

Modelado y estimación de la volatilidad estocástica: aplicación a la inflación

Modeling and estimation of stochastic volatility: application to inflation

Juan Carlos Abril¹  y María de las Mercedes Abril¹ 

¹Universidad Nacional de Tucumán y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Av. Independencia 1900, San Miguel de Tucumán, Tucumán, Argentina.

Correspondencia: jabril@herrera.unt.edu.ar;
mabrilblanco@hotmail.com

Recepción: 28 de agosto de 2024 - **Aceptación:** 15 de octubre de 2024 - **Publicación:** 28 de octubre de 2024

Resumen

Este trabajo presenta un análisis detallado sobre la modelización de la volatilidad estocástica (MVE) aplicada a la serie de inflación del Gran Buenos Aires, Argentina, cubriendo el período de enero de 1943 a mayo de 2019. Utilizando modelos de espacio de estado y técnicas de estimación basadas en el filtro y suavizador de Kalman, se propone un enfoque alternativo y más flexible que los tradicionales modelos ARCH-GARCH, ya que en los MVE la volatilidad depende de sus propios valores pasados y no de los retornos de la serie. Se desarrolla un modelo específico que capta las características clave de la serie inflacionaria, incluyendo componentes no observables que son estimados y modelados a lo largo del tiempo. Este enfoque permite descomponer la inflación en componentes estructurales como la inflación núcleo (que excluye sectores volátiles como alimentos y energía) y la inflación tendencia (que incluye la inflación núcleo más el resto de los sectores). Asimismo, se realiza un análisis en profundidad del período 2004-2015, cuando el Instituto Nacional de Estadística y Censos (INDEC) estuvo intervenido políticamente, mostrando cómo la intervención afectó la precisión y fiabilidad de los datos. Los resultados obtenidos demuestran que el modelo MVE es capaz de capturar las dinámicas de volatilidad en una serie de tiempo económica compleja como la inflación, proporcionando mejores estimaciones y predicciones que los modelos ARCH-GARCH en contextos de alta variabilidad y cambios estructurales. En particular, el enfoque de espacio de estado permite estimar la volatilidad estocástica de los errores, revelando información clave sobre los ciclos de inflación y los errores sistemáticos en los datos reportados por el INDEC. Además, se discuten las implicaciones teóricas de estos hallazgos en la economía argentina y su relevancia para la modelización de series temporales económicas volátiles.

Palabras claves: volatilidad estocástica, inflación, modelos de espacio de estado, filtro de Kalman, intervención política, inflación núcleo, inflación tendencia, series de tiempo.

Abstract

This study presents a detailed analysis of the modeling of stochastic volatility (SVM) applied to the inflation series of Greater Buenos Aires, Argentina, covering the period from January 1943 to May 2019. Using state-space models and estimation techniques based on the Kalman filter and smoother, the paper proposes an alternative and more flexible approach than traditional ARCH-GARCH models. In SVMs, volatility depends on its own past values rather than on the returns of the series. A specific model is developed that captures the key characteristics of the inflation series, including unobservable components that are estimated and modeled over time. This approach allows for the decomposition of inflation into structural components such as core inflation (which excludes volatile sectors like food and energy) and trend inflation (which includes core inflation plus the remaining sectors). Additionally, an in-depth analysis of the 2004-2015 period is conducted, when the National Institute of Statistics and Censuses (INDEC) was politically intervened, demonstrating how the intervention impacted the accuracy and reliability of the reported data. The results show that the SVM model is capable of capturing volatility dynamics in a complex economic time series like inflation, providing better estimates and forecasts than ARCH-GARCH models in contexts of high variability and structural changes. In particular, the state-space approach enables the estimation of the stochastic volatility of errors, revealing key insights into inflation cycles and systematic errors in the data reported by INDEC. Furthermore, the theoretical implications of these findings for the Argentine economy and their relevance for modeling volatile economic time series are discussed.

Keywords: stochastic volatility, inflation, state-space models, Kalman filter, political intervention, core inflation, trend inflation, time series.

1. Introducción

El estudio del fenómeno de la volatilidad se ha desarrollado principalmente a partir del análisis de las series de tiempo referidas a la economía. Sin embargo, debe enfatizarse que cualquier serie de tiempo puede estar sujeta a la presencia de volatilidad, no solamente las que serán parte del objeto de nuestro estudio.

Numerosas series de tiempo económicas no tienen una media constante y en situaciones prácticas, frecuentemente vemos que la varianza del error observacional, condicional al conocimiento pasado, está sujeta a una sustancial variabilidad a través del tiempo. Ese fenómeno es conocido como *volatilidad*.

Para tomar en cuenta la presencia de la volatilidad en una serie económica es necesario recurrir a modelos conocidos como *modelos heterocedásticos condicionales*. En estos modelos, la varianza de una serie en un dado instante de tiempo, depende de la información pasada y de otros datos disponibles hasta aquel

instante de tiempo, de modo que se debe definir una varianza condicional, que no es constante y no coincide con la varianza global de la serie observada.

Una característica importante de las series de tiempo financieras es que ellas no son en general serialmente correlacionadas, pero sí dependientes. De este modo modelos lineales como aquellos pertenecientes a la familia de los modelos ARMA pueden no ser apropiados para describir estas series.

Existe una variedad muy grande de modelos no lineales en la literatura, útiles para el análisis de series de tiempo económicas con volatilidad. Una clase importante de ellos son los modelos de tipo ARCH introducidos por Engle (1982) y sus extensiones. Estos modelos son no lineales en lo que se refiere a la varianza.

Los modelos de la familia ARCH o GARCH suponen que la varianza condicional (volatilidad) depende de los retornos pasados. En otras palabras, si σ_t^2 es la volatilidad, la familia ARCH-GARCH supone que la misma depende de la serie y_j para $j < t$. Por otra parte, el *modelo para volatilidad estocástica* o MVE (o SVM según sus siglas en inglés), propuesto por primera vez por Taylor (1980, 1986) no parte de este supuesto. Este modelo tiene como premisa el hecho de que la volatilidad σ_t^2 depende de sus valores pasados (σ_j^2 para $j < t$) pero es independiente del pasado de la serie bajo análisis (y_j para $j < t$).

La inspección superficial de series como la que presentamos en este trabajo, sugiere que las mismas no tienen una media y una varianza constante. Una variable estocástica en donde la varianza es constante se dice que es *homocedástica*, en contraposición a lo que sería una variable *heterocedástica*. Para aquellas series en donde haya volatilidad, la varianza no condicional puede ser constante aun cuando la varianza condicional en algunos períodos sea inusualmente grande y en otros pequeña.

Como aplicación, se analiza la serie de inflación desde enero de 1943 hasta mayo de 2019 del Gran Buenos Aires. Los datos son los publicados oficialmente por el Instituto Nacional de Estadística y Censos (INDEC). Si bien es un período muy largo en el cual sucedieron cambios de base, cambios de canasta e intervenciones en el INDEC, se puede apreciar que ciertos patrones generales de comportamiento han persistido en el tiempo, lo que nos permite admitir que el estudio está adecuadamente basado en la información disponible.

Si bien el término "inflación núcleo", también conocida como "inflación subyacente", goza de un uso común generalizado, parece no tener una definición clara. Se puede decir que inflación núcleo es el cambio en los precios de los bienes y servicios excluyendo aquellos sectores de los alimentos, bebidas y energía. Se excluye a esos sectores porque sus precios suelen ser los que presentan mayor volatilidad. En general, cuando las personas usan el término parecen referirse al componente a largo plazo o persistente de la serie. Pero una definición clara de la inflación núcleo requiere necesariamente un modelo de cómo se puede explicar el proceso generador de la serie a estudiar. En economía, cualquier estructura formal de este tipo no es fácil de formular. Pero el enfoque de espacio de estado que aplicaremos en este caso nos permite definir el componente correspondiente y estimarlo adecuadamente, lo que haremos para el caso de la inflación en la Argentina estudiada en este trabajo.

La inflación tendencia es usualmente la tendencia de la infla-

ción que se suele dar en las noticias. Incluye la inflación núcleo más los cambios en los precios de los sectores de los alimentos, bebidas y energía, y otras variables que expliquen el proceso inflacionario a largo plazo. Nuevamente, el enfoque de espacio de estado que aplicaremos en este caso nos permite definir el componente correspondiente y estimarlo adecuadamente, lo que haremos para el caso de la inflación en la Argentina estudiada en este trabajo.

Una discusión importante sobre la inflación núcleo y la inflación tendencia puede verse en Bryan y Cecchetti (1994), Bryan et al. (1997) y Stock y Watson (2015).

2. Modelado de series de tiempo económicas y financieras

La idea básica de una serie de tiempo es muy simple, consiste en el registro de cualquier cantidad fluctuante medida en diferentes puntos del tiempo.

Concretamente, una *serie de tiempo* es un conjunto de observaciones y_1, \dots, y_n ordenadas en el tiempo. El modelo básico y general que se utiliza para representar cualquier serie de tiempo es el modelo aditivo, dado por

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \dots, n, \quad (1)$$

donde μ_t es un componente que cambia suavemente en el tiempo llamado *tendencia*, γ_t es un componente periódico llamado *estacionalidad* y ε_t es un componente irregular llamado *error*. Como podemos ver, la característica común de todos los registros que pertenecen al dominio de las *series de tiempo* es que ellos están influenciados, aunque sea parcialmente, por fuentes de variación aleatoria.

La principal razón para modelar una serie de tiempo es permitir la predicción de sus valores futuros. La característica distintiva de un modelo de serie de tiempo, opuesto, por ejemplo, a un modelo econométrico, es que no se realiza ningún intento para formular una relación de comportamiento entre la serie de tiempo considerada y otras variables explicativas. Los movimientos de la serie son explicados solamente en términos de su propio pasado, o por su posición en relación al tiempo o por su estructura. Las predicciones, son realizadas mediante extrapolación.

Numerosas series de tiempo económicas no tienen una media constante y en muchos casos se observan fases en donde reina una relativa tranquilidad seguido de períodos de importantes cambios. Una gran parte de la investigación actual en series de tiempo y econometría se concentra en extender la metodología clásica y comúnmente usada de Box y Jenkins para analizar este tipo de comportamiento. Ahora bien, existe una característica presente en las series de tiempo que se refieren a activos financieros (o directamente series de tiempo financieras) y otras series referidas a actividades económicas y es lo que se conoce como *volatilidad*, que puede ser definida de varias maneras, pero que no es directamente observable. Para tomar en cuenta la presencia de grupos de volatilidad en una serie financiera o económica es necesario recurrir a modelos conocidos como *heterocedásticos condicionales*. En estos modelos, la varianza (o volatilidad) de una serie en un dado instante de tiempo, depende de su pasa-

do y de otras informaciones disponibles hasta aquel instante de tiempo, de modo que se debe definir una *varianza condicional*, que no es constante y no coincide con la varianza global o no condicional de la serie observada.

3. La volatilidad

La volatilidad se define como la varianza de una variable aleatoria, condicional a toda la información pasada. Como la volatilidad no puede ser medida directamente, la misma puede manifestarse de varias maneras en una serie de tiempo.

Sea y_t la serie bajo estudio cuya dimensión $p = 1$. Definimos

$$\mu_t = E(y_t | Y_{t-1}) = E_{t-1}(y_t), \quad (2)$$

$$\begin{aligned} \sigma_t^2 &= \text{var}(y_t | Y_{t-1}) = E \{ (y_t - \mu_t)^2 | Y_{t-1} \} \\ &= E_{t-1}(y_t - \mu_t)^2 = \text{var}_{t-1}(y_t), \end{aligned} \quad (3)$$

como la media y la varianza condicionales de y_t dada la información hasta el instante $t - 1$ contenida en Y_{t-1} , respectivamente

La media y la varianza incondicional de y_t se denotarán como $\mu = E(y_t)$ y $\sigma^2 = \text{var}(y_t)$, respectivamente, y sea G la distribución de y_t . Es claro que (2), (3) y F determinan μ , σ^2 y G , pero no lo contrario. mayores detalles sobre esta formulación pueden verse en Abril M. (2014).

