

42 Jornadas Nacionales de Administración Financiera
Septiembre 22 y 23, 2022

Modelos fraccionalmente integrados aplicados al mercado de renta variable

Daniel Miliá

Ramiro M. Pérez

Universidad de Buenos Aires

SUMARIO

1. Introducción
2. Metodología ARIMAX
3. Metodología ARFIMAX
4. Comparación de pronósticos
5. Conclusiones

Para comentarios:
daniel@economicas.uba.ar
ramiromperez@hotmail.com

1. Introducción

Existe una vasta variedad de fenómenos que se encuentran condicionados por la expectativa generada por movimientos relativos a dichos sucesos ocurridos en el pasado o incluso que son su consecuencia inmediata. Los mercados financieros y sus vaivenes son un claro ejemplo de este tipo de fenómenos. A menudo se mira este mercado para intentar entender qué es lo que la generalidad de los actores descuenta sobre sucesos del pasado como si sus variables tuvieran memoria de sí mismas en el momento inmediatamente anterior. Resulta evidente entonces que este tipo de variables poseen un comportamiento que en el momento actual responden a una expectativa generada sobre el valor de cambio producido en el momento precedente, es decir, a un valor condicionado por la varianza del período anterior.

La utilización de modelos econométricos a fin de realizar pronósticos sobre el comportamiento futuro de la variable en estudio tiene, históricamente, dos aproximaciones. La primera está constituida por modelos econométricos causales o “estructurales” los cuales se construyen y estiman en base a un marco teórico que resulta relevante desde el punto de vista económico. Éstos intentan dar cuenta de las relaciones estructurales que existen entre variables. En segundo lugar, existen los modelos de series temporales, los que se construyen sobre la premisa de que las series de tiempo en estudio poseen una historia estadística recurrente y particular, la cual puede ser modelada y explotada a fines de hacer pronósticos.

En este trabajo se desarrollan dos tipos de modelos que combinan ambas metodologías: ambos incluyen términos rezagados de la variable dependiente y del término de error, con los que se pretende modelar la serie en base a su historia estadística y una estructura de variables exógenas independientes, sustentadas por la teoría económica, cuya anexión contribuye a mejorar los procesos de predicción. La principal diferencia entre uno y otro será el grado de diferenciación o integración que se le aplicará a la serie para adaptarla a fin de lograr su correcto estudio. Utilizando la metodología desarrollada por Box y Jenkins ¹ se lleva a cabo el estudio de la variable utilizando ambos modelos para luego comparar los pronósticos de ambos y analizar qué modelo ajusta mejor a la serie empírica, bajo la hipótesis de que para procesos que poseen memoria larga, el segundo se ajustará de mejor manera.

Este estudio encuentra motivación en la creciente volatilidad de los mercados bursátiles. El rechazo a la incorporación de innumerables variables dummy como caracterización de determinados contextos de volatilidad ya sea por cuestiones políticas (por ej., la lucha entre el gobierno y el sector agropecuario por las retenciones móviles en 2008), de economía externa (crisis económica mundial de 2008), combinaciones de ambas o incertidumbre local (expectativas de devaluación), sumado a la falta de un modelo clásico ya sea dentro de la escuela estructuralista o de series de tiempo que dé cuenta de manera correcta de las variaciones en los rendimientos de distintos activos son las principales razones para optar por un modelo cuyo orden de integración sea fraccionado de manera tal de no subestimar ni sobreestimar las consecuencias de los distintos shocks que aparecen a lo largo de la serie.

En cuanto a la variable en estudio, la misma será la serie de precios históricos para el activo de renta variable local BMA, acciones del Banco Macro.

¹ Box & Jenkins, *Times Series Analysis: Forecasting and Control*, 1970.

En la sección 2 se lleva a cabo el desarrollo y aplicación del primer modelo en estudio, siguiendo la metodología desarrollada por Box y Jenkins para modelos ARIMA con componente estructural (metodología ARIMAX). La sección 3 comienza con una breve descripción de procesos con memoria larga y su formalización ARFIMA. Se trata de modelos autorregresivos fraccionalmente integrados de medias móviles, a los que se agrega la componente estructural (metodología ARFIMAX); se efectúa a continuación similar aplicación a la de la sección anterior. En la sección 4 se comparan los pronósticos realizados por ambos modelos y respecto a los valores de la serie observada. La sección 5 detalla las conclusiones.

2. Metodología ARIMAX

2.1 El modelo

En el enfoque del dominio del tiempo, una de las maneras más usuales de representar un proceso generador de datos es mediante una combinación lineal que incluya términos del pasado de la variable en estudio y componentes puramente aleatorios. Este tipo de modelos se conoce como modelo autorregresivo de medias móviles o ARMA (p,q) . Sin embargo, la principal restricción de éstos es su aplicabilidad a series de tiempo estacionarias².

En la práctica, la mayoría de las series económicas dista de poseer esta característica por lo que resulta necesaria una nueva clase de modelos que permita representar procesos no estacionarios homogéneos. Este tipo de procesos pueden ser convertidos en estacionarios a partir de la aplicación de filtros lineales y transformaciones no lineales, dando origen a los modelos Autorregresivos Integrados de Medias Móviles o ARIMA (p,d,q) .

Se definen los modelos ARIMA como:

$$\Phi_p(L)\Delta^d Y_t = \Phi_p(L)(1-L)^d Y_t = \Theta_q(L)e_t \quad \text{Ec 1}$$

Donde $\Delta^d Y_t = (1-L)^d Y_t$ es un proceso estacionario; d es un número entero positivo, L es el operador retardo, $\{e_t\}$ es un proceso ruido blanco con distribución $N(0, \sigma^2)$, y los términos $\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ y $\Theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$, son polinomios en L de orden p y q , que corresponden a la parte autorregresiva y media móvil, respectivamente.

La parte autorregresiva (AR), de gran aplicación en econometría, especialmente en la especificación de ecuaciones dinámicas, permite caracterizar a la variable dependiente como explicada por valores pasados de ella misma. El orden (p) del modelo ARIMA (p,d,q) viene dado por el desfase máximo de la variable dependiente que se incluye como explicativa. La parte de medias móviles (MA) se construye a partir de una combinación lineal de shocks aleatorios, referidos a sucesivos momentos t .

En economía, esta estructura resulta de gran utilidad para describir fenómenos reflejados en diversos indicadores que se ven afectados por perturbaciones cuyos efectos no son instantáneos, sino que persisten durante períodos subsiguientes. El orden (q) del modelo ARIMA (p,d,q) viene dado por el desfase máximo de la variable aleatoria o perturbación. El orden (d) del

² De aquí en más entenderemos estacionariedad en el sentido débil del concepto (ver Anexo A).

modelo, viene dado por el orden de integración o diferenciación necesario para convertir la serie original, en estacionaria.

Finalmente, el modelo se completa con una o más variables explicativas que resultan de carácter estructural, es decir, que además de una fundamentación estadística, poseen sustento teórico en cuanto a su capacidad de explicar a la variable dependiente. El modelo ARIMA(p, d, q) resulta entonces:

$$\Phi_p(L)\Delta^d Y_t = \Phi_p(L)(1-L)^d Y_t = X_t + \Theta_q(L)e_t \quad Ec 2$$

Donde $X_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_s x_{st}$ es el polinomio de orden s de variables explicativas en el momento t .

2.2. Metodología de Box y Jenkins

Los modelos autorregresivos y de medias móviles eran conocidos por los estadísticos desde los años 1920 pero no fue hasta la década de 1960 que la posibilidad del cálculo electrónico hizo posible su implementación. Por esos años, G.E.P. Box y G.M. Jenkins lograron desarrollar una metodología generalizada y bien estructurada, complementada con soporte computacional, para la elaboración de modelos ARIMA.

Este desarrollo significó una ruptura significativa con la metodología econométrica clásica, la cual recurre a la teoría económica para especificar modelos causales. Por el contrario, el análisis de series de tiempo utiliza observaciones históricas de la propia serie como única fuente de información para la formulación, especificación y elección del modelo. En otras palabras, la forma en la que se comporta la serie empírica y su correspondiente análisis resultan suficientes para modelar una serie temporal.

