

Predicción de inflación en Argentina con métodos econométricos clásicos y *machine learning*

Rafael Aguilar

Universidad de Buenos Aires (UBA)

Resumen:

Argentina está en un régimen de alta inflación. Para los *policy makers* y el resto de los agentes económicos, es crucial contar con predicciones confiables de la inflación. Un hándicap de los modelos predictivos clásicos es que al aumentar la cantidad de variables, también aumenta la varianza de las predicciones. El desarrollo de técnicas de *machine learning* permite modelar con decenas o centenas de variables y aprovechar la amplitud de datos disponible. En base a trabajos de D'Amato et al. (2018) y Silva Araujo y Piazza Gaglianone (2023), desarrollamos un set de modelos para predecir la inflación argentina entre 2016 y 2024 y comparamos su poder predictivo en distintos horizontes temporales. El ejercicio incluye (i) un modelo ARIMA de serie de tiempo univariado, (ii) modelos de autorregresión vectorial VAR y (iii) regresiones LASSO y ELASTIC NET con elementos de *machine learning*. Utilizamos como *benchmark* la mediana de la encuesta REM del BCRA. Encontramos un mejor desempeño de los modelos con *machine learning* en ambos horizontes. Entre las variables seleccionadas por las regresiones LASSO y ELASTIC NET se destacan la inflación rezagada, el tipo de cambio oficial, los salarios nominales, el ITCRM-BCRA y agregados monetarios como los depósitos y préstamos del sector privado.

1. Introducción

Argentina está en un régimen de alta inflación¹. La tasa mensual alcanzó un pico de 25,5% en diciembre de 2023. Desaceleró a 4,0% en julio de 2024, una baja significativa, pero que todavía implica una tasa anualizada de 60,7%. La reaparición de inflaciones anuales de dos dígitos o más ocurrió en la década del 2000. Sin embargo, debido a la intervención del INDEC entre 2007 y 2015, no contamos con estadísticas fiables para aquellos años. Por eso, este trabajo abarca el período que va de junio de 2016 a marzo de 2024.

Vivir en alta inflación eleva el costo de tomar decisiones económicas, como fijar precios, firmar contratos estipulados en valores nominales o invertir en ciertos activos, en base a predicciones erróneas de la inflación. Para los *policy makers* en el BCRA y el Ministerio de Economía, es útil conocer en tiempo real los efectos de sus políticas. El propósito de este trabajo es desarrollar distintos modelos econométricos, con y sin *machine learning*, para predecir la inflación mensual y comparar sus resultados entre sí y con las predicciones de bancos, consultoras y centros de estudio recogidas en la encuesta REM del BCRA².

1.1. La literatura existente sobre predicción de la inflación en Argentina

El BCRA ha producido varios trabajos. D'Amato et al. (2008) comparan el desempeño de modelos ARIMA, VAR y Curva de Phillips para predecir la inflación entre 1993 y 2006. Encuentran que los ARIMA se destacan en horizontes cortos, pero ningún modelo se impone en todos los horizontes. Por eso recurren al *forecast pooling*, que combina los resultados de los modelos para mejorar la predicción. Una ventaja que D'Amato et al. le encuentran al *pooling* es que permite complementar distintos ángulos teóricos para explicar la inflación. En su caso, un VAR que toma como insumos los agregados monetarios (un enfoque vinculado al monetarismo) con un modelo de Curva de Phillips, que conecta la inflación al *output gap*, la brecha entre la producción potencial y efectiva de la economía.

D'Amato et al. no encuentran literatura anterior sobre predicción de la inflación en Argentina. Toman como referencia dos trabajos estadounidenses. Uno es Stock y Watson (1999), en el que comparan distintos modelos de Curva de Phillips para predecir la inflación estadounidense entre 1970 y 1996. Atkenson y Ohanian (2001) disputan los resultados de Stock y Watson, señalando que para un período similar, de 1984 a 1999, los modelos de Curva de Phillips no superan el desempeño de un modelo "naif" autorregresivo.

D'Amato et al. (2018) actualizan el *paper* de 2008 para el período de 2004 a 2018. Incluyen modelos VAR bayesianos (BVAR) y VAR con parámetros que varían en el tiempo (TVP-VAR), sin utilizar *forecast pooling*. En horizontes cortos, no encuentran diferencias significativas entre estos

¹ Heymann y Leijonhufvud (1995) definen el régimen de alta inflación como aquel en que la inflación mensual desplaza la anual como dato relevante y lo cuantifican como una tasa mensual de entre 5 y 50%. Aunque recientemente haya perforado este piso, tomando la tasa mensual promedio del año, Argentina cumple con este criterio a partir de 2022 (5,7%).

² Esta tesis se empezó a escribir en el verano de 2023. A partir de la aceleración inflacionaria que siguió a las devaluaciones de agosto y diciembre de 2023, muchos analistas consideran que la inflación semanal se volvió una variable más relevante que la mensual para Argentina. Es una observación atendible, pero modelar la inflación semanal presenta varios inconvenientes. En primer lugar, no existen datos oficiales. Hay estimaciones privadas, que en general no son del IPC-INDEC sino de canastas acotadas de alimentos, bebidas y otros bienes y servicios de venta *online*. En segundo lugar, la mayoría de los contratos importantes de la economía (salarios, alquileres, tarjetas de crédito, cuotas de educación y salud, etc.) todavía se negocian a plazo mensual. Para actualizar montos acorde a la inflación o transformarlos de términos nominales a reales, la tasa mensual sigue siendo el indicador más relevante.

modelos y los VAR tradicionales, Curva de Phillips o ARIMA. En horizontes más largos, se destacan los VAR tradicionales y Curva de Phillips. En este trabajo adaptamos dos de los VAR utilizados en D'Amato et al. (2018), eligiéndolos por su flexibilidad y buen poder predictivo.

En otro trabajo del BCRA, Garegnani y Gómez Aguirre (2018) encuentran que los VAR bayesianos superan al ARIMA en horizontes más largos, para el período de 2004 a 2017.

En los últimos años, creció el interés por predecir la inflación en Argentina. Por fuera del BCRA, Frenkel y Friedheim (2017) construyen un modelo neoestructuralista para explicar la suba de precios entre 2003 y 2016, con foco en los salarios, el tipo de cambio y las tarifas de servicios públicos. Marinozzi (2023) elabora modelos VAR, VEC y de largo plazo para predecir la inflación entre 2004 y 2019. Encuentra un mejor desempeño general de modelos estructurales como el VEC. Con un enfoque novedoso, Aromí y Llada (2020) buscan predecir la inflación entre 2012 y 2019 a partir de la cantidad de mensajes que incluyen las palabras “inflación” o “inflacionario” en la red social Twitter/X.

A pesar de la cuantía y la calidad de los trabajos mencionados, al momento de escribir esta tesis no existen publicaciones que utilicen *machine learning* para predecir la inflación en Argentina. Es una herramienta que se ha vuelto muy popular en el campo a nivel internacional, con trabajos publicados sobre la predicción de variables macroeconómicas con estas técnicas en EEUU, Europa, Brasil, Turquía y otros países. Esta tesis, que compara modelos tradicionales con dos de la familia del *machine learning*, LASSO y ELASTIC NET, puede servir como contribución inicial para el desarrollo del campo en Argentina.

1.2. El aporte de los modelos de *machine learning*

Un problema central a la hora de construir modelos econométricos es escoger las variables. Un buen modelo incluye solo las relevantes y excluye el resto. Este es un proceso discrecional, para el cual el econometrista se apoya en la teoría económica. El *machine learning* ayuda a reducir la discrecionalidad: según Hall (2018), funciona automatizando parte del proceso de selección de variables con dos elementos centrales, un método de aprendizaje y un algoritmo, que optimizan el modelo sin supuestos derivados de la teoría.