4. Modelos para volatilidad estocástica

Diremos que la serie y_t sigue un MVE, si

$$y_t = \sigma_t \varepsilon_t, \quad (4)$$

$$\sigma_t = e^{\frac{h_t}{2}}, \quad (5)$$

donde ε_t es una serie estacionaria con media igual a cero y varianza uno, y h_t es otra serie estacionaria con densidad de probabilidad dada por una función $f(h)$.

Como se puede ver en (5), h_t no es igual a la volatilidad σ_t^2 como suele ser la notación en los modelos de la familia ARCH-GARCH.

La formulación más simple del modelo supone que el logaritmo de la volatilidad, h_t , está dado por

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad (6)$$

donde η_t es una serie estacionaria, Gaussiana, con media cero, varianza σ_η^2 e independiente de ε_t . De esto se sigue que debemos tener $|\alpha_1| < 1$.

4.1. Otras formulaciones de MVE

Otras formulaciones de MVE pueden encontrarse en la literatura, dentro de las cuales destacamos las siguientes:

1. Forma canónica de Kim et al. (1998). En este caso el MVE

se escribe como

$$y_t = \beta e^{\frac{h_t}{2}} \varepsilon_t, \quad (7)$$

$$h_t = \mu + \alpha_1 (h_{t-1} - \mu) + \sigma_\eta \eta_t, \quad (8)$$

con

$$h_t \sim N \left(\mu, \frac{\sigma_\eta^2}{1 - \alpha_1^2} \right), \quad (9)$$

donde ε_t y η_t son ambas $N(0, 1)$ e independientes entre sí. Si $\beta = 1$, entonces $\mu = 0$.

2. La forma de Jaquier, Polson y Rossi (1994) del MVE es igual a

$$y_t = \sqrt{h_t} \varepsilon_t, \quad (10)$$

$$\log(h_t) = \alpha_0 + \alpha_1 \log(h_{t-1}) + \sigma_\eta \eta_t, \quad (11)$$

donde ε_t y η_t son ambas $N(0, 1)$ e independientes entre sí.

5. Propiedades de los MVE

Regresemos al modelo definido en las ecuaciones (4), (5) y (6). Supongamos que $\{\varepsilon_t\}$ constituya una sucesión de variables aleatorias independientes tal que $\varepsilon_t \sim N(0, 1)$, entonces $\log(\varepsilon_t^2)$ tiene una distribución llamada "log chi cuadrada", de tal manera que

$$E\{\log(\varepsilon_t^2)\} \approx -1, 27, \quad (12)$$

$$\text{var}\{\log(\varepsilon_t^2)\} = \pi^2/2. \quad (13)$$

De (4), (5) y (6) obtenemos

$$\log(y_t^2) = \log(\sigma_t^2) + \log(\varepsilon_t^2), \quad (14)$$

$$h_t = \log(\sigma_t^2) = \alpha_0 + \alpha_1 h_{t-1} + \eta_t. \quad (15)$$

Llamando $\xi_t = \log(\varepsilon_t^2) - E\{\log(\varepsilon_t^2)\} \approx \log(\varepsilon_t^2) + 1, 27$, tenemos que $E(\xi_t) = 0$, $\text{var}(\xi_t) = \pi^2/2$ y

$$\log(y_t^2) = -1, 27 + h_t + \xi_t, \quad \xi_t \sim \text{iid}(0, \pi^2/2), \quad (16)$$

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad (17)$$

donde *NID* significa que las variables son normales e idénticamente distribuidas. Se supone también que ξ_t y η_t son independientes entre sí en todo momento de tiempo.

A partir de las ecuaciones (4), (5) y (6) vamos a calcular algunos momentos del MVE. Tomando esperanza de (4) tenemos

$$E(y_t) = E(\sigma_t \varepsilon_t) = E(\sigma_t)E(\varepsilon_t) = 0, \quad (18)$$

dado que σ_t y ε_t son independientes.

La varianza de y_t es

$$\text{var}(y_t) = E(y_t^2) = E(\sigma_t^2 \varepsilon_t^2) = E(\sigma_t^2)E(\varepsilon_t^2) = E(\sigma_t^2). \quad (19)$$

Puesto que suponemos que $\eta_t \sim N(0, \sigma_\eta^2)$ y que h_t es esta-

cionario con

$$E(h_t) = \frac{\alpha_0}{1 - \alpha_1} = \mu_h, \quad (20)$$

y con

$$var(h_t) = \frac{\sigma_\eta^2}{1 - \alpha_1^2} = \sigma_h^2, \quad (21)$$

tenemos que

$$h_t \sim N\left(\frac{\alpha_0}{1 - \alpha_1}, \frac{\sigma_\eta^2}{1 - \alpha_1^2}\right). \quad (22)$$

Como h_t es normal o Gaussiana, $\sigma_t^2 = e^{h_t}$ es log-normal, luego tenemos

$$var(y_t) = E(y_t^2) = E(\sigma_t^2) = e^{\mu_h + \sigma_h^2/2}. \quad (23)$$

No es difícil mostrar que

$$E(y_t^4) = 3e^{2\mu_h + 2\sigma_h^2}, \quad (24)$$

de lo cual obtenemos que la kurtosis de y_t es

$$K(y_t) = \frac{3e^{2\mu_h + 2\sigma_h^2}}{e^{2\mu_h + \sigma_h^2}} = 3e^{\sigma_h^2} > 3, \quad (25)$$

como es de esperar, es decir, se tienen colas pesadas para los MVE.

La función de autocovarianzas de la serie y_t está dada por

$$\gamma_y(s) = E(y_t y_{t+s}) = E(\sigma_t \sigma_{t+s} \varepsilon_t \varepsilon_{t+s}) = 0, \quad (26)$$

puesto que ε_t y η_t son independientes. Entonces y_t es serialmente no correlacionada pero no independiente ya que existe correlación en $\log(y_t^2)$. Denotemos como $z_t = \log(y_t^2)$, entonces la función de autocovarianzas de la serie z_t está dada por

$$\gamma_z(s) = E[(z_t - E(z_t))(z_{t+s} - E(z_{t+s}))]. \quad (27)$$

Como el primer término entre paréntesis de (27) es igual a $h_t - E(h_t) + \xi_t$ y h_t es independiente de ξ_t , obtenemos

$$\begin{aligned} \gamma_z(s) &= E[(h_t - E(h_t) + \xi_t)(h_{t+s} - E(h_{t+s}) + \xi_{t+s})] \\ &= E[(h_t - E(h_t))(h_{t+s} - E(h_{t+s}))] + E(\xi_t \xi_{t+s}), \end{aligned} \quad (28)$$

y llamando $\gamma_h(s)$ y $\gamma_\xi(s)$ respectivamente a las autocovarianzas del segundo miembro de (28), tenemos que

$$\gamma_z(s) = \gamma_h(s) + \gamma_\xi(s), \quad (29)$$

para todo s .

Como estamos suponiendo que (6) se cumple, o sea un modelo AR(1), obtenemos

$$\gamma_h(s) = \alpha_1^s \frac{\sigma_\eta^2}{1 - \alpha_1^2}, \quad s > 0. \quad (30)$$

Por otra parte $\gamma_\xi(s) = 0$ para $s > 0$. Luego, $\gamma_z(s) = \gamma_h(s)$ para

todo $s \neq 0$. Con esto podemos escribir la función de autocorrelación de $z_t = \log(y_t^2)$ como

$$\rho_z(s) = \frac{\gamma_z(s)}{\gamma_z(0)} = \frac{\alpha_1^s \sigma_\eta^2 / (1 - \alpha_1^2)}{\gamma_h(0) + \gamma_\xi(0)}, \quad s > 0, \quad (31)$$

de donde obtenemos

$$\rho_z(s) = \frac{\alpha_1^s}{1 + \frac{\pi^2}{2\sigma_\eta^2}}, \quad s > 0,$$

lo que tiende a cero exponencialmente a partir del rezago (lag) $s = 2$, y esto indica que $z_t = \log(y_t^2)$ puede ser modelada mediante un modelo AR(1).

En la práctica obtenemos valores de α_1 próximos de uno, lo que implica la aparición de altas correlaciones para la volatilidad y consecuentes grupos de volatilidades en la serie.

Un MVE general puede ser obtenido si se admite un modelo AR(p) para h_t , esto es

$$y_t = \sigma_t \varepsilon_t, \quad (32)$$

$$\sigma_t = e^{\frac{h_t}{2}}, \quad (33)$$

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) h_t = \alpha_0 + \eta_t, \quad (34)$$

donde el operador de rezago se define como $B^j h_t = h_{t-j}$, los supuestos sobre las innovaciones ε_t y η_t son los mismos que se hicieron anteriormente, pero ahora suponemos que las raíces del polinomio $(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)$ están fuera del círculo unitario.

Los MVE se han ampliado para incluir el hecho de que la volatilidad tiene memoria larga, en el sentido de que la función de autocorrelación de $z_t = \log(y_t^2)$ decae lentamente, aunque como vimos en este caso, los y_t no tienen correlación serial.

6. Estimación de los MVE

Los modelos MVE son difíciles de estimar. Podemos usar el enfoque propuesto por Durbin y Koopman (1997a, 1997b, 2000, 2001, 2012) que consiste en utilizar un procedimiento de cuasi máxima verosimilitud por medio del filtro y suavizador de Kalman. En este caso, el modelo definido en las ecuaciones (4), (5) y (6) puede ser reexpresado de la forma

$$y_t = \sigma \varepsilon_t e^{\frac{h_t}{2}}, \quad (35)$$

$$\sigma_t = \sigma e^{\frac{h_t}{2}}, \quad (36)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (37)$$

en donde $\sigma = \exp(\alpha_0/2)$ es un factor de escala, α_1 es un parámetro, y η_t es un término de disturbio el que en el modelo más simple es no correlacionado con ε_t . Revisiones bibliográficas de este modelo fueron realizadas por Shephard (1996, 2005) y Ghysels et al. (1996). Este MVE tiene dos atractivos principales. El primero es que el mismo es un análogo natural (Euler) en tiempo discreto del modelo en tiempo continuo usado en trabajos sobre precios de opciones, tal como el de Hull y White (1987). El segundo es que sus propiedades estadísticas son fáciles de de-

terminar. La desventaja con respecto a los modelos de varianza condicional del tipo GARCH es que la estimación basada en la verosimilitud puede solamente ser realizada mediante técnicas intensivas de computación tales como las descritas en Kim et al. (1998) y Sandmann y Koopman (1998). Sin embargo un método de cuasi máxima verosimilitud es relativamente fácil de aplicar y es usualmente razonablemente eficiente. El método se basa en escribir (35), (36) y (37) en la siguiente forma equivalente

$$\log(y_t^2) = \kappa + h_t + \xi_t, \quad (38)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (39)$$

donde $\xi_t = \log(\varepsilon_t^2) - E\{\log(\varepsilon_t^2)\}$ y $\kappa = \log(\sigma^2) + E\{\log(\varepsilon_t^2)\}$.

Las ecuaciones (38) y (39) se encuentran expresadas en la forma de espacio de estado, como se lo puede ver en Abril y Abril (2018). La fórmula (38) recibe el nombre de *ecuación de observación* o *ecuación de medición* y la fórmula (39) recibe el nombre de *ecuación de estado* o *ecuación de transición*. El proceso de estimación se realiza, en consecuencia, utilizando el filtro y suavizador de Kalman de la misma manera que los desarrollado en la referencia anterior.