El gran aporte de Box y Jenkins fue el de diseñar la metodología general para el tratamiento de las series. “La pregunta del millón es obvia: al ver una serie de tiempo ¿cómo sabemos si sigue un proceso AR puro (de ser así, ¿cuál es el valor de p), un proceso MA puro (de ser así, cual es el valor de q), un proceso ARMA (de ser así, cuáles son los valores de p y q) o un proceso ARIMA, en cuyo caso se deben conocer los valores de p , d , y q ?” (Gujarati & Porter, 2010, 777).

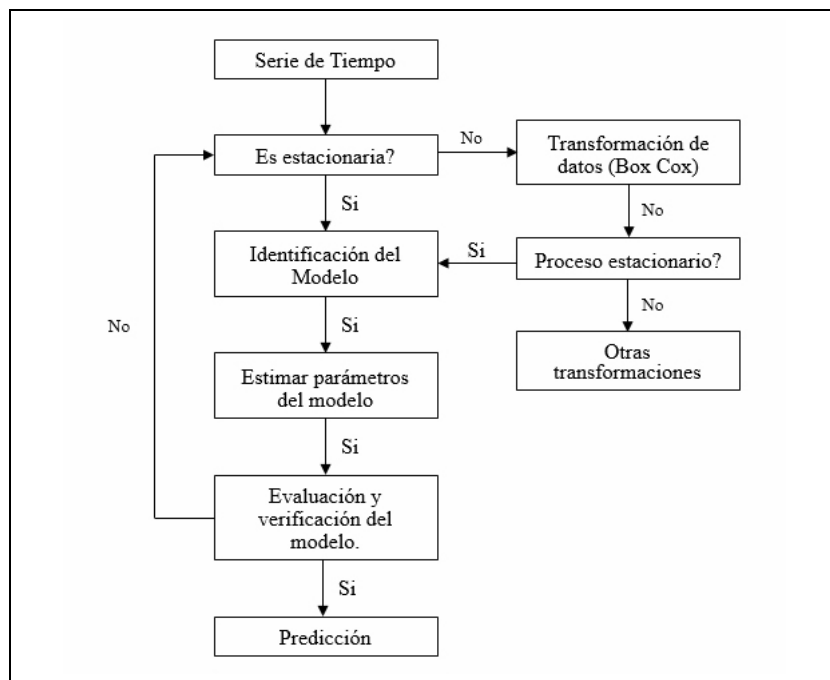
La metodología Box-Jenkins, de ahora en más, B-J, nos ayuda a encontrar respuestas. “Si bien en teoría existe una familia infinita de modelos ARIMA, empíricamente con valores bajos de los parámetros p , d y q , se logran modelos que ajustan a las observaciones con elevado nivel de confiabilidad” (Urbisaia & Brufman, 2000, 79). En la ilustración 1 se muestra la secuencia en etapas de dicha metodología, incluyendo un análisis sobre la condición de estacionariedad.

Si la serie no es estacionaria se realizan distintos tipos de transformaciones hasta lograr su estabilidad. Box y Cox definieron una serie de transformaciones que se definen a continuación:

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln Y_t & \text{si } \lambda = 0 \end{cases}$$

Cuando $\lambda = 1$ la transformación es nula, por lo que equivale a trabajar con la serie original.

Ilustración 1: Etapas en la metodología Box-Jenkins



Una vez estabilizada la serie en media y varianza se puede pasar a la siguiente etapa. Lo primero es identificar el tipo y orden del modelo a desarrollar. Para dicha etapa se examinan las Funciones *empíricas* de Autocorrelación y Autocorrelación Parcial (FACE y FAPE). A continuación, se estiman los coeficientes del modelo mediante la aplicación de algún método como Mínimos Cuadrados Ordinarios o Máxima Verosimilitud. Estimados los parámetros, se pasa a verificar si el modelo está correctamente identificado. Esta etapa sirve para juzgar la calidad del modelo estimado y se lleva a cabo mediante el análisis de residuos: si el modelo no es apropiado, se vuelve a la etapa de identificación procediendo nuevamente con toda la metodología. Finalmente, se realizan y evalúan pronósticos y comparan con modelos alternativos.

2.3 Aplicación

Esta sección propone determinar cuál es el modelo que mejor ajusta a nuestra variable en estudio como réplica del proceso generador de datos. La serie a estudiar refiere a los precios de cierre diarios del activo de renta variable BMA la cual llamaremos Y_t , acciones del Banco Macro, comenzando el 3 de enero de 2005 y finalizando el día 2 de mayo de 2018.

Como puede apreciarse en la ilustración 2, la serie presenta tendencia y tampoco existen indicios de varianza constante por lo que pareciera ser una serie no estacionaria. Realizaremos el test de Dickey-Fuller de raíz unitaria a modo de confirmación.

Como se observa en la salida de datos (tabla 1) no se logra rechazar la hipótesis nula de raíz unitaria a ningún nivel de confianza por lo que debemos concluir que la serie no es estacionaria.

Tabla 1: Test D-F para la serie de precios históricos
BMA 2005-2018

Dickey-Fuller test for unit root		Number of obs. = 3263		
		Interpolated Dickey - Fuller		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	1.599	-3.43	-2.86	-2.57
MacKinnon approximate p-value for Z(t) = 0.9979				

Ilustración 2: Precio de cierre BMA período 2005-2018



1) Análisis de estacionariedad

Como se vio en la sección anterior, la aplicación del modelo $ARIMAX(p,d,q)$ implica una transformación de la serie original en caso de que ésta no sea estacionaria. Dicha transformación consistirá en tomar los *rendimientos porcentuales entre observaciones consecutivas* a fin de intentar adaptar la serie para su correcto estudio. Creamos así una nueva variable:

$$Z_t = (Y_t - Y_{t-1})/Y_{t-1}$$

En la ilustración 3 se observa que la serie en rendimientos parecería, ahora sí, mostrar estacionariedad tanto para la media como para la varianza. A modo de confirmación realizamos una vez más el test de raíz unitaria de Dickey-Fuller.

La salida de datos (tabla 2) muestra que la hipótesis nula de raíz unitaria queda rechazada con un nivel de significatividad muy alto por lo que ahora sí podemos comenzar a hacer un análisis del modelo para la serie de rendimientos.

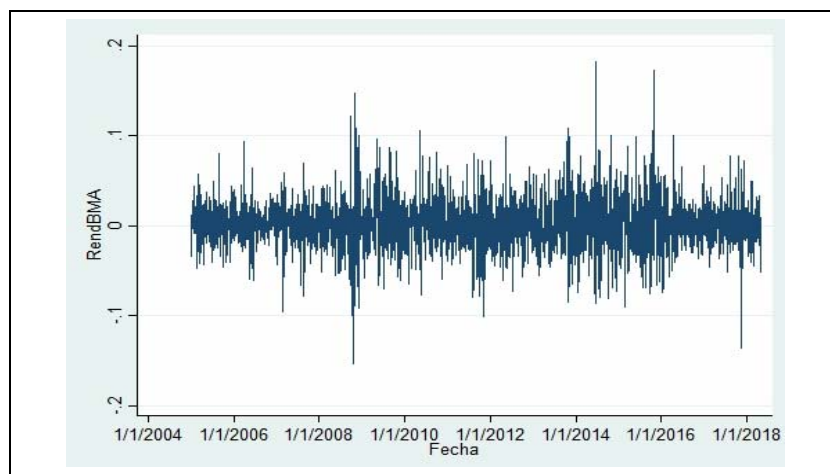
2) Identificación

Estudiando las funciones empíricas de autocorrelación y autocorrelación parcial se pueden ensayar distintos modelos para diferentes valores de rezagos p y q a fin de determinar cuál es el que mejor ajusta. A este estudio se le pueden sumar diferentes variables explicativas tales

Tabla 2: Test D-F para la serie de rendimientos del activo BMA 2005-2018.

Dickey-Fuller test for unit root		Number of obs. = 3262		
		Interpolated Dickey - Fuller		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-53.874	-3.43	-2.86	-2.57
MacKinnon approximate p-value for Z(t) = 0.000				

Ilustración 3: Rendimientos para el activo BMA período 2005-2018.



como el volumen operado durante dicha jornada, el rendimiento de otro activo del mismo sector, el rendimiento del mercado en general, el nivel o rendimiento del tipo de cambio o nivel de la tasa de interés BADLAR y finalmente, el nivel de la base monetaria.

Observando los gráficos de las funciones empíricas de autocorrelación y autocorrelación parcial (ilustraciones 4 y 5) se está en condiciones de decir que los rezagos significativos van de 1 a 3 para la componente MA y de 1 a 2 para la componente AR. La metodología B-J tiene su parte artesanal por lo que observando ambos gráficos no es posible determinar cuál es exactamente el mejor modelo. Es por eso que se requiere un estudio adicional a fin de obtener dicho modelo.

