**23 de enero de 2025**

**Prueba Gerente Modelador (a) de Riesgo de Crédito**

El objetivo de esta prueba es conocer con mayor detalle la forma en que la candidata o el candidato aborda un problema, así como la congruencia y claridad del hilo conductor que se plantea para dar respuesta a las interrogantes de este ejercicio.

Se pide a quien aplica a la vacante que proporcione el material suficiente para que quien evalúa la prueba pueda entender el racional y los argumentos utilizados.

Es importante mencionar que este ejercicio no tiene respuestas únicas y aunque se proporcionan los elementos básicos para el análisis, no se restringe al participante el deseo de incluir elementos adicionales que a su consideración pudieran nutrir o robustecer sus respuestas. Tampoco se requiere algún lenguaje de programación en particular.

Finalmente, se solicita enviar las respuestas a más tardar el miércoles a las 9 am, y así dar agilidad para la revisión y transmisión de la retroalimentación correspondiente.

**Planteamiento del problema**

Dentro del paquete enviado para el ejercicio se incluye una base con series macroeconómicas, así como indicadores de crédito al consumo que serán el principal insumo para la actividad. Estas series se encuentran disponibles para distintas ventanas de información, por lo que, en primer lugar**, el participante tendrá que definir la ventana que considere adecuada para realizar el ejercicio[[1]](#footnote-1).**

Esta prueba se centra en **el análisis de la variación anual del Índice de Morosidad (IMORA\_YoY)** de la cartera consolidado de consumo para la Banca Múltiple, a partir de su propio comportamiento, así como de los factores externos que pudieran -o no- influir sobre cambios y posibles valores futuros de la serie. **El objetivo final consiste en ofrecer un pronóstico de los valores de la serie para los siguientes dos años.**

En este sentido, cada participante deberá abarcar los siguientes puntos de manera cualitativa y cuantitativa y dar respuesta a algunos cuestionamientos:

* Apartado introductorio que detalle la metodología y los pasos seguidos para el desarrollo del ejercicio; (**Toda la información sobre este análisis la dejo anexada en el documento *Respuesta Humberto Ortiz Espinosa.ipynb* )**
  + **R- La metodología utilizada para este ejercicio consistió en calcular la variación anual del Índice de Morosidad (IMORA) usando Python con la función pct\_change para obtener su variación porcentual. Posteriormente, se gestionaron los outliers y se aplicaron transformaciones como logaritmo, Box-Cox y raíz cuadrada para suavizar los datos. Se analizó la tendencia de la información y se construyó una matriz de correlación para evaluar la relación de variables externas con el IMORA. Finalmente, se desarrollaron modelos de predicción, como ARIMA, Random Forest y LSTM, para predecir el IMORA en los años 2025 y 2026, optimizando los modelos con las variables más relevantes**
* Análisis exploratorio de la serie objetivo (IMORA\_YoY), así como de las series explicativas que los aspirantes seleccionen como determinantes para el ejercicio;
  + ¿Qué criterios se consideraron para esta decisión?

**R- Para el análisis exploratorio, se consideraron los siguientes criterios: Correlación con IMORA, se analizó la relación de cada serie explicativa con el IMORA utilizando una matriz de correlación; Disponibilidad y calidad de los datos: Se priorizaron las series que contaban con datos completos y consistentes a lo largo del tiempo, ya que la calidad de los datos es crucial para entrenar modelos precisos; Relevancia económica o financiera: Se seleccionaron variables externas que tuviesen un impacto directo sobre la morosidad, tales como tasas de interés, inflación, desempleo; Estabilidad temporal: Se consideraron series explicativas que tuvieran una evolución temporal coherente con el IMORA, es decir, aquellas que mostraran patrones estacionales o tendencias alineadas con el comportamiento histórico del IMORA.**

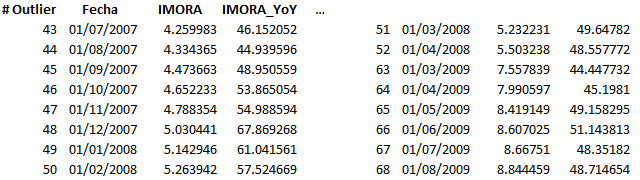
* Identificación de periodos de mayor adversidad económica (ventanas de información explícitas) para cada una de estas series (ej. La caída del Producto