Aquí es importante hacer algunas observaciones:

1. Cuando α_1 en (39) es próximo a 1, el ajuste de un MVE es similar al de un modelo GARCH(1, 1) con la suma de sus coeficientes próxima a 1.
2. Cuando $\alpha_1 = 1$ en (39), h_t es un camino aleatorio y el ajuste de un MVE es similar al de un modelo IGARCH(1, 1).
3. Cuando algunas observaciones son iguales a cero, lo que puede ocurrir en la práctica, no se puede hacer la transformación logarítmica especificada en (38). Una forma de evitar este problema es restando la media general de la serie y_t de cada una de las observaciones y tomando a este resultado como la serie a trabajar; o sea tomando como serie de trabajo a

$$y_t - \bar{y}, \quad t = 1, \dots, n, \quad (40)$$

donde $\bar{y} = n^{-1} \sum_{t=1}^n y_t$. Otra solución, sugerida por Wayne Fuller y analizada por Breidt y Carriquiry (1996), es hacer la siguiente transformación basada en una expansión de Taylor

$$\log(y_t^2) = \log(y_t^2 + cS_y^2) - \frac{cS_y^2}{y_t^2 + cS_y^2}, \quad t = 1, \dots, n, \quad (41)$$

donde S_y^2 es la varianza muestral de la serie y_t y c es un número pequeño. Versiones anteriores a la 8.3 del programa STAMP desarrollado por Koopman et al. incorporaban la transformación definida en (41) con $c = 0,02$ como una operación previamente especificada que podía ser utilizada si se la necesitaba. A partir de la versión 8.3 de ese programa (ver Koopman et al., 2010) se dejó de lado esa transformación como un elemento previamente especificado, pudiéndose realizar esa u otras transformaciones, como la definida en (40), mediante el uso de la calculadora o de la disponibilidad de Algebra dentro de ese programa de acuerdo a los requerimientos y necesidades del usuario.

Como se mostró en Harvey et al. (1994), la forma de espacio de estado dada por las ecuaciones (38) y (39) brinda las bases para la estimación por cuasi máxima verosimilitud vía el filtro y suavizador de Kalman y también permite construir estimaciones suavizadas del componente h_t de la varianza y realizar predicciones. Uno de los atractivos del enfoque de cuasi máxima verosimilitud es que puede ser aplicado sin un supuesto sobre una distribución particular para ε_t . Otro de los atractivos en utilizar un procedimiento de cuasi máxima verosimilitud por medio del filtro y suavizador de Kalman para estimar los MVE es que puede llevarse a cabo directamente usando paquetes estándares de computación tal como el STAMP de Koopman et al. (2010). Esto es una gran ventaja comparado con los métodos basados en simulaciones que requieren mayor trabajo.

Shephard y Pitt (1997) propusieron el uso de muestreo ponderado (“importance sampling”) para estimar la función de verosimilitud en el caso no Gaussiano.

Como el MVE es un modelo jerárquico, Jaquier, Polson y Rossi (1994) propusieron un análisis bayesiano del mismo. Véase también Shephard y Pitt (1997) y Kim et al. (1998). Una reseña del problema de estimación del MVE está hecha por Motta (2001).

7. Serie con errores que siguen un MVE con componentes estructurales

El MVE básico dado en (35), (36) y (37) captura solamente las características salientes de la heterocedasticidad condicional cambiante en una serie de tiempo. En algunos casos el modelo es más preciso cuando la serie y_t es modelada incorporando componentes estructurales, variables explicativas y otras características que expliquen su comportamiento, todo esto hecho mediante un esquema de espacio de estado con errores que siguen un MVE con componentes estructurales, por ejemplo con estacionalidad. Basándonos en lo expuesto anteriormente, para una serie y_t univariada, lo antes dicho puede ser formulado como

$$y_t = \mathbf{Z}_t \beta_t + \nu_t, \quad (42)$$

$$\beta_t = \mathbf{T}_t \beta_{t-1} + \mathbf{R}_t \omega_t, \quad \omega_t \sim N(\mathbf{0}, \mathbf{Q}_t), \quad (43)$$

con

$$\nu_t = \sigma e^{\frac{\gamma_t^{(v)}}{2}} e^{\frac{h_t}{2}} \varepsilon_t, \quad (44)$$

$$\sigma_t = \sigma e^{\frac{\gamma_t^{(v)}}{2}} e^{\frac{h_t}{2}}, \quad (45)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (46)$$

donde β_t es el vector de estado de orden $m \times 1$, ω_t son disturbios serialmente independientes, independientes entre sí e independientes de ν_t en todo momento de tiempo. Las matrices de sistema \mathbf{Z}_t , \mathbf{T}_t , \mathbf{R}_t y \mathbf{Q}_t tienen dimensiones $1 \times m$, $m \times m$, $m \times m$ y $m \times m$ respectivamente, y si existen en ellas elementos desconocidos, son incorporados al vector ψ de hiperparámetros el cual es estimado por máxima verosimilitud. A (42) se la denomina ecuación de medición o ecuación de observación y a (43)

ecuación de transición o ecuación de estado. Las ecuaciones (42) y (43) definen un modelo de espacio de estado con todas las características y propiedades de los mismos presentadas en Abril y Abril (2018). En efecto, allí se puede tener tendencia, estacionariedad, ciclos, variables explicativas y otras características importantes que expliquen el comportamiento del proceso $\{y_t\}$. Las ecuaciones (44), (45) y (46) definen un MVE con componentes estructurales (estacionalidad en este caso) para los errores del modelo de espacio de estado dado antes, donde ε_t es una serie estacionaria con media igual a cero y varianza uno, y η_t es una serie estacionaria, Gaussiana, con media cero, varianza σ_η^2 e independiente de ε_t en todo momento de tiempo.

Las estimaciones de los modelos de espacio de estado se las realiza usando paquetes estándares de computación tal como el STAMP de Koopman et al. (2010), que es el utilizado en este trabajo.

Un camino para capturar una estructura estacional determinística es mediante un conjunto de funciones senos y cosenos. Permitiendo que ellas sean estocásticas nos lleva a la forma trigonométrica de la estacionalidad estocástica

$$\gamma_t^{(v)} = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t}^{(v)}, \quad (47)$$

donde s es la longitud del período estacional y $[x]$ significa la parte entera de x , $\gamma_{j,t}^{(v)}$ es generado por

$$\begin{bmatrix} \gamma_{j,t}^{(v)} \\ \gamma_{j,t}^{*(v)} \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1}^{(v)} \\ \gamma_{j,t-1}^{*(v)} \end{bmatrix} + \begin{bmatrix} \omega_{j,t}^{(v)} \\ \omega_{j,t}^{*(v)} \end{bmatrix}, \quad (48)$$

$j = 1, \dots, \lfloor \frac{s}{2} \rfloor$

donde $\lambda_j = 2\pi j/s$ es la frecuencia en radianes, $\omega_t^{(v)}$ y $\omega_t^{*(v)}$ son dos disturbios del tipo ruido blanco mutuamente no correlacionados con media cero y varianza común $\sigma_{\omega^{(v)}}^2$ para $t = 1, \dots, n$. Para s par $\lfloor s/2 \rfloor = s/2$, mientras que para s impar $\lfloor s/2 \rfloor = (s-1)/2$. Para s par, el componente para $j = s/2$ colapsa en

$$\gamma_{j,t}^{(v)} = \gamma_{j,t-1}^{(v)} \cos \lambda_j + \omega_{j,t}^{(v)}. \quad (49)$$

El componente estacional de (44) y (45) es el dado en (47) y (48).

Sin los términos de disturbio este modelo estacional dará la misma estructura determinística que el modelo estacional con variables ficticias. De cualquier manera, es un mejor modelo de estacionalidad estocástica porque permite que el componente estacional evoluciones con mayor suavidad. Puede mostrarse que la suma de las estacionalidades sobre el año anterior sigue un modelo MA($s-2$) en lugar de ser un ruido blanco.

En este caso, la volatilidad total es igual a

$$\sigma_t^2 = \sigma^2 e^{\gamma_t^{(v)}} e^{h_t}. \quad (50)$$

Por otra parte, la volatilidad básica es e^{h_t} y la volatilidad estacional es $e^{\gamma_t^{(v)}}$.

El tratamiento práctico en estos casos es como sigue: dada una serie $\{y_t\}$ se identifican los componentes lineales que pueden

explicar el comportamiento de su media, incluyendo las variables explicativas que pudieran corresponder, de tal manera de definir explícitamente el modelo de las ecuaciones (42) y (43). Con ello se realiza primero el filtrado y luego el suavizado de Kalman, obteniéndose el estimador suavizado $\hat{\beta}_t$ del vector de estado β_t . Este estimador permite calcular los residuos suavizados como

$$\hat{\nu}_t = y_t - \mathbf{Z}_t \hat{\beta}_t, \quad t = 1, \dots, n. \quad (51)$$

Estos residuos suavizados estiman a los disturbios ν_t . Los valores de $\hat{\nu}_t$ sirven de base para testar la hipótesis nula de falta de correlación serial de ν_t . Si se acepta esta hipótesis se podría decir que el modelo dado en las ecuaciones (42) y (43) fue adecuadamente identificado, definido y estimado. Por otra parte, si $\log(\hat{\nu}_t^2)$ muestra correlación serial con picos en s , se puede decir que los errores ν_t siguen un MVE de la forma dada en (44), (45) y (46) junto con (47) y (48). En consecuencia, se toma a $\hat{\nu}_t$ como la serie observada y se estima el siguiente modelo de espacio de estado

$$\log(\hat{\nu}_t^2) = \kappa + h_t + \gamma_t^{(v)} + \xi_t, \quad (52)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (53)$$

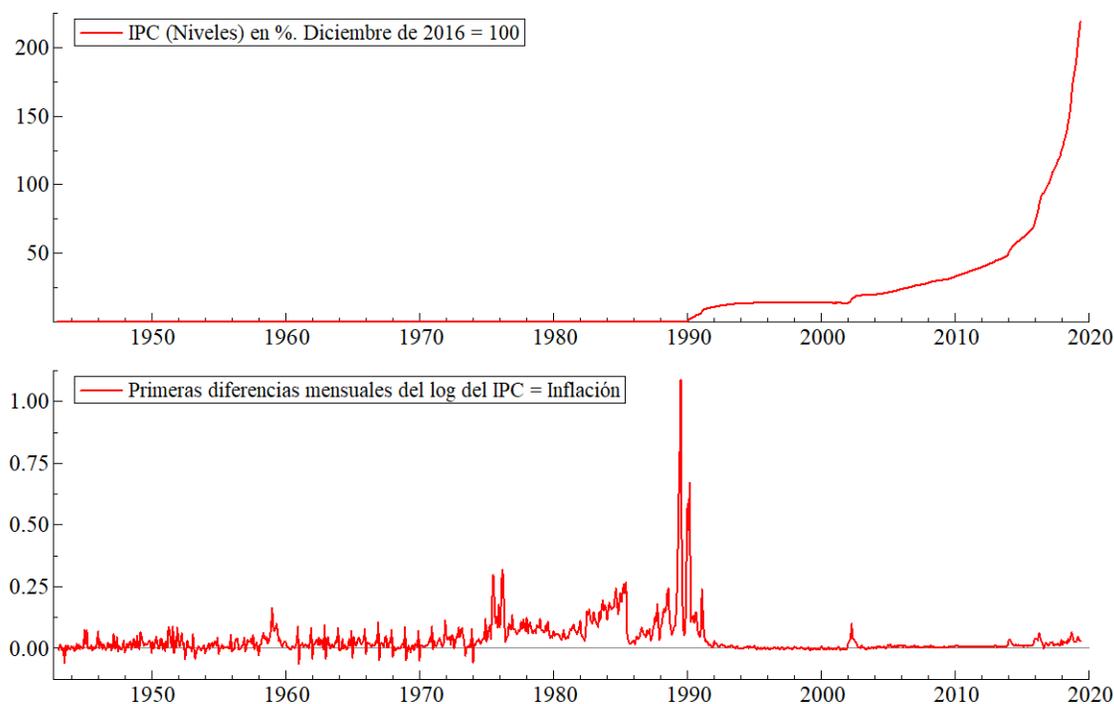
donde $\xi_t = \log(\varepsilon_t^2) - E\{\log(\varepsilon_t^2)\}$, $\kappa = \log(\sigma^2) + E\{\log(\varepsilon_t^2)\}$, $\gamma_t^{(v)}$ está definido en (47) y (48), ε_t es una serie estacionaria con media igual a cero y varianza uno, y η_t es una serie estacionaria, Gaussiana, con media cero, varianza σ_η^2 e independiente de ε_t en todo momento de tiempo. El proceso de estimación de (52), (53), (47) y (48) se realiza, por ello, utilizando el filtro y suavizador de Kalman de la misma manera que los desarrollamos anteriormente.