Para ello haremos un análisis en función de los criterios de información de Akaike y Bayesiano, además de estudiar los residuos que ellos producen. Los mismos se pueden encontrar en el Anexo B para distintos modelos tales como ARIMAX(2,0,3), ARIMAX(2,0,2), ARIMAX(2,0,1), ARIMAX(1,0,3), ARIMAX(1,0,2), ARIMAX(1,0,1), ARIMAX(0,0,3), ARIMAX(0,0,2), ARIMAX(0,0,1), ARIMAX(2,0,0) y ARIMAX(1,0,0) que a su vez incorporan diversas variables explicativas.

Dados los criterios antes mencionados, nos inclinamos por elegir un modelo ARIMAX(1,0,3) que incluye el primer rezago para la parte AR mientras que para la parte MA serán incluidos el primer y tercer rezago. A esto se le suman como variables explicativas el rendimiento del mercado en general, representado por el rendimiento del índice Merval para la misma rueda, llamado RendIMV y el rendimiento del Banco Galicia, como un activo alternativo, RendGGAL. Matemáticamente:

Ilustración 4: Función de autocorrelación - Serie de rendimientos BMA

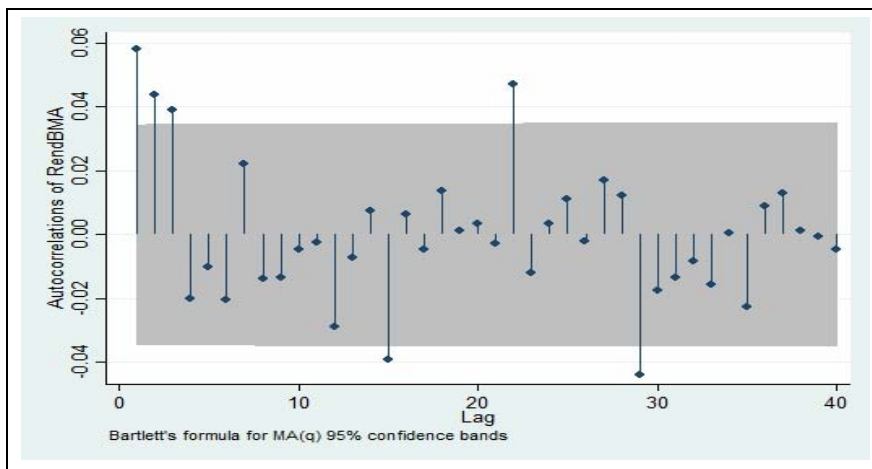
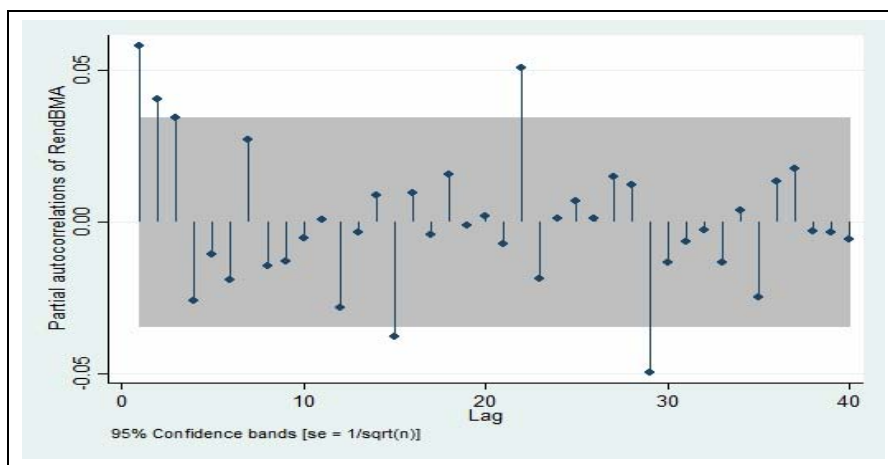


Ilustración 5: Función de autocorrelación parcial - Serie de rendimientos BMA



$$\widehat{\text{RendBMA}}_t = \beta_0 + \beta_1 \text{RendGGAL}_t + \beta_2 \text{RendIMV}_t + \phi_1 \widehat{\text{RendBMA}}_{t-1} + \theta_1 e_{t-1} + \theta_2 e_{t-3} + e_t$$

La salida de datos y valores para los criterios de elección del modelo se exponen en la sección siguiente.

3) Estimación

En la tabla 3 se puede ver la salida de datos para la estimación del modelo por el método de máxima verosimilitud. Como puede observarse, la variable RendIMV resulta altamente significativa. Esto es de esperar dado que el activo en estudio forma parte de este índice y por tanto su rendimiento repercute en el rendimiento del índice y viceversa. Por otro lado, se sabe que el sector financiero es el primero en reaccionar frente a una suba o baja generalizada del mercado, incluso cuando ésta proviene de otro sector como podría ser una suba o baja del petróleo por lo

que es de esperar que el primero posea una alta sensibilidad al rumbo general del segundo. En cuanto a la variable RendGGAL es de esperar que su correlación sea positiva. Esto es así porque ambos activos pertenecen al mismo sector (financiero) y por tanto las noticias buenas y malas deben afectar al sector en una misma dirección. Por otro lado, tampoco es perfecta ($\rho = 1$) dado que cada activo tiene sus particularidades.

Tabla 3: Estimación del modelo ARIMAX(1,0,3) para RendBMA por máxima verosimilitud

ARIMA regression						
Sample: 03jan2005 - 02may2018			Number of obs. = 3263			
Log likelihood = 8476.593			Wald chi2 (5) = 8783.94			
			Prob. > chi2 = 0.000			
OPG						
RendBMA	Coef.	Std. Err.	z	P > (z)	[95% Conf. Interval]	
RendBMA						
RendGGAL	0.4117643	0.0130158	31.64	0.000	0.3862537	0.4372748
RendIMV	0.5131937	0.0190117	26.99	0.000	0.4759315	0.5504558
_cons	0.0004153	0.0003227	1.29	0.205	-0.0002271	0.0010577
ARMA						
ar						
L1.	-0.8335322	0.0661135	-12.61	0.000	-0.9631122	-0.7039521
ma						
L1.	0.8413778	0.0657929	12.79	0.000	0.712426	0.9703296
L3.	0.0277749	0.0124792	2.23	0.026	0.0033161	0.0522336
/sigma	0.0180117	0.0001429	126.03	0.000	0.0177315	0.0182918

Por su parte los rezagos tanto para la parte autorregresiva como de medias móviles resultan significativos a un nivel de confianza del 95% y de signos opuestos. La constante, en cambio, no es significativa.

La tabla 4 expone los valores para los criterios de Akaike y Bayesiano. Entre todos los modelos estudiados estos valores no resultan ser los más bajos, sin embargo, la diferencia es pequeña y el modelo elegido tiene la particularidad de generar residuos con una estructura mucho más parecida a la del “ruido blanco” que sus competidores.

Tabla 4: Criterios de información AIC y BIC para ARIMAX(1,0,3)

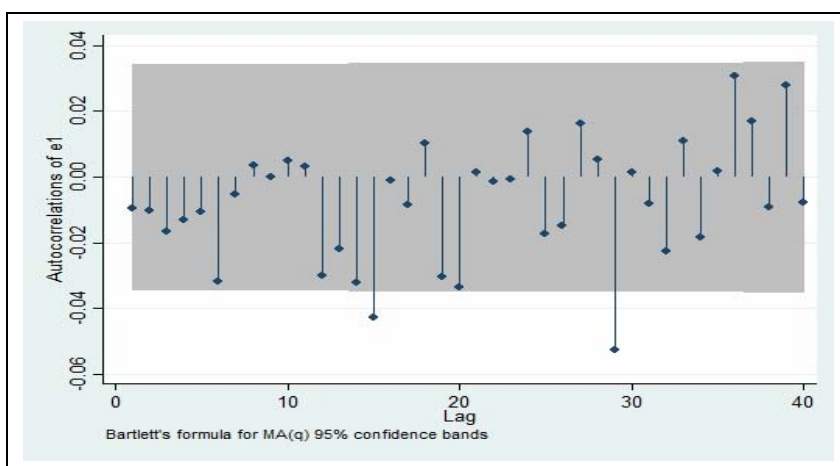
Akaike's information criterion and Bayesian information criterion						
Model	obs.	ll (null)	ll (model)	df	AIC	BIC
.	3263	.	8476.593	7	-16939.19	-16869.55

4) Verificación

Continuando con la metodología B-J se procede a realizar un análisis tanto a nivel individual como a nivel global de los residuos de estimación. Como se sabe, en todo modelo ARIMAX(p,d,q): $\Phi_p(L)(1-L)^d Y_t = X_t + \Theta_q(L)e_t$ la componente aleatoria e_t es un proceso “ruido blanco”, lo que implica que sus coeficientes de autocorrelación ρ_k para $k = 1,2, \dots$ sean

nulos. Si el modelo está correctamente especificado, entonces debe transformar las observaciones Y_t en residuos $\hat{\epsilon}_t = [\Phi_p(L)(1 - L)^d Y_t - X_t] \Theta_q(L)^{-1}$ para los que se espera un comportamiento similar al del ruido blanco.

