



**“Comparación del desempeño predictivo de modelos ARIMA, MARS e híbridos ARIMA-MARS
para el pronóstico de la inflación argentina: evidencia bajo distintos esquemas de validación”**

Colaso, Matías Nicolás

Morel Niveyro, Lautaro

Facultad de Ciencias Económicas

Universidad Nacional de Misiones

Licenciatura en Economía

Dr. Dip Juan Antonio

Mayo de 2026

ÍNDICE

1. Introducción	5
2. Aporte.....	8
3. Antecedentes	8
4. Modelos teóricos para pronósticos.....	12
4.1. ARIMA	12
4.2. MARS	14
4.3. Modelo híbrido.....	16
5. Metodología	18
5.1. Base de datos.....	18
5.2. Tratamiento de valores atípicos	20
5.3. Estadísticos descriptivos y test.....	21
5.3.1. Test de Diebold-Mariano (DM)	21
5.4. Análisis de propiedades estadísticas	22
5.4.1. Propiedades de la serie	22
5.4.2. Test de multicolinealidad VIF (Variance Inflation Factor)	23
5.4.3. Test de Ljung-Box	24
5.4.4. Test de Shapiro-Wilk	25
5.4.5. Prueba ARCH-LM	25
5.5. Especificación de los modelos	26
5.5.1. ARIMA	26
5.5.2. MARS	27
5.5.3. MARS con poda y dummies estacionales.....	27
5.5.4. MARS con validación cruzada (cv)	27
5.5.5. MARS con rezagos de primer y segundo orden	27
5.5.6. MARS sin valores atípicos.....	28
5.5.7. Modelo híbrido.....	28
5.6. Métricas de evaluación.....	28
5.7. Partición de datos y estrategia de validación	30
5.7.1. Partición de datos	30
5.7.2. Validación estadística	31
6. Resultados	32
6.1. Propiedades estadísticas y test de multicolinealidad.....	32
6.1.1 Estimación de modelos ARIMA y distintos tipos de MARS.....	34
6.2. Desempeño predictivo.....	42

6.3. Modelo MARS_CV	45
6.4. Comparación estadística entre modelos. Test de Diebold-Mariano (DM)	47
6.5. Diagnóstico de residuos de MARS_CV	49
6.6. Robustez temporal.....	52
7. Limitaciones	54
8. Conclusiones	55
9. Referencias.....	57
10. Anexo	60
10.1. Especificación del modelo MARS_CV_lag	61
10.2. Diagnóstico de residuos	63
10.3. Estimaciones de los distintos modelos.....	65
10.3.2. MARS sin poda.....	65
10.3.3. MARS con poda y dummies estacionales.....	66
10.3.4. MARS con rezagos	67
10.3.5. MARS sin valores atípicos.....	67
10.3.6. Modelo híbrido.....	68
10.3.7. MARS_CV_lag.....	69

Resumen

Este trabajo analiza la capacidad predictiva de distintos enfoques econométricos para poder estimar la inflación mensual en Argentina, en un contexto caracterizado por alta volatilidad y posibles cambios de régimen. El objetivo central consiste en comparar el desempeño predictivo y la estabilidad temporal de los modelos autoregressive integrated moving average (ARIMA), multivariate adaptive regression splines (MARS) e híbridos ARIMA-MARS para los pronósticos de la inflación mensual argentina bajo distintos regímenes inflacionarios y esquemas de validación fuera de muestra, utilizando como criterio principal el error cuadrático medio (RMSE).

Metodológicamente, se comparan tres especificaciones, un modelo ARIMA como *benchmark* lineal, un modelo MARS como alternativa no paramétrica con capacidad de capturar relaciones complejas y un modelo híbrido ARIMA-MARS. La evaluación se realiza mediante una estrategia de validación fuera de muestra, combinando partición estática y una validación dinámica, junto con métricas complementarias como error absoluto medio (MAE) y error porcentual absoluto medio (MAPE).

Los resultados evidencian que el desempeño relativo depende críticamente del esquema de validación. Bajo un *hold-out* estático en el período de desinflación reciente, los modelos MARS superan al ARIMA con significancia estadística (*test* de Diebold-Mariano con corrección de Harvey-Leybourne-Newbold). Bajo validación dinámica que abarca la aceleración inflacionaria de 2022-2023, el pico de fines de 2023 y la desaceleración posterior, la jerarquía se invierte: el ARIMA exhibe mayor estabilidad temporal, mientras que los modelos MARS deterioran su precisión al extrapolar fuera del dominio histórico de las explicativas. La evidencia sugiere que, en una economía con quiebres estructurales recurrentes como la argentina, no existe un modelo uniformemente superior sino un trade-off entre flexibilidad, donde MARS_CV es preferible para pronóstico puntual en regímenes estables, y robustez temporal, donde ARIMA resulta más adecuado para monitoreo en entornos de cambio de régimen.

Palabras clave: MARS, ARIMA, modelo híbrido, inflación, desempeño predictivo.

1. Introducción

La predicción de la inflación ocupa un lugar central en la toma de decisiones económicas, especialmente en países con alta nominalidad e inestabilidad macroeconómica como Argentina. Contar con herramientas que permiten anticipar la evolución mensual de los precios resulta relevante tanto para formulación de política económica como para la coordinación de expectativas, la negociación de salarios, la planificación financiera y la evaluación de escenarios futuros.

Este trabajo tiene como objetivo evaluar el desempeño predictivo de distintos enfoques econométricos y de aprendizaje automático para pronosticar la inflación mensual en Argentina durante el período 2014-2025. Particularmente se comparan tres estrategias: un modelo ARIMA como *benchmark* tradicional de series temporales, distintas especificaciones del modelo MARS como alternativa no lineal, con y sin poda, entendiendo por poda el procedimiento de eliminación de términos redundantes que reduce el sobreajuste del modelo, y un enfoque híbrido ARIMA-MARS, orientado a combinar la dinámica lineal capturada por ARIMA con posibles componentes no lineales remanentes.

La contribución central de la investigación consiste en analizar no sólo qué modelo presenta menor error predictivo en una partición estática fuera de muestra, sino también si dicha superioridad se mantiene bajo una estrategia dinámica de validación. Es importante destacar esto debido a que en economías sujetas a episodios de alta volatilidad y cambios de régimen, un modelo puede mostrar buen desempeño en un período específico y, sin embargo, perder estabilidad cuando se evalúa en distintos contextos macroeconómicos.

En Argentina, la dinámica inflacionaria reciente refuerza la importancia de este problema. La inflación, medida como variación interanual del IPC, se ubicó en torno al 50% en 2021, se aproximó al 100% en 2022 y superó el 200% en 2023. Si bien durante 2024 se observó una desaceleración, los niveles de inflación continuaron siendo elevados: en junio, el IPC registró una variación interanual de 271,5% y una variación acumulada de 79,8% en el primer semestre. Este régimen de inflación alta y persistente incrementa la incertidumbre macroeconómica y amplifica los costos asociados a pronósticos imprecisos.

Desde una perspectiva histórica, el pronóstico de inflación se apoyó inicialmente en modelos estructurales vinculados a la Curva de Phillips y la tasa de desempleo no aceleradora de la inflación (o tasa NAIRU por sus siglas en inglés) especialmente durante las décadas de 1970 a 1990 (Arcusa, 2020). Si bien estos enfoques resultaron importantes para interpretar la relación entre inflación, actividad económica y mercado laboral, también presentaron limitaciones frente a episodios de inestabilidad macroeconómica y cambios estructurales. En este sentido, Diebold (1998) destaca que los grandes modelos macroeconómicos perdieron credibilidad durante la década de 1970 al no lograr explicar ni anticipar transformaciones profundas en la dinámica económica, fenómeno asociado históricamente con el episodio de estanflación.

Contemplando estas limitaciones, la literatura comenzó a otorgar mayor importancia a modelos estadísticos de series temporales, como los elaborados por Box y Jenkins, entre los que se destacan las especificaciones ARIMA (Box et al., 2015). Estos modelos permiten realizar pronósticos a partir de la información contenida en la propia serie, sin marcar una estructura económica rígida, y continúan siendo utilizados como referencia en ejercicios predictivos.

Posteriormente, el desarrollo de modelos multivariados, como los VAR, permitió incorporar la interacción dinámica entre distintas variables macroeconómicas. Sin embargo, en economías sujetas a cambios de régimen, la hipótesis de parámetros constantes puede resultar restrictiva, ya que una misma especificación estimada sobre períodos largos puede mezclar dinámicas heterogéneas y generar pronósticos promedio que no representan adecuadamente ningún régimen en particular.

En los últimos años, la literatura sobre pronóstico de inflación ha incorporado de manera creciente enfoques de machine learning, debido a su capacidad para modelar relaciones no lineales, efectos umbral e interacciones entre variables que pueden ser difíciles de capturar mediante especificaciones lineales tradicionales (Kilic, 2025). En este marco, el modelo MARS resulta especialmente relevante, ya que permite aproximar relaciones complejas sin imponer una forma funcional única para toda la muestra. Esta característica es particularmente útil en contextos donde la inflación puede presentar no linealidades,

cambios de pendiente e interacciones entre determinantes macroeconómicos (Medeiros, Montes Schütte y Soussi, 2025)

La elección de los modelos ARIMA, MARS y ARIMA-MARS responde a esta evolución de la literatura. El ARIMA se utiliza como modelo de referencia debido a su capacidad para capturar la dinámica temporal de la inflación a partir de sus propios valores pasados. MARS se incorpora como alternativa no lineal, con el objetivo de evaluar si una mayor flexibilidad funcional mejora la capacidad predictiva frente al modelo tradicional. Finalmente, el enfoque híbrido ARIMA-MARS se fundamenta en la idea de que una serie temporal puede contener un componente lineal y otro no lineal, siguiendo la lógica propuesta por Zhang (2003), por lo que resulta pertinente analizar si la modelización de los residuos del ARIMA mediante MARS permite mejorar el pronóstico final.

No obstante, la evidencia empírica expone que no se debe asumir de forma automática la superioridad de los modelos más flexibles. El desempeño predictivo puede depender del periodo evaluado, del régimen inflacionario vigente, de la cantidad de observaciones disponibles y del esquema de validación utilizado. Por esa razón, la presente investigación compara los modelos no solo mediante métricas tradicionales de error fuera de muestra, como RMSE, MAE y MAPE, sino también considerando la estabilidad de los resultados bajo distintos esquemas de validación.

En este contexto, la pregunta de investigación se plantea como: ¿Cuál es el desempeño predictivo relativo de los modelos ARIMA, MARS y ARIMA-MARS para pronosticar la inflación mensual argentina bajo distintos esquemas de validación fuera de muestra?

Puntualmente, se busca analizar si la incorporación de no linealidades mediante MARS mejora la capacidad predictiva respecto de un modelo ARIMA tradicional y si esta mejora se mantiene cuando la evaluación se realiza no solo mediante una partición estática, sino también a través de una estrategia dinámica de validación. Esta distinción permite evaluar el trade-off entre precisión puntual y robustez temporal, aspecto especialmente relevante en una economía caracterizada por alta volatilidad y cambios recurrentes en el régimen inflacionario.

2. Aporte

Este trabajo busca contribuir a la literatura sobre pronóstico de inflación en Argentina mediante la comparación de tres enfoques alternativos: un modelo ARIMA como *benchmark* parsimonioso, un modelo no lineal MARS y una especificación híbrida ARIMA–MARS. El aporte central consiste en evaluar si la incorporación de no linealidades mejora la capacidad predictiva frente a un enfoque puramente autorregresivo, en un contexto macroeconómico como el argentino, atravesado por shocks recurrentes y posibles cambios estructurales. Asimismo, la investigación incorpora una doble estrategia de validación, estática y *rolling window*, que permite no solo comparar la precisión predictiva, sino también analizar la estabilidad temporal de los modelos en un entorno de actualización constante. Los resultados aportan evidencia de que la regularización desempeña un rol clave en el desarrollo de MARS: sin poda, el modelo tiende al sobreajuste, mientras que con poda mejora su capacidad de generalización y logra resultados competitivos. Sin embargo, también se observa que el desempeño relativo de los modelos depende del esquema de validación adoptado, ya que MARS con poda puede destacarse en una evaluación estática, mientras que ARIMA muestra una mayor estabilidad promedio en la validación dinámica.

3. Antecedentes

Históricamente, el pronóstico de inflación se ha apoyado en modelos clásicos de series temporales, entre los cuales se destacan las especificaciones autorregresivas y de promedios móviles, en adelante, AR, MA, ARMA y ARIMA. Estos enfoques permiten modelar la dependencia temporal de la inflación a partir de su trayectoria histórica y han sido ampliamente utilizados para la predicción del Índice de Precios al Consumidor (IPC). Asimismo, cuando el objetivo ha sido incorporar información proveniente de otros determinantes macroeconómicos, se han empleado modelos multivariados como VAR, VECM y BVAR, estos permiten captar interrelaciones dinámicas entre inflación, tipo de cambio, actividad económica, salarios o agregados monetarios. En su conjunto, estos modelos han sido el fundamento

clásico de las predicciones inflacionarias, en particular en situaciones en las que la serie es relativamente estable y los precios pueden aproximarse mediante patrones lineales.

Sin embargo, en las últimas décadas, el uso de técnicas de machine learning para la predicción económica ha sido motivado por la volatilidad macroeconómica creciente, el surgimiento de cambios de régimen y la posibilidad de relaciones no lineales entre la inflación y sus determinantes. En este contexto, diversas investigaciones han demostrado que modelos más flexibles pueden superar las estrategias tradicionales en escenarios en los que la dinámica inflacionaria es más compleja.

En el caso de Argentina, Aguilar (2024) compara modelos ARIMA y VAR con métodos de aprendizaje automático, como LASSO y Elastic Net, para predecir la inflación mensual. Los resultados muestran que los modelos de machine learning obtienen un mejor desempeño predictivo que ARIMA en horizontes de uno y tres meses, mientras que Elastic Net presenta los mejores indicadores de ajuste y de error fuera de muestra. Entre las variables con mayor peso predictivo se destacan la inflación rezagada, el tipo de cambio oficial, los salarios nominales, el ITCRM-BCRA y diversos agregados monetarios. Este trabajo aporta evidencia de que, en un contexto de alta nominalidad e inestabilidad como el argentino, los enfoques más flexibles pueden capturar mejor la dinámica inflacionaria de corto plazo.

Un antecedente distinto, aunque complementario, es el trabajo de Garegnani y Gómez Aguirre (2018). En su estudio, los autores examinan si la incorporación de una mayor cantidad de información macroeconómica en un esquema BVAR mejora los pronósticos de inflación en comparación con un *benchmark* autorregresivo univariado. Sus resultados muestran que las predicciones resultan más precisas cuando se emplea un enfoque multivariado que integra la interacción entre diversas variables macroeconómicas relevantes.

Por su parte, el trabajo de Graña Colella (2020) cumple una función distinta dentro de los antecedentes, ya que su objetivo no es comparar la precisión predictiva entre modelos, sino estudiar las causas de la inflación argentina mediante un VECM para el período 2003-2019. En el largo plazo, los resultados asignan un papel predominante a la puja distributiva y al tipo de cambio, mientras que en el corto y

mediano plazo resultan significativos la inercia inflacionaria, la variación del tipo de cambio, la emisión monetaria y el término de corrección del error. El autor subraya que la inflación argentina responde a determinantes distintos según el horizonte de análisis, lo que justifica comparar modelos alternativos cuando el objetivo es pronosticarla. En conjunto, esta evidencia sugiere que la inflación argentina combina inercia, multicausalidad y cambios en la intensidad de sus determinantes, rasgos que vuelven especialmente relevante contrastar un modelo clásico como el ARIMA con un enfoque más flexible como el MARS.

En el plano internacional, la literatura sobre el pronóstico de la inflación muestra una transición progresiva desde modelos clásicos de series temporales hacia metodologías de machine learning y deep learning. Sin embargo, esa transición no implica una superioridad automática de los modelos más complejos. De hecho, la evidencia comparativa sugiere que el desempeño depende del indicador elegido, del horizonte de pronóstico, de la disponibilidad de información y de la estructura de cada economía. En este sentido, Ülke, Sahin y Subasi (2018), para el caso de Estados Unidos, comparan modelos de series temporales y modelos de machine learning para distintas medidas de inflación y distintos horizontes. Sus resultados muestran una evidencia mixta: los modelos de machine learning obtienen mejores pronósticos en siete de dieciséis escenarios, mientras que los modelos de series temporales resultan superiores en nueve; además, los modelos multivariados superan a los univariados en catorce de dieciséis casos. Este trabajo es especialmente relevante porque introduce una idea central para la presente tesis: no existe un modelo universalmente dominante, sino que la precisión predictiva depende de las características del problema empírico.

En América Latina, Riofrío et al. (2020) analizan el caso de Ecuador con una base de 174 observaciones mensuales entre 2005 y 2019 y destacan que el pronóstico de la inflación se vuelve especialmente desafiante debido al cambio de tendencia observado desde mediados de 2016. Para enfrentar ese problema, comparan enfoques clásicos, como SARIMA y el suavizamiento exponencial, con métodos más flexibles, como SVR, LSTM y Facebook Prophet, en un ejercicio de pronóstico a 12 meses. Este

estudio es importante porque muestra que, incluso en muestras relativamente pequeñas y con cambios marcados en la dinámica inflacionaria, la literatura internacional ya recurre a comparaciones entre enfoques lineales y no lineales para evaluar qué metodología se adapta mejor a una serie compleja.