Como se mostró en Harvey et al. (1994), la forma de espacio de estado dada por las ecuaciones (52) y (53) brinda las bases para la estimación por cuasi máxima verosimilitud vía el filtro y suavizador de Kalman y también permite construir estimaciones suavizadas del componente h_t de la varianza y realizar predicciones. Uno de los atractivos del enfoque de cuasi máxima verosimilitud es que puede ser aplicado sin un supuesto sobre una distribución particular para ε_t . Otro de los atractivos en utilizar un procedimiento de cuasi máxima verosimilitud por medio del filtro y suavizador de Kalman para estimar los MVE es que puede llevarse a cabo directamente usando paquetes estándares de computación tal como el STAMP de Koopman et al. (2010). Esto es una gran ventaja comparado con los métodos basados en simulaciones que requieren mayor trabajo.

8. Análisis preliminar de la serie bajo estudio

Los datos que manejamos pertenecen a un campo muy especial dentro de la ciencia estadística que es el de las series de tiempo. La característica común de todos los registros que pertenecen al dominio de las series de tiempo es que los mismos están influenciados, aunque sean parcialmente, por componentes no observables que contienen variaciones aleatorias, es decir, la ocurrencia de sucesos no planificados.

Figura 1. Niveles mensuales del IPC en el Gran Buenos Aires desde enero de 1943 hasta mayo de 2019



Nota. Parte superior: Niveles mensuales del Índice de Precios al Consumidor desde enero de 1943 hasta mayo de 2019. Parte inferior: Primeras diferencias del logaritmo del nivel del Índice de Precios al Consumidor desde enero de 1943 hasta mayo de 2019.

La serie que estudiaremos está constituida por los niveles mensuales del Índice de Precios al Consumidor (o *IPC* según sus siglas) en el gran Buenos Aires desde enero de 1943 hasta mayo de 2019. Estos datos fueron obtenidos del Instituto Nacional de Estadística y Censos (INDEC) perteneciente al Ministerio de Hacienda y Finanzas Públicas de la Nación.

Es necesario destacar que si bien este es un período muy largo para analizar; en donde se registraron numerosos cambios como ser de base, canasta, e intervenciones en el mismo INDEC, es posible realizar un estudio muy interesante en donde se aprecian las principales características de la serie.

En primer lugar, procedemos a graficar la misma. Dentro del estudio de una serie, los métodos gráficos son una excelente manera de comenzar una investigación para luego poder sumergirnos en un estudio pormenorizado de la temática bajo consideración. Entre las funciones que cumplen las tablas y los gráficos que presentaremos se encuentran las siguientes:

1. Hacen más visibles los datos bajo estudio, los sistematizan y sintetizan.
2. Ponen de manifiesto sus variaciones y su evolución histórica o espacial.
3. Pueden evidenciar las relaciones entre los diversos elementos de un sistema o de un proceso y mostrar indicios de la futura correlación entre dos o más variables.

Además, la aplicación de estos métodos sugiere nuevas hipótesis de investigación y permite la posterior implementación de

modelos estadísticos que van desde los más simples hasta aquellos que son mucho más refinados, lográndose así un mejor análisis de los datos y de las fluctuaciones de los mismos a lo largo del tiempo.

En el apartado superior de la Figura 1 se muestran los niveles mensuales del Índice de Precios al Consumidor, a su vez, en el apartado inferior de la misma figura podemos ver las primeras diferencias del logaritmo del nivel mensual del IPC. Esto es lo que popularmente se conoce con el nombre de inflación y será la serie con la cual realizaremos nuestro trabajo.

Realizando una detenida inspección de este último apartado, podemos ver que existen períodos donde la volatilidad es baja y se puede confundir con la presencia de estacionalidad en especial dentro del período que va desde enero de 1943 hasta finales del año 1974. Luego se inicia un claro período de volatilidad medianamente importante que va hasta finales del año 1977, el cual se vuelve a repetir con similares características entre 1983 y finales de 1985. Posteriormente, entre 1987 y finales de 1992 tenemos un período de alta volatilidad. Desde fines del año 2001, la volatilidad en la serie bajo estudio es casi nula, a partir de ese momento se registra una escasa presencia de ella y la misma casi desaparece a partir del año 2009. Luego, a partir del año 2015 se observa un aumento de la misma.

En la Figura 2 se muestra en la parte superior izquierda la serie de inflación para el período comprendido entre enero de 1943 y mayo de 2019, en la parte superior derecha se muestra la autocorrelación estimada para la serie de inflación bajo estudio, en la parte inferior izquierda se puede ver la función de autocorre-

lación parcial estimada, y por último, en la parte inferior derecha tenemos la función de densidad estimada la cual está representada por una línea de color rojo, comparada con la función de densidad normal la cual se encuentra representada por una línea verde.

Del estudio de esta última Figura vemos que la serie no es estacionaria presentando un nivel estocástico, tiene componentes estacionales, autorregresivos, cambios de nivel y observaciones atípicas (outliers) y su distribución es diferente a la de una variable normal, por lo que es posible que debamos realizar sus correspondientes estimaciones haciendo uso del método de cuasi máxima verosimilitud.

9. Modelado de la serie de inflación bajo estudio

Se probaron diferentes modelos alternativos para la serie y_t de inflación en la Argentina para el período comprendido entre enero de 1943 y mayo de 2019, y se determinó que la misma debe ser modelada incorporando componentes estructurales, variables explicativas y otras características que expliquen su comportamiento, todo esto hecho mediante un esquema de espacio de estado con errores que siguen un MVE con componentes estructurales. En ese sentido, el modelo estimado es

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{13} \theta_j y_{t-j} + \theta_{14} x_{1t} + \theta_{15} x_{2t} + \nu_t, \quad (54)$$

$$\mu_t = \mu_{t-1} + \varphi_t, \quad (55)$$

$$\gamma_t = \sum_{j=1}^6 \gamma_{j,t}, \quad (56)$$

donde $\gamma_{j,t}$ es generado por

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \text{sen} \lambda_j \\ -\text{sen} \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad j = 1, \dots, 6. \quad (57)$$

es una variable que captura una observación atípica que sucede en Agosto de 1989, o sea que

$$x_{1t} = \begin{cases} 1, & \text{cuando } t \text{ es igual a Agosto de 1989,} \\ 0, & \text{de otra manera,} \end{cases} \quad (58)$$

x_{2t} es una variable que captura un cambio de nivel que sucede a partir de mayo de 1990, o sea que

$$x_{2t} = \begin{cases} 0, & \text{cuando } t \text{ es menor que mayo de 1990,} \\ 1, & \text{cuando } t \text{ es mayor o igual que mayo de 1990,} \end{cases} \quad (59)$$

la ecuación (55) representa un nivel estocástico y es lo que se denominó anteriormente la inflación núcleo. Por otra parte, en

(54), la inflación tendencia es

$$\mu_t + \gamma_t + \sum_{j=1}^{13} \theta_j y_{t-j} + \theta_{14} x_{1t} + \theta_{15} x_{2t}. \quad (60)$$

Las ecuaciones (56) y (57) corresponden a un componente estacional estocástico, φ_t de (55) es normal con media cero y varianza σ_φ^2 , ω_t y ω_t^* son dos disturbios del tipo ruido blanco mutuamente no correlacionados con media cero y varianza común σ_ω^2 , todo para $t = 1, \dots, n$. Los errores ν_t , φ_t , $\omega_{j,t}$ y $\omega_{j,t}^*$ son independientes entre sí en todo momento de tiempo.

Para la volatilidad, tenemos que ν_t satisface

$$\nu_t = \sigma e^{\frac{\gamma_t^{(v)}}{2}} e^{\frac{h_t}{2}} \varepsilon_t, \quad (61)$$

$$\sigma_t = \sigma e^{\frac{\gamma_t^{(v)}}{2}} e^{\frac{h_t}{2}}, \quad (62)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (63)$$

Las ecuaciones (61), (62) y (63) definen un MVE con componentes estructurales (estacionalidad en este caso) para los errores del modelo de espacio de estado dado en las ecuaciones (54), (55), (56) y (57), donde ε_t es una serie estacionaria (usualmente Gaussiana) con media igual a cero y varianza uno, y η_t es una serie estacionaria, Gaussiana, con media cero, varianza σ_η^2 e independiente de ε_t en todo momento de tiempo. Además

$$\gamma_t^{(v)} = \sum_{j=1}^6 \gamma_{j,t}^{(v)},$$

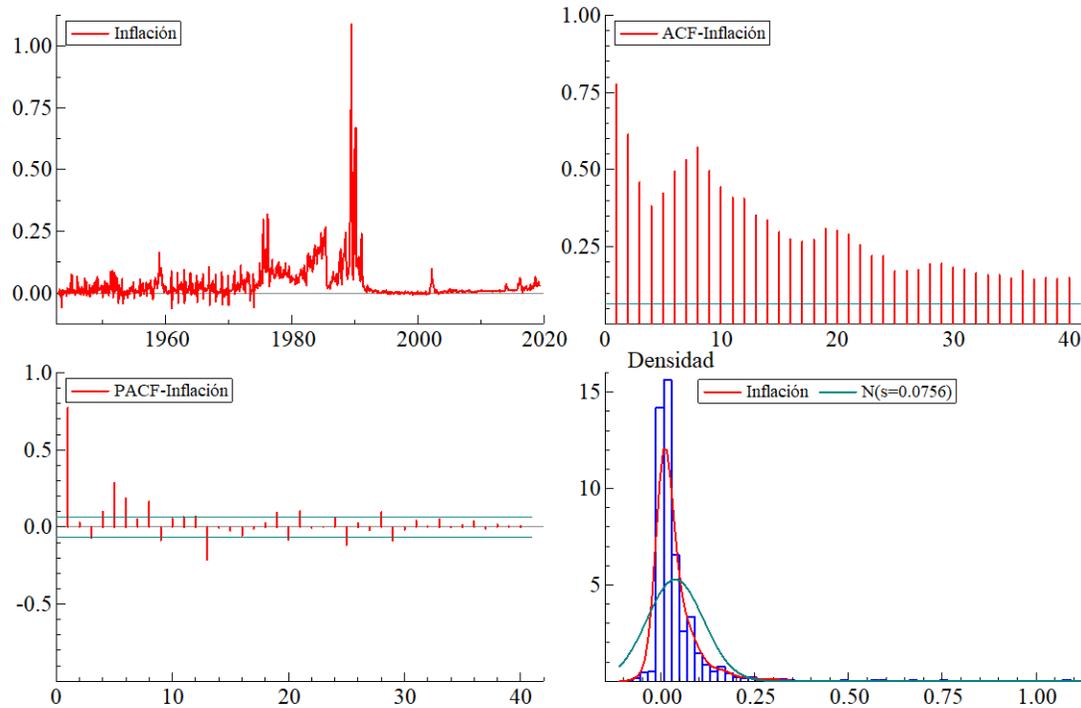
donde $\gamma_{j,t}^{(v)}$ es generado por

$$\begin{bmatrix} \gamma_{j,t}^{(v)} \\ \gamma_{j,t}^{*(v)} \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \text{sen} \lambda_j \\ -\text{sen} \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1}^{(v)} \\ \gamma_{j,t-1}^{*(v)} \end{bmatrix} + \begin{bmatrix} \omega_{j,t}^{(v)} \\ \omega_{j,t}^{*(v)} \end{bmatrix}, \quad j = 1, \dots, 6.$$

donde $\lambda_j = 2\pi j/12$ es la frecuencia en radianes, $\omega_t^{(v)}$ y $\omega_t^{*(v)}$ son dos disturbios del tipo ruido blanco mutuamente no correlacionados con media cero y varianza común $\sigma_{\omega^{(v)}}^2$ para $t = 1, \dots, n$. Al tomar logaritmos de (61) se lo pone de una forma similar a (52) y se lo estima como un modelo de espacio de estado, de igual manera a lo señalado anteriormente.