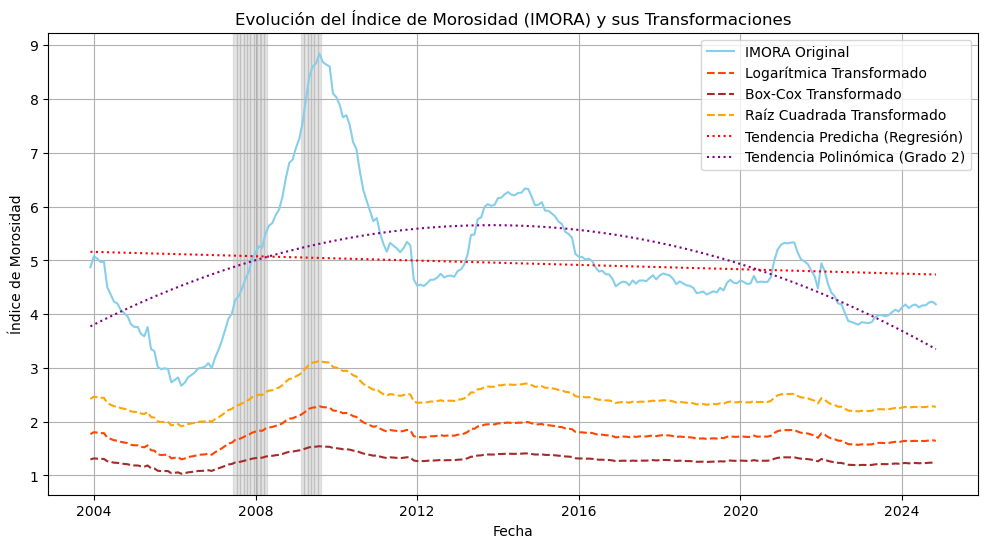
Interno Bruto crisis sanitaria por COVID19);

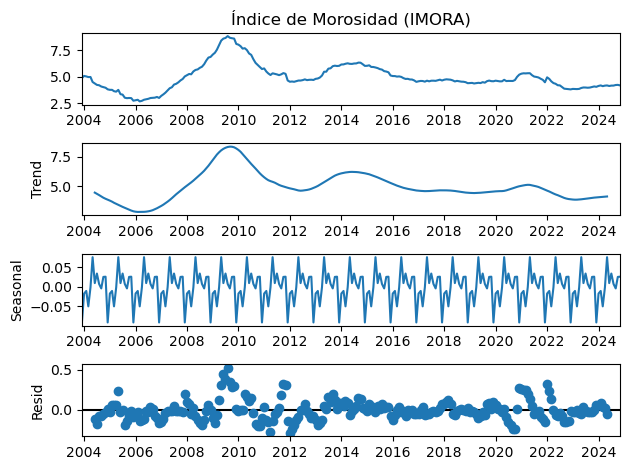
* + ¿Qué aspectos se tomaron en cuenta para esta esta decisión? R1
  + ¿Se realizó alguna transformación sobre las variables para dicho análisis? R2

**R1- Para identificar los periodos de mayor adversidad económica, se tomaron en cuenta los *outliers* en el Índice de Morosidad Anual (IMORA) durante eventos económicos clave, como la crisis de 2008 y la crisis sanitaria por COVID-19. En particular, la crisis de 2008 resultó en *outliers* extremadamente fuertes en el IMORA, tanto que estos valores distorsionan la capacidad de ver el impacto de otros eventos en años posteriores. Los valores fuera de los límites establecidos por el rango intercuartílico (IQR) fueron identificados como *outliers*, y estos marcaron un aumento tan drástico en el IMORA que no permitieron que el análisis futuro reflejara con claridad los efectos de otras crisis económicas o eventos similares.**

**R2- Se aplicaron transformaciones de suavización (logarítmica, Box-Cox y raíz cuadrada) para mitigar el efecto de estos *outliers* extremos y facilitar el análisis de tendencias a largo plazo. A pesar de estas transformaciones, los *outliers* de la crisis de 2008 siguen siendo tan representativos que afectan la distribución del IMORA, impidiendo que la serie de datos posterior capture el impacto de otras crisis con la misma intensidad. Además, se implementaron regresiones lineales y polinómicas para modelar la tendencia general del IMORA, destacando cómo las crisis impactaron la morosidad, especialmente durante el periodo de 2008.**