Nguyen et al. (2023) comparan cuatro metodologías para Estados Unidos: MARS, MLR, SVR y ARDL, utilizando datos de 2017 a 2022. El estudio reporta que todos los modelos alcanzan altos niveles de precisión, pero que MARS obtiene la mayor exactitud en la fase de prueba. Esto muestra que MARS no solo puede funcionar como un modelo competitivo frente a enfoques clásicos y otros métodos de aprendizaje automático, sino también como una herramienta particularmente adecuada para modelar la inflación cuando existen relaciones no lineales entre la inflación y variables como el precio del petróleo, el precio del oro y la tasa de fondos federales.

Un antecedente metodológico relevante es Zhang (2003), quien sostiene que, en series temporales complejas, un único modelo puede no ser suficiente para captar toda la dinámica de los datos. En particular, el autor argumenta que los modelos lineales como ARIMA son útiles para representar la estructura autorregresiva, pero pueden dejar sin modelar relaciones no lineales relevantes. A partir de ello, propone una estrategia híbrida en la que el componente lineal se estima con ARIMA y el componente no lineal se modela sobre los residuos. Aunque en el trabajo se implementa esta segunda etapa mediante redes neuronales, su planteo resulta especialmente útil como fundamento metodológico para extender esta lógica a otras técnicas no lineales flexibles, como MARS, y evaluar si una combinación híbrida mejora la capacidad predictiva respecto de enfoques puramente lineales o puramente no lineales. La evidencia más reciente, además, sugiere que la literatura internacional avanza hacia modelos híbridos, en los cuales se combinan arquitecturas capaces de capturar dependencias temporales con algoritmos que explotan no linealidades y relaciones complejas entre variables. En esa línea, Gur (2024) evalúa para Estados Unidos modelos como LSTM, MARS, XGBoost, LSTM-MARS y LSTM-XGBoost sobre una muestra larga, desde 1974 hasta 2023. Sus resultados indican que el mejor desempeño corresponde al

modelo LSTM-XGBoost, lo que sugiere que los esquemas híbridos pueden ofrecer ganancias adicionales de precisión frente a modelos individuales.

Al mismo tiempo, otros antecedentes recuerdan que los enfoques clásicos siguen siendo relevantes. En Albania, Gjika, Puka y Zaçaj (2018) plantean una estrategia que combina modelos de series temporales sobre subcomponentes del IPC con una regresión múltiple para obtener el pronóstico agregado, subrayando la utilidad de incorporar información externa y desagregada. En Kenia, Lidiema (2017) encuentra que un modelo SARIMA supera al método de Holt-Winters según medidas como MASE, MAE y MAPE. En conjunto, estos trabajos muestran que la discusión internacional no debe plantearse como una oposición simplista entre “modelos viejos” y “modelos nuevos”, sino como una comparación empírica en la que la estructura de la serie, la calidad de los datos y el horizonte de predicción condicionan cuál enfoque resulta más conveniente.

4. Modelos teóricos para pronósticos

4.1. ARIMA

Según Box et al. (2015), los modelos ARIMA (*Autoregressive integrated moving average*) son herramientas usadas con regularidad en análisis y pronósticos de series temporales económicas, debido a que suelen ser precisas y son bastante prácticas dado que captura la dinámica de una variable a partir de sus valores históricos. Tres elementos componen la estructura matemática: Un término autorregresivo (AR), uno de integración o diferenciación (I) y un término de medias móviles (MA). El modelo ARIMA estacional, denotado como SARIMA (p,d,q)(P,D,Q)_s, constituye una generalización del modelo ARIMA clásico que incorpora explícitamente patrones de dependencia temporal tanto de corto plazo como de naturaleza estacional. Este tipo de especificación resulta particularmente adecuado para series económicas mensuales, como la inflación, en las que la estacionalidad de período $s=12$ es estructural (Medeiros et al., 2025).

La formulación del modelo se construye a partir del operador de rezago L , definido como $Ly_t = y_{t-1}$, y de los operadores de diferenciación

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1} \quad (\text{Dif. regular}) \quad (1)$$

$$\Delta_s y_t = (1 - L^s)y_t = y_t - y_{t-s} \quad (\text{Dif. estacional}) \quad (2)$$

Siendo “s” el valor correspondiente a cada mes del año, las ecuaciones 1 y 2 representan los operadores que permiten transformar la serie original en una serie estacionaria, eliminando tendencias estocásticas y componentes estacionales no estacionarios. De manera general, el modelo SARIMA puede expresarse como:

$$\Phi_p(L^s)\phi_p(L)\Delta^d\Delta_s^D y_t = \theta_q(L^s)\theta_q(L)\varepsilon_t \quad (3)$$

Donde $\varepsilon_t \sim WN(0, \sigma^2)$ representa ruido blanco

Si se expanden de manera general la ecuación 3, los polinomios autorregresivos quedan:

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \quad \text{AR regular} \quad (4)$$

$$\Phi_p(L^s) = 1 - \phi_1 L^s - \phi_2 L^{2s} - \dots - \phi_p L^{ps} \quad \text{AR estacional} \quad (5)$$

Y los polinomios de medias móviles como:

$$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \quad \text{MA regular} \quad (6)$$

$$\Theta_Q(L^s) = 1 + \theta_1 L^s + \theta_2 L^{2s} + \dots + \theta_Q L^{Qs} \quad \text{MA estacional} \quad (7)$$

Por lo tanto el modelo SARIMA (p,d,q)(P,D,Q)s quedaría escrito:

$$\Phi_P(L^s)\phi_p(L)(1 - L)^d(1 - L_s)^D y_t = \theta_q(L^s)\theta_q(L)\varepsilon_t \quad (8)$$

Aquí: $(1 - L)^d$ elimina tendencia, $(1 - L_s)^D$ elimina estacionalidad, $\phi_p(L)$ y $\theta_q(L)$ capturan la dinámica de corto plazo y $\Phi_P(L^s)$ y $\Theta_Q(L^s)$ capturan la dependencia estacional

Para la determinación de los parámetros del modelo, se evalúa todos los modelos en una grilla de parámetros y selecciona el que minimiza el criterio de información de Akaike corregido (Cavanaugh, 1997):

$$AICc = -2l(\hat{\theta}) + 2k + \frac{2k(k+1)}{T-k-1} \quad (9)$$

Donde, $l(\hat{\theta})$ es el log-verosímil maximizado, $\hat{\theta}$ es el máximo valor de la función de verosimilitud para el modelo estimado, $k=p+q+P+Q+d+D$ es el número de parámetros libres y T es el tamaño muestral. El término $\frac{2k(k+1)}{T-k-1}$ es una penalización adicional, importante para muestras pequeñas.

Al ser series donde se observa persistencia inflacionaria, se justifica la utilización de modelos como el SARIMA que priorizan la dinámica temporal propia del proceso. Estos modelos son adecuados puesto que permiten capturar la inercia inflacionaria a través de sus componentes autorregresivos e integrados, sin prestar atención a las relaciones estructurales entre la inflación y variables explicativas que podrían presentar inestabilidad a través del tiempo. En la tesis, se estimarán modelos estacionarios y no estacionarios, los cuales serán contrastados según criterios de información.

4.2. MARS

El modelo MARS (*Multivariate adaptive regression splines*) introducido por Friedman, 1991, es un modelo de regresión no paramétrico que captura relaciones complejas y no lineales. A diferencia de otros modelos tradicionales, MARS no determina ex-ante una forma funcional específica, sino que permite que los datos determinen la relación específica entre las variables. Este modelo se basa en par funciones base de tipo *splines*, las cuales dividen a la serie en tramos, proporcionando un conjunto de datos de entrenamiento a cada uno.

Par de funciones base:

$$h(x-t) = \max(0, x-t) = (x-t) \quad (10)$$

$$h(t - x) = \max(0, t - x) = (t - x)$$

Dentro de cada tramo de la ecuación 10 es donde se determinan las relaciones entre las variables, que son aproximadas mediante funciones simples. Generalmente, las relaciones suelen ser lineales pero el modelo permite que los coeficientes, y por lo tanto las pendientes, cambien cuando la serie cruza determinados puntos de corte, llamados nudos. Estos nudos son determinados de manera endógena a partir de los datos.

Esta característica implica que el modelo es capaz de capturar efectos no lineales, asimetrías e interacciones entre variables sin tener que especificarlas previamente. Lo que resulta especialmente relevante y útil para capturar relaciones entre variables que pueden sufrir alteraciones según el nivel que adopten o donde el comportamiento es diferenciado a lo largo del dominio.

En términos generales, el modelo MARS puede especificarse como:

$$\hat{f}(x) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} h(s_{km}(x_{v(k,m)} - t_{km})) \quad (11)$$

Donde β_m son los coeficientes estimados por mínimos cuadrados ordinarios, K_m es el grado de interacción en el m-ésimo término, s_{km} indica la dirección de la función y $\in \{-1, +1\}$, $v(k, m)$ es el índice de la variable usada y t_{km} es el punto de quiebre seleccionado de los datos

La estimación del modelo se puede desarrollar en dos fases. En primer lugar, la fase *forward* que es la construcción del modelo. Se comienza con el modelo con constante $\hat{F} = \beta_0$, y en cada paso se agrega un par de funciones base que minimiza la suma de los residuos al cuadrado. Este proceso se repite hasta alcanzar el máximo de términos permitido el cual es determinado por $M_{max} = \min(200, 2 * k + 1)$, siendo k el número de variables.

Luego, se puede llevar a cabo la fase *backward* o de poda, donde se eliminan secuencialmente los términos menos importantes según el criterio de Validación Cruzada Generalizado (GCV). En cada

iteración, elimina una función y calcula el error GCV. Si este último mejora, prosigue quitando otros términos.

El modelo óptimo es aquel que minimiza el GCV:

$$GCV(M) = \frac{RSS(M)/T}{(1 - \underline{C}(M)/T)^2} \quad (12)$$

Siendo $RSS(M) = \sum_{i=1}^T (y_i - \hat{F}_M(x_i))^2$ la suma de los cuadrados residuales del modelo con M términos, $\underline{C}(M) = C(M) + dM$ es una penalización efectiva, $C(M)$ es el número de coeficientes lineales (β) y $d \approx 3$ es un factor de penalización por los *knots*.

Los modelos MARS entonces, son versátiles para relaciones aditivas o con pocas variables. Ponderan las variables de entrada con impacto significativo en los parámetros de salida, lo que permite que el modelo se pueda aplicar a conjuntos de datos de diferentes tamaños (Nguyen, et al., 2023). Una característica muy importante de los modelos MARS es que no están diseñados para interpretar los coeficientes o realizar inferencia causal, son modelos específicos para realizar predicciones, especialmente en relaciones altamente no lineales y difíciles de modelar.

4.3. Modelo híbrido

El enfoque híbrido entre ARIMA y MARS se fundamenta en la descomposición aditiva de la serie de la inflación en componentes lineales y no lineales, siguiendo la propuesta de Zhang (2003). La idea central es que los modelos ARIMA capturan adecuadamente la dependencia lineal y la estructura estocástica de baja complejidad, mientras que los métodos no paramétricos como MARS permiten modelar relaciones no lineales, umbrales y efectos de interacción que no pueden ser representados dentro del marco lineal clásico. Siguiendo a Zhang (2003), podemos escribir formalmente: $y_t = L_t + N_t$ con L_t es el componente lineal y N_t es el componente no lineal.

La construcción del modelo se realiza en tres pasos. En primer lugar, se estima el modelo ARIMA a la serie de la inflación, obteniendo así el pronóstico lineal y sus residuos expresado por la ecuación 13:

$$\hat{e}_t = y_t - L_t \quad (13)$$

Tales residuos exponen la parte de la serie que no fue explicada por el bloque lineal. En el esquema híbrido, la interpretación económica y estadística de esta etapa es que eliminada la inercia inicial y la estructura estacional encontrada por ARIMA, la información restante puede contener relaciones de tipo no lineal, variaciones de intensidad en los efectos o interacciones entre variables macroeconómicas que no son recogidas adecuadamente por una especificación lineal.

Luego, se realiza la construcción del modelo MARS sobre los residuos del ARIMA ($\hat{e}_t$) para capturar las no linealidades de la variable macroeconómica x_t que no fueron capturadas por ARIMA. El modelo MARS construye la relación entre la variable dependiente y los predictores a partir de funciones base *spline* seleccionadas de forma adaptativa. Esto le permite capturar no linealidades, umbrales, cambios de pendiente e interacciones entre variables sin necesidad de imponer de antemano una forma funcional única para toda la muestra. En este trabajo, el bloque MARS utiliza como predictores un conjunto de variables macroeconómicas x_t entre ellas el tipo de cambio nominal, el precio del petróleo, el precio del oro y la tasa de interés nominal anual para depósitos a 30 días en entidades financieras. Formalmente, esta segunda etapa puede expresarse como:

$$\hat{e}_t = g(x_t) + \varepsilon_t = \beta_0 + \sum_{m=1}^M \beta_m B_m(x_t) + \varepsilon_t \quad (14)$$

x_t son las variables que utiliza el modelo MARS (TC nominal, precio del petróleo y oro, tasa y mes).

Por último, se realiza un pronóstico combinado, sumando ambos componentes.

$$\hat{y}_{T+h} = \hat{L}_{T+h} + \hat{N}_{T+h} = \hat{y}_{T+h}^{ARIMA} + \hat{g}(x_{T+h}) \quad (15)$$

Así el modelo híbrido busca aprovechar simultáneamente la capacidad de ARIMA para representar la persistencia y la dinámica lineal de la inflación, y la capacidad de MARS para capturar relaciones no lineales vinculadas a variables macroeconómicas relevantes.

5. Metodología

5.1. Base de datos

Para la especificación del modelo se seleccionaron variables explicativas conforme la teoría macroeconómica como la evidencia empírica reciente en el contexto argentino, siguiendo algunos trabajos como los de Forte (2024), Garegnani y Gomez Aguirre (2018) y Aguilar (2024). También se tomaron en cuenta economías emergentes y con precios volátiles y en menor medida, economías desarrolladas (Álvarez et al., 2009).

En primer lugar, en este trabajo se utiliza la variación del **IPC** (Índice de Precios del Consumidor) como variable que refleja la tasa de inflación. La variable IPC es tomada del Instituto Nacional de Estadísticas y Censos (INDEC) de la Argentina, quien calcula el índice mediante una cesta de consumo representativa. La variable utilizada para este trabajo son variaciones porcentuales mensuales, la variable no es tomada como índices en niveles, y toma valores entre 0,2 y 25,5 en toda la serie. Esta decisión es consistente con la metodología empleada en otros trabajos para reflejar la inflación.

En segundo lugar, variables como la **tasa de interés** suelen ser incluidas por su rol indicador de la política monetaria, en la literatura se reconoce que las tasas de interés presionan sobre el nivel de general de precios de manera indirecta, afectando de manera directa el costo del crédito, la formación de expectativas y en los incentivos al ahorro y el gasto, lo cual es coherente con enfoques que integran factores no lineales para la predicción. En este trabajo, se llamará tasa de interés al promedio mensual de la tasa nominal anual para depósitos a 30 días en entidades financieras, obtenida del Banco Central de la República Argentina (BCRA).

Además, se incluyeron los precios de ciertas *commodities* internacionales, como el **precio de cierre promedio de los futuros del petróleo Brent (LCOU6)** y el **precio de cierre de los futuros del oro (GCQ6)**, son utilizados como indicadores de shocks externos y de incertidumbre global que pueden transmitirse a la inflación local, especialmente en economías sensibles a las condiciones internacionales.

Esto coincide con trabajos que han estudiado las relaciones entre las *commodities* y la inflación. Ambas series de precios son en dólares estadounidenses y fueron obtenidas de Investing.com.

Por otra parte, se incorporó un **componente estacional mensual**, incorporando la variable mes que toma valores entre 1 y 12 como una variable determinística con el objetivo de capturar patrones recurrentes en la inflación asociados al calendario, permitiendo que dichos efectos interactúen de manera no lineal con otras variables macroeconómicas dentro del marco del modelo MARS. Los datos fueron tomados desde enero de 2014 hasta julio de 2025, representando un total de 139 observaciones.

Por último, el **tipo de cambio nominal** oficial resulta ser un determinante fundamental de la inflación en Argentina debido a que existe cierta indexación de facto en la economía, con alta dolarización de precios relativos y la transmisión del comportamiento del dólar a los precios minoristas.

Una cuestión fundamental es que la disponibilidad de series mensuales de estos indicadores en periodos amplios permite a los modelos MARS identificar patrones no lineales, efectos de umbral e interacciones entre las variables sin imponer relaciones funcionales rígidas, lo que puede ser apropiado al combinar información internacional y doméstica para la predicción de la inflación en un contexto de alta volatilidad estructural como el Argentino.

