

**UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
ESCUELA DE POSGRADO**



**TRABAJO DE POSGRADO
PREDICCIÓN DE REMESAS E IMPACTO EN FINANZAS PÚBLICAS DE EL
SALVADOR USANDO MACHINE LEARNING**

**PARA OPTAR AL GRADO DE
MAESTRO(A) EN ADMINISTRACIÓN FINANCIERA**

**TRABAJO PRESENTADO POR
LICENCIADA ESTELY DOLORES MÉNDEZ MORENO
LICENCIADO RAFAEL ANÍBAL CALDERÓN MELARA**

**DOCENTE ASESOR
MAESTRO DANIEL EDUARDO HURTADO HERRERA**

**DICIEMBRE, 2025
SANTA ANA, EL SALVADOR, CENTROAMÉRICA**

**UNIVERSIDAD DE EL SALVADOR
AUTORIDADES**



**ING. JUAN ROSA QUINTANILLA QUINTANILLA
RECTOR**

**DRA. EVELYN BEATRIZ FARFÁN MATA
VICERRECTORA ACADÉMICA**

**M.Sc. ROGER ARMANDO ARIAS ALVARADO
VICERRECTOR ADMINISTRATIVO**

**LICDO. PEDRO ROSALÍO ESCOBAR CASTANEDA
SECRETARIO GENERAL**

**LICDA. ANA RUTH AVELAR VALLADARES
DEFENSORA DE LOS DERECHOS UNIVERSITARIOS**

**LICDO. CARLOS AMILCAR SERRANO RIVERA
FISCAL GENERAL**

**FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
AUTORIDADES**



**M.Ed. ROBERTO CARLOS SIGÜENZA CAMPOS
DECANO**

**DR. JOSÉ GUILLERMO GARCÍA ACOSTA
VICEDECANO**

**LICDO. JAIME ERNESTO SERMEÑO DE LA PEÑA
SECRETARIO**

**M.Ed. MIGUEL ANGEL CRUZ
DIRECTOR DE LA ESCUELA DE POSGRADO**

Agradecimientos

Agradecemos, en primer lugar, a Dios y a la vida por la posibilidad de culminar este proceso y este trabajo de forma exitosa. También expresamos nuestro agradecimiento a la institución y a sus autoridades por brindarnos el espacio adecuado para desarrollar no solo nuestros estudios de posgrado, sino también — en nuestro caso particular— los de grado que realizamos años atrás. Podemos decir con orgullo que somos hijos de la Minerva por segunda vez.

Agradecemos a nuestro asesor de trabajo de grado, a la directora de la maestría y a todos los profesores que nos acompañaron a lo largo de este proceso formativo, pues sin su apoyo y dedicación nada de esto habría sido posible.

De manera especial, agradecemos a nuestras familias, y en particular a nuestros padres, por todo el apoyo que nos brindaron. Sin ellos, no podríamos estar cerrando esta etapa tan significativa de nuestras vidas.

Finalmente, un agradecimiento muy especial para Santiago, quien, sin ser del todo consciente, ha sido la luz más brillante que nos acompañó durante este proceso.

*Una dedicación especial a la memoria de nuestro querido amigo Jorge con quien
hace casi 10 años hablábamos sobre nuestra área de especialización*

ÍNDICE

Introducción.....	ix
Capítulo I: Planteamiento del problema.....	11
1.1 Justificación	11
1.2 Hipótesis	12
1.3 Resumen Ejecutivo.....	13
1.4 Definición del problema	14
Capítulo II: Marco conceptual	17
2.1 Remesas familiares	17
2.2 Factores económicos y no económicos que influyen en los flujos de remesas	18
2.3 Impacto de las remesas	18
2.4 Ingresos Tributarios Netos	19
2.5 Gastos Corrientes.....	20
2.6 Índice de Precios al Consumidor (IPC) e Inflación.....	20
2.7 Índice de Volumen de Actividad Económica	21
2.8 Modelos de series de tiempo y aplicaciones en los flujos de remesas	22
2.9 Modelos ARDL.....	25
2.10 Modelos predictivos modernos.....	30
2.11 Inteligencia Artificial	31
2.12 Machine Learning	32
2.13 Deep Learning	34
2.14 Aprendizaje Automático y las redes de comunicación.....	36
2.15 Áreas de Aplicación del Aprendizaje Automático en Redes	37
2.16 Machine Learning y predicción económica	38
2.17 Árboles de decisión	39
2.18 Random Forest (bosques aleatorios).....	40
2.19 Evaluación de un modelo Random Forest	43
2.19.1 Uso de Out-of-Bag (OOB) Samples	43
Capítulo III: Metodología de la investigación.....	48

3.1 Enfoque de la investigación.....	48
3.2 Diseño de la investigación.....	48
3.3 Tipo de estudio	48
3.4 Población y muestra	49
3.5 Recolección de datos	51
3.6 Procesamiento de los datos	51
Capítulo IV: Análisis y resultados de la investigación	52
4.1 Análisis de resultados del modelo de predicción de los flujos de remesas familiares.....	52
4.2 Análisis de resultados de los modelos ARDL	67
Conclusiones.....	85
Recomendaciones	87
Referencias	88
Anexos	92

Índice de figuras

Figura 1: ejemplo gráfico de las predicciones de la serie CO2 de tiempo con un modelo ARIMA	25
Figura 2: ilustración de cómo se relacionan la Inteligencia artificial, Machine Learning y Deep Learning	33
Figura 3: Línea de tiempo del desarrollo de la inteligencia artificial	36
Figura 4: ejemplo de visualización de un árbol de decisión	40
Figura 5: Ilustración del funcionamiento del algoritmo de un modelo de tipo Random Forest.....	43
Figura 6: Variables seleccionadas de los países receptores para el modelo de flujos de remesas	54
Figura 7: Variables seleccionadas de Estados Unidos para el modelo de flujos de remesas.....	55
Figura 8: serie de la variable inmigrantes expulsados por USA para El Salvador sin imputar.....	56
Figura 9: serie de la variable de inmigrantes expulsados por USA para El Salvador después de imputación	56
Figura 10: comportamiento bajo RMSE de los modelos candidatos.....	57
Figura 11: salida del proceso de optimización de parámetros con Optuna	58
Figura 12: resultados del modelo para los Folds de entrenamiento.....	59
Figura 13: Resultados finales del modelo en conjunto de validación	59
Figura 14: Predicción del modelo en el conjunto Holdout	60
Figura 15: Visualización de los errores absolutos de las predicciones del modelo	61
Figura 16: Importancia de las variables predictoras en el modelo de flujos de remesas.....	64
Figura 17: Visualización de las predicciones del modelo para toda la base de datos	65
Figura 18: Visualización de las predicciones del modelo para Honduras	65
Figura 19: Predicciones del modelo para República Dominicana	66
Figura 20: Predicciones del modelo para El Salvador.....	66
Figura 21: Diagrama de dispersión de las predicciones del modelo de flujos de remesas.....	67
Figura 22: Relación Ingresos tributarios vs Ingresos corrientes totales	69
Ilustración 23: Relación entre IVA y los ingresos tributarios netos.....	69

Figura 24: Comparación de los ingresos corrientes y los gastos corrientes	70
Figura 25: Comportamiento histórico del déficit fiscal	71
Figura 26: Comportamiento de los crecimientos de los ingresos y gastos	71
Figura 27: Diagrama de dispersión de los ingresos vs los gastos	72
Figura 28: Análisis de la correlación para los ingresos y gastos anuales	72
Ilustración 29: Resultados modelo ARDL para los ingresos tributarios.....	74
Ilustración 30: resultados del modelo ARDL que predice el gasto corriente	75
Figura 31: Modelo que relaciona remesas con Iva	76
Ilustración 32: resumen de prueba de causalidad de Granger de las remesas sobre el Iva	77
Ilustración 33: Gráfico de impulso-respuesta de las remesas sobre los ingresos tributarios netos	78
Ilustración 34: Gráfico de impulso-respuesta de las remesas sobre el gasto corriente.....	79
Figura 35: gráfico de las series de tiempo del ingreso neto sobre el IVAE anualizado real vs predicción	82
Figura 36: gráfico de las series de tiempo del gasto corriente sobre el IVAE anualizado real vs predicción	82
Figura 37: proyecciones del indicador de ingresos neto con un shock de +10% de remesas.....	84
Figura 38: proyecciones del indicador de gastos corrientes con un shock de +10% de remesas.....	84

Introducción

El Salvador ha tenido una expansión demográfica muy grande en los últimos 50 años, como se presenta en el informe del Fondo de Población de las Naciones Unidas (UNFPA, 2023): “La cantidad de habitantes en el país ha pasado de 3.89 en 1973 a 6.36 millones en 2023. De este modo, hay 2.47 millones de habitantes más de los que había en 1973” (p. 8).

En este mismo informe del UNFPA, se establece que el número de personas que han migrado superó los 1.6 millones para 2020, lo que es equivalente al 25.4% de la población que vive en el territorio de El Salvador, siendo Estados Unidos el país de destino principal para los migrantes salvadoreños. Esto de forma paralela al deterioro económico cada vez más agudo de las últimas décadas y a los problemas asociados a la inseguridad.

Es así, como las remesas familiares han llegado a representar cerca de una cuarta parte del PIB de El Salvador, convirtiéndose en una importante fuente de ingresos para las familias salvadoreñas, como lo menciona Manuel Zuleta en su estudio publicado por FUSADES: “La intensidad de recepción de remesas es más alta en departamentos que han experimentado altos niveles de migración, como La Unión (48.8% de los hogares reciben remesas), Chalatenango (41.3%) y Cabañas (40.1%)” (p. 1, 2021).

Dada la evidente importancia de las remesas y sus flujos, existen múltiples estudios sobre su impacto en diferentes áreas de la economía de El Salvador, su

relación con diferentes factores y en algunos casos predicciones de los flujos de remesas anuales.

Las técnicas más utilizadas en las predicciones (modelos predictivos) de los flujos de remesas, son modelos econométricos basados en series temporales (Forecasting), estos modelos estadísticos comúnmente aceptados tienen buenos resultados haciendo predicciones en el corto plazo, además, tienen ventajas en su desarrollo e interpretación relativamente sencillos, sin embargo, también tienen algunas desventajas, una de las cuales es una fuerte dependencia de una propiedad de estacionariedad estadística, además, dependen casi completamente de variables internas y un mal funcionamiento para relaciones no lineales entre sus variables.

Actualmente existe la posibilidad de explorar técnicas que son capaces de gestionar de mejor manera las variables tanto internas como externas, las relaciones complejas no necesariamente lineales y los datos no necesariamente estructurados y heterogéneos. Con todo esto surge la posibilidad de desarrollar un modelo que aplique múltiples técnicas de aprendizaje supervisado o no supervisado, esto permitiría encontrar relaciones entre variables económicas, demográficas e incluso sociales, analizando su impacto dentro de las finanzas públicas, generando propuestas de posibles decisiones que se podrían tomar desde la administración pública y así generar una mayor eficiencia.

Capítulo I: Planteamiento del problema

1.1 Justificación

Las remesas familiares han llegado a representar casi una cuarta parte del Producto Interno Bruto (PIB) del país, por lo que predecir sus flujos de forma eficiente, estudiar su impacto en las finanzas públicas y su relación con diferentes variables económicas y no económicas se vuelve muy importante.

Actualmente, el estudio de los flujos de las remesas familiares se realiza por medio de modelos econométricos tradicionales, basados en series temporales, que tienen un desarrollo e interpretación relativamente sencillos, pero presentan limitaciones frente a relaciones complejas no lineales. Por otra parte, con el auge de metodologías o técnicas de predicción y análisis relativamente nuevas, como los modelos de aprendizaje automático (Machine Learning), que permiten manejar datos heterogéneos y relaciones complejas con mayor eficacia, surge la oportunidad de generar nuevos conocimientos que permitan entender y explicar de mejor manera los flujos de remesas y su impacto.

Esta investigación es pertinente porque busca no solo mejorar la precisión en la predicción de los flujos de remesas, sino también, identificar los factores económicos y no económicos que los afectan. Esto permitirá comprender mejor cómo las fluctuaciones de las remesas impactan en áreas clave del presupuesto público y proponer estrategias para mitigar dichos efectos.

Además, los resultados de esta investigación tienen el potencial de influir directamente en la toma de decisiones estratégicas en la administración de las finanzas públicas, generando una mejor sostenibilidad y minimizando los posibles impactos negativos de las fluctuaciones de las remesas familiares.

1.2 Hipótesis

Hipótesis General:

Los flujos de remesas familiares tienen un impacto significativo en la estabilidad y sostenibilidad de las finanzas públicas de El Salvador, influyendo en áreas clave del presupuesto público.

Hipótesis Específicas:

- Los modelos de Machine Learning alcanzan niveles de precisión aceptables en la predicción de los flujos de remesas familiares.
- Las variables económicas internacionales (como el crecimiento económico de EE.UU. y las tasas de empleo) tienen una influencia significativa en los flujos de remesas familiares.
- Las variables no económicas (como la inseguridad y la migración) son determinantes en el comportamiento de los flujos de remesas familiares.
- Las fluctuaciones en los flujos de remesas familiares afectan principalmente áreas del presupuesto público como salud, educación y seguridad social.

- La implementación de estrategias financieras y administrativas basadas en predicciones de Machine Learning puede mitigar el impacto negativo de las fluctuaciones de remesas en las finanzas públicas.

1.3 Resumen Ejecutivo

Las remesas familiares son una de las principales fuentes de ingresos para muchas familias salvadoreñas y, por lo tanto, tienen un impacto directo o indirecto en la economía de El Salvador y en sus finanzas públicas, por lo que existen muchos estudios analizando su impacto y comportamiento.

En este trabajo se busca analizar qué factores influyen en los flujos de remesas familiares, considerando factores económicos y no económicos, desarrollando un modelo predictivo eficiente y evaluando su efecto en la sostenibilidad de las finanzas públicas, es decir, se abordan dos grandes objetivos, construir un modelo predictivo que alcance un nivel de precisión bueno y responder a cómo y en qué se ven afectadas las finanzas públicas por los flujos de remesas.

Metodología:

1. Identificación de variables: Se seleccionó un conjunto amplio de indicadores con cobertura temporal suficiente.
2. Modelado predictivo: Construcción de una base de entrenamiento para generar un modelo con alta predictibilidad (se mencionará al final qué modelo de Machine Learning fue seleccionado).

3. Impacto en las finanzas públicas: Análisis cuantitativo basado en teoría económica y métodos estadísticos para medir la relación entre remesas y finanzas públicas.

1.4 Definición del Problema

Preguntas de Investigación

Para la problemática: impacto en las finanzas públicas de El Salvador de los flujos de remesas familiares y su relación con factores económicos y no económicos, surgen las siguientes preguntas de investigación:

Pregunta General.

- ¿Cuál es el impacto de los flujos de remesas familiares en las finanzas públicas de El Salvador?

Preguntas Específicas.

- ¿Cuál modelo de Machine Learning es el que logra mejores resultados prediciendo los flujos de remesas familiares?
- ¿Cuáles factores económicos nacionales e internacionales influyen en los flujos de remesas familiares?
- ¿Cuáles factores no económicos son los que influyen en los flujos de remesas familiares?

- ¿Cuáles estrategias se pueden adoptar para disminuir el impacto de las fluctuaciones de los flujos de remesas familiares en las finanzas públicas de El Salvador?