En un modelo predictivo, hay dos tipos de error según su fuente: los de sesgo y los de varianza. Los primeros se deben al uso de supuestos incorrectos sobre la relación entre datos. Los segundos, a la sensibilidad del modelo a perturbaciones en estos datos. Se suele presentar la relación entre errores de sesgo y varianza como un *trade-off*, en el que complejizar un modelo reduce el primer tipo de error y aumenta el segundo. Al sumar variables al modelo, pequeñas perturbaciones en estas variables tendrán mayor efecto sobre la que se busca predecir.

Mediante la regularización, las técnicas de *machine learning* “disciplinan” a los modelos. Penalizan la inclusión de variables irrelevantes y la contribución de cada variable, la magnitud de los coeficientes. Así, reducen el *trade-off* entre sesgo y varianza, lo que permite sumar variables sin generar un impacto desmedido en los resultados (Hall, 2018).

Con esto, no queremos decir que la teoría sea un lastre a la hora de construir modelos predictivos. Es un debate que viene de largo. Diebold (1998) escribió que “los rumores de la muerte de la predicción con modelos macroeconómicos de gran escala *no* son exagerados”, refiriéndose al descrédito en el que habían caído los modelos estructurales keynesianos a finales de los setenta y la aparición de modelos nuevos, no-estructurales, que en muchos casos conseguían un mejor desempeño predictivo que sus predecesores.

Estos modelos incluyen el ARIMA de Box y Jenkins y sobre todo el VAR de Sims, que explota la correlación entre series de tiempo sin los supuestos restrictivos de la teoría keynesiana. Sims apuntó contra uno de los pilares de la econometría keynesiana, la especificación “una-ecuación-a-la-vez”, que desarrollamos en la sección siguiente. Sin embargo, el VAR se ha utilizado en trabajos que recuperan aspectos teóricos, como los del BCRA mencionados en la sección previa.

Sin profundizar demasiado en la historia del pensamiento económico, queremos resaltar como la teoría se nutre y se refresca con la incorporación de herramientas estadísticas³ que ponen en duda u hacen caso omiso de los supuestos y restricciones que impone la teoría.

Automatizar la selección de variables⁴ funciona como una “caja negra” que vía *machine learning* estudia los patrones entre todas las variables y nos devuelve aquellas de mayor correlación con la dependiente, proceso desprovisto de supuestos teóricos. De esta “caja negra” pueden surgir variables que no hubiéramos considerado en base a la teoría económica, aunque luego volvamos a la teoría para analizar los resultados del modelo y el sentido de las variables seleccionadas.

1.3. La literatura existente sobre predicción de inflación con *machine learning*

Uno de los trabajos seminales es Terasvista et al. (2005), que utiliza modelos NEURAL NETWORK (NN) para predecir variables macroeconómicas en países del G7. Uno de los colaboradores del *paper* es Marcelo Medeiros de la PUC-RJ, quién luego desarrolló una investigación amplia sobre la predicción con *machine learning*. Medeiros y Mendes (2016) y Medeiros et al. (2021) consiguen resultados superadores para predecir la inflación estadounidense con modelos LASSO.

García et al. (2017)⁵ es un ejercicio de predicción de la inflación brasileña entre 2003 y 2015 con *machine learning* en el que también participa Medeiros, que encuentra un desempeño de la familia LASSO (sobre todo la adaptación “adaLASSO”) superior al de los modelos ARIMA, VAR bayesianos e incluso la encuesta FOCUS del Banco Central de Brasil, análoga al REM-BCRA.

Un trabajo subsidiario de Medeiros es Silva Araujo y Piazza Gaglianone (2023). Compara modelos clásicos como el ARIMA, VAR y Curvas de Phillips con un set de modelos *machine learning* que incluye LASSO y ELASTIC NET para predecir la inflación brasileña entre 2004 y 2018 y encuentra que estos últimos obtienen mejores resultados. Silva Araujo y Piazza Gaglianone eligen modelos clásicos similares a los de D’Amato et al. (2018), una de las razones por la cual utilizamos su trabajo como referencia.

Özgür y Akkoç (2022) es otra aplicación de *machine learning* para predecir la inflación de una economía emergente, en este caso Turquía entre 2007 y 2019, obteniendo desempeño superior a los modelos tradicionales. El caso turco nos es de particular interés porque se trata de un país con inflación alta, encima de los dos dígitos anuales a partir de 2017. Además, el período abarcado incluye un quiebre estructural, la devaluación de la lira turca en 2018. Es un caso más cercano al de Argentina, que sufrió *shocks* similares entre 2016 y 2024, que el de los otros países abarcados en la literatura de predicción de inflación con *machine learning*.

³ Incluso las que tienen origen en otras disciplinas, como el LASSO, que nació en las ciencias médicas.

⁴ Por supuesto, no existe una automatización total. La construcción del *dataset* amplio del cual los modelos seleccionan las variables significativas conlleva discrecionalidad.

⁵ El código en R utilizado en este trabajo fue adaptado de García et al. (2017) y se encuentra disponible en: <https://github.com/gabrielrvsc/HDeconometrics/tree/master>

2. Modelos utilizados

Construimos cuatro tipos de modelo: ARIMA, VAR, LASSO y ELASTIC NET. El primero es univariado y el resto son multivariados. D'Amato et al. (2018) señalan que a pesar de que los modelos univariados no captan la interacción entre variables económicas diferentes, en muchos casos tienen desempeño predictivo similar al de modelos más complejos.

2.1. ARIMA

El ARIMA es un modelo lineal univariado que asume que las observaciones futuras son determinadas en parte por las observaciones pasadas. El acrónimo refiere a sus tres componentes, autorregresivo (AR), integrado (I) y de medias móviles (MA). El ARIMA genérico toma la forma:

$$X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (1)$$

Donde X_{t-p} son los valores rezagados hasta el orden p de la variable a predecir y ε_{t-q} sus errores rezagados hasta el orden q . Seleccionamos estos parámetros según el Criterio de Akaike, que premia a los modelos que generan predicciones cercanas a la data observada utilizando la menor cantidad de parámetros posible, mediante la fórmula:

$$AIC = 2k - 2\ln(\mathcal{L}) \quad (2)$$

El ARIMA fue planteado por Box y Jenkins (1970). Es un *benchmark* común en la literatura por su simpleza y poder predictivo. En ausencia de *shocks*, es razonable suponer que la inflación de meses pasados alimenta la futura.

2.2. VAR

El VAR es un modelo multivariado de autorregresión vectorial que representa cada variable como una función de sus propios rezagos y los rezagos de las demás variables. Asume un tratamiento simétrico de las variables endógenas, donde todas se retroalimentan entre sí (D'Amato et al., 2018). Un VAR de orden p toma la forma:

$$X_t = v + A_1 X_{t-1} + \dots + A_p X_{t-p} + \varepsilon_t \quad (3)$$

Donde X_t es un vector $n \times 1$ de variables endógenas, v un vector fijo $n \times 1$ de términos constantes, A una serie de matrices de coeficientes $n \times n$ y $\varepsilon_t \sim N(0, \Sigma)$ un vector $n \times 1$ de *shocks* exógenos, que toma la forma de ruido blanco. Una ventaja de los VAR es que no imponen supuestos restrictivos sobre la estructura ni la distribución de los datos. Surgen de la Crítica de Sims, desarrollada en una serie de *papers* (Sims 1972, 1980a, 1980b) que contribuyeron al cambio de paradigma respecto de la econometría keynesiana que ocurrió en los setenta, desde un enfoque distinto a la Crítica de Lucas⁶.

