch, G. D. *Long memory and persistence in aggregate output*. Journal of Monetary Economics, 24, 189-209 (1989).
- [4] Domínguez, Melina.; *Modelos ARFIMA: Una alternativa para modelar Series de Series de Memoria Larga*. Tesis de Licenciatura. Facultad de Ciencias Matemáticas de la UNMSM (2000).
- [5] Geweke, J.; Porter - Hudak. *The estimation and application of long memory times series models*. Journal of Time Series Analysis, 4, pp. 221 - 38 (1983).
- [6] Granger, C. W. J. *Long memory relationships and the aggregations of dynamic dynamic model*. Journal of Econometric, 14, 227 - 238 (1980).
- [7] Granger, C. W. J.; Joyeux, R. *An introduction to long memory time series models and fractional differencing*. Journal of Time Series Analysis, 1, 15 - 29 (1980).
- [8] Hipel, K. W.; Mcleod, A. I. *Preservation of the rescaled adjusted range. Simulation the studies using Box and Jenkins models*. Water Resour. Res 14, 509 - 516 (1978).
- [9] Hosking, J. R. M. *Fractional differencing*. Biometrika, 68, 165 - 76 (1981).
- [10] Li, W. K. and Mcleod, A. I. *Fractional time series modelling*. Biometrika, 73 (1986).
- [11] Ray, B. K. *Modeling Long Memory proceses for optimal long range prediction*. Journal of time Series Analisis 14, 511 - 525 (1993).
- [12] Pieris, M. S. *A note on the predictors differenced sequences*. Australian Journal of Statistics 29, 42 - 48 (1987).
- [13] Pieris, M. S. and B. J. C. Perera. *On prediction with fractionally differenced ARMA models*. Journal of time Series Analisis 14, 511 - 525 (1988).