Tabla 1

Listado de variables utilizadas

Nombre	Unidad de medida	Fuente
Inflación	Variación porcentual mensual	INDEC
Tipo de cambio	Pesos argentinos (\$) por dólar estadounidense (US\$)	BCRA
Precio del oro	Dolar estadounidense	Investing.com
Precio del petróleo	Dolar estadounidense	Investing.com
Tasa de interés	Unidades porcentuales	BCRA

Elaboración propia. Variables objetivo y predictoras, con su unidad de medida y su fuente.

La exclusión de otros determinantes documentados en la literatura argentina, en particular los agregados monetarios y el resultado fiscal (Graña Colella, 2020), responde a tres criterios de diseño. En primer lugar, el objetivo del trabajo es comparar el desempeño predictivo de especificaciones alternativas bajo un conjunto de información común y parsimonioso, de modo que las diferencias observadas sean atribuibles a la flexibilidad funcional de cada método y no al contenido informativo de conjuntos de variables distintos. En segundo lugar, dado el tamaño muestral disponible ($n=139$), la ampliación más variable (más candidatos) incrementaría el riesgo de sobreajuste y de extrapolación fuera del dominio de la serie histórica, problema que, como se menciona en la sección 6.6, constituye la principal debilidad de los modelos basados en splines en contextos de cambio de régimen. Por último, en un régimen de alta inflación los agregados nominales presentan tendencias compartidas y elevada colinealidad; la tasa de interés incluida ya opera como indicador del sesgo de la política monetaria, mientras que el tipo de cambio nominal captura el principal canal de transmisión nominal. Por su parte, el tipo de cambio paralelo se excluye por la ausencia de una serie oficial homogénea y porque su contenido informativo es dependiente del régimen cambiario, siendo la brecha aproximadamente nula en el tramo 2016-2019 de la muestra; su incorporación se propone como extensión en la sección 7.

5.2. Tratamiento de valores atípicos

Con el objetivo de evaluar la sensibilidad de los modelos a los valores extremos de la serie, se identificó las observaciones atípicas de la inflación a través del criterio estándar de rango intercuartílico (IQR). Definiendo como valores extremos las observaciones que se encuentran por debajo de $Q1 - 1,5 * IQR$ o por encima de $Q3 + 1,5 * IQR$ donde $Q1$ y $Q3$ representan el primer y tercer cuartil de la distribución, respectivamente.

De esta forma se logró identificar un total de 11 observaciones extremas, aquellas que corresponden al periodo comprendido entre abril de 2023 y abril de 2024. Estos valores coinciden con una etapa de aceleración inflacionaria significativa en la Argentina.

5.3. Estadísticos descriptivos y test.

5.3.1. Test de Diebold-Mariano (DM)

El *test* de Diebold-Mariano (Diebold y Mariano, 1995), se emplea para evaluar si existen diferencias estadísticamente significativas en la precisión predictiva de dos modelos. Se basa en la comparación de una función de pérdida cuadrática aplicada a los errores de pronóstico de cada modelo:

$g(e_{i,t}) = e_{i,t}^2$ función de pérdida cuadrática (equivalente a RMSE)

$d_t = g(e_{1,t}) - g(e_{2,t}) = e_{1,t}^2 - e_{2,t}^2$ diferencial de pérdida cuadrática

Donde, $e_{1,t} = y_t - \hat{y}_{1,t}$ y $e_{2,t} = y_t - \hat{y}_{2,t}$ son los errores del modelo 1 y del modelo 2 respectivamente.

Se construye un estadístico a partir del promedio de la diferencia de pérdida entre ambos modelos, ajustado por una estimación de su varianza que corrige por autocorrelación (Bartlett), adecuado a que $h_{test}=14$.

Estadístico: $DM = \frac{\underline{d}}{\sqrt{\hat{V}(\underline{d})}} \sim N(0,1)$

Siendo $\underline{d} = \frac{1}{h} \sum_{t=1}^h d_t$ y $\hat{V}(\underline{d})$ es una estimación consistente de la varianza de larga corrida (HAC) del proceso $\{d_t\}$.

$$\hat{V}(\underline{d}) = \frac{1}{h} [\hat{\gamma}_0 + 2 \sum_{k=1}^{h-1} w_k \hat{\gamma}_k] \quad (16)$$

Donde: $\hat{\gamma}_k = \frac{1}{h} \sum_{t=k+1}^h (d_t - \underline{d})(d_{t-k} - \underline{d})$ es la autocovarianza muestral de orden k y $w_k = 1 - \frac{k}{h}$ es el peso del kernel de Bartlett, que pondera más los rezagos cortos.

La hipótesis nula establece que ambos modelos tienen igual capacidad predictiva, mientras que la alternativa indica superioridad de uno sobre otro. Esta prueba es particularmente relevante en evaluaciones fuera de muestra y permite complementar métricas como RMSE o MAE con inferencia estadística formal.

5.4. Análisis de propiedades estadísticas

5.4.1. Propiedades de la serie

Con el objetivo de determinar las propiedades estocásticas de la serie de inflación mensual, se aplican las pruebas de Dickey-Fuller Aumentado (ADF), de Phillips-Perron (PP) y de Kwiatkowski-Phillips-Schmidt-Shin (KPSS).

Dickey y Fuller (1979), el *test* de ADF busca determinar si la serie presenta raíz unitaria, evidenciando tendencia estocástica, lo que invalida el supuesto de estacionariedad en covarianza. Para ello parte del proceso autorregresivo de orden p :

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d. (0, \sigma^2)$$

Si $\rho = 1$ la serie tiene raíz unitaria (*random walk*). Si $|\rho| < 1$ es estacionaria.

La prueba ADF reescribe el proceso en primeras diferencias para determinar directamente si $\delta = \rho - 1 = 0$:

$$\Delta y_t = \alpha + \beta_t + \delta y_{t-1} + \sum_{j=1}^k \gamma_j \Delta y_{t-j} + \varepsilon_t \quad (17)$$

Donde $\Delta y_t = y_t - y_{t-1}$, α es una constante, β_t es una tendencia determinística y los k rezagos de Δy_{t-j} corrigen la autocorrelación serial de los errores.

El estadístico de prueba (ecuación 18) no sigue una distribución de Student. Sus valores críticos fueron tabulados por Dickey y Fuller (1979) mediante simulación de Monte Carlo, depende del tamaño muestra T y de si se incluye constante y tendencia.

$$\tau = \frac{\hat{\delta}}{SE(\hat{\delta})} \quad (18)$$

Para la determinación del número de rezagos, la función del R implementa la regla de Schwert: $k_{max} = [12 * (T/100)^{1/4}]$. Con $T = 125$ observaciones de entrenamiento, $k_{max} \approx 13$. El número final se determina por AIC.

Por otro lado, Phillips y Perron (1988) presentan el *test* de Phillips-Perron que es una alternativa al ADF pero que corrige autorregresión y heterocedasticidad sin agregar rezagos. Parte de la misma ecuación diferenciada en primer grado de ADF y presenta la misma hipótesis nula pero ajusta el estadístico t de δ usando correcciones no paramétricas sobre la varianza de largo plazo del error:

$$Z_t = t_\delta - \text{corrección HAC}$$

HAC (*Heteroskedasticity and Autocorrelation Consistent*, representado por la ecuación 19), corrige la estimación de la varianza bajo el supuesto de heterocedasticidad:

$$\hat{\omega}^2 = \delta_0 + 2 \sum_{k=1}^q \left(1 - \frac{k}{q+1}\right) \hat{\delta}_k, \quad (19)$$

donde $1 - \frac{k}{q+1}$ es el kernel de Bartlett, $\hat{\delta}_k$ es la autocovarianza de los residuos y q es el *bandwidth*

Ahora bien, como complemento de estas pruebas, se emplea el *test* de KPSS. Propuesta por Kwiatkowski et al. (1992), la prueba KPSS determina si una serie es estacionaria alrededor de una media o una tendencia determinística y presenta como hipótesis nula que la serie es estacionaria. Para esto se descompone la serie de manera matemática como:

$$y_t = \delta_t + r_t + \varepsilon_t \quad (20)$$

Donde: r_t es *random walk* y ε_t es el error estacionario

Luego se construye un estadístico basado en la suma acumulada de los residuos de la regresión:

$$KPSS = \frac{1}{T^2} + \sum_{t=1}^T S_t^2 / \omega^2 \quad (21)$$

Donde $S_t = \sum_{i=1}^t \hat{\varepsilon}_i$ y ω^2 es mismo estimador HAC que se utiliza en Phillips-Perron

5.4.2. Test de multicolinealidad VIF (*Variance Inflation Factor*)

Con el objetivo de evaluar la presencia de multicolinealidad entre las variables explicativas, se realizó un *test* de multicolinealidad VIF. De acuerdo con Backhaus et al. (2018), el VIF (ecuación 22) es una

medida diagnóstica de multicolinealidad entre variables explicativas en modelos de regresión. Se calcula como:

$$VIF = \frac{1}{1-R_j^2} \quad (22)$$

Donde R_j^2 proviene de una regresión auxiliar de la variable explicativa “j” sobre el resto de los regresores. El VIF deriva de comparar la varianza del estimador de OLS en presencia de multicolinealidad con la varianza en un caso ideal de variables ortogonales:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{S_{jj}(1-R_j^2)} = \frac{\sigma^2}{S_{jj}} * VIF_j \quad (23)$$

Donde $S_{jj} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ es la suma de cuadrados de x_{ij} . Por lo tanto, el VIF es el factor multiplicativo por el cual la varianza se infla con respecto al caso ortogonal.

En la práctica, valores de VIF superiores a 5 o 10 indican problemas potenciales de multicolinealidad. Su relevancia radica en que altos niveles de colinealidad inflan la varianza de los coeficientes estimados, dificultando la inferencia y la interpretación económica (Backhaus et al., 2018).

5.4.3. Test de Ljung-Box

El test de Ljung-Box evalúa la presencia de autocorrelación conjunta en una serie de residuos hasta un determinado número de rezagos (Ljung y Box, 1978). Se basa en una suma ponderada de autocorrelaciones muestrales al cuadrado:

$$Q_{LB}(m) = T(T+2) + \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k} \quad (24)$$

Siendo: T el tamaño de la muestra, $\hat{\rho}_k = \frac{\sum_{t=k+1}^T \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^T \hat{e}_t^2}$ es la autocorrelación de orden k y m es el número de rezagos evaluados ($\log(T)$)

La distribución del estadístico para los parámetros AR y MA estimados ($\hat{p} + \hat{q}$)

$$Q_{LB}(m) \sim X^2(m - \hat{p} - \hat{q})$$

Para los modelos MARS, como no hay parámetros autorregresivos que descontar, se utiliza $X^2(m)$ directamente.

La hipótesis nula establece que no existe autocorrelación entre los residuos, mientras que la alternativa indica presencia de dependencia serial. Es fundamental en el diagnóstico de modelos de series de tiempo, ya que permite verificar si la dinámica temporal ha sido capturada correctamente.

5.4.4. Test de Shapiro-Wilk

Propuesta por Shapiro y Wilk (1965), la prueba demuestra si una serie presenta una distribución normal. Se basa en la estimación entre los datos ordenados y los valores esperados bajo normalidad, utilizando estimadores derivados de la distribución normal.

$$W = \frac{(\sum_{i=1}^T a_i e(i))^2}{\sum_{i=1}^T (e_i - \underline{e})^2} \quad (25)$$

Donde: $e(i)$ son los residuos ordenados de menor a mayor, a_i son coeficientes óptimos que dependen de T y provienen de los momentos y covarianzas de los estadísticos de orden de una distribución normal estándar y $\underline{e}$ es la media de los residuos.

El estadístico “W” toma valores entre 0 y 1, siendo 1 cuando la distribución es completamente normal. La hipótesis nula establece que los datos siguen una distribución normal, mientras que la alternativa indica desviaciones respecto a la normalidad. Es especialmente adecuado para muestras pequeñas y resulta relevante para evaluar la validez de intervalos de confianza y supuestos clásicos de inferencia.

5.4.5. Prueba ARCH-LM

La prueba ARCH (*Autoregressive Conditional Heteroskedasticity*), propuesta por Engle (1982), implementado en R a través de la prueba LM de McLeod-Li se utiliza para detectar heterocedasticidad condicional en los residuos de un modelo. Consiste en estimar una regresión auxiliar de los residuos al

cuadrado sobre sus propios rezagos, utilizando mínimos cuadrados ordinarios (MCO), y evaluar la significancia conjunta de los coeficientes mediante un estadístico tipo multiplicador de Lagrange.

Proceso ARCH $\varepsilon_t = \sigma_t z_t, \quad z \sim i.i.d., (0,1) \quad \sigma_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2$

Donde σ_t^2 es la varianza condicional en el tiempo t. Si $\alpha_1 = \alpha_2 = \dots = \alpha_q = 0$, la varianza es constante, por lo tanto la serie es homocedástica.

Entonces, la regresión de los residuos al cuadrado queda:

$$\hat{\varepsilon}_t^2 = \gamma_0 + \gamma_1 \hat{\varepsilon}_{t-1}^2 + \dots + \gamma_q \hat{\varepsilon}_{t-q}^2 + \mu_t \quad (26)$$

Si existen efectos ARCH, los cuadrados de los residuos están autocorrelacionados, por lo tanto, los periodos de alta volatilidad se agrupan.

La hipótesis nula establece ausencia de efectos ARCH (varianza constante), mientras que la alternativa indica presencia de heterocedasticidad condicional.

Estadístico: $LM = T * R^2 \sim X^2(q)$, R^2 es es coeficiente de correlación y T el número de residuos.

Esta prueba es particularmente relevante en series macroeconómicas volátiles, donde la varianza puede cambiar en el tiempo.

5.5. Especificación de los modelos

5.5.1. ARIMA

En este trabajo se estima un modelo ARIMA, el cual permite capturar tanto la persistencia como los shocks transitorios presentes en la serie. El procedimiento de selección utilizado fue basado en AIC.

Con el objetivo de evaluar la adecuación de un modelo híbrido (ARIMA + MARS) como representación de la dinámica de la inflación, se analizan los residuos de ARIMA para comprobar indicios de autocorrelación, normalidad y heterocedasticidad condicional y poder descartar que los residuos conformen ruido blanco.

El modelo híbrido se incluye como un ejercicio empírico exploratorio, con el fin de evaluar si los residuos del ARIMA contienen algún componente no lineal remanente que pueda ser capturado por MARS.

5.5.2. *MARS*

En esta primera especificación, el modelo se estima sin poda, lo que implica que el procedimiento de selección de términos se detiene únicamente al alcanzar el número máximo permitido, sin aplicar ningún criterio de penalización por complejidad. Asimismo, se permite un grado de interacción de hasta orden dos, habilitando la inclusión de efectos no lineales e interacciones entre variables.

Este es un modelo altamente parametrizado, con 17 términos seleccionados sobre un total posible de 17, incluyendo múltiples funciones base e interacciones entre variables como el tipo de cambio nominal, la tasa de interés, el precio del petróleo, el precio del oro y la estacionalidad mensual. En particular, se logra identificar numerosos nudos en variables clave, lo que sugiere la existencia de cambios en la relación entre inflación y sus determinantes a distintos niveles de estas variables.

5.5.3. *MARS con poda y dummies estacionales*

La estimación del modelo MARS con poda basado en validación cruzada generalizada (GCV) y la incorporación de *dummies* estacionales para cada mes del año permite evaluar explícitamente el *trade-off* entre capacidad de ajuste y parsimonia. Este modelo es comparado con el MARS sin poda en términos predictivos.

5.5.4. *MARS con validación cruzada (cv)*

La estimación del modelo MARS utilizando validación cruzada (10 veces replicado 5 veces) introduce un criterio de selección más exigente basado en desempeño fuera de muestra, en contraste con los enfoques tradicionales basados en ajuste interno como el GCV. Dado que el algoritmo de validación cruzada implica particiones aleatorias de los datos, se fijó una semilla aleatoria (`set.seed(42)`) antes de cada estimación con `pmethod=cv` para garantizar la reproducibilidad de los resultados. La misma semilla se aplicó en el *rolling window* y en la estimación de `MARS_CV_lag`.

5.5.5. *MARS con rezagos de primer y segundo orden*

La incorporación de rezagos de primer y segundo orden de la variable dependiente en el modelo MARS permite introducir explícitamente la dinámica autorregresiva dentro de una estructura originalmente

estática. Este ajuste resulta particularmente relevante en el caso de la inflación argentina, caracterizado por una elevada persistencia temporal (Forte, 2024), donde los valores pasados del índice de precios constituyen determinantes fundamentales de su evolución presente.

5.5.6. *MARS sin valores atípicos*

La estimación del modelo MARS sobre una muestra depurada de valores atípicos permite evaluar la sensibilidad de los resultados a episodios extremos, particularmente relevantes en el caso argentino, donde shocks discrecionales (cambiarior, monetarios o políticos) pueden dominar la dinámica de corto plazo. A diferencia de las especificaciones previas, este modelo se entrena excluyendo valores atípicos, pero se evalúa sobre el mismo conjunto de prueba, lo que garantiza la comparabilidad en términos de desempeño predictivo.