⁶ Lucas (1976) se enfocó en cómo las expectativas racionales subvierten el impacto de las decisiones de política. Sims se preocupó por una cuestión más amplia; como las expectativas afectan la identificación en los modelos, es decir, su conexión con la realidad. Apuntó contra la especificación “una-ecuación-a-la-vez” o incremental de los modelos keynesianos, con la que se buscaba aislar las curvas de oferta o demanda para un sector de la economía, aplicando restricciones de manera que expresaran siempre una variable

La retroalimentación entre variables es relevante a nuestro trabajo, por ejemplo, en la relación entre inflación y tipo de cambio. Una devaluación tiene *pass-through* a precios, y una aceleración inflacionaria puede aumentar las expectativas de devaluación. Al definir tanto inflación como tipo de cambio variables endógenas del VAR, capturamos ambas relaciones, dejando lugar para shocks exógenos en el término ε_t .

Elegimos las variables siguiendo a D'Amato et al. (2018), donde construyen cuatro versiones del VAR. Todas utilizan la inflación (ipc) y el tipo de cambio oficial (arsusd). El VAR-1 suma la actividad económica (emae) y la tasa de plazos fijos (badlar) como variables, mientras que el VAR-2 agrega los salarios nominales del sector privado formal (ripte).

Desde un punto de vista teórico, el VAR-1 modela un canal tradicional de transmisión de la política monetaria, por el cual la suba de tasas de interés encarece el crédito, reduciendo la inversión de las empresas y el consumo de las familias, de manera que se contrae la demanda agregada y la presión sobre los precios (o viceversa). El VAR-2 busca reflejar la interacción entre inflación, tipo de cambio y salarios. D'Amato et al. lo vinculan con el "*cost-push channel of monetary policy*"⁷.

Tabla 2.2. Variables de los modelos VAR desarrollados en este trabajo

Variables utilizadas en cada modelo		Modelo		
Variable	Código	VAR-1	VAR-2	VAR-3
Índice de precios al consumidor	ipc	X	X	X
Tipo de cambio oficial	arsusd	X	X	X
Actividad económica	emae	X		X
Tasa de interés en pesos, PF 30-59 días	badlar	X		
Salario nominal privados formales	ripte		X	X

El VAR-3 y el VAR-4 de D'Amato et al. incluyen los precios regulados como variable adicional. Sin embargo, en el paper de 2018 estos últimos dos modelos obtienen peores resultados, por lo cual no los replicamos. En su lugar, creamos un nuevo VAR-3 con tipo de cambio, actividad y salarios, una síntesis de los dos primeros modelos. Los tres VAR planteados se detallan en la tabla 2.2.

endógena en función de otras exógenas. Para Sims, estas restricciones podían ser útiles para describir comportamientos microeconómicos en equilibrio parcial, pero los modelos macro que resultaban de la acumulación de estas ecuaciones simplemente no eran creíbles.

En Sims (1980b), ejemplifica con la demanda de café brasileño de los consumidores estadounidenses. En un modelo donde oferta y demanda se especifican por separado, cada una tiene sus propias variables exógenas: el clima en Brasil para la oferta, los gustos y modas en Estados Unidos para la demanda. Sin embargo, incorporando el rol de las expectativas, si los consumidores estadounidenses leen sobre una helada en Brasil, tal vez salgan a comprar café para anticiparse a una suba de precios. El punto de Sims es que en muchos casos, definir que ciertas variables endógenas afectan solo a un lado del mercado es una construcción artificial, que va en contra de la identificación creíble en los modelos.

⁷ Tadeu Lima y Setterfield (2010) definen esta teoría como la "canónica" para explicar la fijación de precios dentro de la economía heterodoxa. En síntesis, una suba de las tasas de interés nominales eleva el costo de capital de las firmas y por ende sus costos marginales de producción, lo que luego se traslada a precios. Como respuesta a la caída del salario real, los trabajadores exigen aumentos nominales, generando una espiral de precios y salarios. En este esquema, una suba de tasas tendría el efecto contrario al del modelo detrás del VAR-1, siendo inflacionaria a corto plazo.

2.3. LASSO

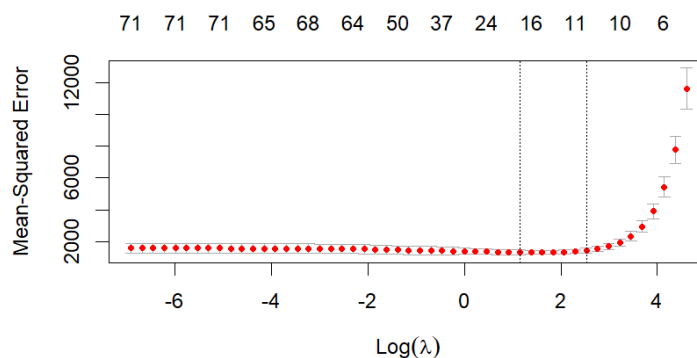
LASSO es un método de análisis de regresión que selecciona entre múltiples variables provistas, reduciendo a cero aquellas que no son útiles para predecir la variable dependiente (Silva Araujo y Piazza Gaglianone, 2023). El operador LASSO toma la forma:

$$\hat{\beta} = \min_{\{\beta_1, \dots, \beta_k\}} \left(\frac{1}{T} \sum_{t=1}^T \left(Y_t - \sum_{j=1}^k X'_{j,t} \beta_j \right)^2 + \lambda \sum_{j=1}^k |\beta_j| \right) \quad (4)$$

Donde β es un vector $k \times 1$ de parámetros, Y_t es la variable dependiente, $\{X'_{1,t}, \dots, X'_{k,t}\}$ es un vector $k \times 1$ de variables independientes y λ es el llamado parámetro de *shrinkage*, que “encoge” a cero los coeficientes que no son útiles. Fue planteada por primera vez en Tibshirani (1996) como método para aislar las medidas clínicas ligadas al cáncer de próstata.

El LASSO estima los coeficientes para las variables independientes y calcula la función de pérdida como la diferencia entre los valores estimados con estos coeficientes y los valores observados. La innovación es que agrega un “término penalizador”, $\lambda \sum_{j=1}^k |\beta_j|$ donde multiplica la suma de los coeficientes por el parámetro de *shrinkage* λ . Cuanto mayor sea λ , mayor es el costo de tener coeficientes distintos a cero. El LASSO elige entre posibles λ , buscando al que minimice el error cuadrático medio, como se ve en el gráfico 2.3.1. Para esto emplea la validación cruzada, una técnica que trabaja sobre un conjunto reducido de la muestra total, llamada de *training*.

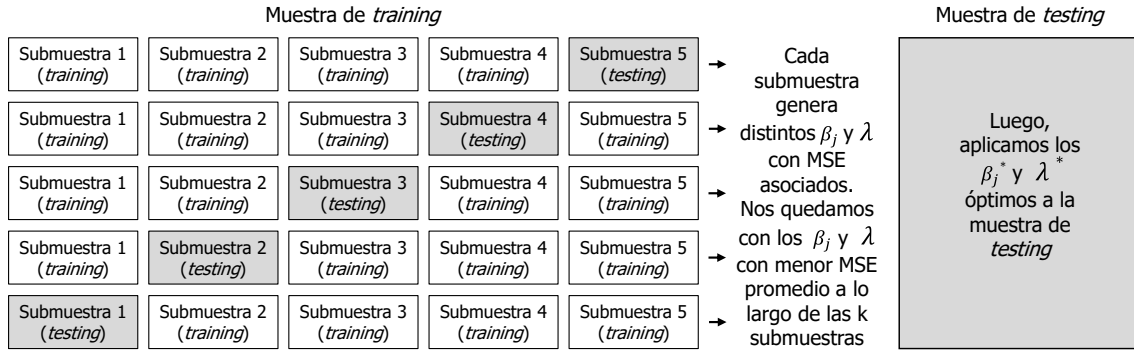
Gráfico 2.3.1.