Objetivos de la Investigación

1. Objetivo General

1.1. Determinar el impacto de los flujos de remesas familiares en la estabilidad y sostenibilidad de las finanzas públicas de El Salvador.

2. Objetivos Específicos

2.1. Desarrollar un modelo que permita predecir los flujos de remesas, logrando establecer las métricas adecuadas para evaluar su desempeño.

2.2. Establecer las variables económicas internas y externas que son significativas para predecir los flujos de remesas familiares.

2.3. Determinar cuáles variables no económicas son significativas en los modelos de predicción de flujos de remesas para mejorar su precisión y entender mejor su comportamiento.

2.4. Establecer las áreas del presupuesto público que se ven afectadas por las fluctuaciones de los flujos de remesas familiares.

2.5. Generar propuestas de posibles estrategias financieras o administrativas que podrían mejorar la salud de las finanzas públicas.

Capítulo II: Marco Conceptual

2.1 Remesas familiares

Las remesas familiares se definen como transferencias de dinero que los migrantes envían a sus países de origen, generalmente a sus familiares, y que representan una fuente importante de ingresos en muchas economías en desarrollo (Banco Mundial, 2022). Según Ratha (2013), las remesas constituyen flujos financieros que, a diferencia de otras entradas de capital, no requieren contraprestación ni generan deuda externa.

En el caso de El Salvador, las remesas han adquirido una relevancia estratégica, alcanzando en algunos años cerca de una cuarta parte del Producto Interno Bruto (PIB) (Banco Central de Reserva de El Salvador [BCR], 2023). Este flujo constante de recursos ha tenido impactos profundos tanto en el nivel de consumo de los hogares como en la estabilidad macroeconómica.

La realidad para el caso de otros países de características similares, tanto económicas, como geográficas o incluso demográficas, es bastante equivalente, por ejemplo, Honduras mostró un crecimiento interanual de 6.2% en 2024, llegando a flujos de USD9,743 millones, y para 2025 mostró un aumento del 17.2% respecto al mismo periodo de 2024 (Banco Central de Honduras, 2025).

Para el caso de República Dominicana, según datos del banco central, durante el primer trimestre de 2024 las remesas recibidas alcanzaron la cifra de USD2,635.6 millones, que representa un aumento de un 6.2% en comparación con el mismo periodo del año anterior, (Banco central de República Dominicana, 2024).

2.2 Factores económicos y no económicos que influyen en los flujos de remesas

Diversos estudios identifican que los flujos de remesas son afectados por factores económicos, como las tasas de desempleo en el país receptor y emisor, el tipo de cambio, el crecimiento económico y las condiciones del mercado laboral (Acosta, Calderón, Fajnzylber & López, 2008). A nivel internacional, también se ha encontrado que las crisis económicas en los países de destino afectan de forma directa las remesas enviadas (Chami, Fullenkamp & Jahjah, 2003).

Por otro lado, factores no económicos como la reunificación familiar, los conflictos sociales, la violencia o los desastres naturales también juegan un rol importante en las decisiones de envío de remesas (Durand, Massey & Parrado, 1996).

2.3 Impacto de las remesas

Las remesas pueden tener efectos tanto positivos como negativos sobre las finanzas públicas. Según Chami et al. (2003), en algunos contextos, las remesas contribuyen a estabilizar el consumo interno, lo cual puede aumentar la recaudación tributaria por medio del IVA y otros impuestos indirectos. No obstante, también se ha advertido que un flujo alto y sostenido de remesas puede generar fenómenos como la "enfermedad holandesa", afectando negativamente la competitividad de las exportaciones (Amuedo-Dorantes & Pozo, 2004).

En El Salvador, diversos análisis locales sostienen que las remesas han permitido suavizar las condiciones fiscales indirectamente, al sostener el consumo privado y reducir las presiones sobre el gasto social (FUSADES, 2021).

Antes de avanzar con la conceptualización de términos más técnicos, se presentan algunos conceptos económicos que se usarán en el desarrollo de este trabajo.

2.4 Ingresos Tributarios Netos

Los **ingresos tributarios netos** son aquellos recursos que el Estado percibe por concepto de impuestos, deducidos de las devoluciones y compensaciones correspondientes. Constituyen la principal fuente de financiamiento del gasto público y son esenciales para mantener la sostenibilidad fiscal (OCDE, 2023).

Este indicador refleja la **capacidad recaudatoria efectiva del gobierno** y su eficiencia en la administración tributaria. En economías dependientes de impuestos indirectos como el IVA, los ingresos tributarios netos se ven influenciados por la actividad económica, el consumo interno y las políticas de control fiscal.

El análisis de su comportamiento frente a variables como las remesas o el gasto público permite evaluar los efectos de la política fiscal sobre el crecimiento y la estabilidad macroeconómica.

2.5 Gastos Corrientes

Los **gastos corrientes** comprenden las erogaciones destinadas al funcionamiento regular del Estado, como salarios, bienes y servicios, intereses de deuda y transferencias corrientes. No generan activos de largo plazo, pero son esenciales para la provisión de bienes públicos y el mantenimiento de la estructura administrativa (Fondo Monetario Internacional [FMI], 2022).

Un nivel elevado de gasto corriente puede reflejar **rigideces presupuestarias**, especialmente cuando se financia con ingresos transitorios o con deuda. Su relación con los ingresos tributarios es un indicador clave de la **sostenibilidad fiscal**: cuando el gasto corriente supera sistemáticamente los ingresos, se genera presión sobre el déficit fiscal y la deuda pública.

2.6 Índice de Precios al Consumidor (IPC) e Inflación

El **Índice de Precios al Consumidor (IPC)** mide la variación promedio de los precios de un conjunto representativo de bienes y servicios consumidos por los hogares. Es el principal indicador para estimar la **inflación**, que representa el aumento sostenido y generalizado del nivel de precios en la economía (Banco Central de Reserva de El Salvador [BCR], 2024).

La inflación tiene múltiples efectos macroeconómicos: reduce el poder adquisitivo, distorsiona la estructura de precios relativos y puede afectar la recaudación tributaria, especialmente en impuestos nominales como el IVA. Además, influye en las decisiones de política fiscal y monetaria, ya que tasas

inflacionarias elevadas pueden erosionar el valor real de las remesas y alterar el consumo doméstico (FMI, 2022).

El control de la inflación a través del monitoreo del IPC es esencial para preservar la estabilidad económica y garantizar la eficacia de la política fiscal.

2.7 Índice de Volumen de Actividad Económica

El Índice de Volumen de Actividad Económica (IVAE) es un indicador que mide la evolución mensual de la producción de bienes y servicios en una economía, expresado a precios constantes. Su objetivo principal es reflejar el comportamiento de la actividad económica en el corto plazo, eliminando el efecto de los precios mediante la deflatación de las series productivas. En el caso de El Salvador, el IVAE es elaborado por el Banco Central de Reserva (BCR) y permite analizar las fluctuaciones económicas con mayor frecuencia que el Producto Interno Bruto (PIB) anual o trimestral.

Según el Banco Central de Reserva de El Salvador (2023), *“el IVAE es un indicador sintético que aproxima la evolución del nivel de actividad económica del país, calculado a precios constantes y ajustado estacionalmente, con base en información mensual de distintos sectores productivos”* (p. 2).

En términos prácticos, el IVAE sirve como una herramienta de análisis coyuntural, ya que capta los cambios en la producción real y permite identificar tendencias o ciclos económicos. Su seguimiento es fundamental para evaluar los

efectos de variables como el consumo, las remesas familiares, la inflación o los ingresos fiscales sobre el desempeño económico del país.

2.8 Modelos de series de tiempo y aplicaciones en los flujos de remesas

La predicción de flujos de remesas ha sido abordada tradicionalmente mediante modelos estadísticos de series temporales como el modelo ARIMA o el modelo de corrección de errores (ECM). Según Gujarati y Porter (2010), estos modelos ofrecen ventajas de interpretación sencilla y requerimientos relativamente bajos de información, pero presentan limitaciones cuando las relaciones entre variables son no lineales o cuando existen cambios estructurales.

Para entender un poco más sobre este tipo de modelos y sus aplicaciones, a continuación, se definen algunos conceptos que se consideran necesarios, y que se han utilizado en este trabajo para determinar el impacto de las remesas en las finanzas públicas.

Serie de tiempo: una serie de tiempo en términos simples es un conjunto de datos que se pueden indexar a momentos específicos en un periodo de tiempo, referentes a una variable en específico, pero de forma más técnica, se puede definir una serie de tiempo como la realización de un proceso estocástico (Ezequiel Uriel Jimenez, 1985).

Proceso Estocástico: para Ezequiel Uriel Jiménez (1985) un proceso estocástico no es más que la realización de una familia de variables aleatorias que

corresponden a momentos sucesivos del tiempo, designado por $Y(t, u)$, donde t es el tiempo y u es la variable aleatoria.

Partiendo de la definición técnica de una serie de tiempo, es posible generar procesos de análisis, modelación, descomposición, entre otros, y su vez aplicarlo a fenómenos reales, como pueden ser los fenómenos económicos o macroeconómicos.

Procesos o Modelos Lineales: estos procesos se caracterizan por que se pueden representar como una combinación lineal de variables aleatorias (Ezequiel Uriel Jiménez, 1985). De estos modelos se presta especial atención los procesos lineales autorregresivos (AR), siendo variantes de estos últimos los que se han usado en este estudio.

Modelos autorregresivos: según Peña, D. (2005), los modelos autorregresivos generalizan la idea de regresión para representar la dependencia lineal entre dos variables aleatorias, por ejemplo, para una serie z_t , se dirá que sigue un proceso autorregresivo de primer orden o AR(1), si ha sido generada por:

$$z_t = c + \phi z_{t-1} + a_t$$

Donde c y $-1 < \phi < 1$ son constantes y a_t es un proceso de ruido blanco. Pudiendo generalizar esta definición para un orden arbitrario p . Se dirá que una serie temporal z_t estacionaria sigue un proceso autorregresivo de orden p si:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t$$

Donde $\tilde{z}_t = z_t - \mu$, siendo μ la media del proceso estacionario y a_t un proceso de ruido blanco.

Todos los modelos que se han presentado radican en procesos estocásticos, y tienen una amplia gama de propiedades y herramientas para su análisis, transformación y aplicación, pero, para todos ellos se asume una propiedad llamada **estacionariedad**, en este documento no se va a profundizar demasiado en estos conceptos, pero sin problemas el lector puede profundizar en ellos para comprender mejor estos modelos.

Sin embargo, en el caso particular de los problemas económicos no siempre se puede asumir que los fenómenos son estacionarios, por tanto, en general se aplican modelos no estacionarios, con un sutil detalle, y es que en general, estos modelos no estacionarios se pueden transformar en estacionarios para así usar la gama de herramientas de estos últimos en su análisis.

Es así como se llega a los modelos no estacionarios que en general son capaces de modelar gran número de fenómenos económicos, a estos modelos se les conoce como modelos ARIMA, que son un tipo de modelo no estacionario y se les llama modelos integrados, porque se pueden transformar en procesos estacionarios mediante la toma de diferencias de un determinado orden, también, estos modelos integrados se definen como los modelos que se obtienen a partir de la suma o de la integración de un modelo estacionario (Ezequiel Uriel Jiménez, 1985).

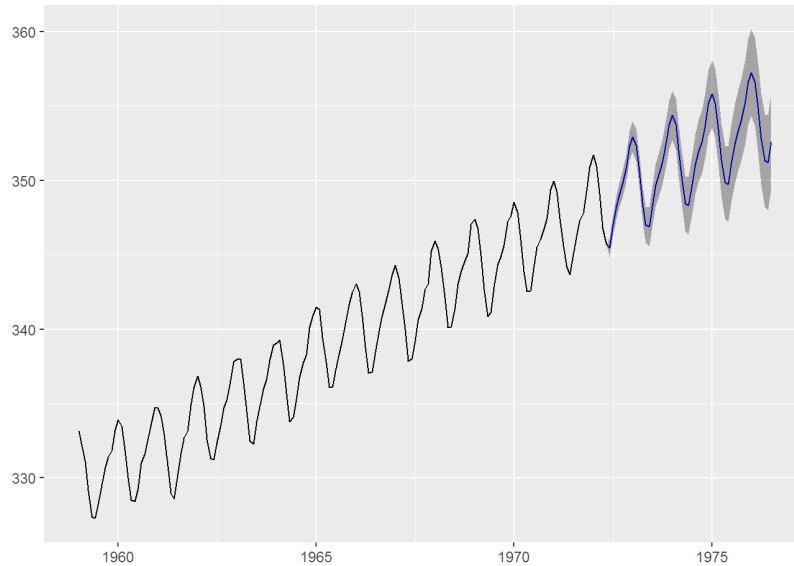


Figura 1: ejemplo gráfico de las predicciones de la serie CO2 de tiempo con un modelo ARIMA

2.9 Modelos ARDL

Para efectos de este trabajo, específicamente para el análisis del impacto de las remesas en las finanzas públicas de El Salvador se optó por el modelo Autorregresivo de rezagos distribuidos (ARDL en nomenclatura anglosajona).

Pesaran, Shin y Smith (2001) desarrollaron el enfoque de **modelos Autorregresivos de Rezagos Distribuidos (ARDL)** como una metodología flexible para analizar relaciones de largo y corto plazo entre variables, especialmente en contextos donde los datos presentan diferentes órdenes de integración, algunos estacionarios en nivel $I(0)$ y otros en primera diferencia $I(1)$. Esta característica

representa una de las principales ventajas del modelo ARDL frente a otros enfoques de cointegración, como el de Johansen, que requiere que todas las series sean $I(1)$.

Los autores formalizan el **modelo ARDL** (p, q_1, q_2, \dots, q_k), en el cual la variable dependiente y las independientes pueden incluir diferentes rezagos, permitiendo captar dinámicas complejas de ajuste temporal. Mediante la **transformación del modelo en una forma de corrección de errores (ECM)**, el enfoque ARDL permite identificar de manera simultánea la **relación de equilibrio de largo plazo** y la **dinámica de corto plazo**.

Dando la siguiente definición formal, El modelo ARDL (p, q_1, q_2, \dots, q_k), se define para una variable independiente y_t y k variables explicativas x_{it} como:

$$y_t = \alpha_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^k \sum_{l=0}^{q_j} \beta_{jl} x_{j,t-l} + \varepsilon_t$$

donde:

- y_t es la variable dependiente,
- $x_{j,t-l}$ son las variables explicativas con sus rezagos,
- ϕ_i son los coeficientes autorregresivos,
- β_{jl} son los coeficientes de los rezagos distribuidos de las variables explicativas, y
- ε_t es el término de error blanco (ruido blanco).

Reformulación en la forma de corrección de errores (ECM)

El modelo ARDL puede reescribirse para separar los efectos de **corto y largo plazo** mediante la diferencia primera:

$$\Delta y_t = \alpha_0 + \lambda \left(y_{t-1} - \sum_{j=1}^k \theta_j x_{j,t-1} \right) + \sum_{i=1}^{p-1} \psi_i \Delta y_{t-i} + \sum_{j=1}^k \sum_{l=0}^{q_j-1} \delta_{jl} \Delta x_{j,t-l} + \varepsilon_t$$

donde:

- Δ representa el operador de diferencia primera (por ejemplo, $\Delta y_t = y_t - y_{t-1}$),
- λ es el **coeficiente de ajuste** o velocidad de convergencia al equilibrio de largo plazo,
- θ_j son los **coeficientes de largo plazo**, calculados como:

$$\theta_j = \frac{\sum_{l=0}^{q_j} \beta_{jl}}{1 - \sum_{i=1}^p \phi_i}$$

- ψ_i y δ_{jl} capturan los **efectos de corto plazo** asociados a los cambios recientes en y_t y x_t .