5.5.7. *Modelo híbrido*

Con el objetivo de capturar posibles estructuras no lineales que no están explicadas por el modelo lineal, se implementó un enfoque híbrido que combina un modelo ARIMA para modelar la dinámica temporal de la inflación, con un modelo MARS aplicado sobre los residuos del ARIMA. Este enfoque permite descomponer la señal en un componente lineal (ARIMA) y uno no lineal (MARS), bajo el supuesto de que los residuos aún contienen información estructurada aprovechable.

5.6. Métricas de evaluación

En la evaluación del desempeño predictivo de los modelos estimados se emplean tres métricas complementarias (Karadzic y Pejovic, 2021): el Error Cuadrático Medio (RMSE), el Error Absoluto Medio (MAE) y el Error Porcentual Absoluto Medio (MAPE). Estas permiten cuantificar la magnitud de los errores de predicción fuera de muestra, proporcionando distintos criterios para comparar la precisión de los modelos.

El RMSE (ecuación 27) se define como la raíz cuadrada del promedio de los errores al cuadrado entre los valores observados y los predichos. Esta métrica penaliza de manera más severa los errores grandes

debido a la elevación al cuadrado, lo que la hace especialmente sensible a desvíos extremos. En contextos como el análisis de la inflación, donde pueden observarse episodios de alta volatilidad, el RMSE resulta útil para identificar modelos que evitan errores significativos en períodos críticos. No obstante, esta misma característica puede sobrerepresentar el impacto de los valores atípicos en la evaluación global. Por su parte, el MAE (ecuación 28) corresponde al promedio de los valores absolutos de los errores de predicción. A diferencia del RMSE, no amplifica los errores grandes, lo que lo convierte en una medida más robusta frente a valores atípicos. Su interpretación es directa, ya que expresa el error promedio en las mismas unidades que la variable analizada (en este caso, puntos porcentuales de inflación mensual), facilitando la comparación entre modelos desde una perspectiva más intuitiva.

Finalmente, el MAPE (ecuación 29) mide el error absoluto en términos relativos, expresándolo como porcentaje del valor observado. Esta métrica permite evaluar la magnitud del error en relación con el nivel de la variable, lo que resulta útil para comparar precisión en distintos rangos de inflación. Sin embargo, presenta limitaciones importantes cuando los valores observados son cercanos a cero, ya que pequeñas variaciones pueden generar errores porcentuales desproporcionadamente elevados. Por este motivo, su interpretación debe realizarse con cautela en períodos de baja inflación.

Estas métricas son calculadas:

$$RMSE = \sqrt{\frac{1}{h} \sum_{i=1}^h (y_{T+i} - \hat{y}_{T+i})^2} \quad (27)$$

$$MAE = \frac{1}{h} \sum_{i=1}^h |y_{T+i} - \hat{y}_{T+i}| \quad (28)$$

$$MAPE = \frac{100}{h} \sum_{i=1}^h \left| \frac{y_{T+i} - \hat{y}_{T+i}}{y_{T+i}} \right| \quad (29)$$

Donde h es el horizonte de pronóstico, y_{T+i} son los valores reales del conjunto del test y $\hat{y}_{T+i}$ son los pronósticos respectivos del modelo.

En conjunto, el uso combinado de estas tres métricas permite obtener una evaluación más completa del desempeño de los modelos. Mientras que el RMSE y el MAE se emplean como criterios principales debido a su estabilidad y consistencia, el MAPE se utiliza como medida complementaria para capturar la dimensión relativa del error, reconociendo sus limitaciones en determinados contextos.

5.7. Partición de datos y estrategia de validación

5.7.1. Partición de datos

Se realizó una división cronológica de la serie para poder evaluar la capacidad predictiva de los modelos, separando las primeras observaciones para el entrenamiento y utilizando las más recientes para el análisis fuera de la muestra. El empleo de esta metodología es apropiado cuando se trabaja con series temporales, porque mantiene el orden cronológico de los datos y previene la utilización de información futura para realizar estimaciones. De acuerdo con lo que afirman Hyndman y Athanasopoulos (2021), segmentar un conjunto de entrenamiento de uno de prueba es una práctica estándar en la evaluación predictiva y el tamaño del bloque de test debe definirse en función de la longitud de la muestra disponible y del horizonte de pronóstico de interés.

En esta investigación, se empleó una división 90%-10% de un total de 139 observaciones mensuales. Así, las primeras 125 observaciones sirvieron para estimar los modelos y las últimas 14 fueron asignadas a la estrategia de validación utilizada (*hold-out* estático). Lo que se hace es reservar un bloque de observaciones finales de la serie, que no forman parte del grupo de entrenamiento de ningún modelo, para utilizarlos exclusivamente en el testeo de la capacidad predictiva fuera de la muestra. La partición cronológica es estática porque se realiza una sola vez y su origen y tamaño no cambian: *train* (enero de 2014 a mayo de 2024) y *test* (junio de 2024 a julio de 2025).

Esta resolución responde a la necesidad de conciliar dos metas: por un lado, mantener una cantidad apropiada de datos para calcular correctamente las especificaciones consideradas y, por otro lado, tener

un bloque de evaluación fuera de muestra con un tamaño mínimo razonable para poder comparar el desempeño predictivo de los modelos. En este sentido, la elección de $h_{\text{test}}=14$, equivalente al 10% de la muestra, resulta más adecuada que una partición más reducida.

No obstante, esta estrategia supone una restricción metodológica, dado que el conjunto de evaluación queda conformado por un número relativamente pequeño de observaciones. En consecuencia, métricas como el RMSE pueden presentar cierta sensibilidad frente a valores atípicos, mientras que pruebas de comparación de precisión predictiva, como el *test* de Diebold-Mariano, pueden perder potencia estadística para detectar diferencias sistemáticas entre modelos. Por ello, los resultados obtenidos a partir de esta partición deben interpretarse como evidencia indicativa, aunque no concluyente, respecto de la superioridad predictiva relativa de las distintas especificaciones.

5.7.2. Validación estadística

Para complementar la evaluación fuera de muestra, que se basa en una partición estática, se implementa una estrategia de validación mediante *rolling window*. Este enfoque permite analizar la estabilidad temporal del desempeño predictivo de los modelos y evaluar su capacidad de adaptación frente a cambios en la dinámica de la serie.

El procedimiento consiste en definir un tamaño inicial de muestra de entrenamiento y, a partir de allí, generar pronósticos recursivos de un paso adelante. En cada iteración, el modelo se reestima incorporando una nueva observación al conjunto de entrenamiento, manteniendo fijo el origen temporal inicial.

Para cada iteración $i = 0, 1, \dots, \tau - 1$:

$\tau_i = \{1, 2, \dots, T_0 + i\}$ Conjunto de entrenamiento creciente

$Test_i = T_0 + i + 1$ Una observación del test

Donde $T_0 = n_{\text{total}} - \text{ventanas} = 139 - 41$ es el tamaño inicial del entrenamiento y τ es el número de ventanas.

De este modo, la ventana de estimación se expande progresivamente, mientras que el conjunto de evaluación se desplaza a lo largo del tiempo. Este esquema replica de manera más realista el proceso de predicción en tiempo real, donde la información disponible crece secuencialmente.

A diferencia del enfoque estático, que evalúa el desempeño en un único bloque temporal, que puede estar influenciado por un régimen específico, el *rolling window* permite obtener una medida promedio del error a lo largo de múltiples contextos macroeconómicos. En particular, resulta especialmente útil en presencia de quiebres estructurales o cambios de régimen, como los observados en la inflación Argentina reciente, ya que exponen la sensibilidad de los modelos a distintas fases del ciclo inflacionario.

La métrica de evaluación se construye a partir del promedio de los errores de predicción acumulados en cada iteración, utilizando específicamente el RMSE (ecuación 30). Esto permite comparar la robustez intertemporal de los modelos, más allá de su desempeño puntual en un período específico. En este sentido, es posible que un modelo que presente un bajo error en el esquema de retención no mantenga dicho desempeño bajo validación rodante, evidenciando problemas de sobreajuste o dependencia de un régimen particular.

$$RMSE_{RW} = \sqrt{\frac{1}{\tau} \sum_{i=0}^{\tau-1} (y_{T_0+i+1} - \hat{y}_{T_0+i+1|\tau_i})^2} \quad (30)$$

En síntesis, la inclusión de la *ventana rodante* como estrategia de validación aporta una dimensión dinámica al análisis comparativo, permitiendo distinguir entre modelos que son óptimos en condiciones específicas y aquellos que exhiben mayor estabilidad y capacidad de generalización a lo largo del tiempo.

6. Resultados

6.1. Propiedades estadísticas y test de multicolinealidad

En primer lugar, el *test* ADF estándar arroja un estadístico de -4,0177 con un p-valor de 0,0107, lo que permite rechazar la hipótesis nula de raíz unitaria al 5%. Este resultado sugiere que la serie está estacionaria en niveles. De esta misma manera, la versión del ADF con selección de rezagos mediante

criterio de Akaike (AIC) también rechaza la hipótesis nula, dado que el estadístico $\tau_2 = -3,4605$ es menor al valor crítico al 5% (-2,88).

Por su parte, el *test* de Phillips-Perron, que resulta más robusto en presencia de heterocedasticidad en los errores, también rechaza la hipótesis nula de raíz unitaria ($p\text{-valor} \leq 0,01$), reforzando la evidencia a favor de la estacionariedad de la serie.

Sin embargo, el *test* KPSS, cuya hipótesis nula es la estacionariedad, arroja un estadístico de 1,3883 con $p\text{-valor} \leq 0,01$, lo que implica rechazar dicha hipótesis. Este resultado contradice parcialmente la evidencia obtenida mediante ADF y PP.

La coexistencia de estos resultados sugiere que la serie no presenta una raíz unitaria en sentido estricto, pero tampoco es estacionaria en un sentido fuerte. En particular, la evidencia es consistente con la presencia de cambios de régimen, heterocedasticidad y episodios de alta volatilidad, característicos de la dinámica inflacionaria argentina reciente, especialmente durante el período 2023-2024.

En este contexto, se interpreta que la serie de inflación mensual es estacionaria en media pero no en varianza, lo que implica que puede ser modelada en niveles sin necesidad de diferenciación previa, aunque incorporando metodologías capaces de capturar no linealidades y cambios de régimen, como MARS o modelos híbridos.

Por otra parte, se estimó el VIF a partir de un modelo lineal auxiliar donde la inflación en niveles se regresa sobre el conjunto de regresores considerados en la especificación MARS original.

Los resultados obtenidos indican valores de VIF relativamente bajos para todas las variables, tipo de cambio nominal (2,55), precio del petróleo (1,15), precio del oro (2,31), tasa de interés (2,41) y variable temporal mensual (1,06). De acuerdo con los criterios convencionales, los valores inferiores a 5 sugieren ausencia de multicolinealidad severa, mientras que los valores por encima de 10 indicarían problemas graves de redundancia entre regresores.

En este sentido, los resultados permiten concluir que no existe evidencia de multicolinealidad relevante entre las variables explicativas, lo que implica que cada una de ellas aporta información diferenciada al modelo y no presenta relaciones lineales fuertes con el resto.

No obstante, esta conclusión debe ser matizada en función de las propiedades estocásticas de las series involucradas. En particular, dado que los *tests* de estacionariedad presentan evidencia mixta, con indicios de estacionariedad en medios pero no en varianza, es posible que existan tendencias compartidas o cambios de régimen que no sean completamente capturados por el VIF. En este contexto, la multicolinealidad podría estar subestimada o manifestarse de forma no lineal, especialmente considerando que el modelo MARS permite capturar interacciones y efectos no lineales entre variables. Por lo tanto, si bien el diagnóstico estándar no evidencia problemas de multicolinealidad severa, se reconoce que la estructura económica está caracterizada por alta volatilidad, shocks comunes y posibles quiebres estructurales, lo que podría generar dependencias más complejas entre las variables, reforzando así la elección de metodologías flexibles como MARS frente a enfoques puramente lineales.

6.1.1 Estimación de modelos ARIMA y distintos tipos de MARS

A continuación, se muestran los resultados de los modelos y tests descriptos en las secciones 5.5 a 5.7.

En primer lugar, se identificó como especificación óptima un **modelo ARIMA (2,1,3)**. Esto implica que la especificación aplica una diferenciación de orden uno ($d=1$) sobre la serie modelada, la tasa de inflación mensual, es decir, la variación porcentual del IPC, y no sobre el índice de precios en niveles, e incorpora dos términos autorregresivos ($p=2$) y tres términos de medias móviles ($q=3$).

La presencia de diferenciación de orden uno resulta consistente con la evidencia empírica mixta obtenida en los tests de estacionariedad (ADF, Phillips-Perron y KPSS), sugiriendo que la serie presenta cambios de régimen o componentes no estacionarios que deben ser capturados para modelar adecuadamente su dinámica.

Los coeficientes estimados muestran una estructura dinámica compleja. Los términos autorregresivos capturan la persistencia inflacionaria, reflejando la inercia típica de la inflación argentina, donde los

valores pasados influyen significativamente sobre los actuales. Por su parte, los términos de medios móviles recogen la respuesta de la serie frente a shocks no anticipados, permitiendo modelar ajustes de corto plazo ante perturbaciones.

En primer lugar, el *test* de Ljung-Box (con lag = 12) arroja un p-valor de 0,5204, lo que implica que no se rechaza la hipótesis nula de ausencia de autocorrelación. Esto indica que el modelo ARIMA logra capturar adecuadamente la dependencia temporal de la serie, eliminando la estructura lineal en los residuos. Entonces, se puede considerar que los errores no están correlacionados en el tiempo, lo que sugiere una correcta especificación dinámica del modelo en los medios condicionales.

En segundo lugar, la prueba de Shapiro-Wilk arroja un p-valor cercano a cero, lo que conduce a rechazar la hipótesis nula de normalidad. Esto indica que los residuos no presentan una distribución normal, probablemente asociada a la presencia de asimetrías o colas pesadas, características encontradas frecuentemente en series macroeconómicas con episodios de alta volatilidad. No obstante, esta desviación de la normalidad no invalida la capacidad predictiva del modelo, aunque sí afecta la validez de los intervalos de confianza basados en supuestos gaussianos.

En tercer lugar, la prueba ARCH-LM (lags = 12) también arroja un p-valor cercano a cero, lo que deja en evidencia la presencia de heterocedasticidad condicional en los residuos al rechazar la hipótesis nula de homocedasticidad. Este resultado indica que la varianza de los errores no es constante en el tiempo, sino que presenta agrupamientos de volatilidad, fenómeno consistente con la dinámica inflacionaria Argentina reciente.

En conjunto, estos resultados permiten concluir que los residuos del modelo ARIMA no constituyen ruido blanco puro. Si bien no presentan autocorrelación, condición necesaria para considerar correctamente especificada la media, sí exhiben no normalidad y heterocedasticidad condicional. Esta evidencia tiene implicancias directas para la justificación del modelo híbrido. No obstante, dado que no se detecta autocorrelación significativa en los residuos y la principal regularidad restante parece asociarse

a heterocedasticidad condicional, es posible que el aporte predictivo de MARS sobre los residuos sea limitado. Donde esta cuestión se evalúa posteriormente en la sección de resultados.

Sin embargo, la presencia de heterocedasticidad condicional sugiere que la principal limitación del modelo ARIMA reside en la especificación de la varianza, no de la media. En este contexto, un enfoque más adecuado para capturar esta característica sería la utilización de modelos de la familia ARCH/GARCH, diseñados específicamente para modelar la dinámica de la volatilidad.

Respecto del ajuste de modelo **MARS**, el mismo presenta un coeficiente de determinación elevado ($RSq = 0,923$), lo que indicaría una alta capacidad para explicar la variabilidad de la inflación dentro de la muestra. Sin embargo, este indicador resulta engañoso si no se considera el coeficiente de determinación generalizado ($GRSq = 0,832$), que penaliza la complejidad del modelo. La brecha significativa entre ambas métricas constituye evidencia clara de sobreajuste, es decir, de que el modelo está capturando no solo la señal sistemática sino también el ruido idiosincrático de la muestra de entrenamiento.

Este comportamiento se refuerza por el hecho de que el modelo retiene la totalidad de los términos generados, lo que implica que no existe ningún mecanismo de selección que discrimine entre relaciones robustas y patrones espurios. En consecuencia, el bajo error *in-sample* obtenido no puede interpretarse como evidencia de superioridad predictiva, sino como resultado de una excesiva flexibilidad del modelo. Desde una perspectiva económica, la proliferación de términos e interacciones dificulta además la interpretación de los resultados, ya que múltiples efectos dependen de umbrales específicos y combinaciones de variables que no necesariamente representan relaciones estructurales estables en el tiempo.

En este sentido, este modelo cumple el rol de ilustrar el *trade-off* entre ajuste y capacidad de generalización. Su inclusión permite evidenciar que un alto desempeño dentro de la muestra no garantiza una buena performance fuera de ella, y justifica la necesidad de aplicar procedimientos de poda y validación cruzada en las especificaciones posteriores.