****





**23 de enero de 2025**

* Identificación de cambios estructurales en el comportamiento de las series;

**R- Si, la identificación de cambios estructurales en el comportamiento de las series implica detectar eventos excepcionales que alteran significativamente su patrón, como la crisis financiera de 2008 o la pandemia de COVID-19. Estos cambios suelen reflejarse en outliers o fluctuaciones extremas que pueden distorsionar las predicciones. Para mitigar su impacto, se aplican transformaciones como logaritmo o Box-Cox, suavizando así los efectos de estos puntos atípicos y permitiendo una mejor identificación de las tendencias a largo plazo**

* Métricas de relación y/o causalidad entre la serie objetivo y las series explicativas;

**R- Para la relación y causalidad entre la serie objetivo (IMORA) y las series externas, se realizó un análisis de correlación mediante un mapa de calor, que muestra la relación con estas variables.**

o ¿Qué variables guardan mayor relación con la serie objetivo**?**

* + ¿Qué variables guardan mayor relación causal con la serie objetivo?
* **R- Índice Nacional de Precios al Productor, Tasa de desocupación, Créditos al consumo (valores en abs)**

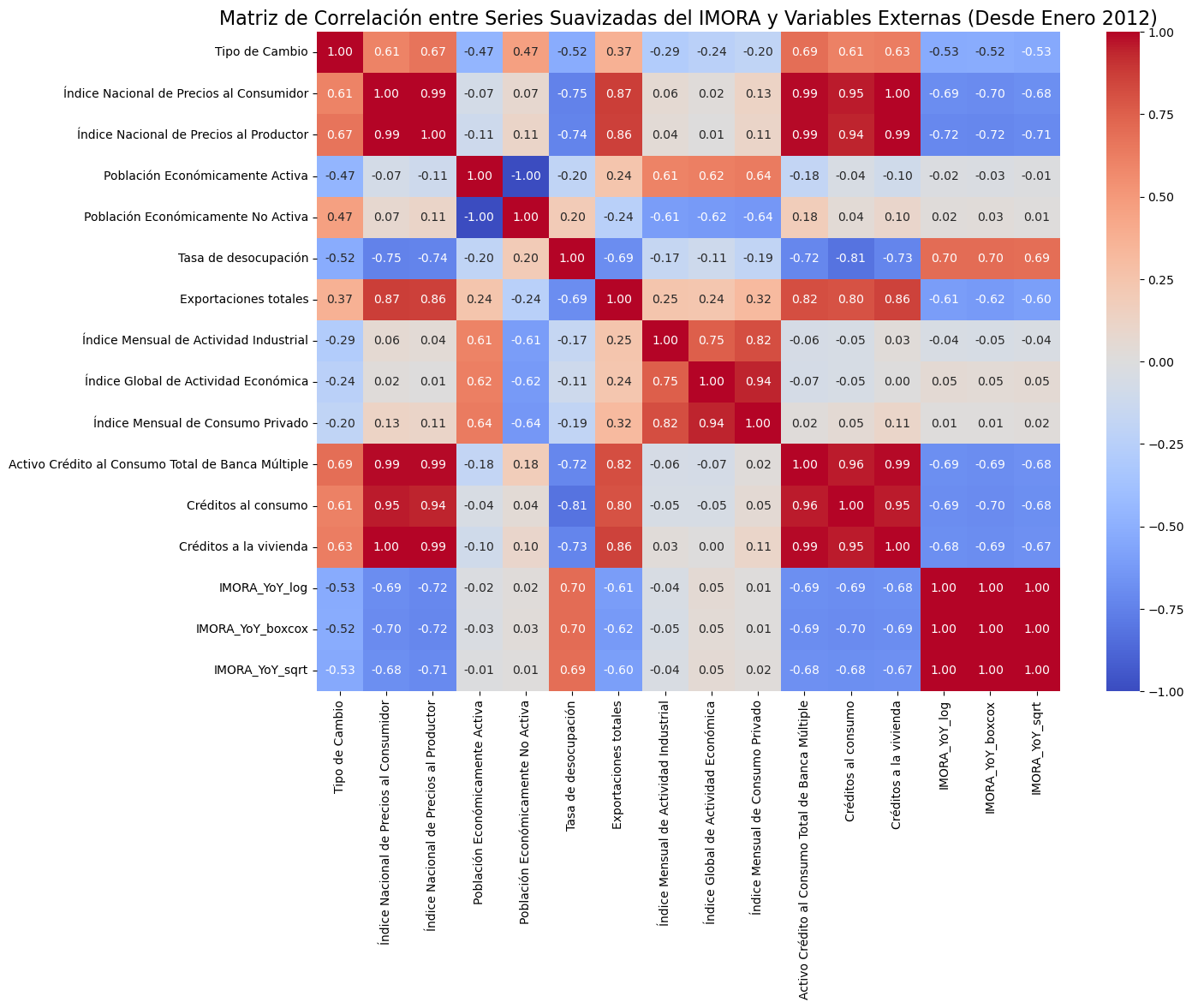
|  |  |
| --- | --- |
| Tasa de desocupación | 0.683729 |
| Índice Global de Actividad Económica | 0.05809 |
| Índice Mensual de Consumo Privado | 0.02042 |
| Población Económicamente No Activa | 0.006042 |
| Población Económicamente Activa | -0.006042 |
| Índice Mensual de Actividad Industrial | -0.030485 |
| Tipo de Cambio | -0.541231 |
| Exportaciones totales | -0.582672 |
| Créditos a la vivienda | -0.663411 |
| Activo Crédito al Consumo Total de Banca Múltiple | -0.668783 |
| Índice Nacional de Precios al Consumidor | -0.670028 |
| Créditos al consumo | -0.674731 |
| Índice Nacional de Precios al Productor | -0.7002 |