Ilustración 6: Función de autocorrelación parcial – Residuos modelo ARIMAX(1,0,3)



Como primera aproximación se utiliza la función de autocorrelación de dichos residuos el cual incluye bandas para el intervalo de confianza al 95% de manera tal de poder analizar uno por uno dichos residuos.

Como puede observarse, los coeficientes para $k = 15$ y $k = 29$ parecerían no estar dentro del intervalo de confianza lo cual es poca evidencia en contra de la especificación del modelo.

A fin de asegurar la correcta especificación del modelo y confirmar el comportamiento de los residuos similar a la de ruido blanco, se procede a realizar el test de Portmanteau para ruido blanco.

Tabla 5: Test de Portmanteau -ruido blanco- residuos del modelo ARIMAX(1,0,3)

Portmanteau test for white noise
Portmanteau (Q) statistic = 50.1819
Prob > chi2 (40) = 0.1298

Del cuadro expuesto no existe evidencia ni a un nivel de confianza del 90% para rechazar la hipótesis nula de que los residuos del modelo se comportan como ruido blanco por lo que podemos concluir que el mismo, se encuentra bien identificado.

5) Predicción

Se busca pronosticar el comportamiento de la variable en estudio de acuerdo con la estructura probabilística subyacente que generó las observaciones. En secciones anteriores, se procuró

identificar el modelo que mejor ajustase al proceso generador de datos, estimarlo y verificar su correcta especificación. En esta sección se exponen los resultados del modelo en comparación con los datos reales.

La ilustración 7 compara los primeros cien rendimientos observados (RendBMA) en contraposición a los estimados por nuestro modelo (RendARIMAX). Si bien el modelo por momentos subestima o sobreestima lo ocurrido en la realidad, pareciera apuntar en la dirección correcta en cuanto al comportamiento de la variable acompañando su trayectoria.

La ilustración 8 compara la serie de precios observados PrecioCierreBMA contra la serie de precios estimados P-ARIMAX, los cuales siguen la siguiente especificación funcional.

$$\hat{P}_t = P_{t-1} * \widehat{RendBMA}_t + P_{t-1}$$

En el mismo gráfico se nota claramente cómo, a medida que la volatilidad del precio comienza a crecer, el modelo sobreestima dicha variación.

Ilustración 7: Comparación primeros 100 rendimientos observados vs. estimados ARIMAX

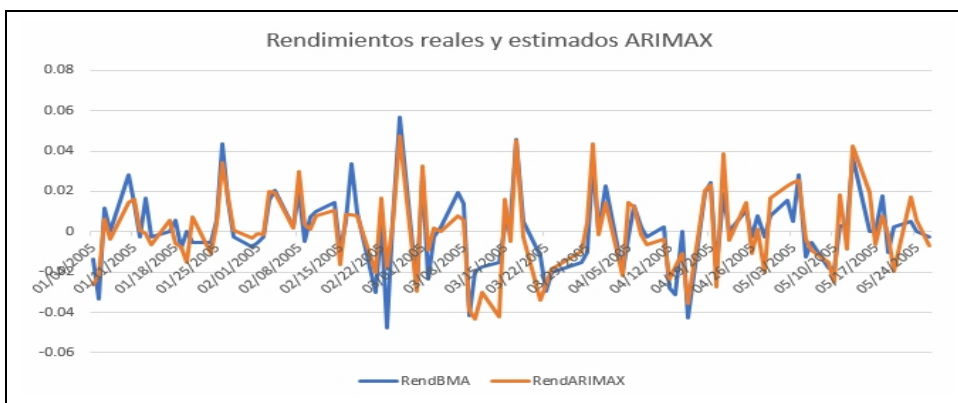
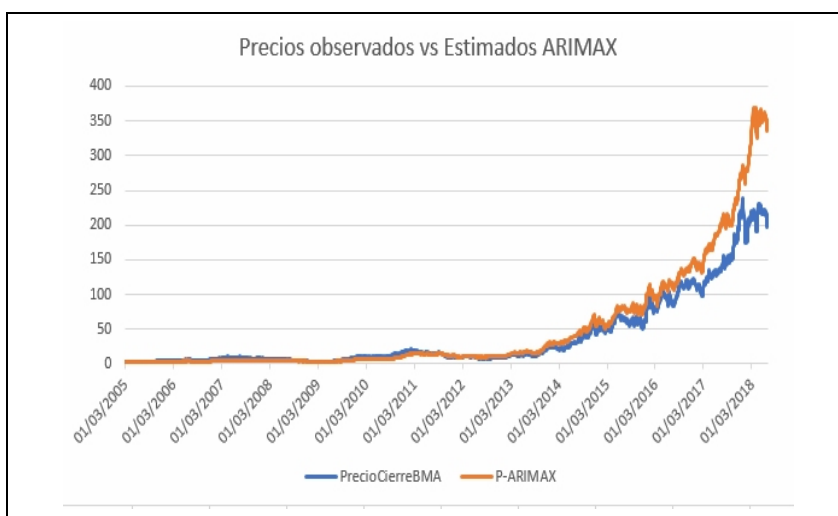


Ilustración 8: Comparación serie de precios observados y serie precios estimados ARIMAX



3. Metodología ARFIMAX

3.1 El modelo

Se dice que una variable Y_t sigue un proceso ARFIMAX(p,d,q) si es solución a la ecuación:

$$\Phi_p(L)\Delta^d Y_t = \Phi_p(L)(1-L)^d Y_t = X_t + \Theta_q(L)e_t \quad \text{Ec 3}$$

donde: $\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ y $\Theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ corresponden, respectivamente, a los polinomios autorregresivos y de medias móviles del proceso ARMA, cuyos ceros están fuera del círculo unidad y no tienen raíces en común.

$$(1-L)^d = \sum_{K=0}^{\infty} \binom{d}{K} (-L)^K = 1 - dL - \left(\frac{1}{2!}\right) d(d-1)L^2 - \left(\frac{1}{3!}\right) d(d-1)(d-2)L^3 \dots$$

es el operador de diferencia fraccional, definido para $d > -1$; d es el parámetro de *diferencia fraccional* y es un número real; X_t es el polinomio de variables independientes estructurales; las e_t son variables aleatorias ruido blanco: independientes e idénticamente distribuidas con media cero y varianza finita σ_e^2 .

En Hosking (1981) queda demostrado que si $d > -0.5$ y las raíces del polinomio $\Theta_q(L) = 0$ caen fuera del círculo unitario, Y_t es un proceso invertible. Además, se demuestra que si $d < 0.5$ y las raíces del polinomio $\Phi_p(L) = 0$ caen fuera del círculo unitario, Y_t es un proceso estacionario. En resumen, el proceso ARFIMAX(p,d,q) resulta estacionario e invertible, si todas las raíces de $\Theta_q(L) = 0$ y de $\Phi_p(L) = 0$ caen fuera del círculo unitario y $-0.5 < d < 0.5$.

En el mismo trabajo se demuestra que cuando $0 < d < 0.5$ existe una constante positiva tal que:

$$\rho_k \approx Ck^{2d-1}$$

Esto significa que la función de autocorrelación decae hiperbólicamente hacia cero y no es absolutamente sumable, es decir, $\sum_{k=-\infty}^{\infty} |\rho_k|$ no converge. La expresión exacta de las autocorrelaciones se deriva de Hosking (1981) (Castaño, 2016, 5).