Todo esto permite la siguiente interpretación económica:

- **Coeficientes de corto plazo (δ_{jl}):** miden el impacto inmediato o transitorio que tiene un cambio en la variable explicativa x_j sobre y_t , antes de que el sistema alcance el equilibrio.
- **Coeficientes de largo plazo (θ_j):** representan el efecto de equilibrio o permanente de una variación sostenida en x_j sobre y_t .

- **Coeficiente de ajuste (λ):** refleja la velocidad con que las desviaciones del equilibrio se corrigen; su valor esperado es negativo y menor que 1 en magnitud.
- **Error correction term (ECT):** el término $\left(y_{t-1} - \sum_{j=1}^k \theta_j x_{j,t-1} \right)$ representa la brecha de desequilibrio entre el valor observado y el valor de equilibrio de largo plazo.

Por otra parte, una de las principales contribuciones metodológicas del artículo de Pesaran, Shin y Smith (2001), es el **test de límites (bounds testing approach)**, que evalúa la existencia de cointegración sin necesidad de conocer el orden exacto de integración de las variables. Este procedimiento se basa en contrastar el estadístico F de significancia conjunta de los coeficientes de largo plazo con valores críticos que delimitan dos zonas: si el estadístico excede el límite superior, existe cointegración; si cae por debajo del límite inferior, no la hay.

Entre las propiedades más destacadas de los modelos ARDL se encuentran:

- **Consistencia y eficiencia** de los estimadores bajo muestras pequeñas, gracias al uso de la técnica de Mínimos Cuadrados Ordinarios (MCO) en cada etapa.
- **Capacidad para modelar dinámicas asimétricas o no lineales**, mediante extensiones como el modelo NARDL (Nonlinear ARDL).

- **Identificación clara de efectos de corto y largo plazo**, lo que facilita la interpretación económica y la formulación de políticas basadas en elasticidades o velocidades de ajuste.

En resumen, Pesaran, Shin y Smith (2001) consolidan el ARDL como un instrumento robusto y versátil para el análisis empírico de series temporales mixtas, destacando su aplicabilidad en estudios macroeconómicos, financieros y de política pública, donde las relaciones entre variables suelen ser dinámicas y de largo horizonte.

Para cerrar lo referente a este tipo modelos se presentan algunos conceptos de cómo se evalúan, en particular se definen dos de los criterios más utilizados para determinar el mejor modelo para un conjunto de datos temporales, cabe mencionar que muchas de las definiciones que se presentan, tienen como base múltiples conceptos muy técnicos, entre ellos la función de máxima verosimilitud, innovaciones, varianzas, entre otros. Todos esos conceptos se utilizan en la construcción teórica de las propiedades de los modelos de series temporales.

Criterio AIC: suponiendo que los parámetros de un modelo se calculan con los datos y que se calcula la esperanza de su verosimilitud respecto a futuras observaciones, se tiene que:

$$AIC = E(-2L(\beta)) = 2k + T \ln \hat{\sigma}_{MV}^2$$

Donde T es el tamaño muestral utilizado en el modelo, $\hat{\sigma}_{MV}^2$ es el estimador MV de la varianza de las innovaciones y k el número de parámetros estimados para calcular las predicciones a un paso.

Criterio BIC: este es un enfoque propuesto por Schwarz en 1978, y es conocido como enfoque de información bayesiana, parte de la idea de la probabilidad de un modelo dada una serie, y se define como:

$$BIC = T \ln \hat{\sigma}_{MV}^2 + k \ln T$$

Siendo T el tamaño muestral, $\hat{\sigma}_{MV}^2$ el estimador MV de la varianza y k el número de parámetros. Ambas definiciones pueden ser revisadas al detalle en Peña, D. (2005).

Para este trabajo (como se presenta en apartados posteriores) ambos criterios han sido utilizados para analizar la eficacia de los modelos ARDL que se han desarrollado, y aunque las definiciones son muy técnicas, se sugiere al lector quedarse con la idea que lo que se busca con ambos criterios es minimizarlos, es decir, *entre más pequeños sean los valores tanto del AIC y BIC (no en magnitud) indicará que un modelo tiene un mejor comportamiento*. En la evaluación de los modelos el otro criterio principal es el p-valor, el cual indica la importancia predictora de las variables de entrada utilizadas (así como sus rezagos).

2.10 Modelos predictivos modernos

Antes de pasar a los modelos predictivos como tal, es importante definir algunos conceptos que son ampliamente utilizados en el contexto de modelación y

la ciencia de datos, pues son claves para entender de mejor manera gran parte de los procesos desarrollados en este trabajo.

Como punto de partida hay que diferenciar entre tres conceptos claves, los cuales, a la fecha de realización de este trabajo son muy utilizados y consultados, estos son: Inteligencia Artificial (IA por sus siglas en inglés), Machine Learning y Deep Learning.

2.11 Inteligencia Artificial

La inteligencia artificial (IA) se entiende como la posibilidad de crear sistemas computacionales capaces de imitar el comportamiento humano y resolver tareas de manera “inteligente”. Aunque el término parece reciente, sus orígenes se remontan al taller de Dartmouth en 1956, donde John McCarthy lo introdujo con el propósito de describir a las denominadas “máquinas pensantes”. En un inicio, la meta era construir sistemas que reprodujeran los procesos mentales humanos y su interacción con el entorno, con la expectativa de que pudieran desempeñar cualquier actividad que realiza una persona e incluso superarla. Esta concepción dio origen a la idea de la Inteligencia Artificial General (IAG). Sin embargo, a pesar del entusiasmo inicial, tales aspiraciones no se cumplieron y ello provocó un periodo de desencanto en el campo, conocido posteriormente como el “invierno de la IA”.

La noción de máquinas inteligentes se remonta a épocas tan antiguas como el 1300 a. C. y continuó desarrollándose hasta el siglo XIX. No obstante, el punto de partida reconocido del campo formal de la inteligencia artificial se sitúa en la Conferencia de Dartmouth de 1956. A partir de ese momento, la disciplina ha

experimentado avances y retrocesos, además de diversificarse en múltiples subcampos. Asimismo, se han realizado esfuerzos por aplicar la IA en áreas como la medicina, las finanzas, la aviación y la industria, alcanzando resultados variados.

A finales de la década de 1990 e inicios de los 2000, los investigadores reconocieron una limitación en su enfoque hacia la inteligencia artificial, para crear una máquina verdaderamente inteligente era necesario comprender primero qué es la inteligencia, algo que aún hoy carece de una definición universal. Ante este desafío, se adoptó una estrategia distinta, en lugar de intentar construir directamente una inteligencia, se buscó desarrollar sistemas capaces de generar la suya propia. Este planteamiento dio origen a un nuevo subcampo de la IA, conocido como aprendizaje automático o *machine learning* (Russell & Norvig, 2010).

2.12 Machine Learning

El aprendizaje automático, o machine learning, constituye una rama de la inteligencia artificial cuyo objetivo es dotar a las máquinas de la capacidad de aprender sin necesidad de una programación explícita. La premisa es que, una vez contruidos, estos sistemas o programas informáticos pueden evolucionar y adaptarse conforme se exponen a nuevos datos.

El principio fundamental de este enfoque es la capacidad del modelo para generalizar a partir de la experiencia. Así, al recibir un conjunto de casos de entrenamiento, el sistema debe elaborar un modelo generalizado que le permita tomar decisiones frente a casos inéditos con un nivel de precisión adecuado (Goodfellow, Bengio, & Courville, 2016).

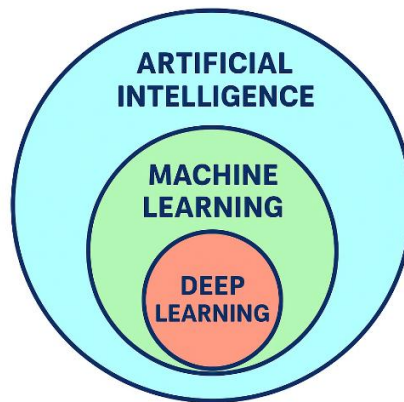


Figura 2: ilustración de cómo se relacionan la Inteligencia artificial, Machine Learning y Deep Learning

Dentro del aprendizaje automático se distinguen tres enfoques principales. En primer lugar, el aprendizaje supervisado consiste en proporcionar al sistema un conjunto de datos etiquetados para que construya un modelo generalizable que pueda aplicarse a nuevos casos. En contraste, el aprendizaje no supervisado trabaja con datos sin etiquetar y busca identificar patrones ocultos en ellos. Finalmente, el aprendizaje por refuerzo se basa en que el sistema ejecute acciones y reciba recompensas en función de sus resultados, lo que le permite aprender qué decisiones generan mejores beneficios en contextos específicos.

La aplicación de estas metodologías ha impulsado significativamente el desarrollo del aprendizaje automático, logrando avances notables en áreas como la visión por computadora y el análisis de texto (Goodfellow, Bengio, & Courville, 2016).

Hacia el año 2010 ocurrieron varios acontecimientos que impulsaron de manera decisiva el aprendizaje automático. El progreso tecnológico facilitó una mayor capacidad de cómputo, lo cual permitió entrenar y evaluar modelos más complejos. Asimismo, el abaratamiento del almacenamiento y del procesamiento de datos generó la disponibilidad de volúmenes crecientes de información para alimentar estos sistemas. Paralelamente, los avances en la comprensión del cerebro humano sirvieron de inspiración para diseñar nuevos algoritmos de aprendizaje.

En conjunto, estos factores dieron lugar al surgimiento de una nueva área dentro del aprendizaje automático conocida como *deep learning* o aprendizaje profundo (Goodfellow, Bengio, & Courville, 2016).

2.13 Deep Learning

El aprendizaje profundo, o *deep learning*, constituye una rama del aprendizaje automático inspirada en el funcionamiento del cerebro humano como mecanismo para adquirir conocimiento. También se le conoce como *deep structured learning* o aprendizaje jerárquico. De acuerdo con Bengio, Courville y Vincent (2013), puede definirse como un subcampo del aprendizaje automático basado en algoritmos que permiten representar múltiples niveles jerárquicos de abstracción para modelar relaciones complejas entre datos. En este esquema, las características de mayor nivel se construyen a partir de las de menor nivel, formando lo que se denomina una arquitectura profunda.

Este enfoque retoma los principios de las redes neuronales artificiales y los amplía mediante la incorporación de un mayor número de capas, lo que habilita a los modelos para procesar grandes volúmenes de datos. A través de este mecanismo, denominado *hierarchical feature learning*, el sistema puede extraer progresivamente características desde los datos en bruto y construir representaciones de mayor nivel de forma automática, con escasa intervención humana (Goodfellow, Bengio, & Courville, 2016).

Una de las particularidades más relevantes del aprendizaje profundo es su capacidad de escalar de manera eficiente: a mayor cantidad de datos disponibles, mejor es su rendimiento. Esto lo diferencia de muchos algoritmos tradicionales de *machine learning*, los cuales tienden a alcanzar un límite en su desempeño. En contraste, los modelos profundos no muestran esta restricción teórica, y en ciertos casos han logrado superar el rendimiento humano, como se ha evidenciado en sistemas modernos de procesamiento de imágenes basados en aprendizaje profundo (Goodfellow et al., 2016).

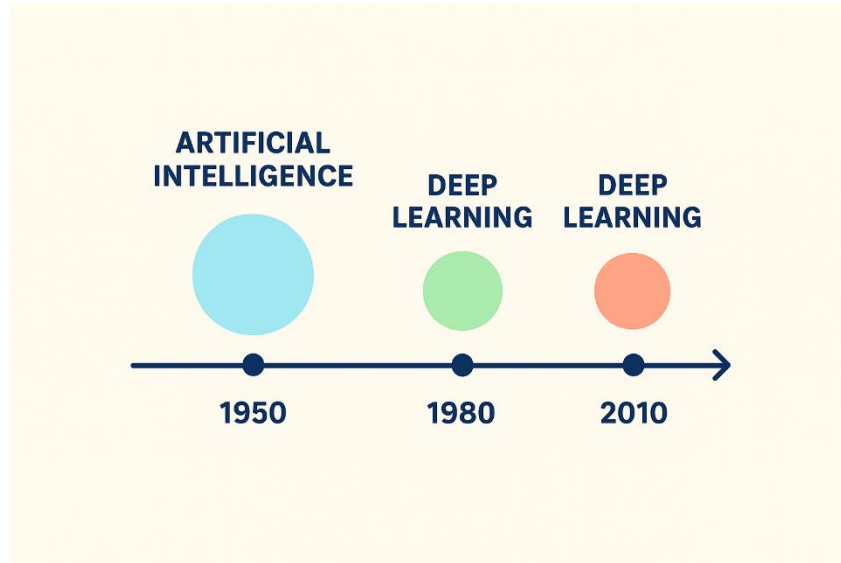


Figura 3: Línea de tiempo del desarrollo de la inteligencia artificial

2.14 Aprendizaje Automático y las redes de comunicación

El aprendizaje automático (Machine Learning, ML) se ha consolidado como una herramienta esencial para la evolución de las redes de comunicación modernas. Su integración permite a las redes adaptarse dinámicamente a cambios en el tráfico, anticipar fallos, optimizar recursos y mejorar la calidad del servicio sin intervención humana constante (Boutaba et al., 2018). Este paradigma representa un cambio fundamental frente a los enfoques tradicionales, basados en reglas estáticas y configuraciones manuales, donde la eficiencia y la seguridad dependían en gran medida de la supervisión humana.

El ML aplicado a redes busca dotarlas de inteligencia, autonomía y capacidad predictiva, transformando la infraestructura de telecomunicaciones en sistemas más

resilientes y proactivos. En este marco, las redes no solo reaccionan eventos, sino que también pueden anticiparlos y adaptarse en tiempo real.

2.15 Áreas de Aplicación del Aprendizaje Automático en Redes

Boutaba et al. (2018) identifican varias áreas estratégicas donde el ML tiene un impacto significativo:

- **Predicción de Tráfico y Optimización de Rutas:** los modelos de ML pueden analizar patrones históricos de tráfico para anticipar congestiones y ajustar rutas de manera proactiva. Esto permite redistribuir recursos de manera eficiente, evitando cuellos de botella y mejorando la calidad del servicio. Ejemplos de algoritmos utilizados incluyen redes neuronales recurrentes (RNN) y modelos de series temporales como ARIMA o LSTM, que permiten capturar patrones secuenciales del tráfico de red.
- **Clasificación y Gestión de Datos:** las técnicas de ML facilitan la categorización de grandes volúmenes de datos de red, priorizando información crítica frente a la no relevante. Por ejemplo, la clasificación de paquetes en tiempo real permite asegurar que aplicaciones de alta prioridad (como videoconferencias) tengan prioridad en el ancho de banda. Algoritmos de clasificación como SVM (Support Vector Machines) y árboles de decisión se utilizan frecuentemente en esta área.
- **Control de Congestión y Asignación de Recursos:** el ML permite detectar tendencias de saturación y ajustar dinámicamente la asignación de recursos, como ancho de banda y capacidad de procesamiento. Esto mejora la

eficiencia operativa y reduce costos. Modelos de aprendizaje por refuerzo se aplican para optimizar decisiones secuenciales en asignación de recursos bajo condiciones cambiantes.