La validación cruzada consiste en dividir esta muestra de *training* en k submuestras. Se utilizan $k - 1$ submuestras para estimar los coeficientes y el parámetro de *shrinkage* λ , y luego se testea esos valores sobre la submuestra restante. Esto se repite k veces, de manera que cada submuestra es utilizada para testear una vez⁸, proceso ejemplificado en el gráfico 2.3.2.

⁸ Por convención, se suele fijar $k = 5$. Así lo hacen los trabajos que tomamos de referencia y así lo hacemos en la regresión LASSO de este trabajo.

Gráfico 2.3.2.



La ventaja de la validación cruzada es que antes de trabajar con la muestra de *testing* (el conjunto que quedó por fuera de la muestra de *training*), ya tenemos una idea de cuan bien se adapta el modelo a información nueva. Los coeficientes β_j y parámetros λ con menor error cuadrático medio en el promedio de las submuestras de *testing*, o las que minimicen la función de pérdida, son las que terminamos seleccionando para el modelo final, que luego se aplica a la muestra de *testing* definitiva.

Para Tibshirani, el LASSO funciona mejor cuando hay un número moderado de variables independientes con efecto significativo, pero menor, en la variable dependiente. Funciona peor cuando hay un número pequeño de variables independientes con efecto determinante en la variable dependiente.

Creemos que el proceso inflacionario en Argentina se corresponde con la primera definición. Hay variables macroeconómicas muy influyentes, como el tipo de cambio o las tasas de interés, pero no se pueden descartar variables “menores” como los precios regulados, precios internacionales o indicadores monetarios menos conocidos.

2.4. ELASTIC NET

El ELASTIC NET es una generalización del LASSO propuesta por Zou y Hastie (2005), que también selecciona variables automáticamente con un parámetro de *shrinkage*. Es capaz de elegir grupos de variables correlacionadas. Toma la forma:

$$\hat{\beta} = \arg \min_{\{\beta_1, \dots, \beta_k\}} \left(\frac{1}{T} \sum_{t=1}^T \left(Y_t - \sum_{j=1}^k x_{j,t}' \beta_j \right)^2 + \lambda P_\alpha(\beta) \right) \quad (5)$$

donde:

$$P_\alpha(\beta) = \sum_{j=1}^k \alpha |\beta_j| + \frac{(1-\alpha)}{2} \beta_j^2 \quad (6)$$

Cuando $\alpha = 1$ es idéntico al LASSO. Para otros valores de $\alpha > 0$, el término penalizador P_α interpola entre la norma l_1 de β y el cuadrado de la norma l_2 de β . El parámetro λ controla la fuerza de la penalización (Silva Araujo y Piazza Gaglianone, 2023). La selección de los parámetros se realiza vía validación cruzada, al igual que en la regresión LASSO.

3. Datos y metodología de evaluación

Las 36 variables seleccionadas se basan en el *dataset* de Silva Araujo y Piazza Gaglianone (2023), adaptadas al contexto argentino. Los datos abarcan de junio de 2016 a marzo de 2024. Para aquellos con frecuencia diaria, se tomó el promedio mensual. Los índices de actividad están desestacionalizados. La lista completa de variables se encuentra en el anexo.

En línea con las normas del FMI, el INDEC se compromete a difundir el IPC mensual con “no más de 10 días hábiles de desfase”, en general entre el día 11 y 16 de cada mes. Aquellas variables que se publican después del IPC se incluyen en el modelo con un rezago adicional.

De acuerdo con la práctica habitual, dividimos la muestra completa en dos períodos. Especificamos los parámetros de los modelos con los datos de junio de 2016 a diciembre de 2021, siguiendo la recomendación de que el período de *training* abarque alrededor de dos-tercios de la muestra total (Silva Araujo y Piazza Gaglianone, 2023). Evaluamos el desempeño de los modelos en un periodo de *testing* que va de enero de 2022 a marzo de 2024, en dos horizontes temporales: un mes y tres meses adelante.

Utilizamos dos métricas estándar para comparar el poder predictivo de los modelos, el R2 y el error cuadrático medio (RMSE). Usamos el test de *forecast encompassing* desarrollado en Harvey et al. (1998) para evaluar la significancia estadística de las diferencias entre los modelos. También empleamos el test no-paramétrico de desempeño predictivo de Pesarán y Timmermann (1992).

Llamando X_t a las variables observadas y \hat{X}_t a las variables estimadas por el modelo, el R2 se calcula como la covarianza entre ambas variables, dividida por el producto de sus respectivos desvíos estándar, elevado al cuadrado. Intuitivamente, el R2 mide cuanto se aproximan las estimaciones a los datos reales. Su valor se ubica entre 0 y 1; cuanto más alto, mayor cercanía entre las series⁹.

$$R2 = \left[\frac{\sum_{t=1}^n (X_t - \bar{X})(\hat{X}_t - \bar{X})}{\sqrt{\sum_{t=1}^n (X_t - \bar{X})^2} \sqrt{\sum_{t=1}^n (\hat{X}_t - \bar{X})^2}} \right]^2 \quad (7)$$

El RMSE es la raíz de la suma del cuadrado de las diferencias entre las variables observadas y estimadas, divididas por el tamaño n de la muestra. Valores más bajos del RMSE indican una menor divergencia entre las estimaciones y los datos reales. A diferencia del R2, sus valores no toman un rango acotado.

$$RMSE = \sqrt{\left[\frac{\sum_{t=1}^n (X_t - \hat{X}_t)^2}{n} \right]} \quad (8)$$

⁹ Un R2 cercano a 1 no asegura que el modelo tenga buen poder predictivo, debido al problema del *overfitting*. Ocurre cuando un modelo se ajusta fuertemente a los datos de la muestra, pero no obtiene resultados similares al ampliar la muestra. La división de la muestra en períodos de *training* y *testing* busca controlar este problema, evaluando el R2 *out-of-sample* de los modelos.

Esto es relevante a nuestro trabajo porque abarca un periodo con múltiples regímenes de política monetaria y cambiaria. Por ejemplo, entre junio de 2016 y agosto de 2019 Argentina tuvo un tipo de cambio único, pero a partir de septiembre de 2019 impuso controles de cambio que dieron lugar a paridades múltiples. Podría ocurrir que un modelo que estime la inflación en función del tipo de cambio oficial obtenga un R2 alto en el período de *training*, pero que estos resultados no se trasladen al período de *testing*.

Ambas métricas cuantifican la distancia entre las predicciones y las observaciones. El R2 lo hace en términos porcentuales y el RMSE en términos absolutos.

El test de *forecast encompassing* de Harvey et al. (1998) compara las predicciones del modelo contra un *benchmark* para ver si contienen información que no esté presente allí. En este trabajo, utilizamos dos *benchmarks*. El ARIMA, por ser el modelo más simple, y la mediana del REM-BCRA, por ser una proyección de calidad elaborada de forma independiente por consultoras, bancos y centros de estudio económicos.

Al combinar dos predicciones \hat{X}_{1t} y \hat{X}_{2t} , cada una con un error $e_{it} = X_t - \hat{X}_{it}$, obtenemos una predicción combinada $\hat{X}_{ct} = (1 - \lambda) \hat{X}_{1t} + \lambda \hat{X}_{2t}$ con $0 \leq \lambda \leq 1$. Tomando ε_t como el error de la predicción combinada, podemos plantear $e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t$.

La predicción combinada debería tener un error menor al de \hat{X}_{1t} , al menos que la covarianza entre e_{1t} y $(e_{1t} - e_{2t})$ sea 0. La literatura sobre *forecast encompassing* considera que un pronóstico \hat{X}_{1t} contiene a otro \hat{X}_{2t} cuando se cumple la hipótesis nula $\lambda = 0$ (Harvey et al., 1998). El estadístico *MCS* toma una distribución normal asintótica y se construye:

$$MCS = \left\{ \frac{(\sum (e_{1t} - e_{2t})^2)}{ns^2} \right\}^{-1/2} n^{1/2} \hat{\lambda} \quad (9)$$

Donde s^2 es la varianza residual de $e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t$ y $\hat{\lambda}$ es el coeficiente del modelo.