* + ¿Qué variables guardan mayor relación causal con la serie objetivo?

**R- las variables como Créditos al consumo y el Índice Nacional de Precios al Productor parecen tener una relación causal más pronunciada, dado su impacto negativo en la morosidad, lo que podría sugerir que el aumento en estas variables genera un incremento en el índice de morosidad.**

* + ¿Los dos puntos anteriores conducen a la misma respuesta? o ¿Existen relaciones entre las variables explicativas? ¿Esto podría generar problemas en el desarrollo del modelo?

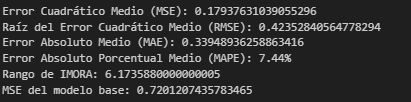
**R- Los resultados de correlación y causalidad no siempre conducen a la misma respuesta, ya que la correlación no implica causalidad. Las relaciones entre variables explicativas también pueden generar problemas de multicolinealidad, afectando el rendimiento de los modelos predictivos. También se tiene que considerar para evitar sesgos en los resultados yno crear sobreajustes.**



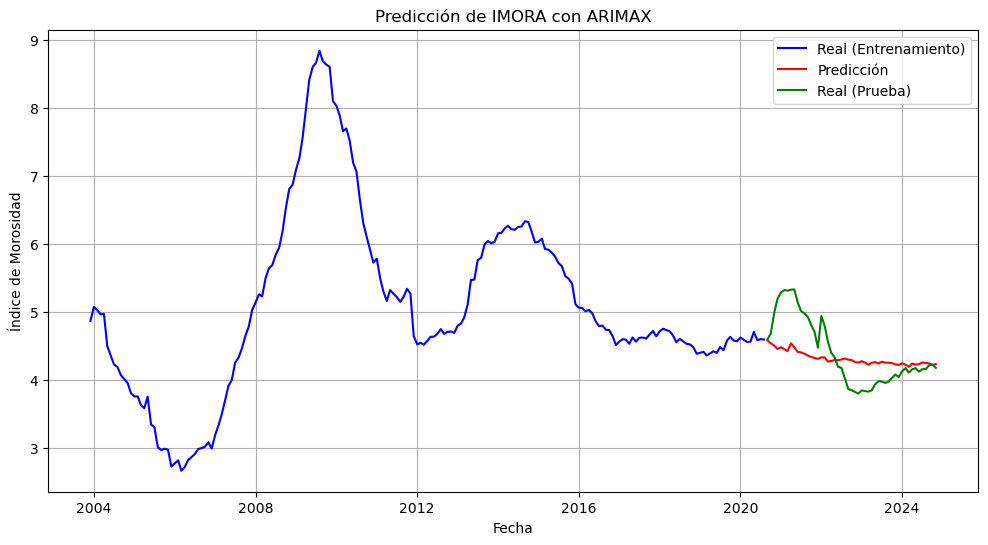
* Pronóstico del valor del IMORA\_YoY e IMORA para los siguientes dos años a partir de las variables explicativas;