En el modelo de **MARS con poda y dummies estacionales**, se observa una mejora sustancial respecto del MARS en el ajuste generalizado (GRSq pasa de 0,831 a 0,908), junto con una reducción del número de términos (de 17 a 11), lo cual indica una menor complejidad efectiva del modelo y una potencial mejora en su capacidad de generalización fuera de muestra. También, se evidencia que los nudos asociados a la variable tipo de cambio nominal difieren respecto de la determinación sin poda, lo cual sugiere que el proceso de selección automática ha identificado umbrales más relevantes desde el punto de vista predictivo. Estos puntos pueden interpretarse como cambios de régimen en la dinámica inflacionaria, posiblemente asociados a episodios de devaluación o saltos discretos en el nivel del tipo de cambio, características de economías con alta nominalidad como la argentina.

Por otro lado, el modelo selecciona únicamente un subconjunto de las variables dummy mensuales (5 de 11), lo cual indica que no todos los meses presentan efectos estacionales estadísticamente relevantes una vez controladas las demás covariables. En particular, los meses seleccionados (febrero, marzo, agosto, octubre y diciembre) podrían estar capturando patrones asociados a dinámicas específicas de formación de precios, tales como ajustes post-vacacionales, recomposición de precios regulados o efectos de cierre de año, lo que sugiere una estacionalidad no uniforme a lo largo del calendario.

Un aspecto relevante del modelo es la presencia de términos de interacción, especialmente entre el tipo de cambio nominal y la función *spline* de la tasa de interés en niveles elevados. El término $TC_{nom} * h(Tasa - 74,7794)$ indica que, en un régimen de tasas de interés altas, el impacto del tipo de cambio sobre la variable dependiente se intensifica. Esto es consistente con un contexto macroeconómico donde condiciones monetarias restrictivas coexisten con presiones cambiarias, amplificando el traspaso a precios.

Finalmente, antes de evaluar formalmente el desempeño predictivo fuera de muestra, es razonable esperar una mejora respecto al modelo sin poda, aunque no necesariamente de magnitud proporcional a la mejora *in-sample*. Esto se debe a que la poda reduce el sobreajuste, pero el incremento en GRSq ya sugiere que parte de la mejora responde a una mejor captación de patrones estructurales y no únicamente a ruido. Sin

embargo, en contextos con alta volatilidad e inestabilidad estructural, como el analizado, la ganancia fuera de muestra suele ser más moderada que la observada en métricas internas.

Por otra parte, el modelo **MARS con validación cruzada (cv)** selecciona únicamente 5 de los 23 términos candidatos, lo cual evidencia que una proporción significativa de los términos incluidos en especificaciones previas no posee capacidad predictiva genuina, sino que responde a sobreajuste sobre la muestra de entrenamiento.

En particular, la reducción drástica en el número de términos sugiere que muchas de las interacciones y no linealidades previamente identificadas no son estables fuera de muestra. Este resultado es consistente con la literatura empírica en contextos macroeconómicos volátiles, donde la alta dimensionalidad tiende a capturar ruido en lugar de señal estructural.

La comparación entre el modelo seleccionado por validación cruzada y el modelo con poda basado en GCV refuerza esta interpretación. Mientras que el modelo con poda *backward* presenta un ajuste *in-sample* superior ($RSq = 0,9417$ y $GRSq = 0,9086$ vs $RSq 0,8943$ y $GRSq 0,8750$), su desempeño fuera de muestra, medido a través del coeficiente de determinación promedio *out-of-fold* ($mean.oof.RSq$), es considerablemente inferior ($\approx 0,43$). En contraste, el modelo con validación cruzada, a pesar de su menor complejidad (5 términos), alcanza un $mean.oof.RSq$ significativamente mayor ($\approx 0,66$), lo que implica una mejora sustancial en la capacidad predictiva genuina.

Este resultado pone en evidencia un caso claro de sobreajuste en el modelo más complejo, el incremento en métricas de ajuste interno no se traduce en una mejora en la predicción fuera de muestra, sino que, por el contrario, la deteriora. En consecuencia, la validación cruzada actúa como un mecanismo disciplinador que penaliza la inclusión de términos espurios.

Ahora bien, el modelo final retiene únicamente tres predictores: el tipo de cambio nominal, la tasa de interés y un componente estacional específico ($Mes_factor2$). La persistencia de estas variables a lo largo del proceso de selección sugiere que constituyen los determinantes más robustos y estables de la dinámica

inflacionaria en el período analizado. En particular, la interacción entre el tipo de cambio y un régimen de tasas elevadas refuerza la hipótesis de que el *pass-through* cambiario se intensifica en contextos de tensión monetaria, mientras que la estacionalidad capturada por el mes de febrero podría reflejar patrones sistemáticos en la formación de precios.

En síntesis, la evidencia empírica sugiere que un modelo más parsimonioso, validado mediante esquemas de validación cruzada, no solo es preferible desde el punto de vista metodológico, sino que además ofrece una representación más confiable de las relaciones económicas, evitando inferencias basadas en patrones espurios.

En términos de importancia de variables en el **modelo MARS con primer y segundo rezago**, se observa que el rezago de segundo orden presenta una mayor relevancia relativa que el rezago inmediato. Este resultado sugiere que la dinámica inflacionaria no responde únicamente a efectos de corto plazo, sino que incorpora componentes de inercia más estructurales, donde los niveles de inflación de períodos anteriores continúan afectando la dinámica actual incluso con cierto rezago. En contextos de alta inflación, esto puede interpretarse como evidencia de mecanismos de indexación o formación de expectativas adaptativas.

Un resultado particularmente significativo es la interacción entre la tasa de interés en niveles elevados y el rezago de inflación, representado por el término $h * (Tasa - 70,3845) * IPC_Lag2$, cuyo coeficiente negativo indica que, en un régimen de tasas altas, el impacto de la inflación pasada sobre la inflación actual se atenúa. Esto sugiere que una política monetaria contractiva podría estar contribuyendo a debilitar los mecanismos de propagación inercial de la inflación, reduciendo su persistencia en el tiempo. Además, el modelo identifica de manera endógena un punto de quietud en el nivel de inflación mensual, ubicado en torno al 7%, a partir del cual la dinámica cambia de régimen. Este umbral aparece en múltiples términos de interacción asociados al rezago de primer orden, lo cual refuerza su relevancia estructural. Por debajo de este valor, la inflación parece responder a una dinámica relativamente más estable,

mientras que por encima del mismo se observa un proceso de aceleración inflacionaria consistente con episodios recientes de alta nominalidad en la economía argentina.

Este cambio de régimen no actúa de manera aislada, sino que interactúa con otras variables del modelo.

En particular, se observa que la dinámica inflacionaria en el régimen de alta inflación depende del nivel del tipo de cambio nominal y de variables externas como el precio del oro, lo que sugiere que, en contextos de mayor inestabilidad, la inflación se vuelve más sensible a factores financieros.

Finalmente, si bien el modelo presenta mejoras sustanciales en términos de ajuste en muestra ($RSq = 0,9677$ y $GRSq = 0,9366$), es necesario señalar que estos resultados no son directamente comparables con especificaciones previas debido a la pérdida de observaciones derivadas de la inclusión de rezagos. Asimismo, la mayor complejidad del modelo incrementa el riesgo de sobreajuste, por lo que su validación fuera de muestra resulta fundamental para evaluar la robustez de estos hallazgos.

En el modelo **MARS sin valores atípicos**, un resultado central es la pérdida de relevancia del tipo de cambio nominal como predictor directo, el cual solo aparece en términos de interacción y no como efecto principal. Esto sugiere que el *pass-through* cambiario no opera de manera lineal ni constante en el tiempo, sino que está fuertemente condicionado por episodios extremos. En otras palabras, el impacto del tipo de cambio sobre la inflación parece concentrarse en momentos de disrupción, mientras que en períodos más “normales” su efecto directo se diluye.

Este hallazgo implica que el tipo de cambio en Argentina no tiene una naturaleza estructural. Es decir, no es un mecanismo que actúa de forma continua y predecible, sino que se activa con mayor intensidad en contextos de estrés macroeconómico, como devaluaciones abruptas o crisis de confianza. Al excluir estos episodios del entrenamiento, el modelo deja de identificar al tipo de cambio como un determinante principal de la inflación.

En contraste, variables como la tasa de interés, el precio del petróleo y el precio del oro adquieren mayor relevancia en esta especificación. Esto sugiere que, en ausencia de shocks extremos, la dinámica

inflacionaria responde en mayor medida a condiciones monetarias internas y factores externos vinculados a precios internacionales y activos de cobertura.

En el **modelo híbrido**, el MARS sobre residuos selecciona 8 términos relevantes, donde las variables con mayor importancia son el precio del oro y ciertos efectos estacionales (particularmente los meses de agosto, noviembre y diciembre), junto con interacciones con la tasa de interés. Este resultado indica que los errores sistemáticos del ARIMA no son aleatorios, sino que se concentran en períodos específicos y bajo determinadas condiciones macroeconómicas.

En particular, la aparición recurrente de meses como agosto, noviembre y diciembre puede asociarse a patrones estacionales propios de la economía argentina, caracterizados por mayor presión cambiaria, cierre de balances fiscales, ajustes de precios regulados y mayor incertidumbre macroeconómica. La interacción con niveles elevados del precio del oro sugiere además un vínculo con contextos internacionales de alta aversión al riesgo, donde el oro actúa como activo refugio.

Un resultado relevante es la ausencia del tipo de cambio nominal en el modelo MARS sobre residuos. Esto indica que el modelo ARIMA ya logró capturar adecuadamente la dinámica asociada a esta variable, dejando como remanente principalmente efectos no lineales vinculados a otras variables. En términos metodológicos, esto refuerza la interpretación de que el componente lineal del modelo absorbe la mayor parte de la señal sistemática del tipo de cambio.

No obstante, al analizar el período de evaluación fuera de muestra, se observa una potencial inestabilidad en las relaciones aprendidas. Durante este período, el precio del oro se mantuvo en niveles históricamente elevados, mientras que la inflación mostró una tendencia descendente. Esto implica una reversión parcial de la relación histórica estimada por el modelo, donde previamente niveles altos de oro se asociaban con mayor inflación.

Esta discrepancia sugiere que el modelo MARS ha capturado relaciones empíricas válidas en el período de entrenamiento, pero que no necesariamente se mantienen estables en contextos de cambio de régimen.

En consecuencia, el problema no radica exclusivamente en una mala especificación del modelo, sino en la naturaleza no estacionaria y altamente cambiante del proceso inflacionario argentino.

En síntesis, el modelo híbrido se encuentra parcialmente justificado, si bien el ARIMA logra capturar adecuadamente la dinámica temporal y parte de la señal estructural (incluido el efecto cambiario), el MARS aporta valor al identificar patrones no lineales y estacionales en los residuos. Sin embargo, la estabilidad de estos patrones fuera de muestra es limitada, lo que introduce restricciones en su capacidad predictiva en contextos de cambio estructural.

6.2. Desempeño predictivo

La evaluación del desempeño predictivo de los modelos se realizó mediante las métricas RMSE, MAE y MAPE, como se muestra en la tabla 2, considerando tanto el ajuste dentro de muestra (*in-sample*) como la capacidad predictiva fuera de muestra (*out-sample*). Los resultados evidencian diferencias sustanciales entre modelos, tanto en términos de precisión como de estabilidad, lo cual se observa con claridad al contrastar las métricas con la dinámica temporal de la inflación presentada en la figura 1.

En congruencia con Forte, 2024 y Aguilar, 2024, donde los modelos de *machine learning* evidencian mejores resultados predictivos, el modelo MARS con validación cruzada (MARS_CV) presenta el menor error fuera de muestra (RMSE = 0,9454), seguido muy de cerca por el modelo MARS con *poda y dummies* (RMSE = 0,9488). Esta diferencia es marginal, lo que sugiere que ambos modelos poseen capacidades predictivas estadísticamente similares. Sin embargo, el primero se destaca por su mayor parsimonia (5 términos frente a 11), lo que indica que captura la señal relevante, evitando el sobreajuste. Esta característica se refleja en el gráfico, durante el período de evaluación (lado derecho de la línea punteada), ambas curvas siguen de cerca la trayectoria descendente de la inflación.

Un resultado que a simple vista podría parecer contraintuitivo es que el RMSE *out-sample* del MARS-CV sea inferior al *in-sample*. Sin embargo, el gráfico muestra claramente la razón, el período de entrenamiento incluye episodios de extrema volatilidad y picos inflacionarios (especialmente durante 2023–2024), mientras que el período de evaluación presenta una fuerte desaceleración. Es decir, el

modelo fue entrenado en un entorno mucho más ruidoso que en el que se evalúa. Por lo tanto, el menor error fuera de muestra no es evidencia de sobreajuste inverso, sino de un cambio en la dificultad del problema predictivo.

Por otro lado, el modelo ARIMA presenta un desempeño significativamente inferior (RMSE fuera de muestra = 4,97). Esto se hace evidente en el gráfico, donde el ARIMA no logra capturar ni los picos inflacionarios ni la posterior caída, mostrando una trayectoria excesivamente suavizada y rezagada respecto a la serie observada. Este resultado coincide con el de Aguilar, 2024, donde el modelo ARIMA mostró predicciones inferiores a las observadas debido a que en la serie fuera de muestra la inflación presenta una marcada aceleración. Este comportamiento confirma su limitación estructural, al ser un modelo lineal, no puede adaptarse adecuadamente a cambios de régimen ni a dinámicas no lineales propias de la inflación argentina reciente. Cabe aclarar que la selección bajo criterio AICc resulta en una especificación ARIMA(2,1,3). Sin embargo, se exploró también la alternativa sin diferenciar, ARIMA(2,0,3), consistente con el resultado de los tests de raíz unitaria ADF y Phillips-Perron. Sin embargo, esa especificación produjo errores estándar no estimables (NaN) en cuatro de los seis coeficientes, indicando una matriz de información casi singular y sugiriendo problemas de invertibilidad en el polinomio MA. Esto es coherente con el diagnóstico mixto de estacionariedad: el rechazo de KPSS atribuible a cambios de régimen, junto con la inestabilidad numérica del modelo en niveles, sugiere que la serie no es bien representada por un proceso ARIMA estándar en ningún esquema (con o sin diferenciación). Esta limitación estructural del *benchmark* lineal motiva metodológicamente el recurso a especificaciones más flexibles como MARS, capaces de adaptarse a quiebres y no linealidades.

El caso del modelo híbrido es aún más problemático (RMSE = 28,47). En el gráfico, esto se observa de manera extrema, las predicciones del híbrido divergen violentamente del comportamiento real en el período de evaluación, incluso generando valores negativos o explosivos. Esto no es ruido, es evidencia directa de cierta inestabilidad estructural. El modelo aprendió relaciones no lineales durante el período de entrenamiento (por ejemplo, asociadas a niveles altos del precio del oro) que se rompen por completo

en el período de prueba. No es un problema menor, invalida el modelo como herramienta predictiva, en contraposición a lo hallado en Zhang 2003, donde el modelo híbrido resulta más preciso a la hora de capturar la variable objetivo.

Por su parte, el modelo MARS con rezagos muestra un patrón clásico de sobreajuste. Aunque presenta un excelente ajuste dentro de la muestra, se observa que sus predicciones fuera de muestra se desalinean fuertemente de la serie real, elevando fuertemente el RMSE. Esto indica que la incorporación de memoria temporal en un contexto de cambios estructurales (como el proceso inflacionario argentino reciente) no estabiliza el modelo, sino que amplifica los errores cuando el régimen cambia.

Finalmente, el modelo MARS-clean, estimado sin valores atípicos, también evidencia una limitación importante. Como en el modelo anterior, sus resultados *out-of-sample* sufren un gran deterioro en la capacidad predictiva, no logrando anticipar adecuadamente los movimientos de la inflación. Esto confirma que eliminar *outliers* en el entrenamiento reduce la capacidad del modelo para adaptarse a escenarios reales en los que esos eventos extremos sí ocurren. En otras palabras, mejora el ajuste “académico”, pero empeora la utilidad práctica.

En síntesis, la evidencia conjunta de métricas y análisis gráfico convergen en una misma conclusión, los modelos MARS con control de complejidad (especialmente MARS-CV) logran el mejor equilibrio entre ajuste y generalización. La parsimonia del MARS-CV no solo reduce el riesgo de sobreajuste, sino que se traduce en una mayor estabilidad predictiva frente a cambios de régimen, lo cual es crítico en contextos macroeconómicos altamente volátiles como el analizado.