- **Detección y Prevención de Fallos:** el ML ayuda a identificar patrones de comportamiento que preceden fallos, permitiendo acciones preventivas. Esto reduce interrupciones en el servicio y disminuye los costos de mantenimiento. Algoritmos de detección de anomalías, como Isolation Forest o K-means, se utilizan para detectar eventos atípicos que pueden indicar fallas inminentes.
- **Seguridad de Redes:** la seguridad es un área crítica donde el ML ha demostrado gran eficacia. Modelos avanzados permiten identificar comportamientos anómalos, ataques de denegación de servicio (DDoS) y patrones de intrusión sin necesidad de reglas predefinidas. Técnicas como aprendizaje supervisado y no supervisado se combinan para crear sistemas de defensa adaptativos capaces de evolucionar frente a nuevas amenazas.

2.16 Machine Learning y predicción económica

El aprendizaje automático (Machine Learning, ML) ofrece herramientas para el modelado predictivo que pueden manejar datos heterogéneos, relaciones no lineales y alta dimensionalidad (Chollet, 2018). Según Hastie, Tibshirani y Friedman (2009), los algoritmos de ML permiten descubrir patrones complejos en los datos que los modelos tradicionales podrían no captar.

Para este trabajo se optó por modelos que capturan relaciones no lineales y que en general no son regresiones lineales o polinómicas, pues se busca obtener

modelos que se fundamenten en técnicas distintas a los modelos de series de tiempo que esencia son modelos lineales.

En el contexto de la predicción de remesas, modelos como Random Forest, XGBoost o métodos de ensamblaje como el Stacking han mostrado ser alternativas competitivas frente a los enfoques clásicos (Makridakis et al., 2020). Estas técnicas permiten integrar múltiples fuentes de información económica y social, mejorando las predicciones y la capacidad de generalización. Dado que en los siguientes apartados se optó por un modelo de tipo Random Forest, no está demás dar algunos conceptos claves de este tipo de modelos.

2.17 Árboles de decisión

son uno de los modelos más sencillos de entender y aplicar, sin embargo, también son de los modelos más capaces con una alta interpretabilidad y fácil visualización (ver figura 1). Para Chollet (2018), los árboles de decisión tienen una estructura similar a la de un diagrama de flujo, que permite clasificar o predecir valores de salida a partir de datos de entrada, comenzando a despertar interés en la década de 2000, y para 2010 ya eran los modelos preferidos sobre los modelos de tipo Kernel (regresiones lineales, logarítmicas y polinómicas).

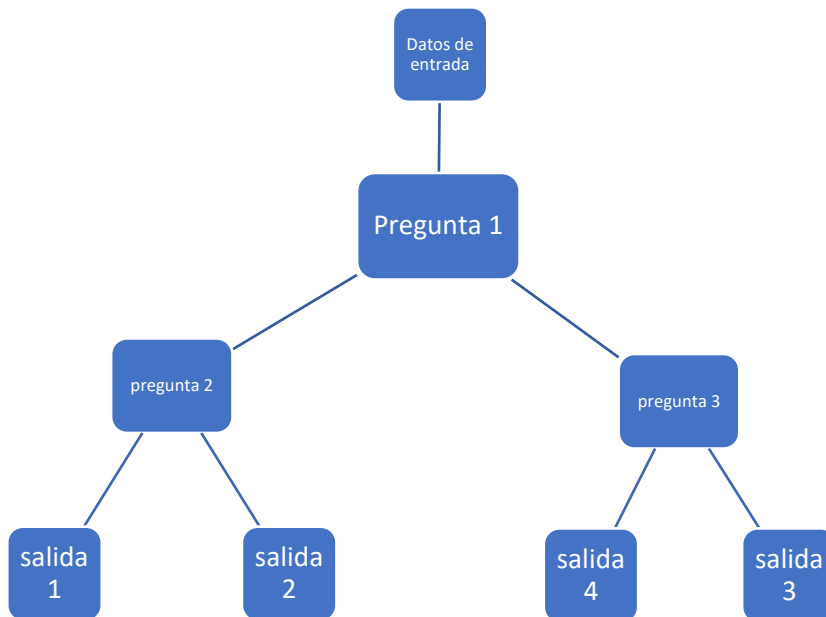


Figura 4: ejemplo de visualización de un árbol de decisión

2.18 Random Forest (bosques aleatorios)

Random Forest es un algoritmo de aprendizaje automático supervisado basado en el concepto de ensambles de árboles de decisión. Su objetivo principal es mejorar la precisión y estabilidad de las predicciones mediante la combinación de múltiples árboles independientes, reduciendo así la varianza y el riesgo de sobreajuste que presentan los árboles de decisión individuales (Becerra, 2023).

En otras palabras, la idea detrás del algoritmo Random Forest como menciona Chollet (2018), no es otra más que construir una gran cantidad de árboles de decisión especializados, y luego ensamblar las salidas de estos árboles en una salida final, dando un modelo mucho más robusto y con una muy amplia gama de

aplicaciones, siendo los modelos favoritos de la comunidad de ciencia de datos del sitio Kaggle hasta que fueron desplazados por los modelos de tipo gradiente en 2014.

El funcionamiento del algoritmo se basa en dos estrategias principales: primero, la generación de múltiples subconjuntos del conjunto de entrenamiento mediante muestreo bootstrap, lo que permite entrenar cada árbol con datos distintos; segundo, la selección aleatoria de atributos en cada nodo de cada árbol, introduciendo diversidad entre los árboles y garantizando que no estén correlacionados (Becerra, 2023). Una vez construidos, los árboles realizan predicciones individuales que se agregan: mediante voto mayoritario para clasificación o promedio para regresión, produciendo un resultado más robusto que cualquier árbol individual.

Muestreo Bootstrap: el proceso de muestreo de los datos con reemplazo se denomina bootstrap. Un tercio de los datos no se usan para el entrenamiento y pueden ser usados para test. Este conjunto se denomina "out of bag (OOB) samples" (Becerra, 2023)

En otras palabras, **bootstrap** consiste en tomar muestras aleatorias del conjunto de entrenamiento con reemplazo para construir cada árbol del Random Forest, permitiendo que cada árbol se entrene con un subconjunto diferente y que algunos datos queden fuera para validación interna (*out-of-bag*).

Entre las ventajas del Random Forest se encuentran su capacidad para reducir el sobreajuste, manejar datos faltantes de manera efectiva, evaluar la

importancia de cada atributo en la predicción y su aplicabilidad a tareas de clasificación y regresión con grandes volúmenes de datos (Becerra, 2023). Sin embargo, el modelo tiene limitaciones, su interpretabilidad es limitada debido a la gran cantidad de árboles que lo componen y requiere un alto consumo de recursos computacionales, especialmente con conjuntos de datos extensos.

Este algoritmo ha demostrado ser útil en diversas áreas, incluyendo la medicina, para diagnóstico y predicción de resultados; en finanzas, para detección de fraudes y evaluación de riesgos; en biología computacional, para clasificación de genes; y en ingeniería, para monitoreo de sistemas y detección de fallos (Becerra, 2023). Por su robustez y versatilidad, Random Forest se ha consolidado como una herramienta esencial en aplicaciones prácticas de aprendizaje automático y constituye **un elemento relevante dentro de los modelos de predicción y análisis de datos.**

Aún ahora después de más de una década del lanzamiento de Kaggle, con el auge de los modelos de inteligencia artificial generativa (genIA) y los modelos de lenguaje natural o modelos procesamiento de lenguaje natural largos (LLM), si se quiere ser más técnico; los modelos Random Forest, modelos de tipo gradiente y árboles de decisión siguen siendo muy utilizados, no solo por su relativa sencillez o su buena interpretabilidad, sino porque siguen siendo muy eficaces.

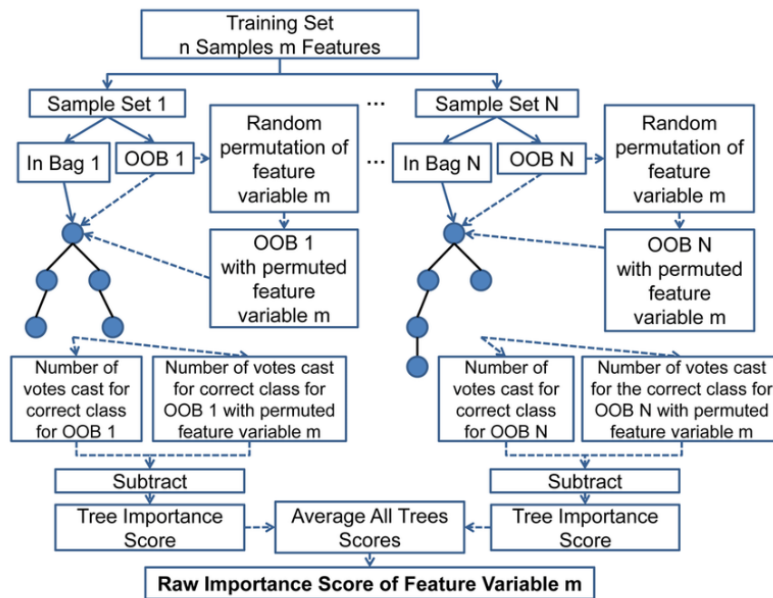


Figura 5: Ilustración del funcionamiento del algoritmo de un modelo de tipo Random Forest

2.19 Evaluación de un modelo Random Forest

La evaluación de un modelo Random Forest es esencial para medir su rendimiento y capacidad de generalización. A continuación, se describen los principales métodos y métricas utilizados en este proceso.

2.19.1 Uso de Out-of-Bag (OOB) Samples

Una característica distintiva de Random Forest es el uso de muestras out-of-bag (OOB) para la validación del modelo. Durante el proceso de entrenamiento, cada árbol se construye utilizando una muestra aleatoria con reemplazo del conjunto de datos original, lo que significa que aproximadamente un tercio de los datos no se utilizan en la construcción de cada árbol. Estos datos no utilizados se denominan muestras OOB y se emplean para evaluar el rendimiento del modelo sin necesidad

de un conjunto de prueba adicional. El error calculado a partir de estas muestras OOB se conoce como error OOB y proporciona una estimación precisa del rendimiento del modelo, comparable al uso de un conjunto de prueba tradicional (Becerra, 2023).

2.19.2 Matriz de Confusión

Para evaluar la precisión de un modelo de clasificación, se utiliza la matriz de confusión, que compara las etiquetas predichas por el modelo con las etiquetas reales del conjunto de prueba. Esta matriz permite identificar el número de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, proporcionando una visión detallada del rendimiento del modelo en cada clase. A partir de la matriz de confusión, se pueden calcular métricas adicionales como la precisión, la sensibilidad y la especificidad, que ofrecen una comprensión más profunda de la eficacia del modelo en diferentes aspectos (Becerra, 2023).

Cabe aclarar que este método de evaluación no aplica para el modelo desarrollado en este estudio pues el modelo es un modelo de predicción y no de clasificación

2.19.3 Métricas de Evaluación

Además de la matriz de confusión, existen otras métricas clave para evaluar el rendimiento de un modelo Random Forest:

- **Precisión (Accuracy):** proporción de predicciones correctas sobre el total de predicciones realizadas. Es una métrica general que indica el porcentaje de aciertos del modelo.
- **Precisión por clase:** Mide la exactitud de las predicciones para cada clase individualmente, calculando la proporción de verdaderos positivos sobre el total de elementos clasificados en esa clase.
- **Sensibilidad (Recall):** Indica la capacidad del modelo para identificar correctamente los casos positivos, calculando la proporción de verdaderos positivos sobre el total de casos reales positivos.
- **Especificidad:** Mide la capacidad del modelo para identificar correctamente los casos negativos, calculando la proporción de verdaderos negativos sobre el total de casos reales negativos.

Estas métricas son fundamentales para comprender el comportamiento del modelo en diferentes escenarios y clases, y deben ser consideradas en conjunto para una evaluación completa (Becerra, 2023).

Para el caso de un modelo de tipo regresor, es decir, que su salida no sea categórica sino numérica, las métricas de evaluación cambian, a continuación, se presenta un resumen de las principales métricas.

2.19.4 Métricas de Evaluación para Regresión

Para cuantificar el rendimiento de un Random Forest Regressor, se utilizan métricas que miden la diferencia entre los valores predichos y los valores reales:

- **Error Cuadrático Medio (Mean Squared Error, MSE):** Calcula la media de los cuadrados de las diferencias entre los valores reales y predichos, penalizando más los errores grandes.
- **Raíz del Error Cuadrático Medio (Root Mean Squared Error, RMSE):** Es la raíz cuadrada del MSE, interpretándose en las mismas unidades que la variable objetivo.
- **Error Absoluto Medio (Mean Absolute Error, MAE):** Calcula la media de los valores absolutos de los errores, ofreciendo una medida menos sensible a outliers.
- **Coefficiente de Determinación (R^2):** Indica la proporción de la variabilidad de la variable dependiente que es explicada por el modelo. Valores cercanos a 1 indican un ajuste más preciso.

Estas métricas permiten comparar modelos y ajustar hiperparámetros para optimizar el desempeño del Random Forest Regressor (Becerra, 2023).

2.19.5 Importancia de las Variables

Una ventaja significativa de Random Forest es su capacidad para estimar la importancia de las variables en la predicción. Este proceso se realiza mediante la permutación aleatoria de los valores de cada variable y la observación del impacto en el rendimiento del modelo. Si la permutación de una variable específica conduce a una disminución significativa en la precisión del modelo, se considera que esa variable tiene alta importancia. Este enfoque permite identificar las variables más

relevantes para el modelo y puede ser útil para la selección de características y la interpretación del modelo (Becerra, 2023).

2.19.6 Validación Cruzada

Aunque Random Forest ofrece una estimación del error mediante las muestras OOB, en algunos casos puede ser beneficioso realizar una validación cruzada para obtener una evaluación más robusta del rendimiento del modelo. La validación cruzada implica dividir el conjunto de datos en varios subconjuntos (o "folds"), entrenar el modelo en algunos de estos subconjuntos y evaluarlo en los restantes. Este proceso se repite varias veces, y los resultados se promedian para obtener una estimación más precisa del rendimiento del modelo. La validación cruzada es especialmente útil cuando se dispone de un conjunto de datos limitado y se desea maximizar la utilización de los datos disponibles (Becerra, 2023).

Capítulo III: Metodología de la investigación

3.1 Enfoque de la investigación

En este apartado se define todo el proceso metodológico que se siguió en el desarrollo del trabajo de investigación, bajo el marco conceptual de **una investigación de tipo cuantitativo**, con alcance explicativo. Por tanto, este trabajo se enfoca en responder y explicar condiciones manifiestas o relacionadas con la variable objetivo (Hernández Sampieri, Fernández Collado, Baptista Lucio, 2010).

3.2 Diseño de la investigación

Para este trabajo se toma un diseño de investigación de tipo **no experimental**, por lo que se podría definir como un estudio donde no se prepara una situación determinada ni controlada (experimento) y sin manipulación de las variables independientes, por lo que sólo se observan los fenómenos en su ambiente natural para después ser analizados (Hernández Sampieri, Fernández Collado, Baptista Lucio, 2010).