Se inicia el estudio estimando el modelo dado en las ecuaciones (54) a la (59). Lo primero que se observa es que el estadístico de normalidad de Doornik-Hansen, cuya distribución bajo la hipótesis nula de normalidad de los errores es una χ_2^2 , da un valor de 1399,2, el cual es muy alto y lleva a rechazar la hipótesis nula. Esto no es de extrañar ya que no hay Gaussianidad (ver Figura 2) y se tienen volatilidad. El test $H(293)$ de heterocedasticidad, que se distribuye como una $F(293, 293)$ bajo la hipótesis nula de presencia de heterocedasticidad en la serie, da un valor de 0,11278, lo que lleva a aceptar la hipótesis de existencia de heterocedasticidad. Por último, el estadístico q de Box-Ljung que en este caso se distribuye como una χ_{28}^2 da un valor igual a 74,908, lo que conduce a rechazar la hipótesis de falta de correlación serial en los residuos.

Figura 2. Fnlación para el período comprendido entre enero de 1943 y mayo de 2019



Nota. Parte superior izquierda: Serie de inflación para el período enero de 1943-mayo de 2019. Parte superior derecha: Función de autocorrelación de esta serie de inflación. Parte inferior izquierda: Función de autocorrelación parcial de la misma serie bajo estudio. Parte inferior derecha: Función de densidad estimada de la serie bajo estudio comparada con la densidad normal (línea verde)

Por otra parte, $\hat{\sigma}_y^2 = 0,00146135$, $\hat{\sigma}_\varphi^2 = 0,000000614929$, $\hat{\sigma}_\omega^2 = 0,000000328550$ y $\hat{\sigma}_\nu^2 = 0,00128796$.

En el Cuadro 1 se muestran los valores estimados de los parámetros de la ecuación (54), los respectivos desvíos estándares, los valores del estadístico t para testar la hipótesis nula que el respectivo parámetro es igual a cero y las probabilidades en las colas correspondientes a ese test de hipótesis. Si esas probabilidades son menores que 0,05 se rechaza la respectiva hipótesis con ese nivel de significación. Como vemos en estas cifras, todos los coeficientes son significativamente diferentes de cero, lo que da una idea de que el modelo propuesto es el adecuado.

Con respecto a los otros componentes del modelo (54), al ser ellos estocásticos sus valores no pueden ser puestos en un cuadro. En ese caso es mejor verlos gráficamente. En la Figura 3 se muestran los componentes estimados de la ecuación (54). Así, en la parte superior izquierda de la Figura se tiene a la serie de inflación (trazo negro) y los componentes definidos en la ecuación (60) (trazo rojo) lo que estima la inflación tendencia, en la parte superior derecha está la estimación del componente nivel definido en la ecuación (55) lo que estima la inflación núcleo o subyacente, en la parte inferior izquierda está la estimación del componente estacional y en la parte inferior derecha, la estimación del componente irregular, o sea $\hat{\nu}_t$. Este último componente estimado es el que se usará para estimar la volatilidad vía los modelos definidos en (61), (62) y (63)

Para estudiar la volatilidad, debido a que los errores ν_t del modelo (54) no son observables, se los estima mediante los re-

siduos del mismo luego de las estimaciones correspondientes y se los denota como $\hat{\nu}_t$. Esta última es la serie con la cual se se trabaja para estimar todo lo relativo a la volatilidad. En la Figura 4 se muestran en la parte superior izquierda los residuos estandarizados del modelo dado en (54), (55) y (56) luego de haber sido estimado, en la parte superior derecha está la función de autocorrelación estimada, luego, en la parte inferior izquierda está la densidad espectral estimada y finalmente en la parte inferior derecha está la función de densidad estimada la cual está representada por una línea de color rojo, comparada con la función de densidad normal la cual se encuentra representada por una línea verde. En esta Figura vemos que los residuos no difieren significativamente de una serie sin correlación serial y aproximadamente normal. Así por ejemplo, las autocorrelaciones estimadas están prácticamente dentro de la banda de confianza, lo que implica que los respectivos parámetros no difieren de cero, y las oscilaciones de la densidad espectral son poco importantes en comparación con su escala. También los estadísticos respectivos avalan estas aseveraciones.

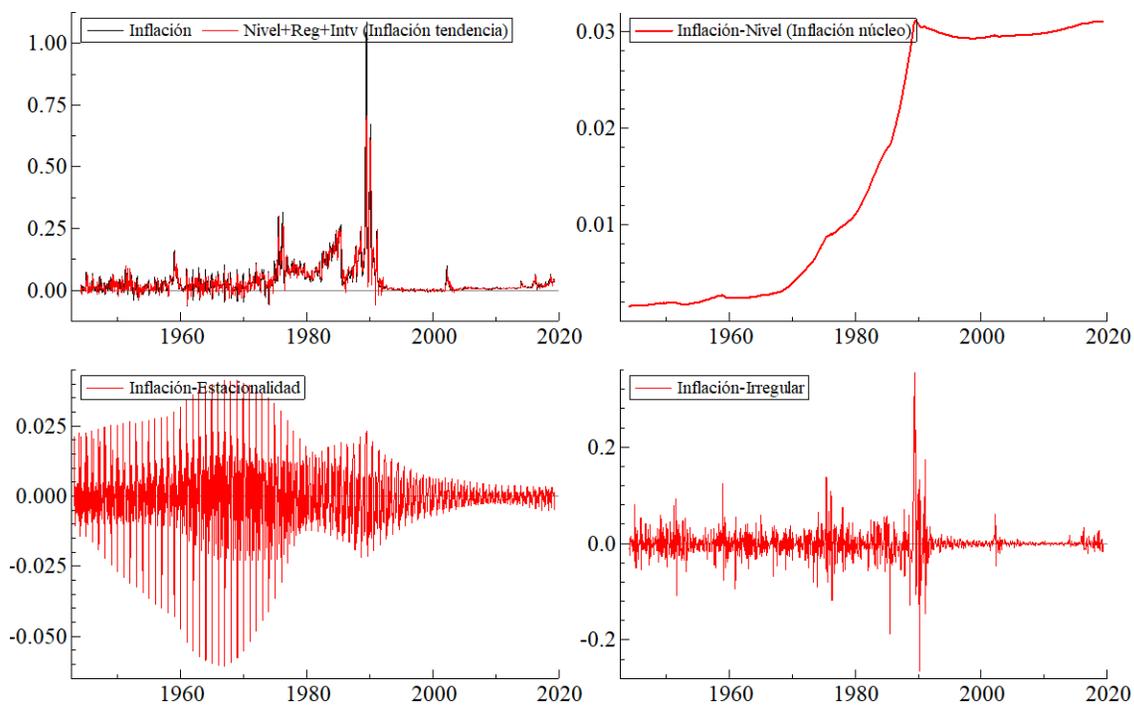
Para aplicar el esquema de espacio de estado y poder hacer las estimaciones respectivas de la volatilidad se debe elevar al cuadrado la serie de residuos y luego tomar logaritmos. Con ello llegamos al modelo dado en (52) y (53). En la Figura 5 se muestran en la parte superior izquierda el logaritmo de los residuos al cuadrado del modelo dado en (54), (55) y (56) luego de haber sido estimado, en la parte superior derecha está la función de autocorrelación estimada, luego, en la parte inferior izquierda está

Cuadro 1. Inflación para el período comprendido entre enero de 1943 y mayo de 2019

Estimador	Valor estimado	Desvío estándar	Valor de t	Probabilidad
$\hat{\theta}_1$	0,92686	0,02908	31,87612	0,00000
$\hat{\theta}_3$	-0,08907	0,03085	-2,88752	0,00398
$\hat{\theta}_4$	-0,15379	0,03636	-4,22971	0,00003
$\hat{\theta}_5$	0,19953	0,03069	6,50182	0,00000
$\hat{\theta}_8$	0,24425	0,03020	8,08768	0,00000
$\hat{\theta}_9$	-0,24926	0,03114	-8,00523	0,00000
$\hat{\theta}_{12}$	0,13866	0,03043	4,55652	0,00001
$\hat{\theta}_{13}$	-0,12824	0,02978	-4,30583	0,00002
$\hat{\theta}_{14}$	-0,65083	0,04613	-14,10890	0,00000
$\hat{\theta}_{15}$	-0,02874	0,00851	-3,37521	0,00077

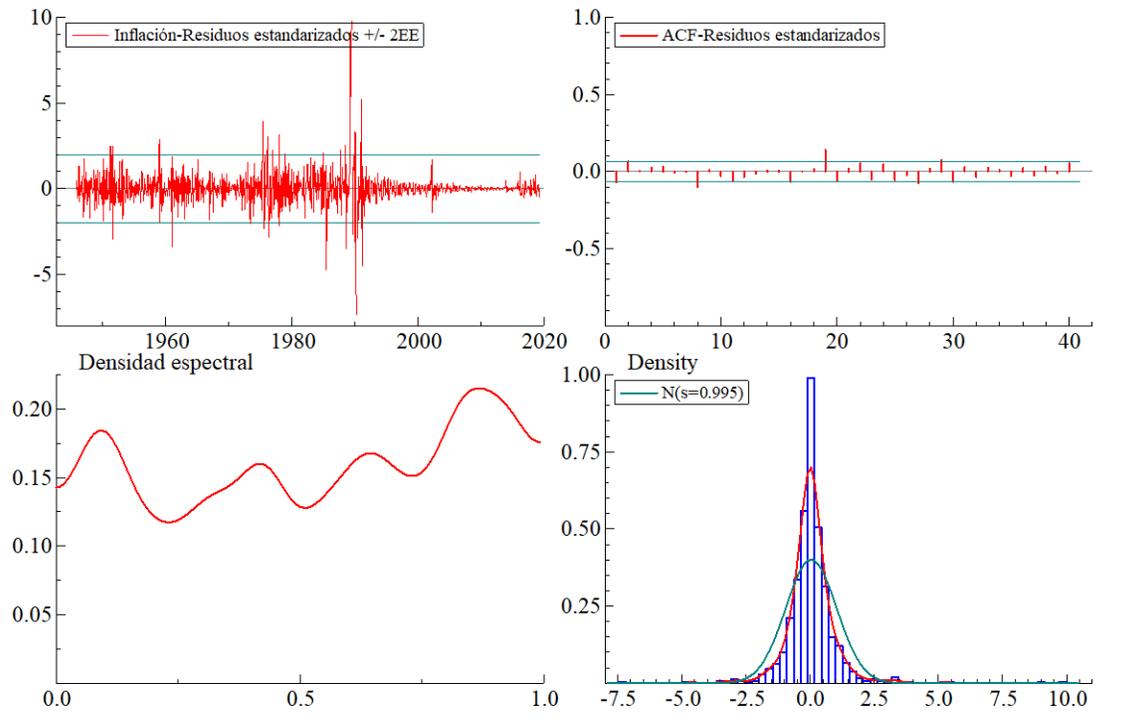
Nota. Valores estimados, desvíos estándares, estadísticos t y sus probabilidades en las colas para el modelo de la serie de inflación para el período desde enero de 1943 a mayo de 2019

Figura 3. Componentes de la serie de inflación entre enero de 1943 y mayo de 2019



Nota. Componentes del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina. Período enero de 1943 a mayo de 2019

Figura 4. Residuos del modelo definido de infalcción entre enero de 1943 y mayo de 2019



Nota. Residuos del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina. Período enero de 1943 a mayo de 2019, su función de autocorrelación estimada, el espectro estimado y la distribución de los mismos

la función de autocorrelación parcial estimada y finalmente en la parte inferior derecha está la función de densidad estimada la cual está representada por una línea de color rojo, comparada con la función de densidad normal la cual se encuentra representada por una línea verde. Allí se ve claramente que el modelo dado en (52) y (53) es el adecuado.