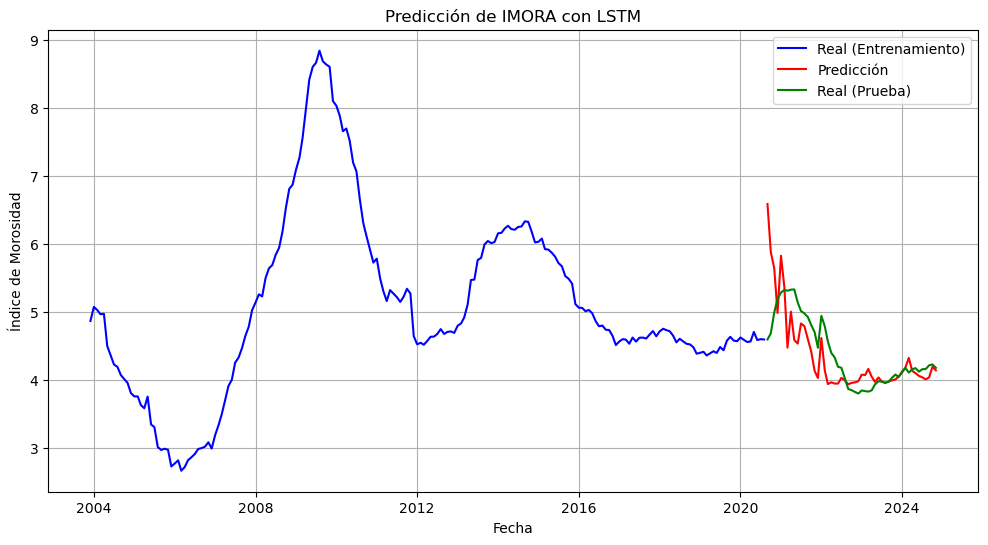
**R- Se utilizaron los Modelos ARIMA y LSTM para evaluar las proyecciones en los próximos dos años, basándonos en las variables explicativas más relevantes identificadas en el análisis de correlación y causalidad**.