3.3 Tipo de estudio

Como menciona Hernández Sampieri, Fernández Collado y Baptista Lucio (2010), los **estudios longitudinales** son aquellos que recaban datos en varios puntos en el tiempo para luego realizar inferencias acerca de su evolución, sus causas y efectos.

Es así que para este trabajo se seleccionó el tipo de estudio **longitudinal** con el que se busca analizar los **flujos de remesas familiares** y las diferentes

variables que pueden o no afectarlos, y de forma similar en el caso de las finanzas públicas las variables principales giran alrededor del **balance primario**, es decir, **ingresos tributarios netos, ingresos corrientes netos y gastos corrientes del gobierno**, realizando un análisis por medio de modelos Auto regresivos con rezagos (ARDL) y modelos de Machine Learning (ML).

3.4 Población y muestra

Unidades de análisis

Según Hernández Sampieri, Fernández Collado y Baptista Lucio (2010), las unidades de análisis también se pueden denominar casos o elementos de estudio, por lo que para este trabajo las unidades de análisis son las siguientes:

- Flujos de remesas familiares mensuales y anuales provenientes de Estados Unidos.
- Indicadores (variables) del Banco Mundial para medir el desarrollo.
- Ingresos tributarios netos del gobierno.
- Gastos corrientes del gobierno.
- Impuesto al valor agregado (IVA) recaudado mensualmente.
- Tasas de homicidios intencionales anuales.
- Datos de inmigrantes admitidos y expulsados en los Estados Unidos.

Población

Partiendo del concepto de **Población o universo**, que es el conjunto de todos los casos que concuerdan con determinadas especificaciones (Hernández Sampieri, Fernández Collado, Baptista Lucio, 2010), la población que se tomó fue:

- Para el primer proceso de análisis se tienen todos los indicadores del banco mundial (250 indicadores) para los 265 países.
- Para el segundo apartado se disponen de todos los flujos de remesas, IVA, Ingresos y gastos del gobierno central para todas las fechas publicados por el Banco Central de Reserva (BCR).

Muestra

Para este trabajo el tipo de muestra elegido es una **muestra no probabilística**, pues su selección se ha hecho con base en la disponibilidad de datos y la veracidad de las fuentes de información.

Para el análisis de los factores que afectan los **flujos de remesas**, de los 265 países de los cuales se disponen datos en las bases del Banco Mundial, se seleccionaron 3: El Salvador, Honduras y República Dominicana, con ellos se consolidó una base de datos, para luego desarrollar un proceso de selección de variables y finalmente, un proceso de modelación con ML.

En el caso del análisis del impacto de los flujos de remesas en las finanzas públicas se tomó un periodo de 15 años de datos, es decir, de las series de tiempo completas (que en algunos variables pueden ir desde 1990 hasta 2025) se tomó el periodo que va de enero de 2010 a junio de 2025.

3.5 Recolección de datos

Para la recolección de los datos se seleccionaron múltiples fuentes de información, siendo todas bases de datos, dentro de las principales fuentes se encuentran:

- Banco Central de Reserva de El Salvador.
- Banco Mundial.
- Oficina de Estadísticas de Seguridad Nacional de Estados Unidos.

3.6 Procesamiento de los datos

Una vez recolectados los datos, se utilizó **Python** como herramienta de limpieza, procesamiento y preparación de datos, definiendo una metodología clara de limpieza de datos, imputación de valores faltantes y de selección de variables.

Capítulo IV: Análisis y resultados de la investigación

En este capítulo se presentan los resultados obtenidos en los procesos de modelación, con el objetivo de poder dar respuesta a qué factores son lo más importantes en la predicción de los flujos de remesas y sobre cuál es el impacto de los flujos de remesas en las finanzas públicas.

4.1 Análisis de resultados del modelo de predicción de los flujos de remesas familiares

Para poder dar respuestas a qué factores tanto económicos como no económicos son los que más impactan en los flujos de remesas, se desarrolló un proceso de modelación completo que permitió obtener un modelo de Machine Learning con un alto nivel de eficacia en las predicciones que además permita analizar la importancia de las variables que se usaron como predictoras.

4.1.1 Proceso de selección de variables

Para este proceso se disponía de 250 variables (indicadores de desarrollo del banco mundial), estas variables van desde indicadores macroeconómicos, variables demográficas, sociales, sociodemográficas, geográficas, entre otras.

Para el proceso de filtrado de variables se utilizó principalmente dos criterios, el primero es **baja varianza** y alto número de **registros faltantes**, esto se hace, porque variables con baja variabilidad estadísticamente son muy poco explicativas y lo mismo ocurre si tienen demasiados registros faltantes, es decir, dejan de ser explicativas. El segundo filtro fue de **alta correlación**, este se aplica pues, si dos

variables están altamente correlacionadas, ambas entregan la misma información explicativa, en otras palabras, en lugar de dar mayor información, estas pueden llegar a causar redundancia en los análisis.

Finalmente, el último filtro que se aplicó fue el denominado **ensamble débil**, este método consiste en crear múltiples modelos “débiles”, por lo general se usan modelos de tipo Random Forest, que se componen de una cantidad pequeña de árboles de decisión poco profundos (de una profundidad de no más de 4 niveles) que en general tendrán poco poder predictivo, de ahí el término débil, luego se agrega un número determinado de variables aleatorias o variables de ruido, y se corre de forma iterativa para ir descartando aquellas variables que muestran un comportamiento por debajo del de las variables de ruido (umbral de ruido).

A continuación, se presentan más detalles de los resultados obtenidos en la aplicación de estos filtros para el caso de las variables disponibles:

- **Filtro por alto número de valores nulos y baja varianza:** en este caso se eliminaron variables que tenían más del 50% de sus registros faltantes o nulos, y todas aquellas que mostraban un coeficiente de variación por debajo de 0.05.
- **Filtro por alta correlación:** el segundo filtro se aplicó usando un umbral de correlación de 0.7, es decir, se eliminaron variables que mostraron una correlación de más del 70% con otras ya seleccionadas.

- **Filtro de ensamble débil:** se tomó un Random Forest con 200 árboles, con una profundidad máxima de 4 y con 5 variables de ruido para cada iteración.

Estos filtros se corrieron para las 215 variables finales que disponía la base de entrenamiento que se construyó con los datos del banco mundial, primero para todos los países receptores de remesas y luego sólo para Estados Unidos, porque la idea de modelación era contar con variables predictoras de los países receptores y variables de USA (ver Figura 6 y Figura 7).

```

Resumen de Importancias:
                                variable  importancia
0      donaciones_sin_cooperacion_tecnica_USD  0.240326
1      tierra_cultivable_hectareas_por_persona  0.185553
2      energia_alternativa_nuclear_pct_uso_total  0.128455
3      valor_agregado_agricultura_PIB  0.110329
4      rendimiento_cereal_kg_hectarea  0.085316
5      empleo_agricultura_mujeres_pct_empleo_femenino  0.042478
6      salidas_transporte_aereo_registradas  0.031133
7      exportaciones_bienes_servicios_pct_PIB  0.026077
8      dinero_amplio_pct_PIB  0.017028
9      servicio_total_deuda_pct_exportaciones  0.015897
10     uso_fertilizantes_kg_por_hectarea  0.015228
11     migracion_neta  0.014900
12     tierra_agricola_pct_total  0.011815
14     crecimiento_PIB_pct_anual  0.010285
15     exportaciones_minerales_metales_pct_exportaciones  0.009874
17     ahorro_bruto_pct_PIB  0.008843
18     inflacion_deflactor_PIB_pct_anual  0.008013
19     deuda_externa_pct_INB  0.006636

```

Figura 6: Variables seleccionadas de los países receptores para el modelo de flujos de remesas

Resumen de Importancias:		
	variable	importancia
0	usa_dinero_amplio_pct_PIB	0.774920
1	usa_valor_agregado_agricultura_PIB	0.082976
2	usa_inflacion_deflactor_PIB_pct_anual	0.061768
5	usa_formacion_bruta_capital_pct_PIB	0.009007
7	usa_crecimiento_PIB_pct_anual	0.006354
8	usa_ingresos_sin_donaciones_pct_PIB	0.002891

Figura 7: Variables seleccionadas de Estados Unidos para el modelo de flujos de remesas

4.1.2 Imputación de valores faltantes

Aunque se pasó un filtro por valores faltantes aún quedaron variables con valores faltantes en un porcentaje menor al 50% del total de registros, para estos casos y a pesar de que existen métodos más sencillos como la imputación por media o mediana, se optó por realizar un proceso de imputación por interpolación usando el método numérico polinómico de grado 3.

Para aquellos casos faltantes que se encontraban en eslabones intermedios de las series y para los valores ubicados en puntos aislados de la serie como al final o al principio o para aquellos donde los valores faltantes eran múltiples y consecutivos (casos donde los métodos de interpolación no funcionan bien) se hizo una imputación por medio de un modelo predictivo que usaba las otras variables para predecir cual sería el valor faltante.

Todo esto se realizó para garantizar que los valores imputados eran lo más realistas posible, y no alteraban el comportamiento de los valores reales de las series de datos de las variables explicativas, en las figuras 8 y 9 se muestra el antes

y después del proceso de imputación de la variable inmigrantes expulsados por EEUU.

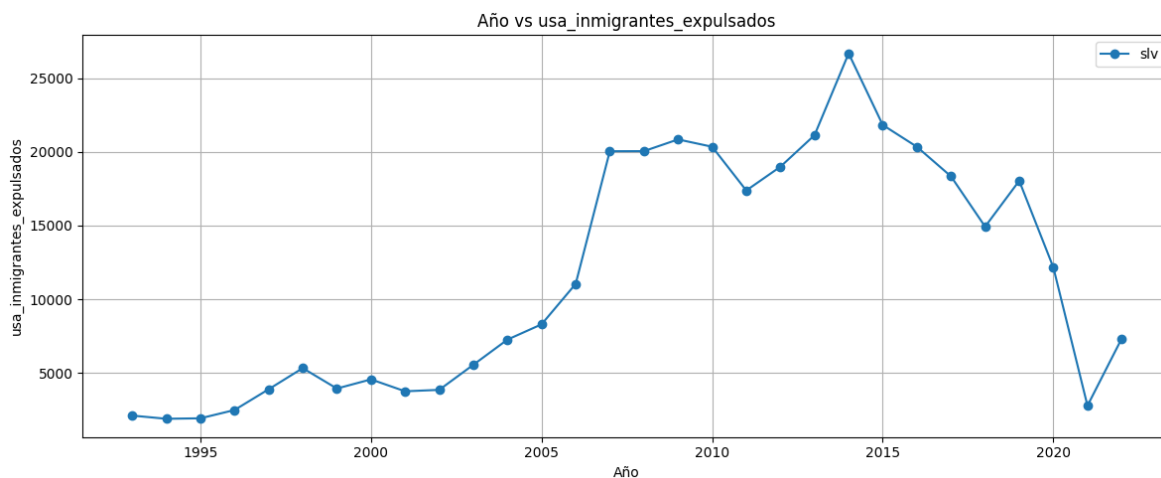


Figura 8: serie de la variable inmigrantes expulsados por USA para El Salvador sin imputar

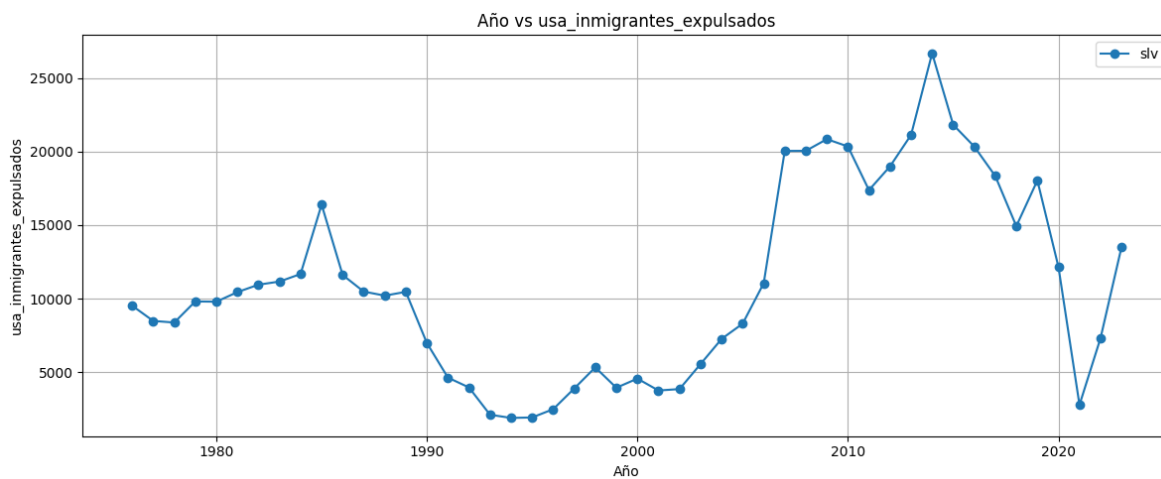


Figura 9: serie de la variable de inmigrantes expulsados por USA para El Salvador después de imputación

4.1.3 Selección del mejor modelo

Usando un proceso de validación cruzada con el método de partición **TimeSeriesSplit** de Python, para 5 conjuntos y usando la métrica RMSE como métrica principal, los resultados se muestran en la figura 10.

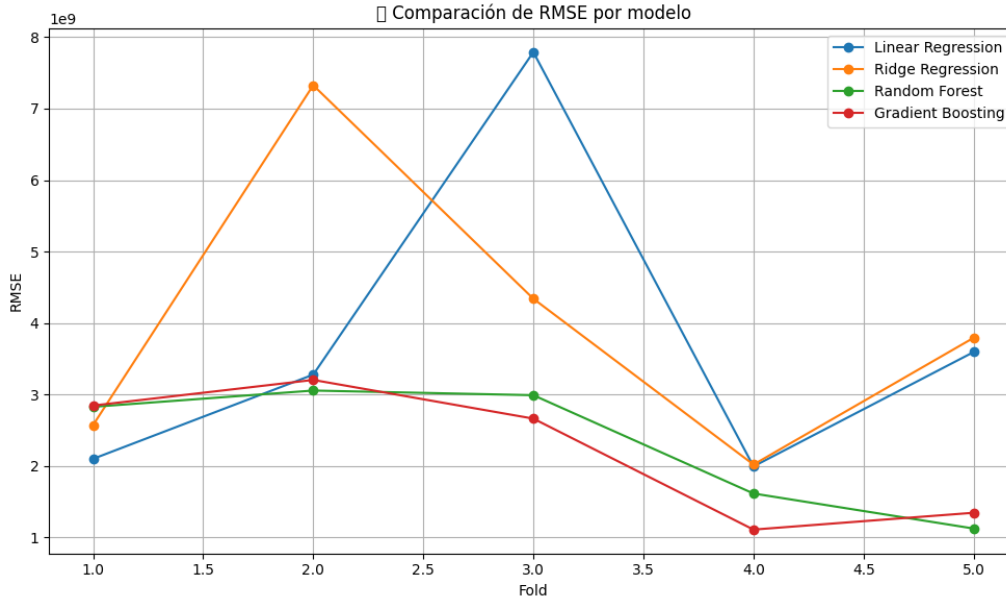


Figura 10: comportamiento bajo RMSE de los modelos candidatos

4.1.4 Modelo Final de predicción de Flujos de Remesas Familiares

El modelo seleccionado fue un modelo Random Forest, para el entrenamiento del modelo, se dispuso de una base de entrenamiento final con 39 variables y 149 registros, que corresponden a los datos de Estados Unidos como país emisor de remesas y *El Salvador, Honduras, y República Dominicana* como países receptores de remesas, la selección final de estos países se hizo después de múltiples pruebas del modelo y por la calidad de sus datos que se reflejó en el desempeño del modelo.