3.2 Procesos con memoria larga

Si Y_t es un proceso estacionario ($d < 0.5$), la función de densidad espectral del proceso ARFIMAX es:

$$f_Y(\omega) = |1 - e^{-i\omega}|^{-2d} f_W(\omega) = \left[2 \sin\left(\frac{\omega}{2}\right)\right]^{-2d} f_W(\omega), 0 \leq \omega \leq \pi \quad \text{Ec 4}$$

donde $f_W(\omega) = \frac{\sigma_a^2}{2\pi} \left| \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})} \right|^2$ es la densidad espectral del proceso ARMA(p,q), $W_t = (1-L)^d Y_t$.

Cuando el proceso es invertible y estacionario ($0 < d < 0.5$), la densidad espectral resulta ser una función decreciente del parámetro ω , no acotada en el origen y está dominada por las frecuencias bajas. Entonces se dice que el proceso tiene memoria larga. Queda expuesta entonces la relación directa que hay entre la persistencia de las correlaciones rezagadas lejanas y la dinámica del espectro en las frecuencias bajas. Si $d < 0$, la densidad espectral se anula en el origen y está dominada por las frecuencias altas. Entonces, se dice que el proceso presenta

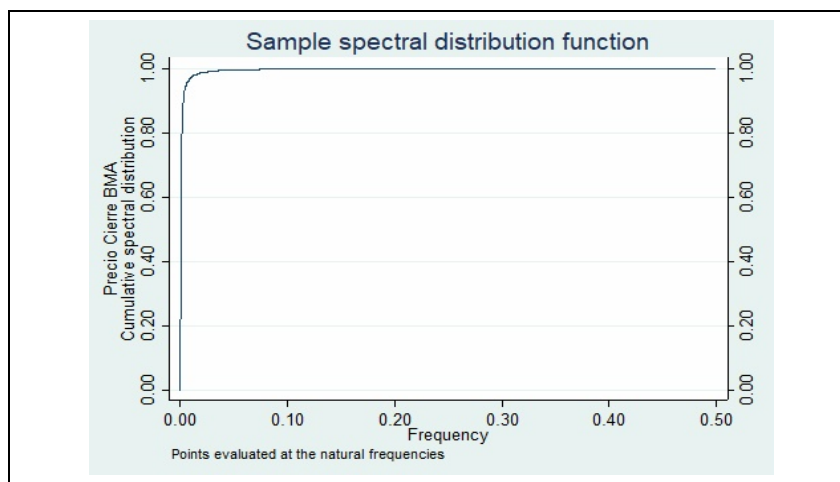
dependencia negativa o antipersistencia, y por tanto tiene memoria corta. Si $d > 0.5$, el proceso es, generalmente, no estacionario. Sin embargo, si $(0.5 < d < 1)$ el proceso no es estacionario, pero tiene reversión a la media, mientras que si $d \geq 1$ el proceso es no estacionario y tampoco tiene reversión a la media (Castaño, 5).

3.3 Aplicación

Para continuar con nuestro activo en estudio, a continuación, se presentan los gráficos de densidad espectral acumulada y el periodograma. Si bien ambos gráficos son sobre la variable PrecioCierreBMA, la cual luego para ser modelada es transformada volviéndola estacionaria, los mismos nos dan una idea acerca de la existencia de memoria larga en la serie original que se quiere estudiar y por tanto la posibilidad de que un proceso ARFIMAX(p,d,q) sea más conveniente que uno ARIMAX(p,d,q) a fin de modelarla.

Como se observa en la ilustración 9 la densidad espectral se acumula casi por completo en las frecuencias más bajas, lo cual nos da indicio del peso de las mismas y por tanto la existencia de memoria larga en el proceso. La ilustración 10 por su parte, confirma dicha idea.

Ilustración 9: Densidad espectral acumulada para la serie de precios del activo BMA

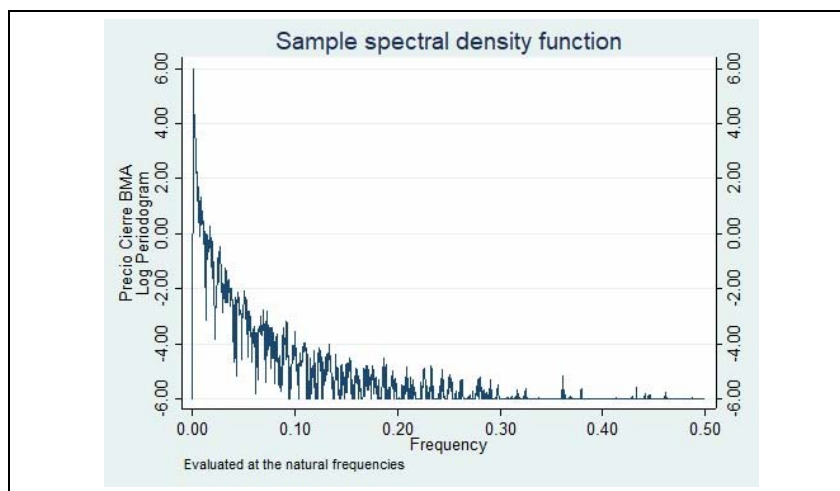


1) Identificación

Esta sección es análoga a la desarrollada para la metodología ARIMAX (p,d,q) en la cual ya se mostró que la variable transformada “RendBMA” es estacionaria y se expusieron sus funciones de autocorrelación y autocorrelación parcial. Las últimas, al igual que en dicha sección, son la hoja de ruta para identificar el modelo.

$$\widehat{\text{RendBMA}}_t^d = \beta_0 + \beta_1 \text{RendGGAL}_t + \beta_2 \text{RendIMV}_t + \phi_1 \widehat{\text{RendBMA}}_{t-1} + \phi_2 \widehat{\text{RendBMA}}_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-3} + e_t$$

Ilustración 10: Periodograma para la serie de precios del activo BMA



2) Estimación

En la tabla 6 se muestra la salida de datos para el modelo ARFIMAX (2,0,3).³ En este caso, al igual que en la metodología anterior, ambas variables estructurales son altamente significativas y, esta vez, la constante también resulta ser significativa. La correlación positiva de ambas variables estructurales se explica de igual manera que en el modelo anterior.

Tabla 6: Estimación del modelo ARFIMAX (2,0,3) para “Rend-BMA” por máxima verosimilitud

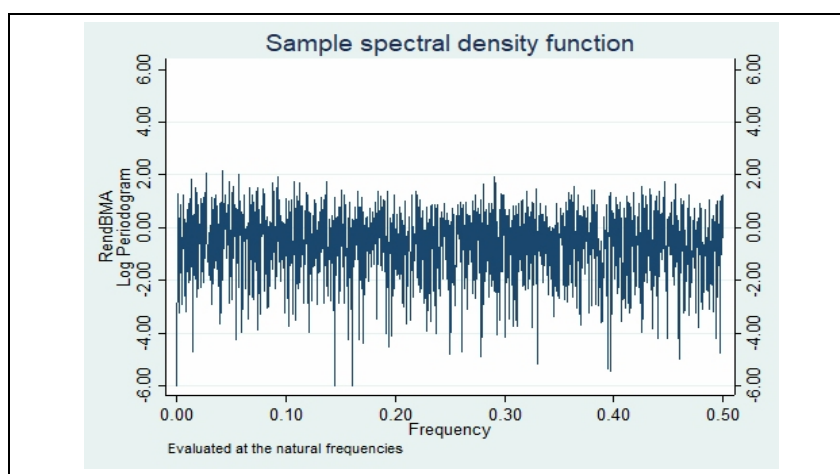
ARIMA regression						
Sample: 03jan2005 - 02may2018			Number of obs. = 3263			
Log likelihood = 8482.0515			Wald chi2 (7) = 4469.49			
			Prob. > chi2 = 0.000			
OPG						
RendBMA	Coef.	Std. Err.	z	P > (z)	[95% Conf. Interval]	
RendBMA						
RendGGAL	0.4136621	0.0184564	22.41	0.000	0.3774882	0.4498361
RendIMV	0.5131119	0.025394	20.21	0.000	0.4633406	0.5628832
_cons	0.0004048	0.0001702	2.38	0.017	0.0000712	0.0007348
ARMA						
ar						
L1.	-0.675333	0.0951752	-7.1	0.000	-0.8618729	-0.4887932
L2.	0.1015365	0.404954	2.51	0.012	0.0221669	0.180906
ma						
L1.	0.7734793	0.0800964	9.66	0.000	0.6164932	0.9304654
L3.	0.0620614	0.020948	2.96	0.003	0.0210039	0.1031188
d	-0.1029726	0.0309948	-3.32	0.001	-0.1637213	-0.042224
/sigma	0.0003233	0.000008	40.39	0.000	0.0003076	0.000339

³ Para conocer los valores de los criterios de información AIC y BIC para distintos modelos ARFIMAX que se compararon, ver Anexo B.