**Anexo los resultados del modelo ARIMA de efectividad:**



**Modelos Entrenamiento comportamiento:**

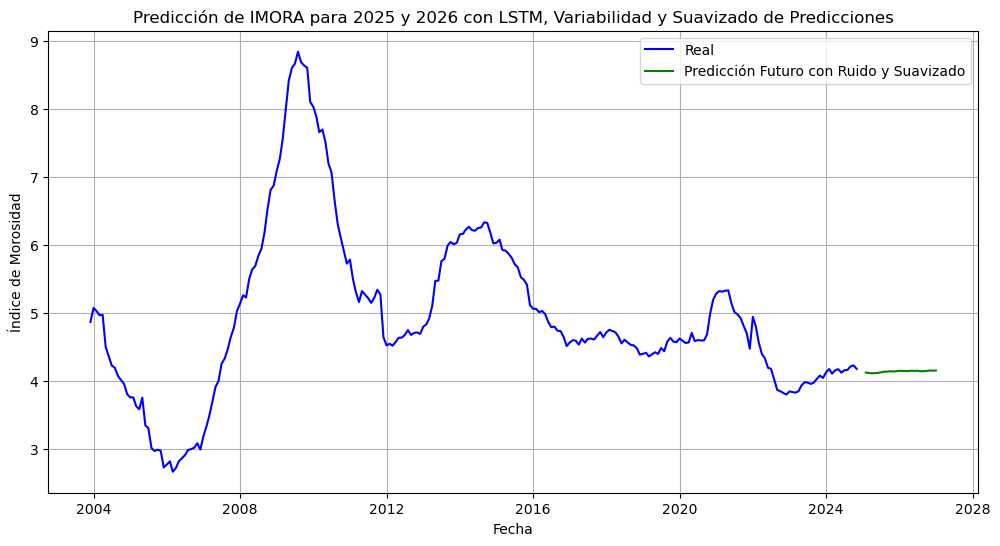


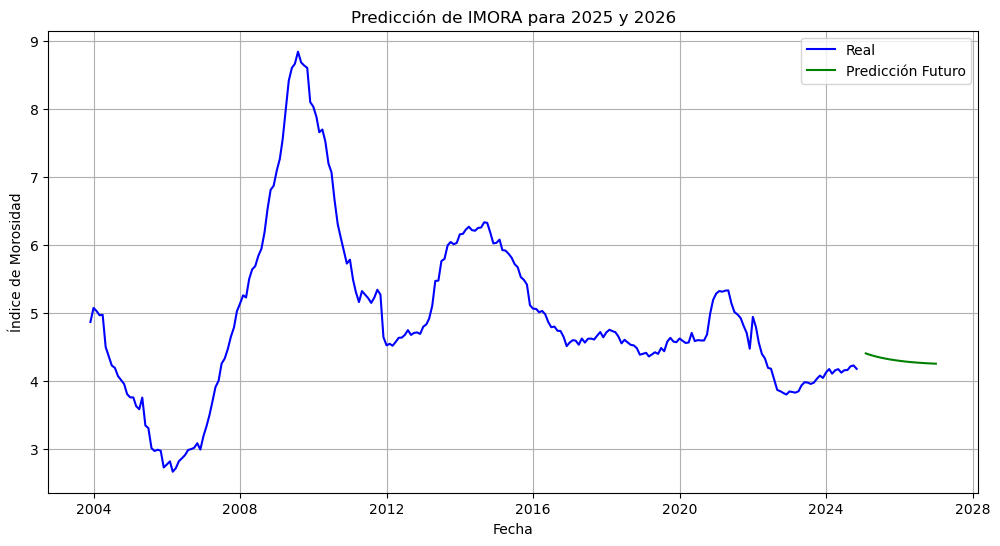


* + ¿Se cumplen con todos los supuestos de la metodología utilizada?

**En el caso del modelo ARIMA, se cumplen los supuestos de estacionariedad y ausencia de autocorrelación en los residuos. Sin embargo, el modelo LSTM no requiere estacionariedad de la serie, por lo que fue adecuado para capturar patrones no lineales a largo plazo. Ambos modelos mostraron un buen ajuste a los datos históricos, lo que sugiere que los supuestos metodológicos se cumplen en el contexto del análisis.**

* + ¿El pronóstico a dos años resulta alentador? **Si, muestra una leve tendencia al alza en el IMORA, lo que podría indicar una posible recuperación de la morosidad**

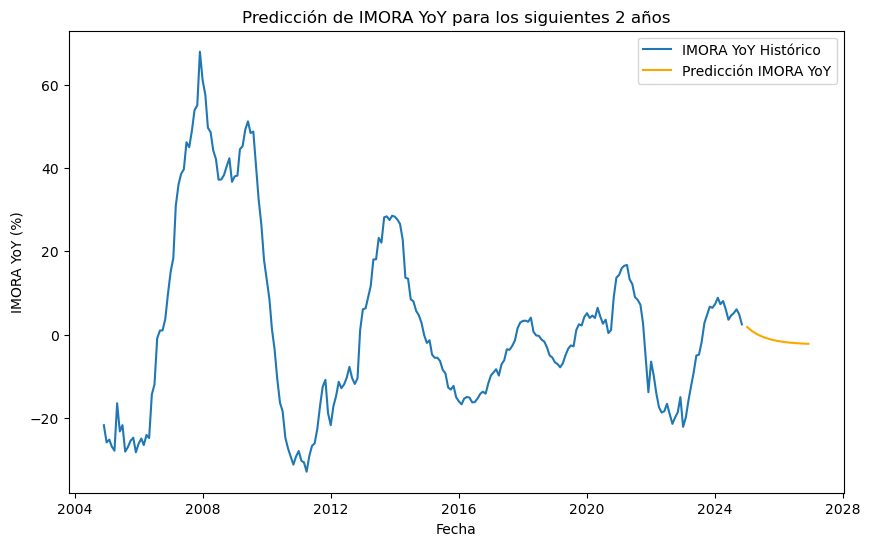




* + ¿Sería posible realizar un pronóstico a 10 años? ¿Sería correcto? **Si bien es posible realizar un pronóstico a 10 años, este podría no ser correcto o confiable debido a la alta incertidumbre a largo plazo. Las condiciones económicas pueden cambiar drásticamente, afectando las variables explicativas que influyen en el IMORA.**