Al estimar el modelo dado en (52) y (53) encontramos que $\hat{\sigma}_{\log(\hat{v}_t^2)}^2 = 5,40649$, $\hat{\sigma}_{\xi}^2 = 4,90710$, $\hat{\sigma}_{\eta}^2 = 3,84443$, el nivel κ es fijo, por lo tanto su varianza es cero, y $\hat{\kappa} = -2,85453$. Se encontró que $\hat{\sigma}_{\omega^{(v)}}^2 = 0,00$, lo que implica que el componente estacional no es estocástico. El estadístico para testar la hipótesis nula de que la estacionalidad no es significativa se distribuye como una χ_{11}^2 y en este caso da un valor de 30,63248, lo que nos lleva a rechazar la hipótesis anterior. Por lo tanto, el componente estacional es significativo y no estocástico. Los valores estimados de los coeficientes estacionales se dan en el Cuadro 2, donde el subíndice indica el respectivo mes del año.

El estadístico de normalidad da un valor de 144,74 el cual es alto. Esto es inevitable porque el modelo transformado (52) y (53) no es Gaussiano. Esto no debe preocuparnos. Por otra parte el estimados de α_1 es $\hat{\alpha}_1 = 0,99133$, el cual es alto como era de esperar.

En la Figura 4 se muestran en la parte superior izquierda el logaritmo de los residuos al cuadrado del modelo dado en (54), (55) y (56) luego de haber sido estimado (línea negra) y

el nivel estimado (línea roja), en la parte superior derecha está el componente estacional estimado cuyos valores están dados en el Cuadro 2, luego, en la parte inferior izquierda está el componente AR(1) estimado cuya ecuación está dada en (53) y finalmente en la parte inferior derecha están las estimaciones del componente irregular de (52).

De (62) se desprende que la volatilidad total es igual a

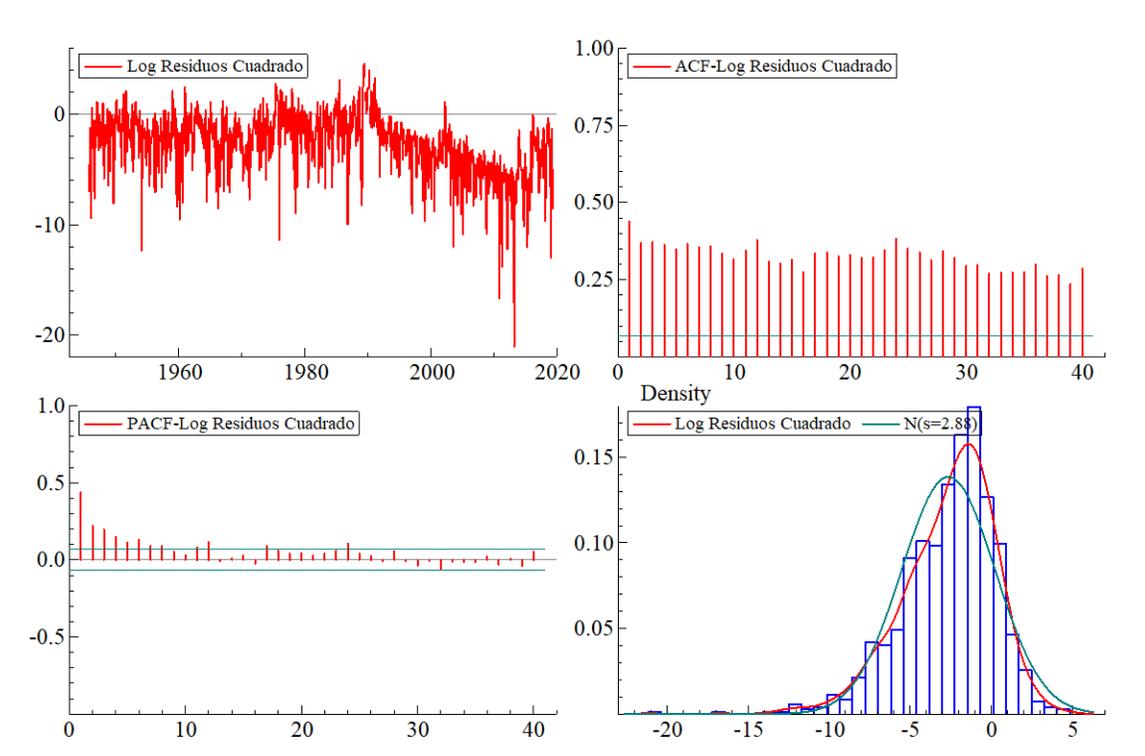
$$\sigma_t^2 = \sigma^2 e^{\gamma_t^{(v)}} e^{h_t}, \quad (64)$$

la cual tiene tres componentes multiplicativos que son: una constante de escala σ^2 , la volatilidad básica e^{h_t} y la volatilidad estacional $e^{\gamma_t^{(v)}}$. En este caso estudiado, el componente estacional resultó ser no estocástico. De estos tres componentes, evidentemente la volatilidad básica es el de mayor importancia ya que los otros dos son o bien una constante multiplicativa o bien un componente periódico no estocástico con período 12. La volatilidad básica se la estima como $e^{\hat{h}_t}$ donde \hat{h}_t es

$$\hat{h}_t = \hat{\alpha}_1 \hat{h}_{t-1},$$

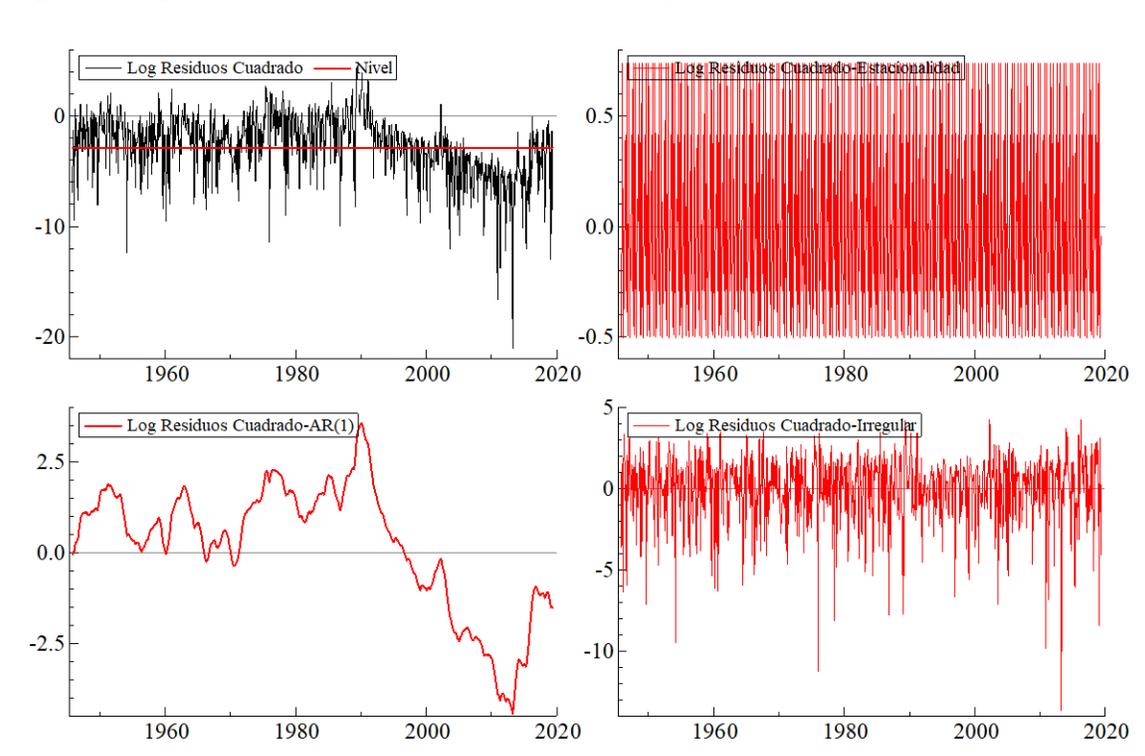
con $\hat{\alpha}_1 = 0,99133$, que es el valor estimado de α_1 . La volatilidad estacional se la estima como $e^{\hat{\gamma}_t^{(v)}}$, donde los valores de $\hat{\gamma}_t^{(v)}$, que corresponden a un componente periódico no estocástico con período 12, están dados en el Cuadro 2. Para estimar a σ^2 , se calcula la serie estimada del componente irregular de (54) corregida

Figura 5. Log de los residuos al cuadrado del modelo definido entre enero de 1943 y mayo de 2019



Nota. Log de los residuos al cuadrado del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina. Período enero de 1943 a mayo de 2019, su función de autocorrelación estimada, su función de autocorrelación parcial estimada y la distribución de los mismos.

Figura 6. Log de los residuos al cuadrado del modelo definido entre enero de 1943 y mayo de 2019



Nota. Log de los residuos al cuadrado del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina. Período enero de 1943 a mayo de 2019, la estimación del nivel, del componente estacional, el componente AR(1) estimado cuya ecuación está dada en (53) y finalmente las estimaciones del componente irregular de (52).

Cuadro 2. Estacionalidad estimada para el modelo de volatilidad estudiado

Estimador	Valor estimado
$\hat{\gamma}_1^{(v)}$	0,74116
$\hat{\gamma}_2^{(v)}$	-0,20224
$\hat{\gamma}_3^{(v)}$	-0,50478
$\hat{\gamma}_4^{(v)}$	-0,11718
$\hat{\gamma}_5^{(v)}$	-0,04209
$\hat{\gamma}_6^{(v)}$	0,17242
$\hat{\gamma}_7^{(v)}$	0,41615
$\hat{\gamma}_8^{(v)}$	-0,29403
$\hat{\gamma}_9^{(v)}$	0,74021
$\hat{\gamma}_{10}^{(v)}$	-0,33793
$\hat{\gamma}_{11}^{(v)}$	-0,49588
$\hat{\gamma}_{12}^{(v)}$	-0,07582

por heterocedasticidad básica y estacional, o sea

$$\tilde{\nu}_t = \hat{\nu}_t \exp \left\{ -\frac{\hat{\gamma}_t^{(v)}}{2} \right\} \exp \left\{ -\frac{\hat{h}_t}{2} \right\}, \quad (65)$$

y luego se computa la varianza $\tilde{\sigma}^2$ de $\tilde{\nu}_t$, la que es una estimación de σ^2 . El valor calculado en nuestro caso es $\tilde{\sigma}^2 = 0,210625$.