Los coeficientes de los rezagos tanto para la parte autorregresiva como para la parte de medias móviles resultan significativos a un nivel del 95% pero esta vez no son totalmente opuestos en signo ya que el primer y segundo rezago de la parte autorregresiva tienen signos opuestos, lo cual otorga a la serie estimada mayor suavidad.

El parámetro d se encuentra en el rango $(-0.5; 0.5)$ por lo que podemos asegurar que la serie es estacionaria e invertible. En cuanto a su signo negativo, esto nos da cuenta de una no-persistencia o memoria corta para la serie transformada. Esto se debe a que la transformación elimina la tendencia y por tanto la componente predominante de frecuencias bajas, dando igual relevancia a las frecuencias altas. Para confirmarlo, resulta relevante observar el periodograma de la serie transformada.

Ilustración 11: Periodograma para la serie transformada - Rendimientos BMA

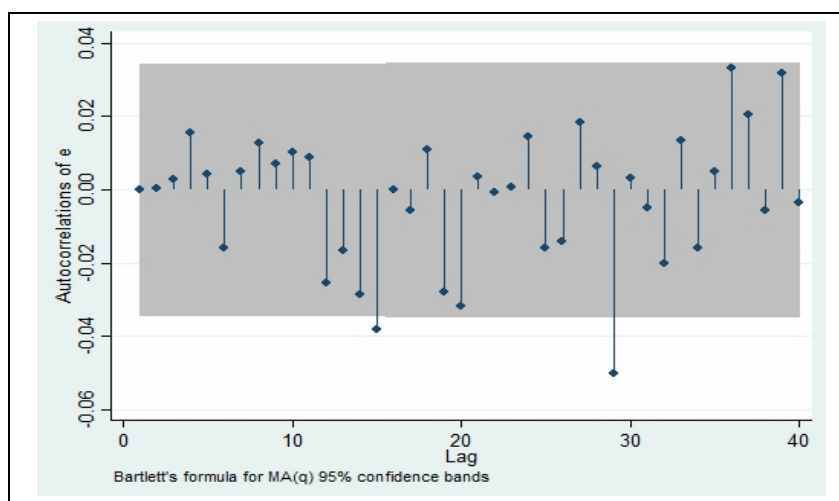


3) Verificación

Aquí nuevamente se busca determinar si el modelo se encuentra correctamente especificado. Para ello se pondrá en estudio, al igual que en la sección anterior, la serie de residuos que genera el modelo.

En primer lugar, se analiza la función de autocorrelación de los residuos a fin de conocer si a nivel individual cada coeficiente de autocorrelación se encuentra dentro del intervalo de confianza a un nivel de 95% para determinar que estos son nulos. Como vemos en la ilustración 12, la función de autocorrelación resulta semejante a la del modelo ARIMAX con sólo dos coeficientes que caen por fuera del intervalo de confianza, para iguales valores de k que en la sección anterior: $k = 15$ y $k = 29$. Tal y como antes, si bien sería deseable que todos los coeficientes a nivel individual estén dentro del intervalo para $\rho_k = 0$, esto no es suficiente evidencia como para descartar el modelo por lo que resulta necesario realizar algún test que evalúe si la serie de residuos tiene un comportamiento “ruido blanco”. Para ello, se realiza, nuevamente, el test de Portmanteau.

Ilustración 12: Función de Autocorrelación Parcial – Residuos modelo ARFIMAX (2,0,3)



A partir de la tabla 7 se deduce que no existe evidencia suficiente para rechazar la hipótesis nula de comportamiento ruido blanco por lo que se considera adecuada la especificación utilizada.

Tabla 7: Test de Portmanteau -ruido blanco- residuos del modelo ARFIMAX (2,0,3)

Portmanteau test for white noise
Portmanteau (Q) statistic = 43.1046
Prob > chi2 (40) = 0.3399

4) Predicción

Finalmente, se expone la comparación entre los primeros cien rendimientos observados en la serie y los primeros cien rendimientos estimados por nuestro modelo, tal y como lo hicimos para la metodología ARIMAX.

En el gráfico de la ilustración 13 la serie de rendimientos estimados parece ser muy similar a la estimada por metodología anterior. Sin embargo, dado que a medida que la serie avanza, la volatilidad aumenta, sería de esperar que esta segunda metodología tenga un mejor desempeño al no sobre ponderar dicha varianza. A continuación, en la ilustración 14, se muestra la serie de precios observada en comparación a la serie estimada, bajo la misma fórmula que se realizó anteriormente.

Como puede observarse, el modelo ARFIMAX ajusta en buena forma a la serie observada, incluso cuando la misma aumenta su volatilidad. Sin embargo, en los momentos previos cuando las observaciones muestran una tendencia menos marcada, el modelo subestima los crecimientos observados como puede observarse para los períodos 2006-2008 y mediados de 2009-fines de 2012.

Ilustración 13: Comparación primeros 100 rendimientos reales vs. estimados ARFIMAX

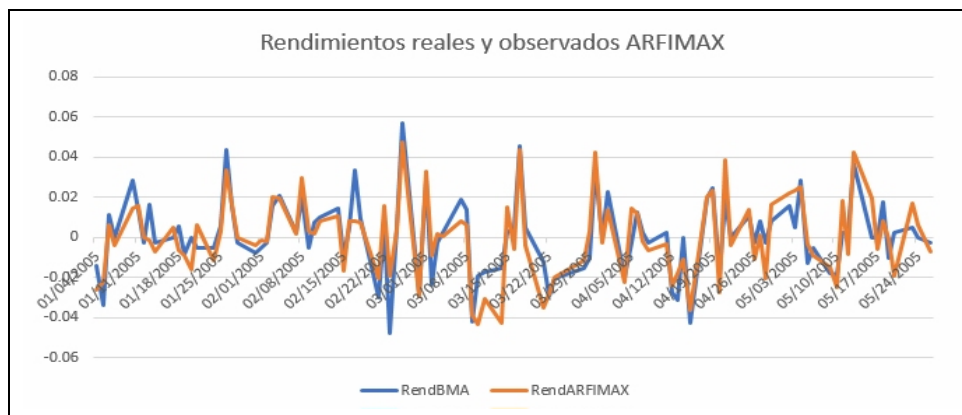
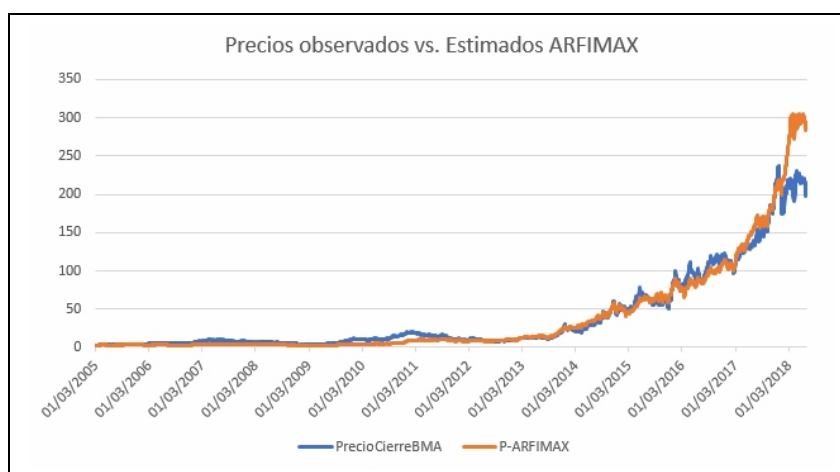


Ilustración 14: Comparación serie de precios observados y serie precios estimados ARFIMAX



Debemos mencionar también que, a partir de noviembre de 2017, la serie de precios observada comienza un período de alta volatilidad y lateralización como consecuencia de la citación a indagatoria de Brito, presidente del Banco Macro, cuyo activo es el que está en estudio. Nuestro modelo no logra captar esta noticia por lo que, al no incorporarla, parece seguir un movimiento de lateralización, pero a un nivel de precio mucho más elevado. Quedará para futuras investigaciones el tratamiento de esta serie con modelos heterocedásticos de la familia ARCH.