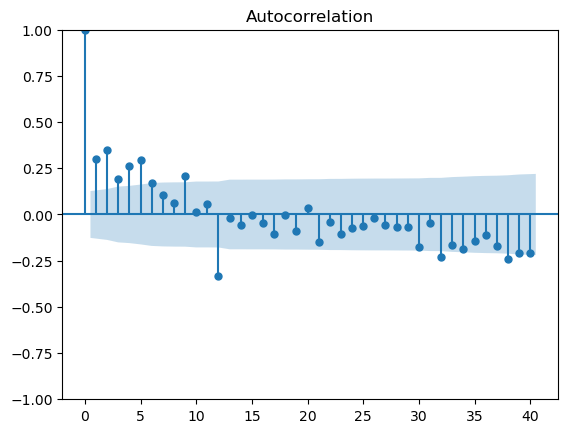
* Pronóstico del valor del IMORA\_YoY e IMORA para los siguientes dos años a partir de su propio comportamiento;

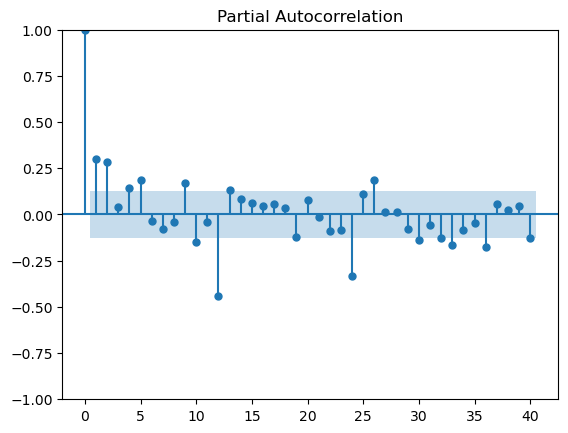
**Utilizando un modelo ARIMA basado únicamente en el comportamiento histórico de IMORA e IMORA\_YoY, se genero pronóstico para los próximos dos años.**



* + ¿Se cumplen con todos los supuestos de la metodología utilizada?

**R-Sí, el modelo ARIMA desarrollado cumple con los principales supuestos: Estacionariedad, Ruido blanco en los residuos y Parámetros significativos**





* + ¿El pronóstico a dos años resulta alentador? **R- no se esperan incrementos abruptos en el índice de morosidad en los próximos dos años.**
  + ¿Sería posible realizar un pronóstico a 10 años? ¿Sería correcto? **R- De igual manera es posible pero no seria correcto.**

* + ¿Existe relación entre el pronóstico hecho a partir de factores externos y el que se desarrolla con base en el comportamiento de la serie?

**R- Sí, existe una relación entre ambos enfoques, ambos modelos capturan patrones generales de estabilización en el IMORA, lo que sugiere que el índice está determinado, en parte, por su comportamiento pasado. Diferencias, el modelo con variables externas puede capturar mejores eventos específicos (como crisis económicas o variaciones en tasas de interés) que un modelo puramente histórico no considera. En este sentido, el enfoque con variables externas podría proporcionar pronósticos más ajustados en entornos dinámicos.**

* + Respecto a los valores observados durante 2024 ¿Qué tan preciso es el pronóstico?

**El pronóstico generado para 2024 muestra un error porcentual absoluto promedio (MAPE) del 2.05%, lo cual indica un nivel de precisión muy alto en la predicción. Esto significa que las diferencias entre los valores observados y los estimados son mínimas**

1. La serie objetivo deberá ser modelada utilizando la información hasta diciembre de 2023, ya que, se busca analizar qué tan precisa es la proyección respecto a los valores observados de 2024. [↑](#footnote-ref-1)