Tabla 2

Resultados de los modelos fuera y dentro de la muestra

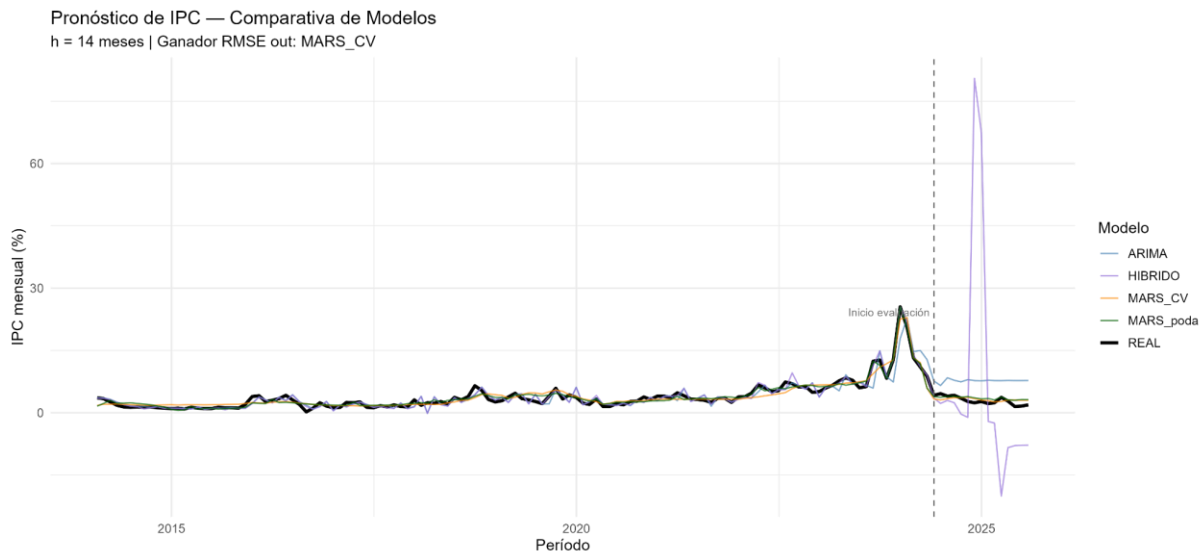
Modelo	RMSE_In	RMSE_Out	MAE_Out	MAPE_Out
ARIMA	1,5239	4,9714	4,8432	202,1819
MARS_sin_poda	0,9909	1,2950	1,1665	50,3703
MARS_poda_	0,8612	0,9488	0,7930	36,7064

dum				
MARS_CV	1,1597	0,9454	0,8227	34,8049
MARS_lag(*)	0,6457	5,8615	5,2868	231,3581
MARS_clean	0,6042	3,9736	2,6563	103,6438
HÍBRIDO	0,9093	28,4667	16,3141	656,6228

Fuente: elaboración propia. Nota. MARS_lag presenta un periodo de evaluación distinto (no es comparable de manera concluyente). La tabla muestra los errores calculados de cada modelo, fuera y dentro de la muestra.

Figura 1

Serie original y pronósticos de la variación IPC por modelo



Fuente: elaboración propia. Nota. El gráfico representa la serie original de la inflación y los pronósticos de los modelos.

6.3. Modelo MARS_CV

Como se muestra en la tabla 3, el modelo selecciona 5 términos, de un máximo posible de 23 y utiliza únicamente 3 predictores relevantes que son, tasa de interés, tipo de cambio nominal y estacionalidad de febrero, lo que refleja un alto grado de parsimonia. El coeficiente de determinación ajustado ($GRSq = 0,875$) y el desempeño fuera de muestra ($mean.oof.RSq = 0,665$) indican una adecuada capacidad predictiva sin evidencia de sobreajuste significativo.

Desde el punto de vista estructural, se identifica un punto de quiebre con respecto a la tasa de interés en torno a 74,78%, a partir del cual la relación con la inflación cambia de forma no lineal.

En particular, el término $h(74,7794 - tasa)$ captura el efecto cuando la tasa se encuentra por debajo del umbral. Su coeficiente negativo (-0,096) indica que, en este rango, reducciones adicionales en la tasa se asocian con aumentos moderados en la inflación, aunque con menor intensidad.

Por otro lado, el término $h(tasa - 74,7794)$ también presenta signo negativo (-0,186), pero con mayor magnitud absoluta, lo que implica que cuando la tasa supera el umbral, su efecto contractivo sobre la inflación es más fuerte. Esto introduce una clara asimetría en la respuesta inflacionaria.

En otro caso, la interacción $TC_{nom} * h(tasa - 74,7794)$ con coeficiente igual a 0,001 aproximadamente indica que, en contextos de tasas elevadas, el efecto del tipo de cambio sobre la inflación se amplifica. Esto sugiere que la política monetaria restrictiva no neutraliza completamente los efectos de *pass-through* cambiario.

Por último, la interacción $h(tasa - 74,7794) * Mes_factor2$ con coeficiente de -0,3104 captura un efecto estacional específico para el mes de febrero en condiciones de tasa elevada, indicando que en dicho período la inflación tiende a ser menor en comparación con otros meses, controlando por el resto de las variables.

En conjunto, el modelo revela que la dinámica inflacionaria no es lineal ni homogénea en el tiempo, sino que depende en particular del nivel de la tasa de interés y de las interacciones con variables nominales y estacionales. La identificación endógena del umbral constituye uno de los principales aportes del enfoque MARS frente a modelos lineales tradicionales.

Tabla 3

Salida del modelo MARS_CV

Call: earth(formula=form_mars_dum, data=train_data, pmethod="cv", degree=2, nfold=10,ncross=5)	
coefficients	

(Intercept)				7,1495032			
h(74.7794-Tasa)				-0,0961954			
h(Tasa-74.7794)				-0,1859353			
TC_nom * h(Tasa-74.7794)				0,0008154			
h(Tasa-74.7794) * Mes_factor2				-0,3103685			
Selected 5 of 23 terms, and 3 of 15 predictors (pmethod="cv")							
GRSq	0,8749857	RSq	0,8943363	mean.oof.RSq	0,6647498	(sd 0,209)	

Fuente: elaboración propia. Nota. Salida del R del modelo MARS_CV

6.4. Comparación estadística entre modelos. Test de Diebold-Mariano (DM)

Tabla 4

Resultados estadísticos de los *test* de Diebold-Mariano

Comparación	DM	DM-HLN _p (HLN, t ₁₃)	Significatividad
MARS_poda vs ARIMA	-9.44 < 0.001	-9.10 < 0.001	Significativo
MARS_CV vs ARIMA	-8.95 < 0.001	-8.63 < 0.001	Significativo
MARS_clean vs ARIMA	-0.72, p=0.483	-0.70, p=0.498	No Significativo
MARS_poda vs MARS sin p	-3.02, p=0.010	-2.91, p=0.012	Significativo
MARS_CV vs MARS_poda	-0.04, p=0.972	-0.03, p=0.973	No significativo
Híbrido vs ARIMA	1.56, p=0.142	1.51, p=0.156	No Significativo

Fuente: elaboración propia. Nota. Resultados obtenidos de la prueba de DM.

Dado el tamaño reducido del conjunto de evaluación ($n=14$), se reporta adicionalmente la corrección de Harvey, Leybourne y Newbold (1997), que ajusta el estadístico por el factor $\sqrt{((n+1-2h+h(h-1)/n)/n)}$ y utiliza la distribución *t* de Student con $n-1$ grados de libertad en lugar de la normal estándar (DM-HLN). Esta corrección es más conservadora y mejora las propiedades de tamaño del *test* cuando es pequeño. El valor estadístico de DM negativo sugiere que el primer modelo listado es más preciso que el segundo. Los resultados de la tabla 4 evidencian que los modelos MARS con poda y con validación cruzada presentan una mejora estadísticamente significativa respecto al modelo ARIMA. En particular, las comparaciones MARS_poda vs ARIMA y MARS_CV vs ARIMA arrojan estadísticas altamente significativas ($p < 0.01$), lo que implica un rechazo robusto de la hipótesis nula de igualdad de precisión

predictiva. Esto confirma que la superioridad observada en las métricas de error fuera de muestra no es producto del azar, sino que refleja una mejora estructural en la capacidad de modelización.

Además, la comparación entre MARS_poda y MARS_sin_poda también resulta significativa, lo que evidencia que la poda no es simplemente un ajuste técnico, sino un mecanismo fundamental para evitar el sobreajuste y mejorar la capacidad de generalización del modelo. En otras palabras, el modelo sin poda captura ruido idiosincrático del conjunto de entrenamiento que no se traduce en capacidad predictiva fuera de muestra.

Por otro lado, no se detectan diferencias estadísticamente significativas entre MARS_CV y MARS_poda ($p > 0.05$), lo que sugiere que ambos modelos presentan un desempeño predictivo estadísticamente equivalente. Sin embargo, el modelo con validación cruzada no solo es más parsimonioso (5 términos y 3 predictores frente a 11 términos y 8 del modelo con poda *backward*), sino que además exhibe una mayor capacidad predictiva genuina fuera de muestra, evidenciada por un mean.oof.RSq de 0,665 frente a 0,584 del *backward*, lo que lo vuelve preferible tanto en términos de robustez como de generalización. En cambio, el modelo MARS_clean no presenta diferencias significativas respecto al ARIMA, lo que indica que la eliminación de valores extremos reduce la capacidad del modelo para capturar episodios relevantes de la dinámica inflacionaria argentina. Esto sugiere que dichos valores no constituyen meros *outliers* estadísticos, sino manifestaciones de cambios de régimen que contienen información económica relevante.

Finalmente, el modelo híbrido no muestra mejoras estadísticamente significativas respecto al ARIMA ($DM = 1,5626$, $p = 0,1422$). El signo positivo del estadístico indica que el híbrido presenta un mayor error que el ARIMA, aunque la diferencia no sea significativa. Este resultado coincide con el diagnóstico de residuos del ARIMA, en el que el *test* de Ljung-Box no detecta autocorrelación ($p = 0,5204$), lo que implica que no hay evidencia sistemática en la media de los residuos que MARS pueda captar. La heterocedasticidad en los residuos del modelo ARIMA no justifica el uso del modelo híbrido, ya que el

modelo MARS modela la media condicional, no la varianza; un modelo GARCH sería adecuado para captar la volatilidad agrupada.

En síntesis, la evidencia empírica indica que los modelos MARS, particularmente aquellos con mecanismos de regularización como poda o validación cruzada, dominan al modelo ARIMA en términos de precisión predictiva. Asimismo, se concluye que la complejidad del modelo debe ser controlada cuidadosamente, ya que el sobreajuste reduce significativamente la capacidad de generalización.

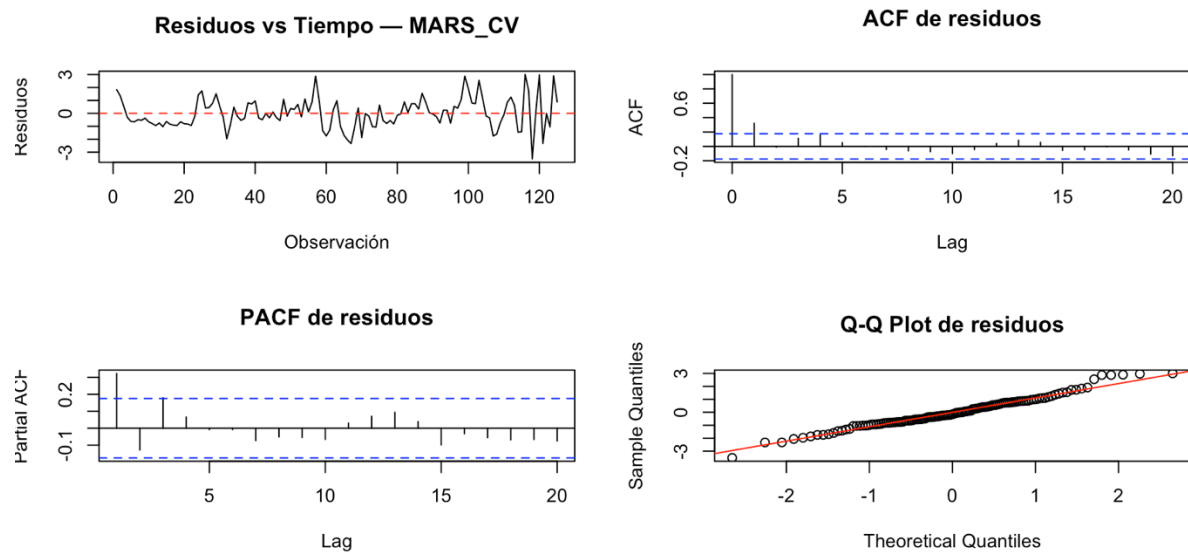
6.5. Diagnóstico de residuos de MARS_CV

Tabla 5

Resultados de las pruebas de diagnóstico de residuos del modelo MARS_CV

Test	H0	p_valor	Conclusión
Ljung-Box lag=12	Sin autocorr.	0,0356	Rechaza H0 (autocorrelación)
Ljung-Box lag=24	Sin autocorr.	0,1139	No rechaza H0
Shapiro-Wilk	Normalidad	0,054	No rechaza H0
ARCH-LM lags=12	Sin ARCH	0,0059	Rechaza H0 (ARCH presente)

Fuente: elaboración propia. Pruebas de diagnóstico de los residuos del modelo MARS_CV

Figura 2*Gráfico de residuos de MARS_CV*

Fuente: elaboración propia. Nota. El gráfico representa las evaluaciones de los residuos del modelo MARS_CV. El diagnóstico de residuos del modelo MARS_CV permite evaluar en qué medida la mejora en capacidad predictiva se logra a costa de incumplir supuestos clásicos de especificación. La evidencia empírica indica que, si bien el modelo captura adecuadamente la media condicional de la inflación, no modela completamente su estructura dinámica ni su distribución.

En primer lugar, el análisis gráfico de los residuos vs tiempo de la figura 2 muestra que éstos se encuentran centrados en torno a cero y no muestran tendencias evidentes, lo que es consistente con una correcta especificación.

Por otra parte, el gráfico de ACF revela la presencia de autocorrelación en los primeros rezagos, lo cual se confirma formalmente mediante el *test* de Ljung-Box a 12 rezagos ($p = 0,0356$), en el que se rechaza la hipótesis nula de ausencia de autocorrelación (tabla 5). Esta evidencia sugiere que el modelo no captura por completo la dependencia temporal a corto plazo. No obstante, al ampliar el horizonte a 24 rezagos ($p = 0,1139$), la autocorrelación deja de ser significativa, lo que indica que la omisión es localizada y no estructural.

El gráfico de PACF refuerza esta interpretación, mostrando picos aislados en rezagos bajos, lo que sugiere que persisten componentes autorregresivos de baja dimensión no capturados por esta especificación. Este resultado es consistente con la naturaleza del modelo, que no incorpora explícitamente una estructura dinámica temporal como sí lo hace un ARIMA.

En relación con la distribución de los errores, el Q-Q plot indica una desviación moderada respecto a la normalidad en las colas, particularmente en los cuantiles superiores, lo cual es coherente con la presencia de episodios extremos en la inflación argentina. Sin embargo, el *test* de Shapiro-Wilk ($p = 0,054$) no permite rechazar la hipótesis de normalidad al 5%, lo que sugiere que los residuos se distribuyen en torno a cero y sin tendencias evidentes, lo cual es consistente con una correcta especificación de la media condicional.

Por otro lado, el *test* ARCH-LM ($p = 0,0059$) indica la presencia de heterocedasticidad condicional, es decir, la varianza de los errores no es constante en el tiempo. Este resultado es esperable en series macroeconómicas como la inflación argentina, caracterizadas por episodios de volatilidad agrupada. Además, la heterocedasticidad implica que los intervalos de confianza del modelo resultan inválidos bajo los supuestos clásicos, aunque no afecta la validez del pronóstico puntual.

En conjunto, estos resultados muestran que el modelo MARS_CV no satisface plenamente los supuestos clásicos de ruido blanco en los residuos. Sin embargo, esto no invalida su capacidad predictiva. De hecho, como se evidenció en la sección anterior, el modelo presenta el menor error fuera de muestra y supera estadísticamente al ARIMA según el *test* de Diebold-Mariano.

Esta aparente contradicción, un modelo dinámicamente mal especificado que predice mejor, se explica por la distinción entre ajuste en muestra y capacidad de generalización en contextos de cambio de régimen. Mientras que el ARIMA logra residuos más cercanos al ruido blanco, su estructura lineal le impide adaptarse a no linealidades y quiebres estructurales, resultando en peor desempeño predictivo. En contraste, el MARS_CV prioriza capturar relaciones no lineales relevantes para la predicción, aun a costa de dejar remanentes de autocorrelación y heterocedasticidad en los residuos.

En consecuencia, el modelo MARS_CV debe interpretarse como un modelo predictivo eficaz, pero no como una representación completa del proceso generador de datos. Esta distinción es fundamental, el objetivo del modelo no es la inferencia estructural, sino la minimización del error de pronóstico en un entorno caracterizado por alta volatilidad y cambios de régimen.

Con el objetivo de abordar la autocorrelación serial identificada en los residuos del modelo MARS_CV, se estimó una versión extendida que incorpora IPC_lag1 como regresor adicional, denominada MARS_CV_lag. Los resultados completos de esta extensión se presentan en el Anexo.

El hallazgo principal es que el algoritmo de validación cruzada descartó el rezago en la selección final de términos, pero su disponibilidad como variable candidata indujo la identificación de dos términos adicionales vinculados a la tasa de interés que capturaron la dinámica temporal remanente de forma indirecta. Como resultado, la autocorrelación serial quedó corregida (Ljung-Box lag = 12: $p = 0,341$, lag = 24: $p = 0,154$), aunque el modelo resultante presenta un RMSE *out-of-sample* superior al original (1,234 vs 0,945), evaluado sobre un período de *test* que difiere en dos observaciones iniciales. En consecuencia, MARS_CV se mantiene como el modelo principal recomendado para predicción puntual de la inflación mensual, mientras que MARS_CV_lag constituye la especificación estadísticamente más completa para aplicaciones que requieran intervalos de confianza válidos

6.6. Robustez temporal

La validación mediante *rolling window* permite evaluar la estabilidad temporal del desempeño predictivo en un contexto más exigente que el *hold-out* estático, al simular un proceso real de predicción secuencial donde el modelo se reestima en cada período utilizando únicamente la información disponible hasta ese momento. En este caso, se utilizó una ventana inicial equivalente al 70% de la muestra (97 observaciones) y se realizó 41 predicciones un paso por delante.