El test no-paramétrico de desempeño predictivo de Pesaran y Timmermann (1992) pone el foco en la cantidad de veces que el modelo predice correctamente la dirección del cambio en la variable dependiente, fijándose solo en el signo y no en la magnitud del cambio. Para nuestro trabajo, si la inflación acelera o desacelera en un mes dado. El estadístico S_n , que toma una distribución normal asintótica, se construye:

$$S_n = \left\{ \frac{P_*(1 - P_*)}{n} \right\}^{-1/2} (\hat{P} - P_*) \quad (10)$$

Donde $P_* = \Pr(y_t x_t > 0)$, siendo y_t la dirección de la variable observada en el momento t , x_t la dirección de la predicción y \hat{P} la proporción de veces que se predice correctamente el signo de y_t .

En nuestra muestra, la inflación mensual aceleró en 16 de 27 meses. Este test es útil para ver cuales modelos captan mejor los períodos de aceleración y desaceleración, al margen de si aciertan el valor absoluto de la variación.

4. Resultados empíricos

4.1 Análisis preliminar de los datos

Evaluamos la estacionariedad de la inflación mensual entre junio de 2016 y diciembre de 2021 utilizando la prueba aumentada de Dickey-Fuller (ADF)¹⁰, con los resultados expuestos en la tabla 4.1. Nuestros resultados coinciden con los de Marinozzi (2023), que señala que con la aceleración

¹⁰ La prueba ADF testea la hipótesis nula de que la muestra de la serie del tiempo presenta una raíz unitaria, es decir, es no-estacionaria y por ende no cumple con las condiciones de media y varianza constantes a lo largo del tiempo, necesarias para el modelo ARIMA.

inflacionaria de los últimos años, el IPC de Argentina se ha vuelto integrado de orden 2, mientras que para la mayoría de los países es de orden 1. Nosotros trabajamos sobre la inflación mensual, que ya tiene una diferencia respecto al IPC.

Tabla 4.1. Prueba de raíz unitaria ADF

Estadístico ADF, p-valores entre parentesis	
Inflación mensual (jun-16 a dic-21)	ADF
Diferencias=0	-2,47 (0,38)
Diferencias=1	-5,62 (0,01)

Notas: hipótesis nula: con p-valor>0,05, la serie de tiempo tiene raíz unitaria

A su vez, el Criterio de Akaike nos arroja que el mejor ARIMA es uno de orden (0,1,2), es decir diferenciado una vez e incluyendo solo rezagos de los términos de error, similar a un Random Walk. Esto también está en línea con los resultados de Marinuzzi sobre la estacionariedad del IPC.

4.2. Predicciones en la muestra out-of-sample

Para predecir la inflación mensual entre enero de 2022 y marzo de 2024, en el horizonte a un mes los modelos con *machine learning* se imponen en la mayoría de las métricas, resumidas en la tabla 4.2.1., con una leve ventaja del ELASTIC NET, que obtiene un R2 de 0,87 y un RMSE de 1,81, sobre el LASSO. De acuerdo con el test de *forecast encompassing*, los resultados del ELASTIC NET marcan una diferencia significativa respecto del benchmark ARIMA, que tiene el peor desempeño entre los modelos. Sin embargo, la distancia entre el ELASTIC NET (o el LASSO) y el benchmark de la mediana del REM-BCRA no es estadísticamente significativa. El REM-BCRA se impone en el test no-paramétrico de dirección.

Tabla 4.2.1. Evaluación (h=1)

Período de testing: ene-22 a mar-24								
	R2	RMSE	S _n	MCS respecto a ARIMA		MCS respecto a REM-BCRA		
ARIMA	0,54	3,42	0,09 (0,46)	-	-	1,56 (0,13)		
VAR-1	0,70	2,79	0,56 (0,29)	0,46 (0,65)		1,53 (0,14)		
VAR-2	0,71	2,71	0,56 (0,29)	-0,15 (0,88)		1,33 (0,20)		
VAR-3	0,71	2,75	0,56 (0,29)	-0,05 (0,96)		1,54 (0,14)		
LASSO	0,84	2,03	1,23 (0,11)	-1,83 (0,08)		-0,97 (0,34)		
ELASTIC NET	0,87	1,81	1,70 (0,04)	-2,79 (0,01)		-0,67 (0,51)		
REM-BCRA	0,77	2,45	2,42 (0,01)	-1,56 (0,13)		-	-	

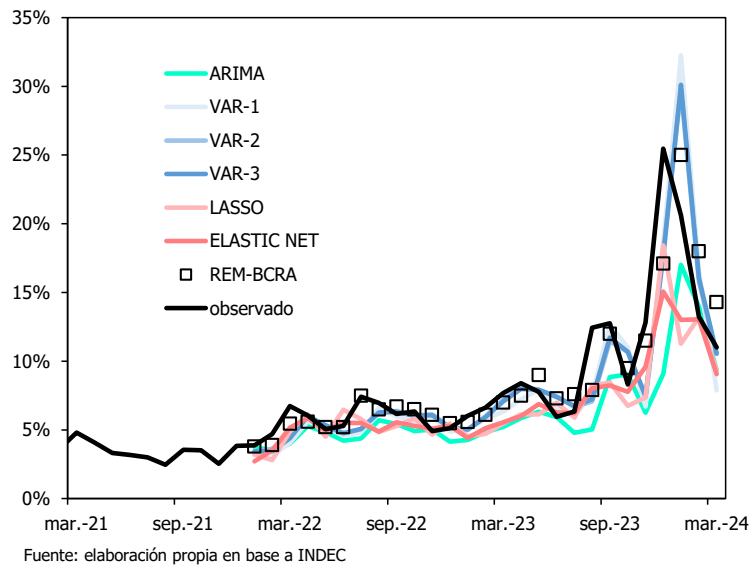
Notas: p-valores entre paréntesis, S_n: Pesaran y Timmermann (1992), MCS: Harvey et al. (1998).

El mejor resultado se marca en negrita. Para MCS, valores negativos indican un desempeño superior al benchmark.

Los VAR utilizan: VAR-1 (ARS/USD, EMAE, BADLAR), VAR-2 (ARS/USD, RIPTE), VAR-3 (ARS/USD, EMAE, RIPTE)

Como muestra el gráfico 4.2.1., el período evaluado contiene dos aceleraciones inflacionarias bruscas en agosto y diciembre de 2023, ambas luego de devaluaciones del tipo de cambio oficial. Un modelo *backward-looking* como el ARIMA tarda en incorporar estos *shocks*. Los modelos con *machine learning* reaccionan a las dos devaluaciones, aunque no las captan en toda su magnitud. En cambio, los modelos VAR sobreestiman el pico inflacionario de diciembre. Esto se puede deber a que el *pass-through* de tipo de cambio a precios minoristas fue menor que en otras ocasiones.

Gráfico 4.2.1. Inflación mensual y predicciones (h=1)



Todos los modelos tienden a subestimar la inflación mensual. Sobre todo el ARIMA, que produce predicciones debajo del valor observado en 89% de los meses del período de *testing*, algo lógico al ser un modelo que predice solo en función de observaciones pasadas en un período en que la inflación se aceleró marcadamente. El LASSO y el ELASTIC NET subestiman la inflación en 85% de los meses de la muestra y los modelos VAR en 67%, en promedio. La mediana del REM-BCRA es la única que no exhibe este sesgo en el horizonte a un mes, con predicciones debajo de la inflación observada en 52% de la muestra.