La Figura 7 muestra el desvío estándar condicional básico estimado, $\exp\{\hat{h}_t/2\}$, para todo el período estudiado en la parte superior y para el período enero de 2004 a mayo de 2019. Es importante observar que a partir del año 2004 el componente más importante de la volatilidad, la denominada volatilidad básica, comienza a decrecer, teniendo un mínimo en Abril de 2013, luego aumenta muy poco y recién retoma niveles importantes a partir de enero de 2016. Esta situación viene a corroborar el hecho de que desde esa fecha no se puedan tener estimaciones de la volatilidad para esta serie, lo cual está en consonancia con el inicio de un período de falta de confianza en las estadísticas oficiales. Estadísticamente, esto dio lugar a una serie extremadamente suave que no coincide con el resto de la misma ni con la realidad vivida. Todo se revirtió a partir de enero de 2016 cuando el INDEC comenzó a trabajar con total libertad.

A fin de detectar posibles relaciones entre la serie estudiada y la volatilidad, se grafica en la Figura 8 el desvío observa ninguna estructura que los relacione. En la Figura 9 se tiene el gráfico de las mismas variables pero para el período enero de 2004 a mayo de 2019, y tampoco se observa relación alguna. Por lo tanto se puede afirmar que las estimaciones realizadas son adecuadas en este caso.

10. Consideraciones finales

La serie estudiada está constituida por las primeras diferencias del logaritmo del nivel mensual del Índice de Precios al Consumidor (o IPC según sus siglas) en el gran Buenos Aires desde enero de 1943 hasta mayo de 2019. Esto es lo que popularmente se conoce con el nombre de *inflación*. Estos datos fueron obtenidos del Instituto Nacional de Estadística y Censos (INDEC) perteneciente al Ministerio de Hacienda y Finanzas Públicas de la Nación.

En una etapa inicial de este proyecto (ver Abril y Abril, 2017) se tomo esta misma serie, pero para el período enero de 1943 a Diciembre de 2013, y se ajustó un modelo de tipo GARCH que capte las principales características de los datos. Vimos que el mismo toma adecuadamente a la volatilidad de la serie, sin embargo, presenta algunas dificultades a la hora de realizar las predicciones. Con ese análisis hemos podido observar que a partir de aproximadamente Octubre de 2004 la volatilidad es igual a cero. Esta situación corrobora el hecho de que a partir de esa fecha no se puedan tener predicciones de la volatilidad para esta serie, lo cual está en consonancia con el inicio de un período de falta de confianza en las estadísticas oficiales. Efectivamente, a partir de esa fecha, las autoridades dieron información falsa sobre la inflación, resultando una serie con muy poca o nula variabilidad condicional. Estadísticamente, esto dio lugar a una serie extremadamente suave que no coincide con el resto de la misma ni con la realidad vivida.

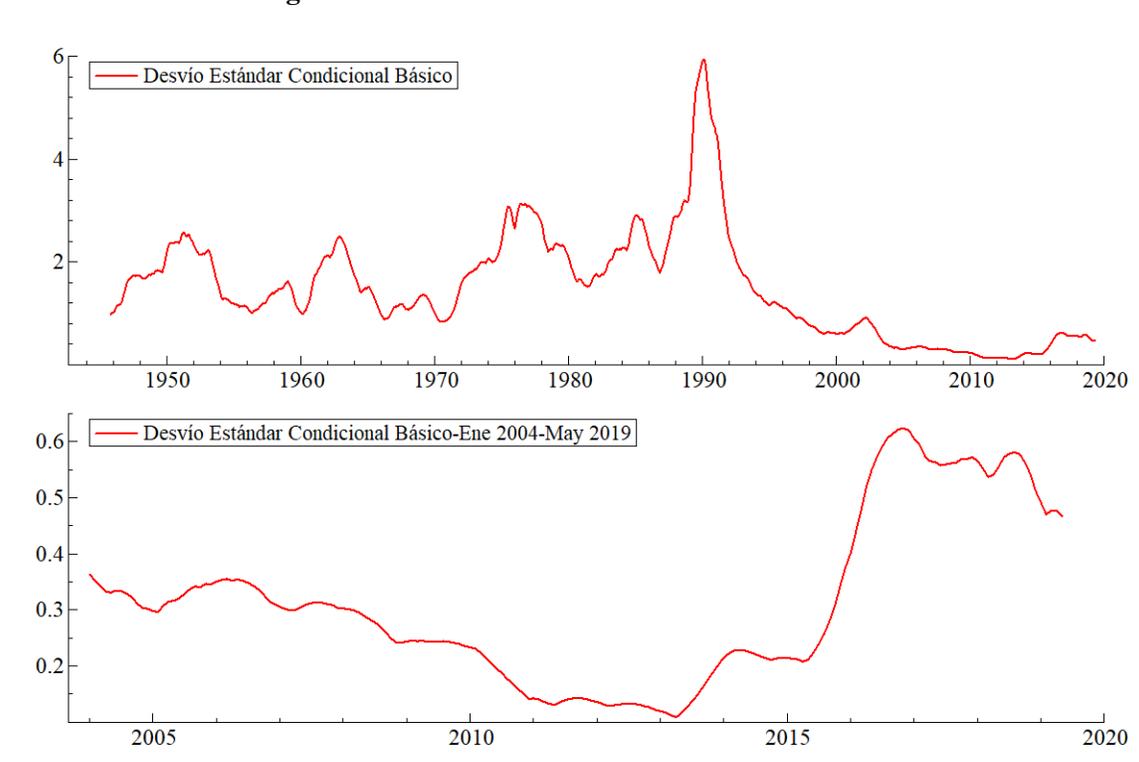
En este trabajo hemos extendido la serie con datos hasta mayo de 2019. Resultó ser altamente adecuada la aplicación de los modelos de espacio de estado para estimar los componentes de la serie de inflación y luego para estimar la volatilidad estocástica de los errores de la misma. Con la aplicación de este método se ha podido estimar adecuadamente la volatilidad total y sus tres componentes multiplicativos: la constante de escala, la volatilidad básica y la volatilidad estacional, siendo esta última, en este caso, no estocástica.

En la etapa inicial de nuestra investigación, nos propusimos analizar métodos para tratar una gran variedad de datos con irregularidades que suceden en las series de tiempo. Los modelos autorregresivos integrados de promedios móviles (o modelos ARIMA según sus siglas en inglés) son frecuentemente considerados como los que proveen la base principal para el modelado de cualquier serie de tiempo. Ahora bien, dado el estado actual del desarrollo de investigación en series de tiempo, puede haber alternativas más atractivas y por sobre todo más eficientes. Numerosas series de tiempo económicas no tienen una media constante y, en muchos casos, se observan fases de relativa tranquilidad seguidas de períodos de cambios importantes, o sea que la variabilidad cambia a través del tiempo. Dicho comportamiento es lo que recibe el nombre de *volatilidad*.

Una característica importante de las series de tiempo económicas es que ellas pueden, en general, no ser serialmente correlacionadas, pero sí dependientes. De este modo, modelos lineales como aquellos pertenecientes a la familia de los modelos ARMA suelen no ser apropiados para describir estas series.

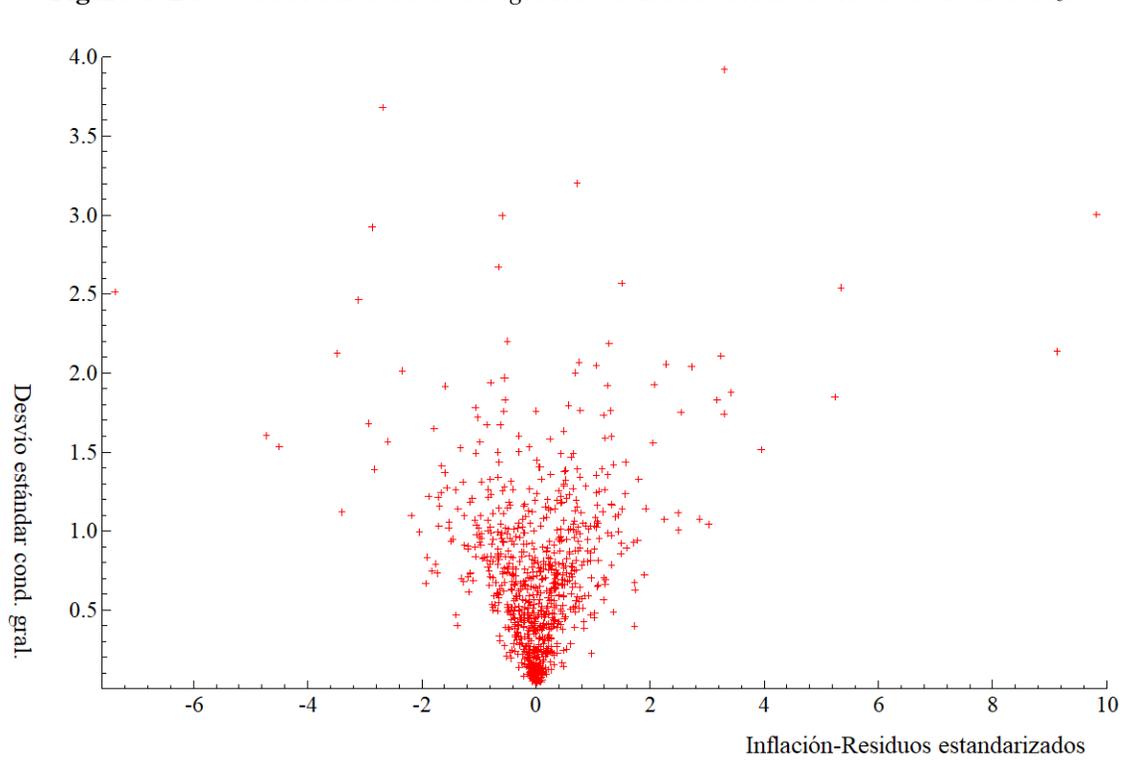
Para subsanar este hecho, tal como lo dijimos anteriormente, y para tomar en cuenta la presencia de la volatilidad en una serie

Figura 7. *Desvío estándar condicional básico estimado*



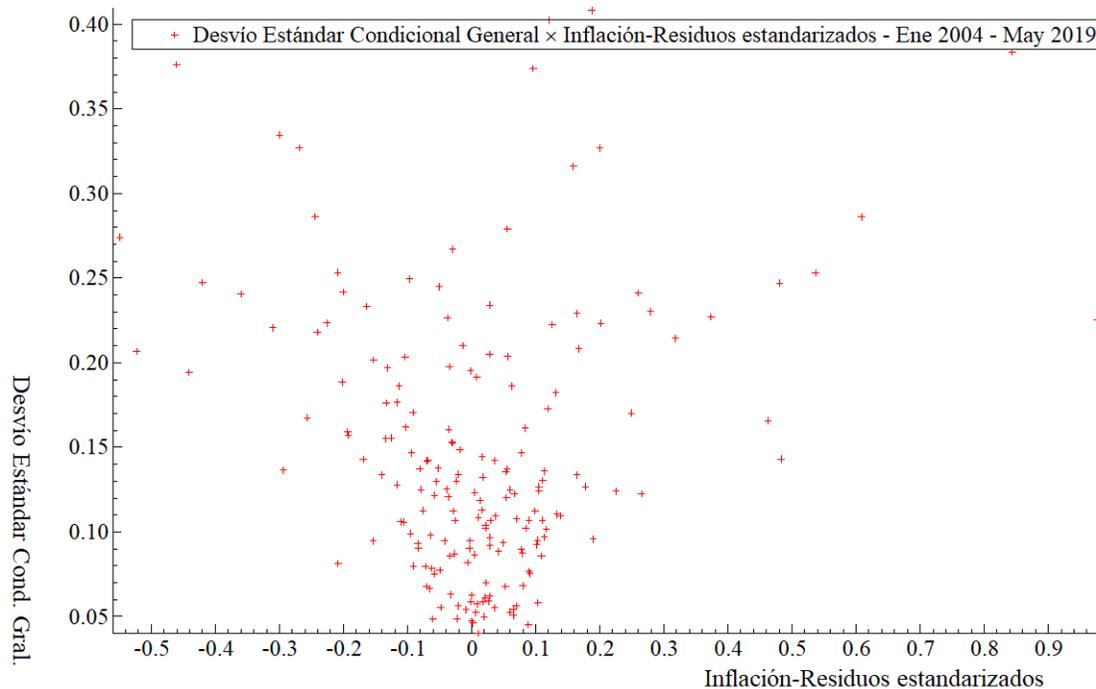
Nota. Desvío estándar condicional básico estimado, $\exp\{\hat{h}_t/2\}$, de los residuos al cuadrado del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina. Período enero de 1943 a mayo de 2019. En la parte superior para todo el período bajo estudio y en la parte inferior para el período enero de 2004 a mayo de 2019.