4. Comparación de pronósticos

En esta sección se exponen los resultados arrojados por ambos modelos, a fin de compararlos entre ellos y contra la serie observada.

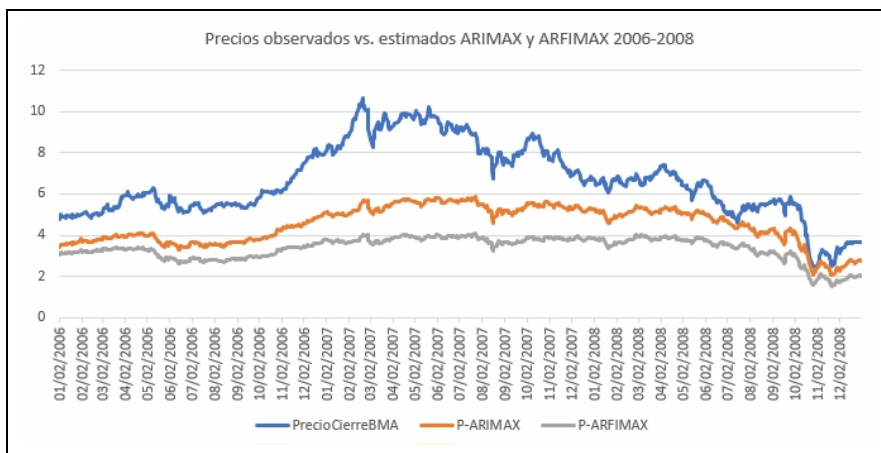
Primero se exponen la serie de precios observada y las series estimadas.

La ilustración 15 logra captar para todo el dominio de la serie los rasgos más característicos de la comparación donde “PrecioCierreBMA” corresponde a la serie de precios observados mientras que P-ARIMAX corresponde a los precios estimados por la metodología ARIMAX y P-ARFIMAX a los estimados por la metodología ARFIMAX. En primer lugar, se puede observar que a partir del año 2013 cuando la serie comienza a tener mayor volatilidad, la metodología ARFIMAX resulta más apropiada que su contraparte ARIMAX en cuanto a la realización de pronósticos.

Ilustración 15: Comparación serie de precios observados y series de precios estimados ARIMAX y ARFIMAX (2005-2018)



Ilustración 16: Comparación serie de precios observados y series de precios estimados ARIMAX y ARFIMAX (2006-2008)



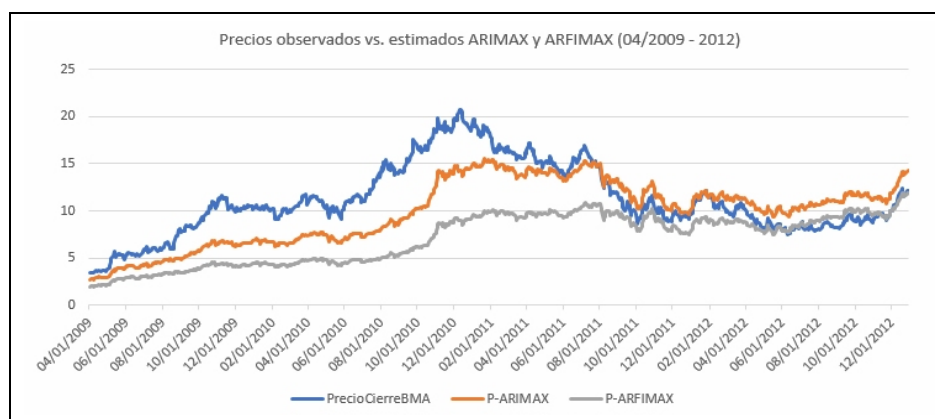
Por otro lado, puede verse que ninguno de los dos modelos logra captar la baja producida en el activo y la volatilidad lateral que le confiere la noticia de la citación a indagatoria del

presidente del Banco, Jorge Brito, y su relación con la causa *Ciccone*.⁴ Por otro lado, la metodología ARIMAX parecería ser más acertada a la hora de realizar pronósticos en momentos en que la tendencia alcista no es tan marcada. Resulta entonces de gran importancia realizar un estudio más detallado de los períodos en los que la diferencia de pronósticos se vuelve más marcada.

Como puede observarse ninguno de los dos modelos logra captar el crecimiento en los rendimientos de manera adecuada cuando estos se producen en forma tan rápida (el activo subió un 60% entre noviembre de 2006 y mediados de febrero de 2007). En estos casos la metodología ARIMAX parecería captar de mejor forma las subas, que su contraparte ARFIMAX la cual suaviza demasiado para estos casos la volatilidad de las estimaciones.

El gráfico en la ilustración 17 resulta aún más elocuente. Se puede apreciar que frente a una suba de los rendimientos, cuando la tendencia de largo plazo aún es lateral, la metodología ARIMAX parece replicar de mejor manera que su contraparte ARFIMAX. Sin embargo, esta última parecería ajustar de mejor manera cuando los precios actúan en la parte baja.

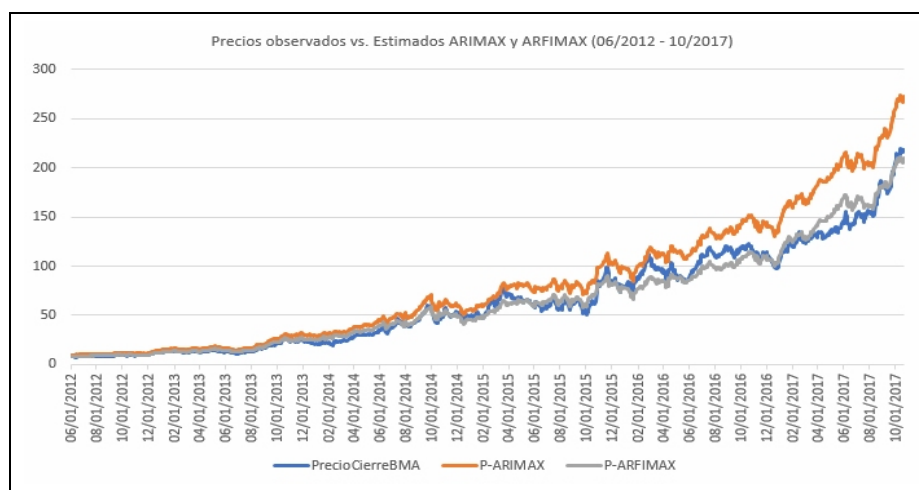
Ilustración 17: Comparación serie de precios observados y series de precios estimados ARIMAX y ARFIMAX (04/2009-2012)



Por otro lado, se puede verificar (ilustración 18) que a partir de junio de 2012, cuando los precios caen para luego continuar con una tendencia de largo plazo alcista más marcada, la metodología ARFIMAX capta de mejor manera los vaivenes en los rendimientos que la metodología ARIMAX.

⁴ Boudou Amado y otro s/cohecho y negociaciones incompatibles, tramitada en Juzgado Nacional en lo Criminal y Correccional Federal N°4.

Ilustración 18: Comparación serie de precios observados y series de precios estimados ARIMAX y ARFIMAX (10/2012-10/2017)



5. Conclusiones

En este trabajo se estudió el desempeño de dos metodologías similares con una única distinción en busca de evidencia a favor de su utilización. La estrategia de identificación, selección, estimación y, finalmente, verificación se desarrolló dentro de los parámetros clásicos de la metodología de Box-Jenkins. La distinción entre ambas metodologías fue el fraccionamiento en el grado de diferenciación.

Se hizo mención a los procesos de memoria larga y su estudio, y la ventaja de fraccionar el orden de diferenciación a fin de captar en forma armónica los shocks que aparecen a lo largo de la serie, de manera tal que estos perduren más que en los procesos no diferenciados (metodología ARMA) pero desaparezcan a lo largo de la serie, a diferencia de la permanencia que tienen cuando los modelos están sobre diferenciados (metodología ARIMA).

Se utilizó la serie de precios históricos para el activo de renta variable BMA (acciones del Banco Macro) con la intención de determinar las principales variables estructurales que la influyen e identificar la estructura autorregresiva y de medias móviles que mejor se ajustan en la modelización de su proceso generador de datos.