4.1.4.1 Variable Objetivo (flujos de remesas anuales)

Para la variable objetivo se realizaron distintas pruebas y al final se concluyó que la versión que mejores resultados daba era la variable sin ningún tipo de tratamiento o escalado, es decir, su versión original en millones de dólares (USD).

4.1.4.2 Cálculos de hiperparámetros del modelo

Para obtener la mejor configuración (con base en el rendimiento) se optó por la librería **Optuna** de Python, tratando de limitar la complejidad del modelo, pensando en evitar un posible sobreajuste (siendo este el principal desafío dado el tamaño limitado de los datos cuando se trabaja con series de tiempo), lo que podría afectar su poder de generalización o predicción, todo el proceso se corrió para un total de 2000 iteraciones, para poder garantizar la convergencia a un configuración final estable.

```
[I 2025-09-07 10:35:17,728] Trial 996 finished with value: 1592954252.6399217 and parameters: {'n_estimators': 99, 'max_depth': 6, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': None}
[I 2025-09-07 10:35:18,295] Trial 997 finished with value: 1663158477.9221323 and parameters: {'n_estimators': 99, 'max_depth': 6, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': None}
[I 2025-09-07 10:35:18,852] Trial 998 finished with value: 1578106773.87338 and parameters: {'n_estimators': 99, 'max_depth': 6, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': None}
[I 2025-09-07 10:35:19,444] Trial 999 finished with value: 1584839512.1618972 and parameters: {'n_estimators': 99, 'max_depth': 6, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': None}
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
✓ Mejor configuración encontrada:
{'n_estimators': 99, 'max_depth': 6, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': None}
✗ Mejor RMSE promedio en CV: 1569235897.8692
```

Figura 11: salida del proceso de optimización de parámetros con Optuna

4.1.4.3 Entrenamiento del modelo

Con el objetivo de garantizar una buena generalización y evitar el sobreajuste se hizo un entrenamiento con validación cruzada para un total de 5 Folds (revisar figura 12), para evaluar los resultados se usaron dos métricas principales:

Coefficiente de determinación: R^2

Error cuadrático medio: RMSE

Luego, finalmente se termina con una prueba con un conjunto de validación (holdout) por fuera de los conjuntos o Folds de entrenamiento, todo esto se muestra en las figuras 13 y 14.

```
Fold 1 ▶ r2: -2.31, RMSE: 4886793187.85
Fold 2 ▶ r2: -63.13, RMSE: 523212931.96
Fold 3 ▶ r2: 0.77, RMSE: 981865928.59
Fold 4 ▶ r2: 0.79, RMSE: 805785686.52
Fold 5 ▶ r2: 0.87, RMSE: 648521754.42

Promedio de métricas con TimeSeriesSplit:
r2      : -12.60
RMSE    : 1569235897.87
```

Figura 12: resultados del modelo para los Folds de entrenamiento

```
Evaluación final en conjunto holdout:
RMSE : 600848037.92
R2  : 0.84
```

Figura 13: Resultados finales del modelo en conjunto de validación

4.1.4.4 Resultados del modelo

El modelo muestra un comportamiento un tanto errático en los primeros Folds, desde valores extremos y negativos para la métrica R2 (-2.31 y -63.13) hasta resultados cada vez más sólidos en los últimos 3 Folds (0.77, 0.79 y 0.87) lo que nos indica que el modelo alcanza a capturar el 87% de la variabilidad de las remesas para el último conjunto de datos de entrenamiento, lo cual es **sobresaliente**.

Para el caso del conjunto de validación o Holdout, el modelo alcanza resultados destacados con un R2 de 0.84 y un RMSE de 600 millones aproximadamente, y aunque este número parezca alto, este valor representa entre

un 7 a 9% de un flujo anual de remesas, lo cual es un error de estimación bastante razonable para estudios macroeconómicos, debemos recordar que los flujos de remesas en general para los países receptores de nuestra región son cantidades expresadas en miles de millones de dólares, alcanzando un máximo para El Salvador de más de 8 mil millones de dólares para años recientes.

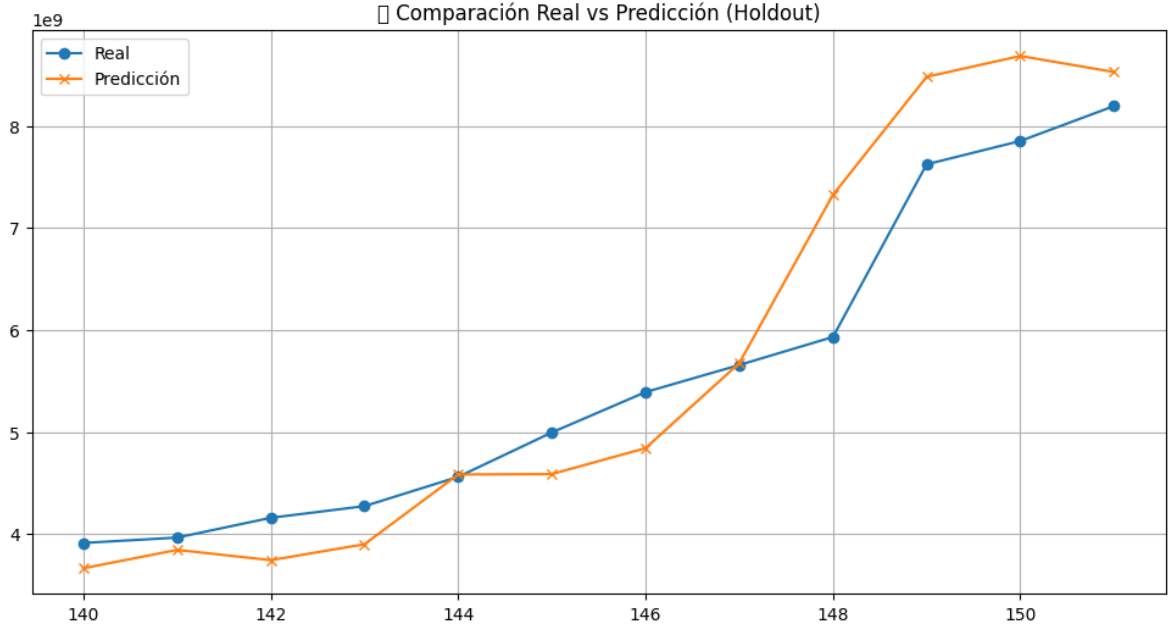


Figura 14: Predicción del modelo en el conjunto Holdout

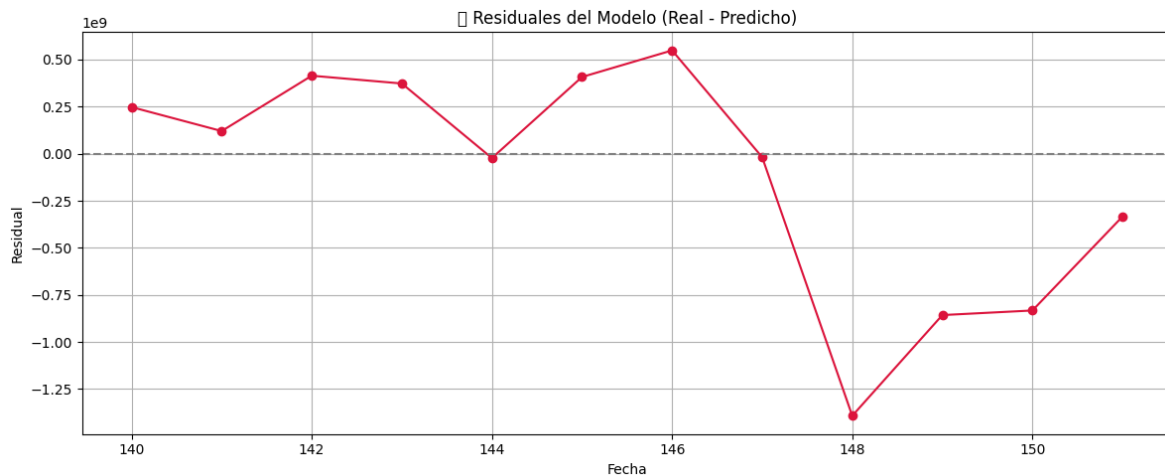


Figura 15: Visualización de los errores absolutos de las predicciones del modelo

Analizando el modelo de forma más profunda se observa que dentro de todas las variables que se le han dado como entrada, casi toda la importancia se centra en 5 variables (ver figura 16), siendo las siguientes 10 las más importantes:

- usa_dinero_amplio_pct_PIB (65.3%)
- usa_valor agregado_agricultura_PIB (16.3%)
- valor_agregado_agricultura_PIB (8.70%)
- usa_inflacion_deflactor_PIB_pct_anual (3.85%)
- exportaciones_bienes_servicio_pct_PIB (0.79%)
- homicidios_intencionales_por_100milhabi (0.60%)
- rendimiento_cereal_kg_hectarea (0.52%)
- dinero_amplio_pct_PIB (0.46%)
- tierra_cultivable_hectareas_por_persona (0.43%)
- usa_inmigrantes_expulsados (0.43%)

Estas 10 variables suman aproximadamente el 88.68% de la importancia dentro del modelo, a continuación, se muestra más información sobre la definición de las variables y que intentan medir.

usa_dinero_amplio_pct_PIB: proporción de la oferta monetaria amplia (M2 o M3, incluye efectivo, depósitos a la vista y de ahorro) respecto al PIB de Estados Unidos, indica el nivel de liquidez financiera disponible en la economía estadounidense en relación con su tamaño.

usa_valor agregado_agricultura_PIB: contribución del sector agrícola al PIB total de EE. UU. Refleja la importancia relativa de la agricultura dentro de la economía estadounidense.

valor_agregado_agricultura_PIB: participación del sector agrícola en el PIB del país receptor de remesas. Da cuenta de cuánto depende la economía doméstica de la agricultura como fuente de producción y empleo.

usa_inflacion_deflactor_PIB_pct_anual: Variación porcentual anual del deflactor del PIB en EEUU. Es una medida de inflación que abarca todos los bienes y servicios de la economía estadounidense, útil para ver el poder adquisitivo y costos relativos.

exportaciones_bienes_servicio_pct_PIB: valor total de las exportaciones de bienes y servicios como proporción del PIB del país receptor de remesas. Refleja el grado de apertura comercial y dependencia de la economía de sus exportaciones.

homicidios_intencionales_por_100milhabi: tasa de homicidios intencionales por cada 100,000 habitantes en el país receptor. Un indicador de violencia e inseguridad, que puede influir en la migración y, por tanto, en las remesas.

rendimiento_cereal_kg_hectarea: producción de cereales (maíz, trigo, arroz, etc.) en kilogramos por hectárea cultivada. Refleja la productividad agrícola del país receptor, importante para ingresos rurales y estabilidad alimentaria.

dinero_amplio_pct_PIB: relación de la oferta monetaria amplia con respecto al PIB del país receptor de remesas. Un indicador de la liquidez y desarrollo financiero interno.

tierra_cultivable_hectareas_por_persona: superficie de tierra cultivable (apta para agricultura) en hectáreas disponible por habitante. Un indicador de la dotación de recursos agrícolas per cápita.

usa_inmigrantes_expulsados: número de inmigrantes expulsados o deportados desde EE. UU. Refleja la política migratoria estadounidense y afecta directamente a la población migrante de los países receptores, lo que influye en el flujo de remesas.

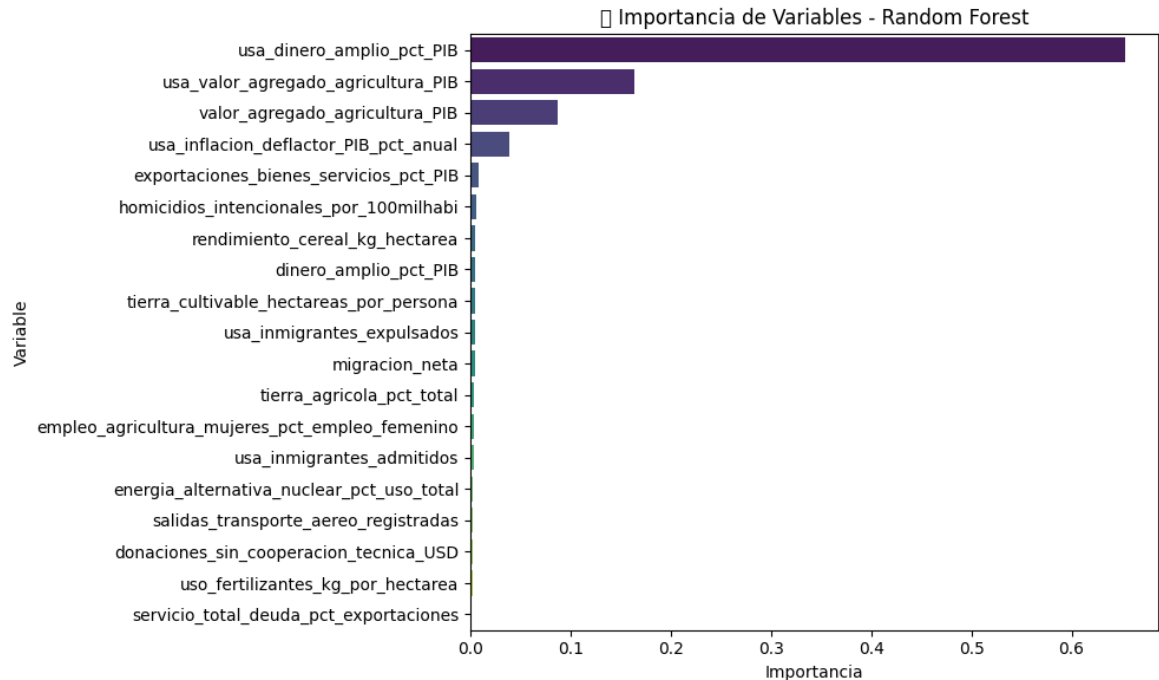


Figura 16: Importancia de las variables predictoras en el modelo de flujos de remesas

Se obtuvieron predicciones para la base completa (es decir, para todos los países receptores), pudiendo visualizar y analizar los resultados en particular para El Salvador, mostrando un desempeño bueno del modelo, teniendo un comportamiento aceptable para los distintos shocks económicos históricos, como, por ejemplo, la crisis financiera de 2008, pandemia COVID y post pandemia, que han impactado de forma importante los flujos de remesas, ver las figuras 17, 18, 19 y 20.