Figura 8. *Desvío estándar condicional general estimado versus los residuos estandarizados*



Nota. Desvío estándar condicional general estimado versus los residuos estandarizados del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina en el período enero de 1943 a mayo de 2019.

Figura 9. Desvío estándar condicional general estimado versus los residuos estandarizados



Nota. Desvío estándar condicional general estimado versus los residuos estandarizados del modelo definido en las ecuaciones (54), (55) y (56) para la inflación en la Argentina en el período enero de 2004 a mayo de 2019

económica, inicialmente se recurrió a modelos conocidos como *modelos heterocedásticos condicionales*. En estos modelos, la varianza de una serie en un dado instante de tiempo, depende de la información pasada y de otros datos disponibles hasta aquel instante de tiempo, de modo que se debe definir una varianza condicional, que no es constante y no coincide con la varianza global de la serie observada.

Entre los primeros modelos desarrollados para tratar la volatilidad se encuentran los de la familia ARCH. Los modelos ARCH o *modelos autorregresivos con heterocedasticidad condicional* fueron presentados por primera vez por Engle en el año 1982 con el objetivo de estimar la varianza de la inflación. La idea básica de este modelo es que y_t no necesariamente se encuentra correlacionado serialmente pero la volatilidad o varianza condicional de la serie depende de los retornos pasados por medio de una función cuadrática. Sin embargo, estos modelos raramente se utilizan en la práctica debido a su simplicidad.

Una buena generalización de este modelo se encuentra en los modelos de tipo GARCH introducidos por Bollerslev (1986). Este modelo es también un promedio ponderado de los residuos cuadráticos pasados, pero es más parsimonioso que los modelos de tipo ARCH y aun en su forma más simple ha probado ser sumamente exitoso en predecir las varianzas condicionales, por lo que decidimos hacer uso de los mismos a la hora de trabajar con nuestros datos.

El viejo refrán “Una pintura vale más que mil palabras” es bastante cierto en el análisis de cualquier conjunto de información. Antes de aplicar cualquier método estadístico a los datos

bajo estudio es fundamental observarlos gráficamente con el propósito de poder familiarizarse con ellos. Esto puede tener numerosos beneficios ya que este proceso nos servirá como indicador de ideas para un estudio posterior más detallado. Este fue el primer paso de nuestro trabajo en el cual pudimos ver las principales características de la serie y nos sirvió para realizar un ajuste adecuado de la misma. Nuevamente es necesario destacar que si bien este es un período muy largo para analizar; en donde se registraron numerosos cambios como ser de base, canasta, e intervenciones en el mismo INDEC, es posible realizar un estudio muy interesante en donde se aprecian las principales características de la serie.

En este trabajo se decidió usar un enfoque de volatilidad estocástica para analizar la serie de inflación para el período enero de 1943 a mayo de 2019, que resultó ser muy útil al captar las principales características de los datos.

Los modelos de la familia ARCH o GARCH suponen que la varianza condicional (volatilidad) depende de los valores pasados. En otras palabras y usando la notación que vimos anteriormente, si σ_t^2 es la volatilidad, la familia ARCH-GARCH supone que la misma depende de la serie y_j para $j < t$. Por otra parte, el *modelo de volatilidad estocástica* o MVE (o SVM según sus siglas en Inglés), propuesto por primera vez por Taylor (1980, 1986, 1994) no parte de este supuesto. Este modelo tiene como premisa el hecho de que la volatilidad σ_t^2 depende de sus valores pasados (σ_j^2 para $j < t$) pero es independiente de los valores pasados de la serie bajo análisis (y_j para $j < t$). Shephard y Pitt (1997) propusieron el uso de muestreo ponderado (“importance

sampling”) para estimar la función de verosimilitud en el caso no Gaussiano. Como el MVE es un modelo jerárquico, Jaquier, Polson y Rossi (1994) propusieron un análisis bayesiano del mismo. Véase también Shephard (2005), Shephard y Pitt (1997), Kim et al. (1998) y Ghysels, Harvey y Renault. (1996). Una reseña del problema de estimación del MVE está hecha por Motta (2001).

Como se mostró en Harvey et al. (1994), la forma de espacio de estado brinda las bases para la estimación por cuasi máxima verosimilitud vía el filtro y suavizador de Kalman y también permite construir estimaciones suavizadas del componente h_t de la varianza y realizar predicciones. Uno de los atractivos del enfoque de cuasi máxima verosimilitud es que puede ser aplicado sin un supuesto sobre una distribución particular para ε_t . Otro de los atractivos en utilizar un procedimiento de cuasi máxima verosimilitud por medio del filtro y suavizador de Kalman para estimar los MVE es que puede llevarse a cabo directamente usando paquetes estándares de computación tal como el STAMP de Koopman et al. (2010). Esto es una gran ventaja comparado con los métodos basados en simulaciones que requieren mayor trabajo. Finalmente, al utilizar un MVE se pudo estimar las diferentes partes de la volatilidad (la constante de escala, la volatilidad básica y la volatilidad estacional).

Con este análisis hemos podido captar que ha partir del el año 2004 el componente más importante de la volatilidad, la denominada volatilidad básica, comienza a decrecer, teniendo un mínimo en Abril de 2013, luego aumenta muy poco y recién retoma niveles importantes a partir de enero de 2016. Esta situación viene a corroborar el hecho de que desde esa fecha no se puedan tener estimaciones adecuadas de la volatilidad para esta serie, lo cual está en consonancia con el inicio de un período de falta de confianza en las estadísticas oficiales. Estadísticamente, esto dio lugar a una serie extremadamente suave que no coincide con el resto de la misma ni con la realidad vivida. Todo se revirtió a partir de enero de 2016 cuando el INDEC comenzó a trabajar con total libertad.

11. Conclusión

El estudio demuestra que los modelos de volatilidad estocástica (MVE) ofrecen una ventaja significativa sobre los modelos tradicionales como ARCH-GARCH, particularmente en contextos de alta variabilidad y cambios estructurales. Al aplicar un enfoque de espacio de estado, los autores logran descomponer la inflación en componentes clave, mejorando la capacidad predictiva y la comprensión de las dinámicas de volatilidad en la serie temporal de inflación en Argentina. Esta metodología resulta especialmente valiosa para entender cómo factores estructurales, como la intervención política del INDEC entre 2004 y 2015, afectaron la calidad y la fiabilidad de los datos inflacionarios.

Referencias

- Abril, J. C. (1999). *Análisis de series de tiempo basado en modelos de espacio de estado*. EUDEBA: Buenos Aires.
- Abril, J. C. (2004). *Modelos para el análisis de las series de tiempo*. Ediciones Cooperativas: Buenos Aires.
- Abril, M. M. (2014). *El enfoque de espacio de estado de las series de tiempo para el estudio de los problemas de volatilidad* (Tesis doctoral). Universidad Nacional de Tucumán, Argentina.
- Abril, J. C., y Abril, M. M. (2017). La heterocedasticidad condicional en la inflación de la Argentina: Un análisis para el período 1943-2013. *Revista de Investigaciones del Departamento de Ciencias Económicas (RINCE)*, Universidad Nacional de la Matanza (UNLaM), **8**. Recuperado de <https://rince.unlam.edu.ar>.
- Abril, J. C., y Abril, M. M. (2018). *Métodos modernos de series de tiempo y sus aplicaciones*. Editorial Académica Española: Saarbrücken, Alemania.
- Baillie, R. T., y Bollerslev, T. (1989). The message in daily exchange rates: A conditional-variance tale. *Journal of Business and Economic Statistics*, **7**, 297–305.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69**, 542–547.
- Bollerslev, T., Chou, R. Y., y Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, **52**, 5–59.
- Bollerslev, T., y Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**, 143–172.
- Box, G. E. P., y Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Revised edition). Holden-Day Inc.: San Francisco.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, **6**, 76–90.
- Bryan, M. F., y Cecchetti, S. G. (1994). Measuring core inflation. En Mankiw, N. G. (Ed.), *Monetary policy* (pp. 195–219). The University of Chicago Press: Chicago.
- Bryan, M. F., Cecchetti, S. G., y Wiggins II, R. L. (1997). Efficient inflation estimation (Working Paper 6183). National Bureau of Economic Research, Cambridge, MA.
- Durbin, J., y Koopman, S. J. (1997a). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, **84**, 669–684.
- Durbin, J., y Koopman, S. J. (1997b). Time series analysis of non-Gaussian observations based on state space models (Preprint). London School of Economics.
- Durbin, J., y Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society, Serie B*, **62**, 3–56.
- Durbin, J., y Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press: Oxford.

- Durbin, J., y Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). Oxford University Press: Oxford.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.
- Engle, R. F., y Gonzalez-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics*, **9**, 345–360.
- Engle, R., y Mezrich, J. (1996). GARCH for groups. *RISK*, **9**, 36–40.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *Computer Journal*, **13**(3), 317–322.
- Fletcher, R. (1987). *Practical methods of optimization* (2nd ed.). John Wiley Sons: New York.
- Ghysels, E., Harvey, A. C., y Renault, E. (1996). Stochastic volatility. En C. R. Rao, y G. S. Maddala (Eds.), *Statistical methods in finance* (pp. 119–191). North-Holland: Amsterdam.
- Goldfarb, D. (1970). A family of variable metric updates derived by variational means. *Mathematics of Computation*, **24**(109), 23–26.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press: Cambridge.
- Harvey, A. C., Ruiz, E., y Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247–264.
- Hsieh, D. A. (1989). Modeling heteroskedasticity in daily foreign exchange rates. *Journal of Business and Economic Statistics*, **7**, 307–317.
- Kim, S., Shephard, N., y Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **85**, 361–393.
- Koopman, S. J., Harvey, A. C., Doornik, J. A., y Shephard, N. (2010). *STAMP 8.3: Structural time series analyser, modeller and predictor*. Timberlake Consultants: London.
- Motta, A. C. O. (2001). Modelos do espaço de estados não-Gaussianos e o modelo de volatilidade estocástica (Dissertação de mestrado). IMECC-UNICAMP.
- Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, **3**, 15–102.
- Shephard, N. (2005). *Stochastic volatility: Selected readings*. Oxford University Press: Oxford.
- Shephard, N., y Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653–667.
- Stock, J. H., y Watson, M. W. (2015). Core inflation and trend inflation (Working Paper 21282). National Bureau of Economic Research, Cambridge, MA.
- Taylor, S. J. (1980). Conjectured models for trend in financial prices tested as forecasts. *Journal of the Royal Statistical Society, Serie B*, **42**, 338–362.
- Taylor, S. J. (1986). *Modelling financial time series*. John Wiley Sons: Chichester.
- Taylor, S. J. (1994). Modelling stochastic volatility. *Mathematical Finance*, **4**, 183–204.
- Weiss, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory*, **2**, 107–131.