En contraposición a los hallazgos de Garegnani y Gómez Aguirre, 2018, donde bajo el esquema de *rolling window*, los modelos de machine learning mostraron un mejor desempeño predictivo. Los resultados expuestos en la Tabla 6 evidencian un cambio sustancial en la jerarquía de su desempeño respecto a la

evaluación estática. El modelo ARIMA presenta el menor error promedio ($RMSE = 3,11$), superando a las especificaciones MARS_CV ($RMSE = 5,01$) y MARS_poda ($RMSE = 5,80$). Por su parte, el modelo híbrido exhibe nuevamente un comportamiento altamente inestable ($RMSE = 30,00$).

Estos resultados requieren una explicación más detallada. Mientras que en la validación *hold-out* estática los modelos MARS mostraron un desempeño notablemente superior ($RMSE = 0,95$ frente al $4,97$ del ARIMA), en la evaluación dinámica su margen de error se incrementa de forma drástica (+512% en MARS_poda y +430% en MARS_CV). En claro contraste, el modelo ARIMA experimenta una mejora en su precisión relativa, reduciendo su error en un 37%.

Una explicación a esta discrepancia radica en la interacción entre la naturaleza de los períodos de evaluación y la matemática de los modelos. La validación *hold-out* estática (de junio de 2024 a julio de 2025) abarca una “fase de desinflación sostenida”, caracterizada por un descenso continuo de la inflación desde niveles elevados hasta valores moderados. En cambio, la ventana móvil (febrero de 2022 a julio de 2025) somete a los algoritmos a múltiples cambios de régimen: la aceleración inflacionaria (2022-2023), el pico de fines de 2023 e inicios de 2024 y la posterior desaceleración, lo que permite evaluar su capacidad de adaptación en tiempo real.

Aunque en el esquema *rolling* los modelos se reestiman en cada iteración con la información más reciente, los splines de MARS pueden hacer extrapolaciones cuando el período siguiente incluye combinaciones de variables explicativas no presentes en los datos de la ventana de entrenamiento. Cuando una variable supera el rango observado, la función bisagra asociada se proyecta linealmente, sin cota, lo que genera predicciones desproporcionadas. Por el contrario, el modelo ARIMA, dada su estructura lineal y su capacidad para absorber la inercia del proceso mediante sus términos autorregresivos, proyecta valores acotados en torno a la media condicional reciente.

El comportamiento observado en las iteraciones individuales corrobora esta interpretación. Por ejemplo, en agosto de 2024 los modelos MARS generan predicciones severamente sobreestimadas (MARS_poda predice 14,88% frente a un valor real de 4,2%), mientras que el híbrido arroja valores carentes de

plausibilidad económica (alrededor de 100% mensual). Esta explosión del híbrido se debe a que el período de *test* contiene niveles de precio del oro y de tipo de cambio fuera del dominio histórico de entrenamiento, en el que el bloque MARS sobre residuos había aprendido relaciones no lineales que no se preservan al extrapolar. El modelo ARIMA, en cambio, aunque resulta menos preciso en escenarios estables, mantiene una trayectoria más consistente ante alteraciones bruscas.

En síntesis, la evidencia empírica revela que los modelos MARS son altamente sensibles a la muestra de entrenamiento y presentan una fuerte propensión a sobreajustar (*overfit*) patrones locales. Esta característica deteriora gravemente su capacidad de generalización en un esquema dinámico y penaliza su rendimiento en la ventana móvil. En consecuencia, la superioridad de los modelos MARS observada en la evaluación estática no resulta robusta ante variaciones metodológicas en la validación. Por el contrario, el modelo ARIMA, pese a su estructura lineal, demuestra una mayor resiliencia y estabilidad predictiva en un entorno de actualización continua de la información. Esta evidencia obliga a matizar la lectura inicial: no existe un modelo uniformemente superior, sino un *trade-off* entre flexibilidad y robustez. Mientras que MARS_CV optimiza la precisión en entornos relativamente estables, el ARIMA ofrece mayor estabilidad frente a cambios de régimen.

Tabla 6

Resultados del *rolling window*

Modelo	RMSE_outOut	RMSE_Rolling	Var %
ARIMA	4,9714	3,1115	-37,4
MARS_poda_dum	0,9488	5,8046	511,8
MARS_CV	0,9454	5,0140	430,3
HÍBRIDO	28,4667	29,9587	5,2

Fuente: elaboración propia. Nota. Comparación de los RMSE de los modelos con *rolling window*.

7. Limitaciones

La selección de variables se realizó considerando la disponibilidad mensual y la actualización sistemática de las mismas, con el fin de poder ser utilizadas para predicciones en tiempo real. Sin embargo, en la

literatura sobre inflación en Argentina existe suficiente evidencia que señala que otras variables como los agregados monetarios, el resultado fiscal, los salarios, los precios regulados y algunos otros juegan un rol importante en la dinámica inflacionaria (Graña Colella, 2020, Aguilar, 2024 y Forte, 2024). Asimismo, considerando además que durante gran parte del periodo analizado existieron restricciones cambiarias, la incorporación del tipo de cambio paralelo o blue podría aportar información adicional a los modelos, extensión particularmente apta para MARS, cuyos nudos endógenos podrían capturar el efecto diferencial de la brecha cambiaria según el régimen vigente.

En este sentido, se propone para futuras investigaciones la ampliación del conjunto de variables predictoras y la evaluación de si su inclusión mejora el desempeño predictivo de estos modelos. Así como también podrían orientarse a profundizar la integración de estructuras dinámicas y modelos no lineales, a modelar explícitamente la varianza condicional mediante enfoques tipo GARCH o a utilizar pruebas de quiebre estructural. Estas extensiones permitirían abordar una limitación central del presente trabajo: la dificultad de captar de forma completa una dinámica inflacionaria caracterizada por persistencia, volatilidad y cambios recurrentes de régimen.

8. Conclusiones

El principal objetivo de este trabajo fue evaluar las capacidades predictivas de modelos lineales y no lineales a la hora de pronosticar la inflación en Argentina, comparando el desempeño predictivo de un modelo ARIMA y dos especificaciones MARS. Los resultados obtenidos a través del análisis empírico permiten establecer que no existe un modelo universalmente superior para el pronóstico de la inflación mensual argentina. Sin embargo, el desempeño relativo de cada especificación depende del régimen inflacionario considerado y del esquema de validación que estemos utilizando.

En el ejercicio de *hold-out* estático (junio 2024 hasta julio 2025), caracterizado por un proceso de desaceleración inflacionaria y menor volatilidad, los modelos MARS_CV (RMSE = 0,9454) y MARS_poda (RMSE = 0,9488) exponen un desempeño claramente superior al ARIMA (RMSE = 4,9714). Eso sugiere que, en entornos relativamente estables, los modelos no lineales capturan con mayor

precisión las relaciones actuales entre la inflación y sus determinantes, aprovechando su flexibilidad funcional.

Por el contrario, cuando se evalúa la capacidad predictiva en un esquema de ventana rodante (febrero de 2022 a julio 2025), el cual incorpora múltiples cambios de régimen, incluyendo episodios de aceleración inflacionaria, picos extremos y posteriores procesos de estabilización. En este contexto, el modelo ARIMA presenta el mejor desempeño ($RMSE \approx 3,11$), superando a MARS_CV ($RMSE \approx 5,01$) y MARS_poda ($RMSE \approx 5,80$). Evidenciando así que, aunque MARS presenta ventajas en una partición estática específica, su superioridad no resulta robusta frente a cambios en el esquema de validación.

Esta divergencia refleja una diferencia estructural entre ambos enfoques. Los modelos MARS tienden a capturar relaciones no lineales y patrones locales presentes en la muestra de entrenamiento, lo que puede mejorar la precisión en determinados regímenes. No obstante, esa misma flexibilidad puede volverlos más sensibles a cambios de régimen o a combinaciones de variables no observadas previamente. En contraste, el ARIMA, al modelar explícitamente la dinámica temporal, ofrece una representación más estable frente a transiciones entre regímenes, aunque a costa de menor precisión en contextos específicos. En este sentido, la evidencia empírica sugiere que la inflación argentina reciente no puede ser adecuadamente modelada bajo un único enfoque, dado su carácter altamente no lineal y sujeto a quiebres estructurales frecuentes. Debido a esto los modelos no lineales capturan mejor la sección transversal del régimen (es decir, las relaciones contemporáneas entre variables), mientras que los modelos lineales capturan mejor la dimensión temporal (persistencia e inercia inflacionaria), pueden resultar más robustos cuando el objetivo es mantener estabilidad predictiva a lo largo de distintos períodos.

Para pronósticos de corto plazo en un contexto de inflación baja y relativamente estable, como el observado en el período más reciente, el modelo MARS_CV resulta preferible por su menor error en el *hold-out* estático, sus residuos presentan evidencia de autocorrelación de corto plazo y heterocedasticidad condicional, por lo que no debe interpretarse como una representación completa del proceso generador de datos. En cambio, para aplicaciones que requieren robustez frente a cambios de régimen, como

sistemas de monitoreo o alerta temprana, el modelo ARIMA constituye una opción más adecuada. De cualquier modo, no se debe descuidar que los resultados fueron en base a una muestra pequeña ($n=139$), dando pie a que exista sensibilidad de las métricas a valores atípicos.

Esto es fundamental para el BCRA u otros organismos que tengan como fin el pronóstico de la inflación, una estrategia en tiempo real bien diseñada para ello debería incorporar mecanismos de detección de cambios de régimen y selección de modelos. Por lo tanto, una configuración de pronóstico flexible, donde se pueda alternar entre modelos lineales y no lineales según convenga, podría resultar más efectiva que la utilización permanente de un único modelo.

El modelo híbrido ARIMA-MARS no mejora el desempeño predictivo y presenta problemas de inestabilidad fuera de muestra, lo que sugiere que las relaciones no lineales aprendidas sobre los residuos del ARIMA no se mantienen estables durante el período de evaluación.

En síntesis, el principal aporte del presente trabajo no radica en la identificación de un modelo “ganador” de manera absoluta, sino en la demostración de que la evaluación predictiva de la inflación argentina es altamente sensible al régimen económico y al esquema de validación empleados. Ignorar esta dimensión puede llevar a conclusiones erróneas sobre la superioridad de un modelo. Por lo tanto, la comparación entre ARIMA y MARS debe interpretarse como un trade-off entre precisión puntual y robustez temporal: MARS_CV mejora el pronóstico en contextos estables mientras que ARIMA ofrece mayor estabilidad frente a cambios de régimen.

9. Referencias

- Aguilar R., (2024). *Predicción de inflación en Argentina con métodos econométricos clásicos y machine learning*. Asociación Argentina de Economía Política: Working Papers 4704, Asociación Argentina de Economía Política.
- Álvarez L.J., Hurtado S., Sánchez I. y Thomas C., (2009). *The impact of oil price changes on Spanish and euro area consumer price inflation*. *Economic Modelling*. Elsevier

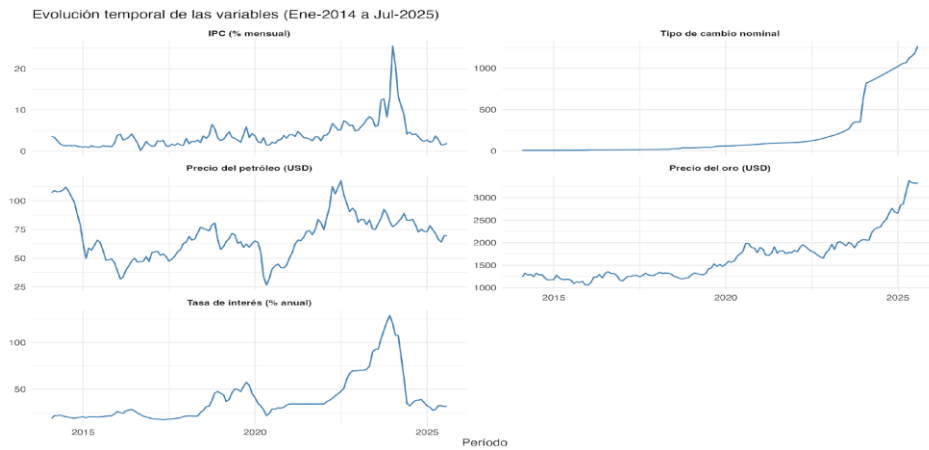
- Arcusa, A. (2020). *Estimación de la Nairu argentina 2003-2017*. [Tesina de Grado]. Universidad Nacional de Mar del Plata Facultad de Ciencias Económicas y Sociales.
- Backhaus, K., Erichson, B., Plinke, W. y Weiber, R., (2018). *Multivariate Analysemethoden*. Eine anwendungsorientierte https://www.jku.at/fileadmin/gruppen/124/AOM/Multivariate-Analysemethoden-Eine-anwendungsorientierte-Einfuehrung_compressed.pdf
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C. y Ljung, G.M., 2015. *Time Series Analysis Forecasting and Control* (5ta ed.). Wiley series in probability and statistics.
- Cavanaugh, J. E., (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*. 33, 201-208.
- Dickey, D. A. y Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*. 74, 427-431
- Diebold, F. X., y Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.2307/1392185>
- Diebold, F. X. (1998). The past, present, and future of macroeconomic forecasting. *Journal of Economic Perspectives*, 12(2), 175–192. <https://doi.org/10.1257/jep.12.2.175>
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
- Forte, F.D., 2024. *Pronóstico de inflación de corto plazo en Argentina con modelos Random Forest* (Documento de trabajo n° 24/10). BBVA Research. <https://www.bbvarsearch.com/wp-content/uploads/2024/09/WP-24-10-Forte-Inflacion-de-corto-plazo-con-random-forest-2.pdf>
- Friedman, J.H., (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19, 1-141.
- Garegnani, L. y Gómez Aguirre, M. (2018). *Forecasting Inflation in Argentina*. Banco Interamericano del Desarrollo.

- Gjika, E., Puka, L. y Zaçaj, O. (2018). *Forecasting consumer price index (CPI) using time series models and multi regression models (Albania case study)*. 10th International Scientific Conference “Business and Management”. Vilnius, Lithuania, Universidad Técnica de Vilna Gediminas.
- Graña Colella, S. (2020). Las causas de la inflación argentina: una estimación empleando la metodología VECM para el periodo 2003-2019. *Revista de la Facultad de Ciencias Económicas y Sociales*, 26 (55), 73-86.
- Gur, Y. E. (2024). Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. *Data Science in Finance and Economics*, 4(4), 469–513. <https://doi.org/10.3934/DSFE.2024020>
- Harvey, D., Leybourne, S., y Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Hyndman, R. J., y Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- Karadzic, V. y Pejovic, B. (2021). Inflation Forecasting in the Western Balkans and EU: A Comparison of Holt-Winters, ARIMA and NNAR Models. *Amfiteatru Economic* 23, 517-532.
- Kiliç, Rehim. (2025). *Linear and nonlinear econometric models against machine learning models: realized volatility prediction*. FEDS 2025-061. <https://doi.org/10.17016/FEDS.2025.061>
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3), 159-178.
- Lidiema, Caspah. (2017). Modelling and Forecasting Inflation Rate in Kenya Using SARIMA and Holt-Winters Triple Exponential Smoothing. *American Journal of Theoretical and Applied Statistics*, 6(3), 161-169. <https://doi.org/10.11648/j.ajtas.20170603.15>
- Ljung, G. M. y Box, G. E. P., (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*. 65, 297-303.

- Medeiros, M, C, Montes Schütte, E, C, y Soussi, T, S, (2025). *Global inflation forecasting*. SSRN N°4145665. <https://doi.org/10.2139/ssrn.4145665>
- Nguyen, T., Nguyen, H., Lee, J., Wang, Y. y Tsai, C. (2023). The consumer price index prediction using machine learning approaches: Evidence from the United States. *Heliyon*, 9(10), e20730. <https://doi.org/10.1016/j.heliyon.2023.e20730>
- Phillips, P. C. B. y Perron, P., (1988). Testing for a unit root in time series regression. *Biometrika*, 75, 335-346
- Riofrío, J., Chang, O., Revelo-Fuelagán, E.J. y Peluffo-Ordóñez, P.H. (2020). Forecasting the Consumer Price Index (CPI) of Ecuador: A Comparative Study of Predictive Models. *International Journal of Advanced Science Engineering and Information Technology*, 10(3), 1078-1084. <https://doi.org/10.18517/ijaseit.10.3.10813>
- Shapiro, S. S. y Wilk, M. B., (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*. 52, 591-611.
- Ülke, V., Sahin, A., y Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: Empirical evidence from the USA. *Neural Computing and Applications*, 30(5), 1519–1527. <https://doi.org/10.1007/s00521-016-2766-x>
- Zhang, G. P., (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)

10. Anexo

Figura 4. Evolución temporal de las variables



Fuente. Elaboración propia

10.1. Especificación del modelo MARS_CV_lag

Con el objetivo de capturar la dependencia temporal de corto plazo detectada en los residuos del modelo MARS_CV original, se estimó una especificación extendida que incorpora el rezago de la variable dependiente, IPC_{t-1} , como regreso adicional dentro del marco MARS. Esta extensión, denominada MARS_CV_lag, mantiene la misma estructura metodológica que el modelo base, diferenciándose únicamente en la inclusión de este componente dinámico.