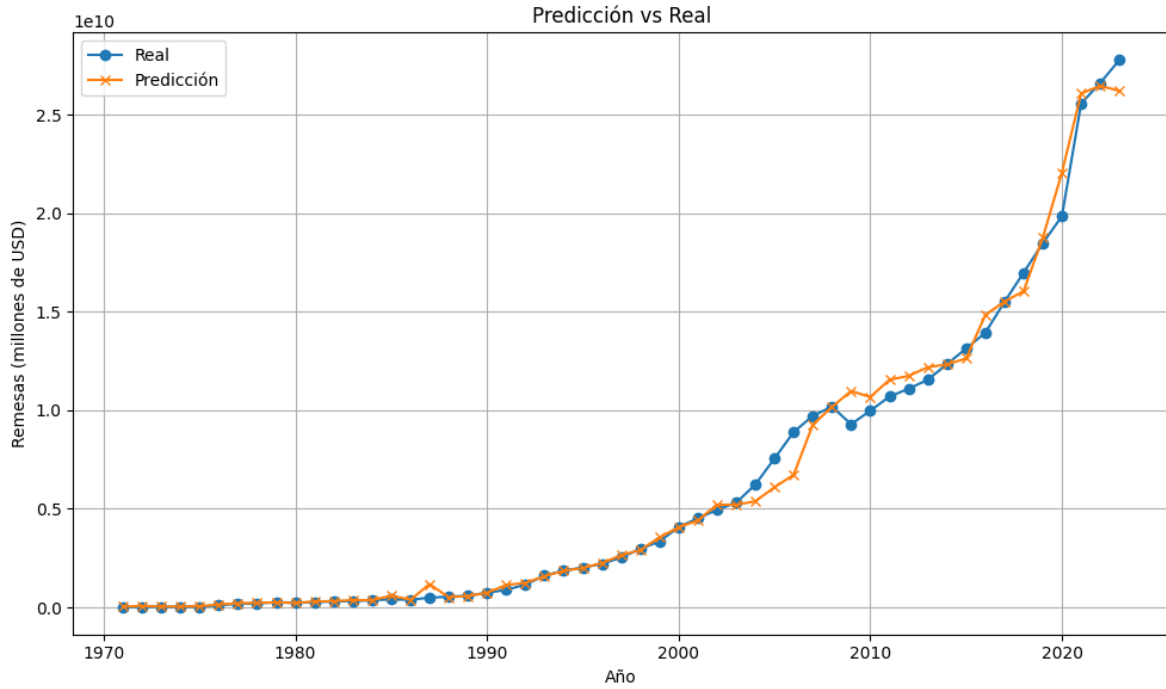


Figura 17: Visualización de las predicciones del modelo para toda la base de datos

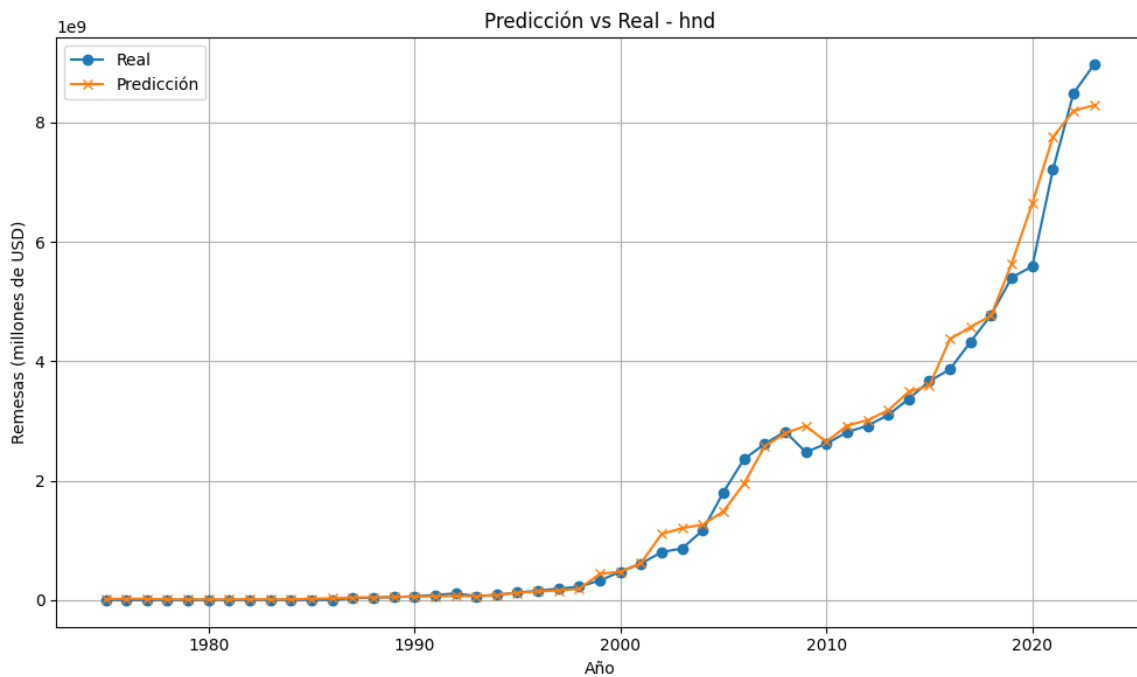


Figura 18: Visualización de las predicciones del modelo para Honduras

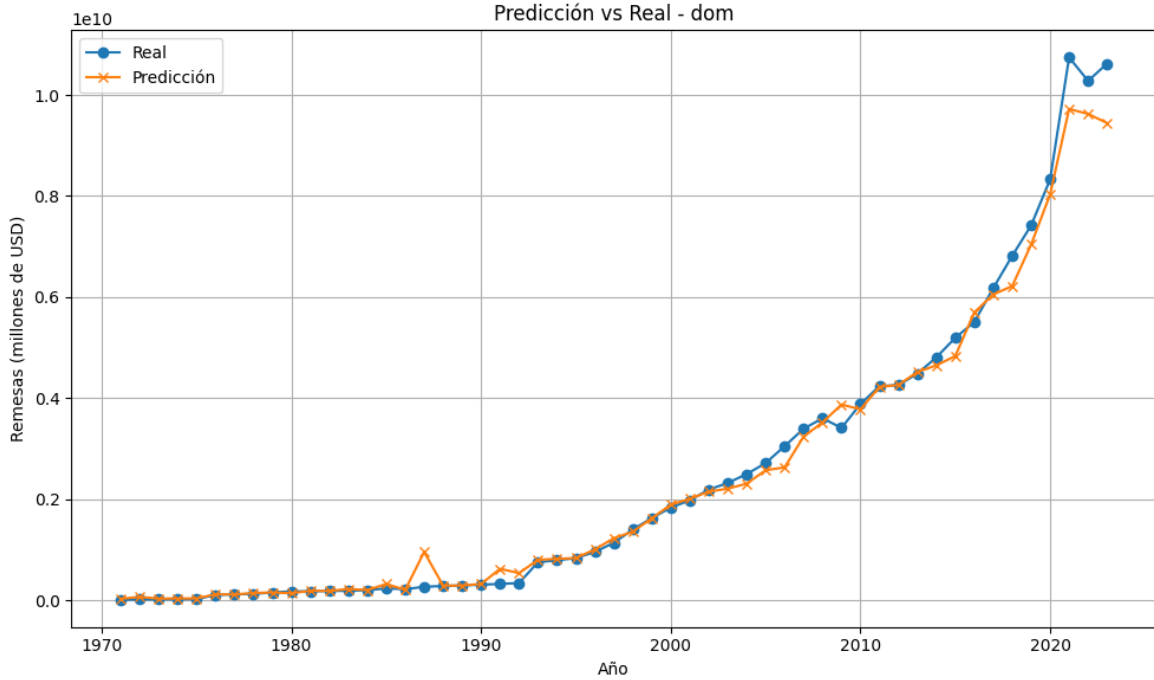


Figura 19: Predicciones del modelo para República Dominicana

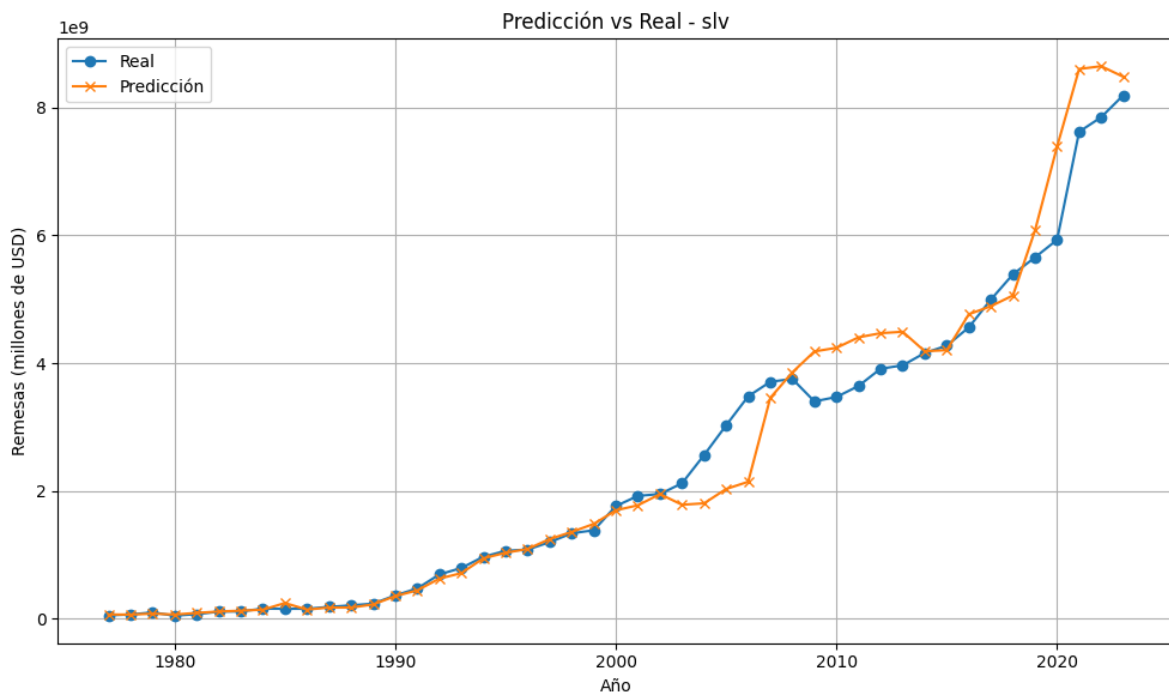


Figura 20: Predicciones del modelo para El Salvador

Para cerrar esta sección, se observa de forma general que el modelo desarrollado tiene un rendimiento muy positivo, con funcionalidades que no se limitan para El Salvador, pues se alimenta de datos de otros países de características similares, como visualización final se deja el diagrama de dispersión de las predicciones, que reafirman ese buen comportamiento, ver figura 21.

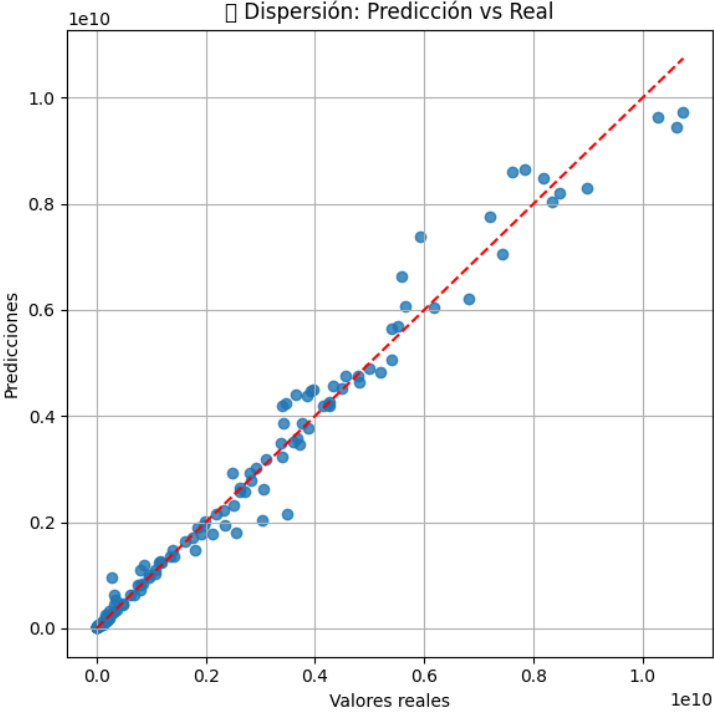


Figura 21: Diagrama de dispersión de las predicciones del modelo de flujos de remesas

4.2 Análisis de resultados de los modelos ARDL

El objetivo de esta sección es poder analizar los modelos de series temporales con rezago ARDL, estos modelos se desarrollaron con el objetivo de

analizar el impacto que tienen los flujos de remesas en las finanzas públicas de El Salvador.

4.2.1 Diagnóstico de las finanzas públicas

Como ya se ha mencionado para esta parte se optó por modelos de tipo ARDL por el tipo de análisis que aportan, tanto para las variables predictoras como para sus niveles de rezago, con esto claro, lo primero es determinar cuál es la naturaleza de las finanzas públicas de El Salvador, analizando la información consolidada publicada por el BCR.

Este primer análisis se realizó principalmente enfocado en dos apartados, los ingresos tributarios netos, que son la fuente de ingresos principal del gobierno y los gastos corrientes, que representan el principal pasivo del gobierno y son la clave para el funcionamiento diario de las distintas instituciones del estado en general.

Lo primero es analizar qué tan representativos son los ingresos tributarios respecto a los ingresos corrientes totales y a su vez, que tanto representa el IVA respecto al ingreso tributario neto, para lo cual se hizo un análisis histórico (ver figuras 22 y 23) y se concluye, que la principal fuente de ingresos corrientes son los ingresos tributarios y que, a su vez, casi todos dependen del IVA, y hay que destacar que en los últimos años se muestra una dependencia cada vez más acentuada.