Luego de la comparación entre ambos modelos y entre ellos para con la serie observada resulta evidente el mejor poder predictivo de la metodología ARFIMAX que el de su contraparte ARIMA. Podríamos atribuir su mejor desempeño a la evidencia en favor de memoria larga en la serie original de precios. Parecería que la mejor captación estaría por el lado de la tendencia de largo plazo, suavizando las estimaciones cuando ésta es lateral y captando de mejor manera la variación cuando la misma se torna ascendente.

Si bien durante la última fase ascendente ninguno de los dos modelos logró captar de manera adecuada el freno en la tendencia y la volatilidad característica de un momento de duda del mercado frente a su continuación, incluso la diferencia de nivel entre la serie estimada y la serie observada fue menor para la metodología ARFIMAX. Quedará a futuro, con mayor cantidad de observaciones determinar si dicho período correspondió a una reversión de tendencia o simplemente a una consolidación. Tal y como se mencionó anteriormente, este período de cambio

de tendencia y aumento de volatilidad se debió a cuestiones no económicas, tal y como fue la incertidumbre en la continuación del presidente del Banco en cuestión. En tal caso, y ya habiendo hecho mención de esto, queda para futuras investigaciones, el tratamiento de dicha serie con modelos heterocedásticos de la familia ARCH.

REFERENCIAS

- Box, G.E.P & Cox, D.R. (1964). *An analysis of transformations*. Journal of Statistical Society Serie B, 26: 211-252
- Box, G.E.P. & Jenkins, G.M. (1976). *Times Series Analysis: Forecasting and Control*. Holden Day
- Castaño Vélez, E. (2015). *Series de tiempo de memoria larga, identificación y aplicaciones*. Seminario de Análisis aplicado de Series de Tiempo (Universidad Nacional de Colombia)
- Castaño Vélez, E. (2016). *Identificación de modelos ARFIMA*. Revista de la Facultad de Ciencias (Universidad Nacional de Colombia), 5 (1): 12-37
- Granger, C.W.J. & Joyeux, R. (1980). *An introduction to long memory time series models and fractional differencing*. Journal of Time Series Analysis, 1: 15-29
- Gujarati, D. & Porter, D.C. (2010). *Econometría*, 5ª Ed. McGraw Hill/Interamericana
- Hosking, J.R.M. (1981). *Fractional differencing*. Biometrika, 68 (1): 165-76
- Pérez, A. y Ruiz, E. (2002). *Modelos de memoria larga para series económicas y financieras*. Investigaciones Económicas, XXVI (3): 395-445
- Robinson, P.M. (2010). *Long memory models*. En *Macroeconometrics and time series analysis. The New Palgrave Dictionary of Economics*, 163-167
- Urbisaia, H. y Brufman, J. (2000). *Análisis de series de tiempo: Univariadas y multivariadas*, 2ª Ed. Ediciones Cooperativas (Buenos Aires)
- Wooldridge, J.M. (2010). *Introducción a la econometría: Un enfoque moderno*, 4ª Ed. Cengage Learning

ANEXOS

Anexo A: Estacionariedad en sentido débil

Este tipo de procesos también son conocidos como estacionarios en covarianza. Un proceso estocástico $\{Y_t, t = 1, 2, \dots\}$ con un segundo momento finito $[E(x_t^2) < \infty]$ es estacionario en covarianza si i) $E(x_t)$ es constante, ii) $Var(x_t)$ es constante y iii) para cualquier $t, h \geq 1$, la $Cov(x_t, x_{t+h})$ depende sólo de h y no de t (Wooldridge 4ta. Edición 2010, 377).

Este tipo de procesos se centran únicamente en los dos primeros momentos del proceso estocástico (media y varianza). Las mismas deben ser constantes en el tiempo y la covarianza entre x_t y x_{t+h} dependerá sólo de la distancia entre los dos términos, h , y no del lugar que ocupa en la serie el periodo inicial, t . Por lo tanto, la correlación entre x_t y x_{t+h} depende también sólo de h .

Si lo que se busca es encontrar algún tipo de relación entre dos o más variables que utilizan el análisis de regresión, se debe tener algún tipo de estabilidad a lo largo del tiempo para dicha

relación. De lo contrario, si se permitiese que ésta cambie en forma arbitraria en cada período, no se podría aprender mucho acerca de cómo un cambio en una variable afecta a la otra si sólo se tiene acceso a una única realización de serie de tiempo. En un nivel más técnico, la estacionariedad simplifica los enunciados de la ley de los grandes números y el teorema del límite central.

Anexo B: Comparación en hoja de salida para modelos ARIMAX y ARFIMAX

En las tablas siguientes, si el modelo tiene la observación de *Modelo no significativo* esto se refiere a que uno o más parámetros del modelo no resultaron ser significativos a nivel individual. Por otro lado, cuanto menor sea el valor del criterio de información (tanto AIC como BIC), mejor será el modelo.

También se incluye el p-value para el factor de correlación número cuarenta de los residuos para el test de ruido blanco de Portmanteau. Cuanto mayor sea este valor, menor es la probabilidad de rechazar la hipótesis nula de que los residuos tienen estructura de ruido blanco y por lo tanto mejor será el modelo. Como última aclaración, el modelo seleccionado es aquel que se encuentra en letra negrita.

ARIMAX

Variabes Estructurales	Orden de "p"	Orden de "q"	AIC	BIC	Prob. > chi2 (40)	Observaciones
RendTC; RendIMV	1, 2	1, 2, 3	-16424	-16478	0.1398	
RendTC; RendIMV	1, 2	1, 3				Modelo no significativo
RendTC; RendIMV	1, 2	1, 2	-16430	-16479	0.1084	
RendTC; RendIMV	1, 2	1				Modelo no significativo
RendTC; RendIMV	1, 2	2	-16432	-16475	0.03	
RendTC; RendIMV	1, 2	3				Modelo no significativo
RendTC; RendIMV	1	1, 2, 3				Modelo no significativo
RendTC; RendIMV	1	1, 3	-16436	-16478	0.08	
RendTC; RendIMV	1	1, 2				Modelo no significativo
RendTC; RendIMV	1	3				Modelo no significativo
RendTC; RendIMV	1	2				Modelo no significativo
RendTC; RendIMV	1	1				Modelo no significativo
RendTC; RendIMV	1	2, 3				Modelo no significativo
RendTC; RendIMV		1, 2, 3				Modelo no significativo
RendTC; RendIMV		1, 2				Modelo no significativo
RendTC; RendIMV		1				Modelo no significativo
RendTC; RendIMV		1, 3				Modelo no significativo
RendTC; RendIMV		2, 3				Modelo no significativo
RendTC; RendIMV	1, 2					Modelo no significativo
RendTC; RendIMV	2					Modelo no significativo
RendTC; RendIMV	1					Modelo no significativo
RendIMV; RendGGAL	1	1	-16903	-16940	0.0754	
RendIMV; RendGGAL	1, 2	1, 3				Modelo no significativo
RendIMV; RendGGAL	1	1, 3	-16896	-16939	0.1298	
RendIMV	1, 2	1, 2, 3	-16429	-16477	0.1291	
RendIMV	1, 2	1, 3				Modelo no significativo
RendIMV	1, 2	1, 2				Modelo no significativo
RendIMV	1	1, 2, 3				Modelo no significativo
RendIMV	1	1, 3	-16441	-16477	0.0821	
RendIMV	1	1, 2				Modelo no significativo
RendIMV	1	1				Modelo no significativo

ARFIMAX

Variabes Estructurales	Orden de "p"	Orden de "q"	AIC	BIC	Prob. > chi2 (40)	Observaciones
RendIMV	1, 2	1, 2, 3				Modelo no significativo
RendIMV	1, 2	1, 3	-16480.42	-16431.7	0.1856	
RendIMV	1, 2	1, 2				Modelo no significativo
RendIMV	1, 2	1				Modelo no significativo
RendIMV	1	1, 2, 3	-16480.35	-16431.62	0.1795	
RendIMV	1	1, 2				Modelo no significativo
RendIMV	1	1, 3				Modelo no significativo
RendIMV	1	1	-16468.45	-16431.91		Residuos no ruido blanco
RendIMV	1	3				Modelo no significativo
RendIMV	1	2				Modelo no significativo
RendGGAL RendIMV	1, 2	1, 3	-16946.1	-16891.29	0.3399	
RendGGAL RendIMV	1	1, 3				Modelo no significativo