El ELASTIC NET también se impone en el horizonte a tres meses, con un R2 de 0,71 y un RMSE de 2,73. La diferencia de los dos modelos *machine learning* es significativa respecto al benchmark ARIMA, de desempeño muy pobre en este horizonte. Nuevamente, la diferencia respecto al benchmark de la mediana del REM-BCRA es demasiado estrecha para ser estadísticamente significativa. El VAR-2 (ipc, arsUSD y ripte) se impone en el test no-paramétrico de dirección, aunque no de forma estadísticamente significativa, como se ve en la tabla 4.2.2.

Tabla 4.2.2. Evaluación (h=3)

Período de testing: mar-22 a mar-24

	R2	RMSE	S _n	MCS respecto a ARIMA	MCS respecto a REM-BCRA
ARIMA	0,15	4,71	-0,70 (0,76)	-	5,43 (0,00)
VAR-1	0,37	4,05	-1,18 (0,88)	0,14 (0,89)	1,83 (0,08)
VAR-2	0,42	3,90	0,59 (0,28)	-0,04 (0,97)	1,74 (0,10)
VAR-3	0,39	3,97	-0,34 (0,63)	0,08 (0,94)	1,77 (0,09)
LASSO	0,71	2,74	0,13 (0,45)	-2,46 (0,02)	-0,67 (0,51)
ELASTIC NET	0,71	2,73	0,13 (0,45)	-2,65 (0,01)	-0,76 (0,45)
REM-BCRA	0,64	3,08	0,13 (0,45)	-5,43 (0,00)	-

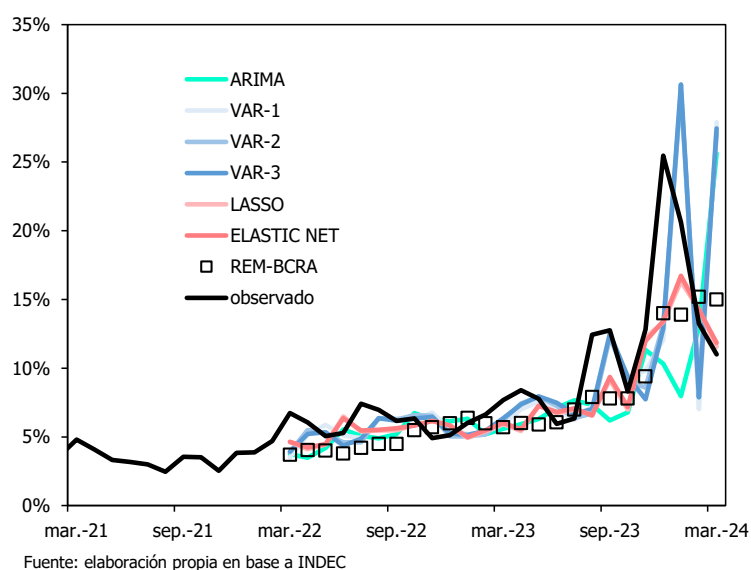
Notas: p-valores entre paréntesis, S_n: Pesaran y Timmermann (1992), MCS: Harvey et al. (1998).

El mejor resultado se marca en negrita. Para MCS, valores negativos indican un desempeño superior al benchmark.

Los VAR utilizan: VAR-1 (ARS/USD, EMAE, BADLAR), VAR-2 (ARS/USD, RIPTTE), VAR-3 (ARS/USD, EMAE, RIPTTE)

El sesgo a subestimar la inflación mensual también se presenta en el horizonte a tres meses. Los modelos generan predicciones debajo del valor observado en 67% de la muestra en promedio, desde 60% para los VAR-2 y VAR-3 a 72% para los modelos de *machine learning*. En este caso el sesgo se extiende al *benchmark* de la mediana del REM-BCRA, que subestima la inflación en 72% de la muestra para este horizonte.

Gráfico 4.2.2. Inflación mensual y predicciones (h=3)



La regresión con ELASTIC NET, la de mejor desempeño general, asigna coeficientes distintos de cero a 25 de las 36 variables del *dataset*. La tabla 4.2.3 exhibe las primeras 10, ordenadas según el valor absoluto del coeficiente. En primer lugar aparecen los índices de precios mayoristas y al consumidor, seguidos por la variación de los depósitos en pesos y dólares del sector privado. Por supuesto, el modelo solo busca correlación tal que permita predecir la inflación y no relaciones de causalidad. Además, algunas de las 36 variables están muy correlacionadas entre sí, o son mediciones ligeramente diferentes o fraccionadas de la misma variable macroeconómica.

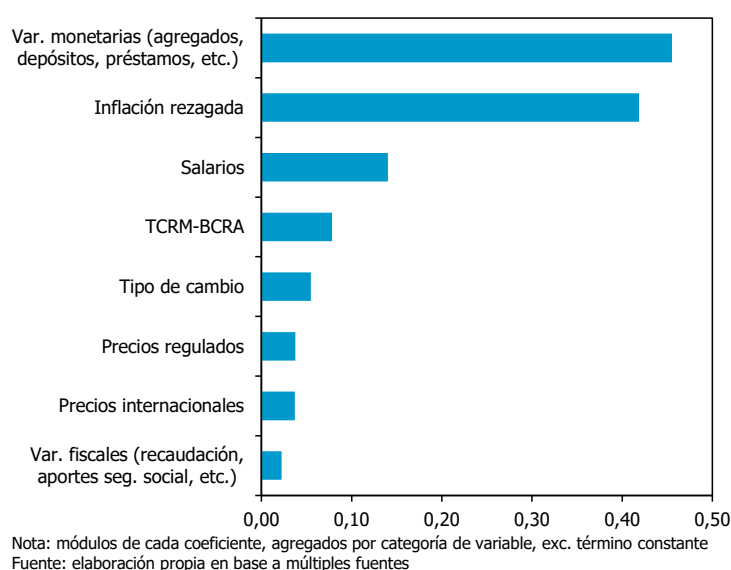
Tabla 4.2.3. Principales coeficientes de la regresión ELASTIC NET

Coefficientes ordenados por valor absoluto, excluyendo término independiente

Variable	Categoría	Código	Coefficiente
Índice de precios mayoristas	Inflación rezagada	sipm	0,25
Índice de precios al consumidor	Inflación rezagada	ipc	0,17
Depósitos del sector privado	Monetaria	depriv	0,11
Tipo de cambio real multilateral	TCRM-BCRA	tcrm	-0,08
M3	Monetaria	m3	-0,07
Salario nominal privados formales	Salarios	ripte	0,06
Préstamos al sector privado, tarjeta de crédito	Monetaria	prestarj	0,05
Préstamos al sector privado	Monetaria	prespriv	-0,05
Tipo de cambio oficial	Tipo de cambio	arsusd	0,05
Depósitos privados en plazo fijo	Monetaria	deprivpf	0,04

En el gráfico 4.2.3 hemos agrupado los coeficientes según el tipo de variable, tomando los módulos para ver su incidencia en la predicción independientemente de su signo. Encontramos que la regresión otorga mayor peso a las variables monetarias (agregados monetarios como M2, M3, base monetaria o circulante, depósitos y préstamos al sector privado), con una ponderación de 37% dentro del total de coeficientes, y a la inflación de períodos anteriores (tanto del índice de precios mayorista como al consumidor), con una ponderación de 34%.

Gráfico 4.2.3. Coeficientes agrupados de la regresión ELASTIC NET



En tercer lugar aparecen los salarios (11%), seguidos por el índice de tipo de cambio real multilateral del BCRA¹¹ (6%) y el tipo de cambio (4%). La regresión otorga un peso menor a los precios regulados, los precios internacionales y variables fiscales como la recaudación tributaria, aportes a la seguridad social, etc., con una ponderación conjunta de 8% en el total.