Formalmente, el modelo puede expresarse como:

$$IPC_t = \beta_0 + \sum_{m=1}^M \beta_m B_m(X_t, IPC_{t-1}) + \varepsilon_t \quad (31)$$

Donde $B_m(.)$ representan funciones base tipo spline (funciones bisagra) generadas automáticamente por el algoritmo MARS, que permiten modelar relaciones no lineales y efectos de interacción entre las variables explicativas. El vector X_t incluye las variables macroeconómicas consideradas en la especificación original (tipo de cambio nominal, precios internacionales del petróleo y oro, tasa de interés y efectos estacionales capturados mediante variables dummy mensuales), a las que se añade IPC_{t-1} como nueva variable.

La estimación se realizó utilizando validación cruzada, 10 veces replicado 5 veces, replicando los hiperparámetros del modelo MARS_CV original con el fin de asegurar comparabilidad metodológica. El grado máximo de interacción se fijó en 2, permitiendo capturar efectos no lineales de segundo orden. El procedimiento de selección automática de términos retuvo 6 funciones base sobre un total de 24 posibles, involucrando únicamente tres predictores, tipo de cambio nominal, tasa de interés y estacionalidad. En este caso, el algoritmo de validación cruzada no retiene explícitamente el rezago de la inflación como predictor en la especificación final, lo que sugiere que su aporte predictivo directo es limitado una vez consideradas las otras variables.

Tabla 7

Comparación MARS_CV vs MARS_CV_lag

Modelo	RMSE_In	RMSE_Out	MAE_Out
MARS_CV (sin lag, test_data)	1,1597	0,9454	0,8227
MARS_CV_lag (con IPC_lag1, test_lag)	1,0770	1,2337	1,0710
Nota:			
test: jun-2024 a jul-2025			
test: ago-2024 a jul-2025 (*)			
(*) test_lag comienza 2 meses después — comparación con cautela			

Fuente: elaboración propia

Siguiendo los resultados de la tabla 7 se observa que, en términos de ajuste, el modelo presenta un coeficiente de determinación $R^2 = 0,9103$ y un $GRSq = 0,8886$, mientras que el $mean.oof.RSq = 0,6506$ mejorando levemente a MARS_CV. No obstante, esta mejora en validación cruzada no se traduce en un mejor desempeño predictivo fuera de muestra: el RMSE *out-of-sample* aumenta a 1,2337 frente a 0,9454 del modelo MARS_CV original (teniendo en cuenta que los períodos de evaluación no son perfectamente comparables). Este deterioro sugiere que la incorporación de dinámica temporal no lineal introduce complejidad adicional sin ganancias en precisión predictiva en el régimen de baja inflación.

10.2. Diagnóstico de residuos

Tabla 8

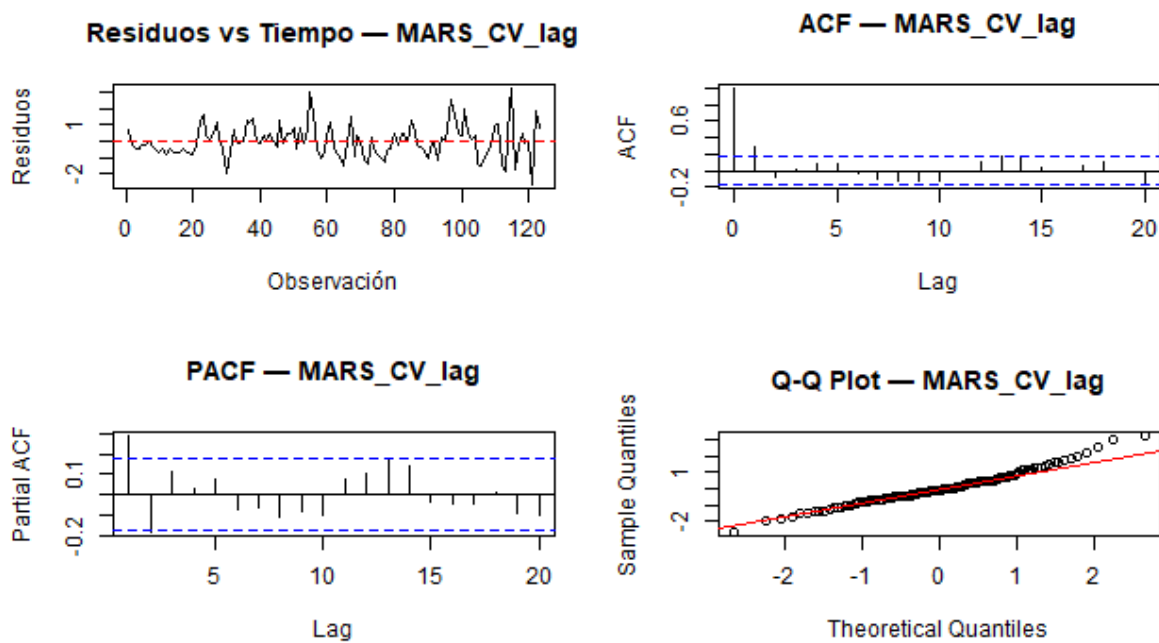
Pruebas sobre residuos de MARS_CV_lag

Test	MARS_CV_p	CV_lag_p	Mejóro
Ljung-Box lag=12	0.0004	0.3407	Si
Ljung-Box lag=24	0.0047	0.1538	Si
Shapiro-Wilk	0.0000	0.0121	Si
ARCH-LM lags=12	0.0406	0.0393	No

Fuente: elaboración propia

Figura 3

Gráfico de residuos de MARS_CV_lag



Fuente: elaboración propia

Desde el punto de vista gráfico (figura 3), los residuos del modelo se distribuyen de manera aproximadamente centrada en cero, sin patrones sistemáticos visibles en el tiempo, lo que indica que la

media condicional está correctamente especificada. Sin embargo, se observan fluctuaciones de distinta magnitud a lo largo de la muestra, lo que anticipa posibles problemas en la variación.

El análisis de autocorrelación muestra una mejora sustancial respecto al modelo original. Tanto la ACF como la PACF presentan coeficientes dentro de las bandas de confianza para la mayoría de los rezagos, sin evidencia clara de estructura AR(1) remanente. Este resultado se confirma con las pruebas de Ljung-Box: para lag 12 ($p = 0,3407$) y lag 24 ($p = 0,1538$) no se rechaza la hipótesis nula de ausencia de autocorrelación (tabla 8), lo que indica que la dependencia serial ha sido efectivamente corregida.

En cambio, la distribución de los residuos continúa alejándose de la normalidad. El QQ plot evidencia desviaciones en las colas, particularmente en los cuantiles extremos, lo que se refleja en el *test* de Shapiro-Wilk ($p = 0,0121$), que lleva a rechazar la hipótesis de normalidad. Este resultado indica que, si bien la dinámica media mejora, la distribución de los errores sigue siendo asimétrica o con colas pesadas, consistente con episodios de shocks inflacionarios.

Por otro lado, el *test* ARCH-LM ($p = 0,0393$) indica la presencia de heterocedasticidad condicional, lo que implica que la varianza de los errores no es constante en el tiempo. Este resultado contrasta con algunas mejoras parciales observadas en versiones intermedias del modelo y sugiere que la inclusión del rezago no logra estabilizar completamente la volatilidad del proceso.

En conjunto, la evidencia muestra que la incorporación del rezago permite corregir de manera efectiva la autocorrelación serial en los residuos pero no resuelve problemas estructurales asociados a la distribución de los errores, como la no normalidad y la heterocedasticidad condicional. Además, esta mejora en términos de especificación estadística se obtiene a costa de un deterioro en la capacidad predictiva fuera de muestra.

Este trade-off refuerza la decisión de mantener el modelo MARS_CV original como especificación principal para predicción puntual, mientras que MARS_CV_lag constituye una alternativa metodológicamente más completa desde el punto de vista econométrico, aunque menos eficiente en términos predictivos bajo el régimen de evaluación considerado.

10.3. Estimaciones de los distintos modelos

10.3.1. Modelo ARIMA

Tabla 9

Modelo ARIMA

Series: ts_ipc						
ARIMA(2,1,3)						
Coefficients:	ar1	ar2	ma1	ma2	ma3	
	-0,72	-0,4341	0,9822	0,1339	-0,5072	
s.e.	0,1697	0,1648	0,2125	0,3424	0,1994	
sigma^2 =	2,439	log likelihood =		-230,4		
AIC=	472,8	AICc=		473,52	BIC=	489,73
Training set error measures:						
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
0,04573309	1,52388	0,9575815	-9,050785	31,43625	0,4998001	0,0203256

Fuente: Elaboración propia. Nota: si bien en la tesis se especifica la teoría de un modelo SARIMA, el ganador fue un modelo ARIMA(2,1,3)

10.3.2. MARS sin poda

Tabla 10

Modelo MARS sin poda

Call: earth(formula=form_mars_orig, data=train_data, pmethod="none", degree=2)	
coefficients	
(Intercept)	5,7039427
h(113,335-TC_nom)	0,0089030
h(TC_nom-113,335)	0,0059605
h(59,5018-Precio_Petroleo)	0,0113618
h(Precio_Petroleo-59,5018)	0,0564228
h(84,4405-Tasa)	-0,0528580
h(Tasa-84,4405)	5,8422007
TC_nom * h(Tasa-84,4405)	0,0009341

Precio_Petroleo * h(Tasa-84,4405)								-0,0224941
Precio_Oro * h(Tasa-84,4405)								-0,0022226
h(Tasa-84,4405) * Mes_num								0,0289561
h(15,5237-TC_nom) * h(84,4405-Tasa)								-0,0015522
h(TC_nom-15,5237) * h(84,4405-Tasa)								-0,0001020
h(113,335-TC_nom) * h(Tasa-34,2495)								-0,0009304
h(113,335-TC_nom) * h(34,2495-Tasa)								-0,0010018
h(Precio_Petroleo-59,5018) * h(Tasa-39,0726)								-0,0003883
h(Precio_Petroleo-59,5018) * h(39,0726-Tasa)								-0,0020537
Selected 17 of 17 terms, and 5 of 5 predictors (pmethod="none")								
GCV	2,174201	RSS	122,7293	GRSq	0,8319011	RSq	0,9228599	

Fuente: elaboración propia. Nota. Salida del R del modelo MARS sin poda

10.3.3. MARS con poda y dummies estacionales

Tabla 11

MARS con poda y *dummies* estacionales

Call: earth(formula=form_mars_dum, data=train_data, degree=2)							
coefficients							
(Intercept)							5,6989621
Mes_factor3							0,7684968
h(11,4278-TC_nom)							-0,4767072
h(Precio_Petroleo-57,6491)							0,0366809
h(74,7794-Tasa)							-0,0688721
h(Tasa-74,7794)							-0,1174340
TC_nom * h(Tasa-74,7794)							0,0006665
h(Tasa-74,7794) * Mes_factor2							-0,2372374
h(Tasa-74,7794) * Mes_factor8							0,0970511
h(Tasa-74,7794) * Mes_factor10							-0,0847233
h(Tasa-74,7794) * Mes_factor12							0,1082760
Selected 11 of 23 terms, and 8 of 15 predictors							
GCV	1,182333	RSS	92,70436	GRSq	0,9085876	RSq	0,9417317

Fuente: elaboración propia. Nota. Salida del R del modelo MARS con poda y dummies estacionales

10.3.4. MARS con rezagos

Tabla 12

Modelo MARS con rezagos

Call :earth(formula=form_mars_full, data=train_lag, degree=2)							
coefficients							
(Intercept)							4,9569832
Mes_factor3							0,6124287
h(TC_nom-113,335)							-0,0047692
h(Precio_Oro-1276,7)							0,0084717
h(1780,9-Precio_Oro)							0,0130414
h(Tasa-61,7745)							-0,4358607
h(70,3845-Tasa)							-0,1644996
h(Tasa-70,3845)							0,4277804
TC_nom * h(Tasa-70,3845)							0,0008834
h(Precio_Oro-1780,9) * Mes_factor7							-0,009405
h(Tasa-70,3845) * Mes_factor11							0,0559184
h(Tasa-70,3845) * IPC_lag2							-0,019925
h(113,335-TC_nom) * h(Tasa-27,6505)							-0,0024848
h(67,7255-TC_nom) * h(7-IPC_lag1)							-0,0057742
h(1279,2-Precio_Oro) * h(7-IPC_lag1)							-0,002268
Selected 15 of 26 terms, and 8 of 17 predictors							
GCV	0,8334734	RSS	51,28911	GRSq	0,9365904	RSq	0,9677541

Fuente: elaboración propia. Nota. Salida del R del modelo MARS con rezagos

10.3.5. MARS sin valores atípicos

Tabla 13

modelo MARS sin valores atípicos

Call: earth(formula=form_mars_dum, data=train_no_out, degree=2)	
coefficients	
(Intercept)	1,37214204
h(Precio_Oro-1316,1)	0,02258353

h(1413,7-Precio_Oro)							0,05732892
h(Precio_Oro-1413,7)							-0,02142953
h(38,124-Tasa)							-0,1658433
TC_nom * Mes_factor3							0,01174225
h(Precio_Petroleo-61,1357) * Mes_factor11							-0,04569163
h(1413,7-Precio_Oro) * Tasa							-0,00094319
h(38,124-Tasa) * Mes_factor8							-0,04978196
h(15,181-TC_nom) * h(38,124-Tasa)							-0,00804817
h(Precio_Petroleo-61,1357) * h(Tasa-30,6924)							0,00212872
h(Precio_Petroleo-61,1357) * h(30,6924-Tasa)							0,00240521
h(1317,9-Precio_Oro) * h(38,124-Tasa)							-0,00219077
Selected 13 of 23 terms, and 7 of 15 predictors							
GCV	0,6886859	RSS	41,61717	GRSq	0,7633653	RSq	0,8723332

Fuente: elaboración propia. Nota. Salida del R del modelo MARS sin valores atípicos

10.3.6. Modelo híbrido

Tabla 14

Modelo híbrido

Call: earth(formula=form_hyb, data=data_res, degree=2)							
coefficients							
(Intercept)							-0,07364647
Mes_factor8							-1,88650432
h(Precio_Oro-1985,9)							-0,01160199
Tasa * Mes_factor8							0,07101084
h(Precio_Oro-1413,7) * Mes_factor3							0,0043053
h(Precio_Oro-1985,9) * Mes_factor3							-0,01894933
h(Precio_Oro-1985,9) * Mes_factor11							0,1164005
h(Precio_Oro-1985,9) * Mes_factor12							0,10135784
Selected 8 of 19 terms, and 6 of 15 predictors							
GCV	1,139148	RSS	103,364	GRSq	0,5168373	RSq	0,6435905

Fuente: elaboración propia. Nota. Salida del R del modelo híbrido

10.3.7. MARS_CV_lag

Tabla 15

MARS_CV_lag

Call: earth(formula=form_cv_lag, data=train_lag, pmethod="cv", degree=2, nfold=10, ncross=5)							
coefficients							
(Intercept)							9,1623492
h(Tasa-61,7745)							-0,1833967
h(70,3845-Tasa)							-0,1476552
TC_nom * h(Tasa-70,3845)							0,0007377
h(Tasa-70,3845) * Mes_factor2							-0,2815289
h(216,556-TC_nom) * h(Tasa-29,0267)							-0,0005899
Selected 6 of 24 terms, and 3 of 16 predictors (pmethod="cv")							
GRSq	0,8886556	RSq	0,9103032	mean,oof, RSq	0,65063	sd	0,199

Fuente: elaboración propia. Nota. Salida del R del modelo MARS_CV_lag

Tabla 16

Estadísticos descriptivos

Variable	N	Media	Mediana	Desv_Est	Minimo	Maximo	Q1	Q3	Asimetría	Curtosis	CV_pcent
IPC	139	3,84	3,1	3,424	0.2	25,5	1,85	4,2	3,268	17,608	89,18
TC_nom	139	197,849	58,531	332,633	7,097	1267,02	15,35	139,458	1,962	5,3	168,12
Precio_Petroleo	139	69,484	68,309	19,879	26,631	117,503	55,4	81,529	0,318	2,656	28,61
Precio_Oro	139	1660,184	1514,8	523,958	1060,2	3374,7	1271,8	1898,1	1,396	4,753	31,56
Tasa	139	38,689	32,231	24,258	17,487	128,99	21,432	44,578	1.9	6,242	62,7

Fuente: Elaboración propia. Nota. Estadísticos descriptivos de las variables.