Las variables elegidas por la regresión LASSO no difieren demasiado de las del ELASTIC NET. A pesar de aplicar un término de penalización más estricto, LASSO solo asignó coeficientes iguales a cero a 2 de 36 variables. De las 10 variables con mayor coeficiente absoluto, 6 coinciden con las del ELASTIC NET, mientras que aparecen como novedades los depósitos privados en cuenta corriente, el salario mínimo, el precio de la soja en dólares y el precio mayorista de la electricidad.

Tabla 4.2.4. Principales coeficientes de la regresión LASSO

Coefficientes ordenados por valor absoluto, excluyendo término independiente

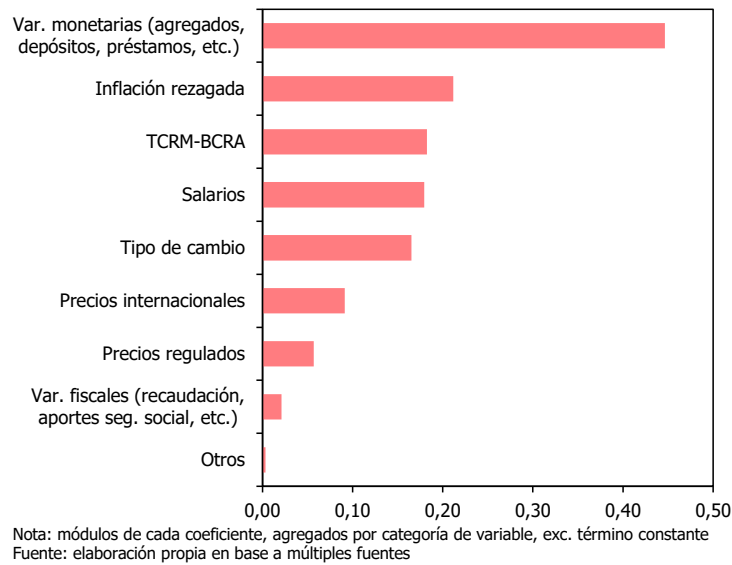
Variable	Categoría	Código	Coefficiente
Índice de precios mayoristas	Inflación rezagada	sipm	0,21
Tipo de cambio real multilateral	TCRM-BCRA	trcm	-0,18
Tipo de cambio oficial	Tipo de cambio	arsusd	0,16
Salario nominal privados formales	Salarios	ripte	0,13
Depósitos del sector privado	Monetaria	deppriv	0,12
Préstamos al sector privado	Monetaria	prespriv	-0,07
Depósitos privados en cuenta corriente	Monetaria	ctacte	0,07
Salario mínimo	Salarios	smvm	0,05
Precio de la soja en dólares	Precios internacionales	sojausd	0,05
Precio mayorista de la electricidad	Precios regulados	pelect	0,05

Notas: las variables resaltadas coinciden con las primeras 10 de la regresión ELASTIC NET

Sin embargo, si agrupamos los coeficientes por el tipo de variable, encontramos una estructura de ponderadores similar a la del ELASTIC NET, donde predominan las variables monetarias y la inflación rezagada, seguidas por el índice de tipo de cambio real multilateral del BCRA, los salarios nominales y el tipo de cambio, visible en el gráfico 4.2.4.

¹¹ Como el índice de TCRM-BCRA clasificaría en parte dentro de inflación rezagada y en parte dentro de tipo de cambio, lo separamos del resto de las variables, en una categoría propia.

Gráfico 4.2.4. Coeficientes agrupados de la regresión LASSO



En ambos modelos, los signos de los coeficientes son mayormente los intuitivos. Al tratarse en su mayoría de variables nominales (inflación rezagada, tipo de cambio, salarios, precios regulados e internacionales) figuran con signo positivo. El signo no distingue entre causa y efecto: puede ser por shocks exógenos (tipo de cambio, precios regulados) que aceleran la inflación, o, en el caso de los salarios, una reacción a la aceleración inflacionaria que se extiende a lo largo de varios meses. Una de las pocas excepciones es el tipo de cambio real multilateral, con signo negativo. Atribuimos esto a su uso repetido como ancla antiinflacionaria, de manera que los períodos de apreciación real coinciden con la desinflación (y viceversa). Los préstamos al sector privado también figuran con signo negativo en ambos modelos de *machine learning*, lo que puede deberse a la paralización del crédito ante aceleraciones súbitas de la inflación.

Tanto ELASTIC NET como LASSO asignan coeficientes iguales o cercanos a cero al índice de actividad económica, las tasas de interés y el tipo de cambio paralelo (ccl). La exclusión de actividad y tasas de interés como variables relevantes es consistente con el mejor desempeño del VAR-2, que toma salarios nominales y tipo de cambio, por sobre el VAR-1 y el VAR-3 que utilizan actividad, tasas y tipo de cambio, aunque esta diferencia no es estadísticamente significativa.

En cuanto a la baja incidencia del tipo de cambio paralelo en la inflación, es posible que nuestros resultados estén afectados por la elección del período de *training*. Hay un debate sobre si el *pass-through* de los paralelos aumentó durante 2022, con el endurecimiento de las restricciones para importar y la disparada de la brecha cambiaria, aunque no es objeto de este trabajo.

En nuestra visión, las regresiones destacan a los depósitos privados porque sirven a la vez de *proxy* de la expansión monetaria, de la masa salarial y del tipo de cambio, al agrupar depósitos en pesos y en dólares medidos en pesos, por lo que *ceteris paribus*, una devaluación del tipo de cambio oficial automáticamente expande el monto total de depósitos medido en pesos. En nuestro régimen de alta inflación, la variación en el corto plazo de la tasa mensual de inflación se determina principalmente por la inflación rezagada, presente en los salarios y los precios de los servicios, y los shocks exógenos del tipo de cambio oficial y precios regulados. Creemos que las variables seleccionadas por el LASSO y el ELASTIC NET captan apropiadamente esta dinámica.

5. Conclusiones

En este trabajo construimos cuatro modelos econométricos para predecir la inflación argentina entre junio de 2016 y marzo de 2024, evaluando su desempeño en el período de muestra entre enero de 2022 y marzo de 2024 contra dos *benchmarks*, el modelo más simple (ARIMA) y la mediana de la encuesta REM-BCRA. En los horizontes a uno y tres meses, los modelos con *machine learning*, LASSO y ELASTIC NET, obtienen resultados superiores al REM-BCRA y el ARIMA, aunque solo con una diferencia estadísticamente significativa respecto al segundo.

El mejor desempeño de los modelos con *machine learning* respecto a los tradicionales para predecir la inflación es consistente con los resultados de García et al. (2017) y Silva Araujo y Piazza Gaglianone (2023) para Brasil, Medeiros y Mendes (2016) y Medeiros et al. (2021) para Estados Unidos o Özgür y Akkoç (2022) para Turquía.

Dentro de las variables seleccionadas por las regresiones ELASTIC NET y LASSO, además de los “sospechosos de siempre”; inflación rezagada, tipo de cambio y salarios, encontramos una amplia variedad de agregados monetarios como depósitos y préstamos del sector privado o medidas de la oferta de dinero como el M3. Sin embargo, creemos que la mayor ponderación otorgada por los modelos con *machine learning* a los depósitos privados se debe a que actúan como *proxy* de los salarios formales y, en cierta medida, el tipo de cambio oficial.

En síntesis, encontramos un desempeño sólido de las técnicas de *machine learning* para predecir la inflación argentina en el período evaluado. Además, nos permiten trabajar con un *dataset* más amplio que los modelos tradicionales y aportar evidencia cuantitativa sobre el rol de las distintas variables macroeconómicas en el proceso inflacionario.

6. Bibliografía

- Aromí, Jose Daniel y Martín Llada (2020), "Forecasting inflation with Twitter", LIV Reunión Anual de la Asociación Argentina de Economía Política.
- Atkenson, Andrew y Lee Ohanian (2001), "Are Phillips Curves useful for forecasting inflation?", Federal Reserve Bank of Minneapolis Quarterly Review, Vol. 25, No. 1, pp. 2-11.
- Box, George, Gwilym Jenkins, Gregory Reinsel y Greta Ljung (1970), Time Series Analysis, Wiley.
- D'Amato, Laura, Lorena Garegnani y Emilio Blanco (2008), "Forecasting Inflation in Argentina: Individual Models or Forecast Pooling?", Economic Research Working Papers, BCRA.
- D'Amato, Laura, Maximiliano Gómez Aguirre, Lorena Garegnani, Ariel Krysa y Luis Libonatti (2018), "Forecasting Inflation in Argentina: A Comparison of Different Models", Economic Research Working Papers, BCRA.
- Diebold, Francis (1998), "The Past, Present and Future of Macroeconomic Forecasting", Journal of Economic Perspectives, Vol. 12, No. 2, pp. 175-192.
- García, Márcio G.P., Marcelo Medeiros y Gabriel F.R. Vasconcelos (2017), "Real-time inflation forecasting with high-dimensional models: The case of Brazil", International Journal of Forecasting, Vol. 33, No. 3, pp. 679-693.
- Hall, Aaron (2018), "Machine Learning Approaches to Macroeconomic Forecasting", Economic Review, Federal Reserve Bank of Kansas City, cuarto trimestre de 2018, pp. 63-81.
- Harvey, David, Stephen J. Leybourne y Paul Newbold (1998), "Tests for forecast encompassing", Journal of Business and Economic Statistics, Vol. 16, No. 2, pp. 254-259.
- Heymann, Daniel y Axel Leijonhufvud (1995), High inflation, Clarendon Press.
- Frenkel, Roberto y Diego Friedheim (2017) "Inflation in Argentina during the 2000s", Journal of Post Keynesian Economics, Vol. 40, No. 1, pp. 43-60.
- Lucas, Robert (1976), "Econometric Policy Evaluation: A Critique", Carnegie-Rochester Conference Series on Public Policy, Vol. 1, pp. 19-46.
- Marinozzi, Tomás (2023), "Forecasting Inflation in Argentina: A Probabilistic Approach", Ensayos Económicos (BCRA), No. 81.
- Medeiros, Marcelo y Eduardo F. Mendes (2016), " ℓ_1 -regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors", Journal of Econometrics, Vol. 191, No. 1, pp. 255-271.
- Medeiros, Marcelo, Gabriel F. R. Vasconcelos, Álvaro Veiga y Eduardo Zilberman (2021), "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods", Journal of Business and Economic Statistics, Vol. 39, No.1, pp. 98-119.
- Özgür, Önder y Uğur Akkoç (2022), "Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms", International Journal of Emerging Markets, Vol. 17, No. 8, pp. 1889-1908.
- Pesaran, Hashem y Allan Timmermann (1992), "A simple nonparametric test of predictive performance", Journal of Economic and Business Statistics, Vol. 10, No. 4, pp. 461-465.

Silva Araujo, Gustavo y Wagner Piazza Gaglianone (2023), "Machine learning methods for inflation forecasting in Brazil: new contenders *versus* classical models", Latin American Journal of Central Banking, Vol. 4, No. 2, 100087.

Sims, Christopher (1972), "Money, income and casualty", American Economic Review, Vol. 62, No.4, pp. 540-552.

Sims, Christopher (1980a), "Comparison of interwar and postwar business cycles: monetarism revisited", American Economic Review, Vol. 70, No. 2, pp. 250-257.

Sims, Christopher (1980b), "Macroeconomics and reality", Econometrica, Vol.48, No. 1, pp. 1-48.

Stock, James y Mark Watson (1999), "Forecasting inflation", Journal of Monetary Economics, Vol. 44, No. 2, pp. 293-335.

Tadeu Lima, Gilberto y Mark Setterfield (2010), "Pricing Behaviour and the Cost-Push Channel of Monetary Policy", Review of Political Economy, Vol. 22, No. 1, pp. 19-40.

Terasvista, Timo, Dick van Dijk y Marcelo Medeiros (2005), "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination", International Journal of Forecasting, Vol. 21, No. 4, pp. 755-774.

Tibshirani, Robert (1996), "Regression shrinkage and selection via the Lasso", J. R. Statist. Soc., Vol. 58, No. 1, pp. 267-288.

Zou, Hui y Trevor Hastie (2005), "Regularization and Variable Selection Via the Elastic Net", Journal of the Royal Statistical Society Series B: Statistical Methodology, Vol. 67, No. 2, pp. 301-320.

Anexo: *dataset* completo utilizado en este trabajo

Suponiendo que el IPC del mes anterior (t-1) se publica el día 15 del mes (t)

	Categoría	Código	Disponibilidad temporal			Fuente
			t-3	t-2	t-1	
1 Índice de precios al consumidor	Inflación rezagada	ipc			X	INDEC
2 Tipo de cambio oficial (ARS/USD)	Tipo de cambio	arsusd			X	Reuters
3 Salario privado formal nominal	Salarios	ripte		X		Sec. de Trabajo
4 Estimador de actividad económica	Actividad	emae	X			INDEC
5 Índice de precios mayorista	Inflación rezagada	sipm		X		INDEC
6 Tasa de interés de plazos fijos a 30-59 días	Monetaria	badlar			X	BCRA
7 Tasa de interés de call interbancario	Monetaria	call			X	BCRA
8 Depósitos del sector privado	Monetaria	deppriv			X	BCRA
9 Depósitos del sector público	Monetaria	deppub			X	BCRA
10 Depósitos privados en cuenta corriente	Monetaria	ctacte			X	BCRA
11 Depósitos privados en caja de ahorro	Monetaria	cajahorro			X	BCRA
12 Depósitos privados en plazo fijo	Monetaria	depprivpf			X	BCRA
13 Salario mínimo nominal	Salarios	smvm			X	Sec. de Trabajo
14 Préstamos al sector privado	Monetaria	prespriv			X	BCRA
15 Préstamos al sector privado, tarjeta de crédito	Monetaria	prestarj			X	BCRA
16 Préstamos al sector público	Monetaria	prestpub			X	BCRA
17 Aportes a la seguridad social	Fiscal	aportes			X	AFIP
18 Recaudación de IVA-DGI	Fiscal	ivadgi			X	AFIP
19 Recaudación fiscal	Fiscal	reca			X	AFIP
20 Tipo de cambio paralelo (CCL)	Tipo de cambio	ccl			X	Reuters
21 Recaudación de derechos de importación	Fiscal	dem			X	AFIP
22 Base monetaria	Monetaria	bm			X	BCRA
23 Circulante	Monetaria	circul			X	BCRA
24 M1	Monetaria	m1			X	BCRA
25 M2	Monetaria	m2			X	BCRA
26 M3	Monetaria	m3			X	BCRA
27 Precio de la soja en dólares	Precios internacionales	sojausd			X	Reuters
28 Precio del maíz en dólares	Precios internacionales	maizusd			X	Reuters
29 Precio del trigo en dólares	Precios internacionales	trigousd			X	Reuters
30 Tipo de cambio real multilateral	TCRM-BCRA	tcrm			X	BCRA
31 Reservas internacionales	Monetaria	rrii			X	BCRA
32 Confianza del consumidor	Actividad	confcons			X	UTDT
33 Precio del petróleo en dólares	Precios internacionales	oilusd			X	Reuters
34 Precio del gas natural en dólares	Precios internacionales	gasusd			X	Reuters
35 Precio mayorista de la electricidad	Precios regulados	pelect		X		CAMMESA
36 Precio mayorista de la electricidad + transporte	Precios regulados	pelecttr		X		CAMMESA