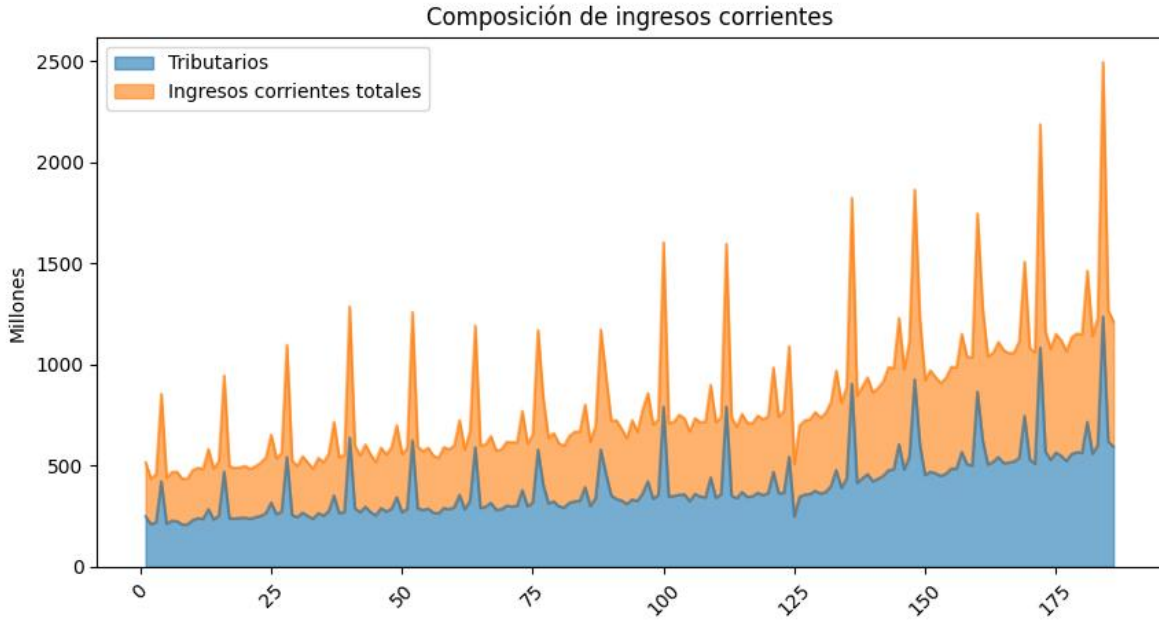


Figura 22: Relación Ingresos tributarios vs Ingresos corrientes totales

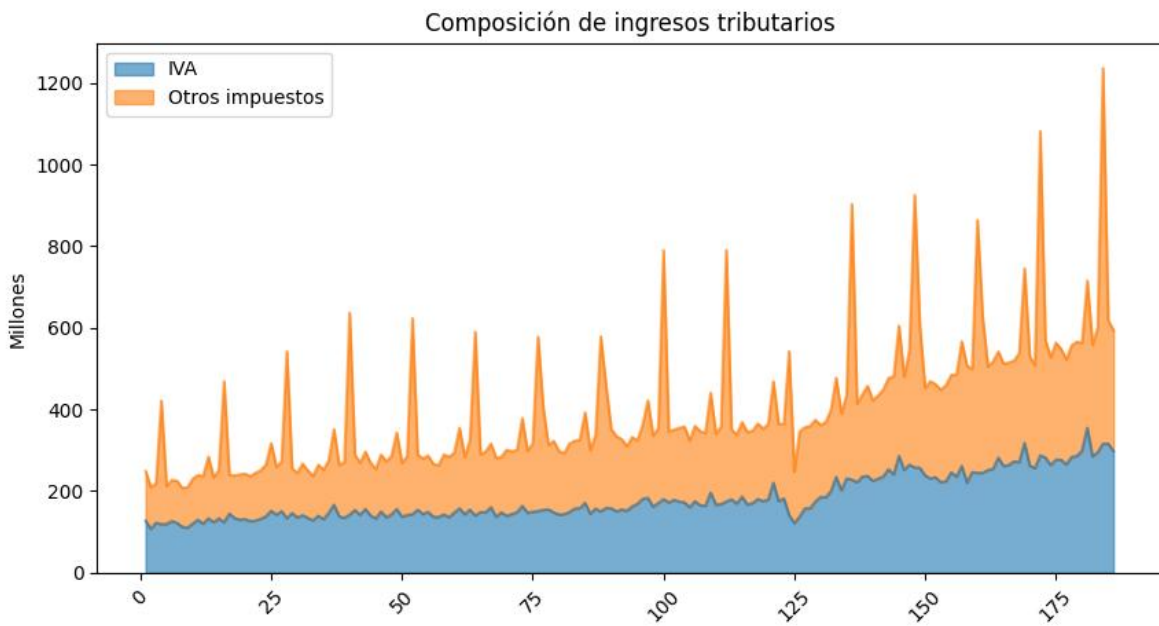


Ilustración 23: Relación entre IVA y los ingresos tributarios netos

Teniendo claro que los ingresos dependen cada vez más de los ingresos tributarios y estos a su vez del IVA, se analizó cómo es la relación entre los gastos corrientes y los ingresos corrientes, mostrando una tendencia creciente ambos, pero

siendo muy superiores los gastos de los ingresos, con lo que se concluye que las finanzas públicas muestran un déficit que no es puntual, sino persistente y no parece que vaya a revertirse en el corto o mediano plazo (ver figura 24).

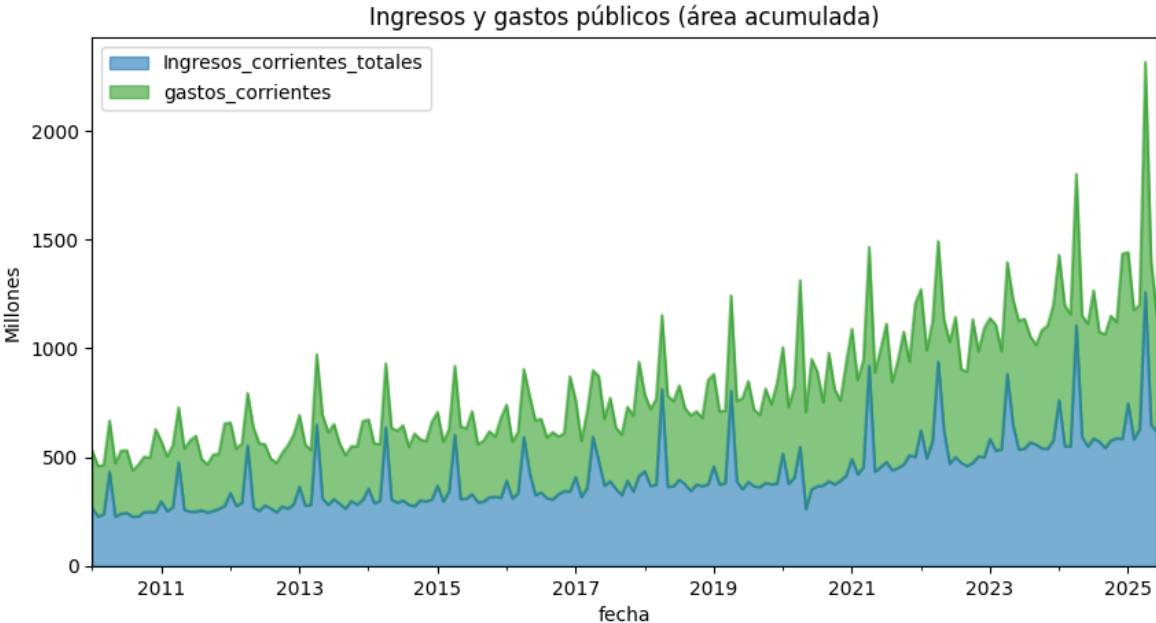


Figura 24: Comparación de los ingresos corrientes y los gastos corrientes

Para profundizar sobre el déficit o superávit y en la relación entre los ingresos y gastos, se hizo un análisis del comportamiento del superávit o del déficit, confirmando que el déficit no es algo ocasional, luego se analizó los crecimientos de los ingresos y gastos, notando que aunque en algunos periodos el crecimiento de los ingresos es mayor al de los gastos para los últimos periodos el crecimiento es muy similar, luego, si se analiza la correlación entre gastos e ingresos, se ve que en general hay una correlación positiva fuerte, lo que indica que si los ingresos crecen también lo hacen gastos, manteniendo así el déficit (ver las figuras 25 a la 28).

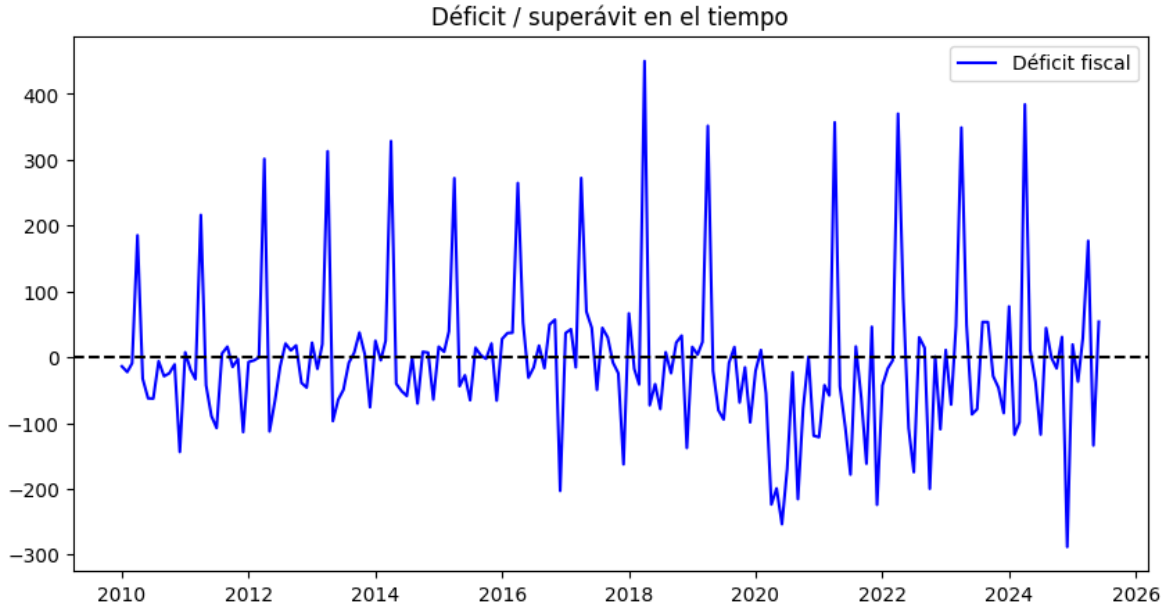


Figura 25: Comportamiento histórico del déficit fiscal

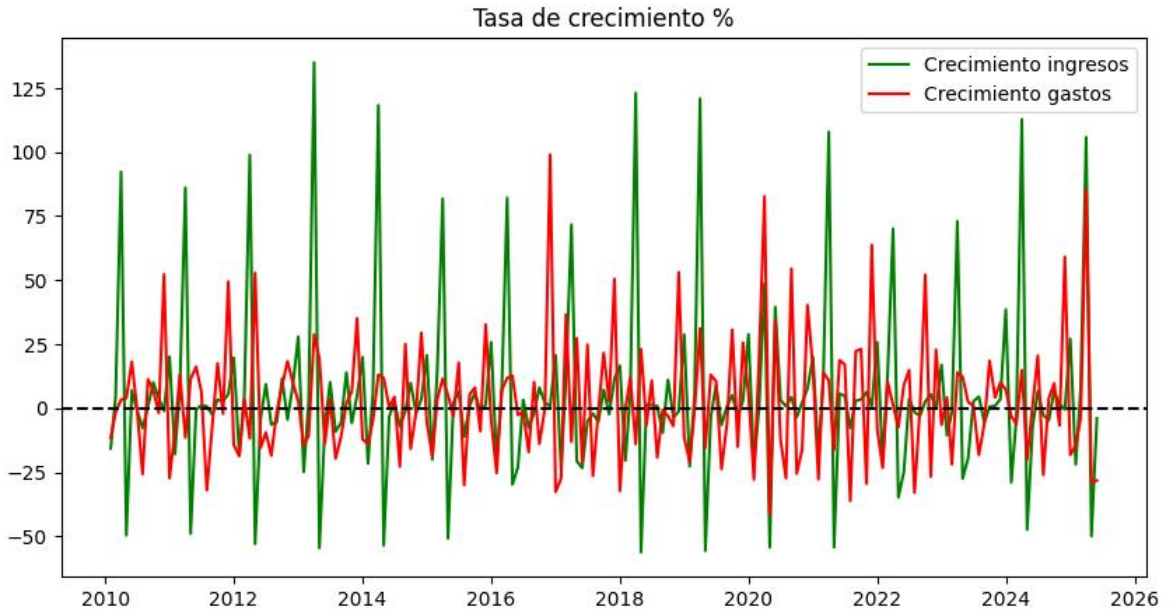


Figura 26: Comportamiento de los crecimientos de los ingresos y gastos

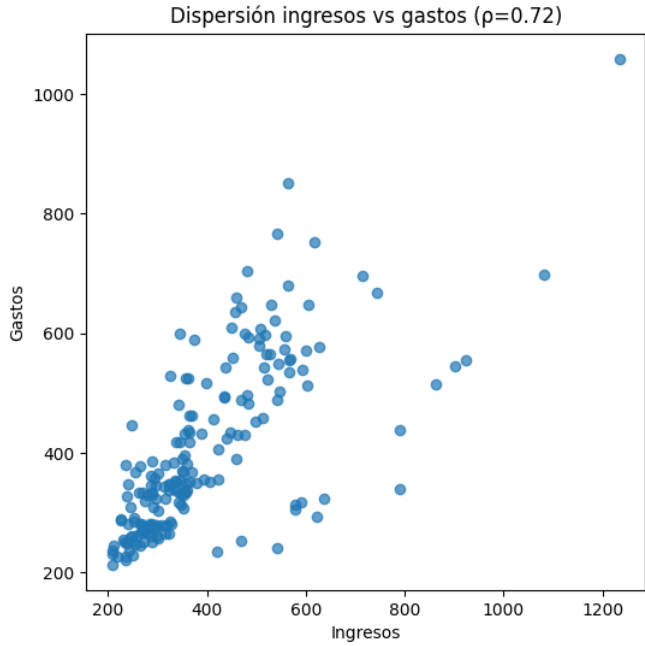


Figura 27: Diagrama de dispersión de los ingresos vs los gastos

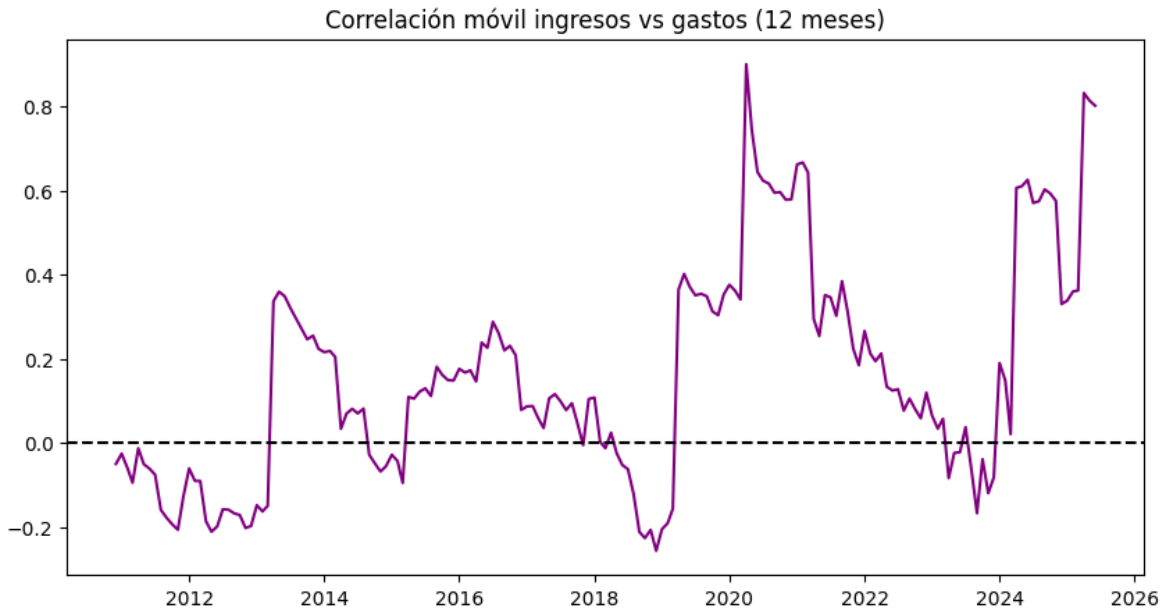


Figura 28: Análisis de la correlación para los ingresos y gastos anuales

4.2.2 Primer modelo ARDL

Para el primer modelo se toma como Y a los ingresos tributarios netos mensuales, las variables predictoras fueron:

- IVA mensual.
- Inflación mensual, la cual se calculó a partir del IPC.
- Remesas familiares mensuales.
- IVA / ingresos tributarios netos.

Todas las variables se dividieron entre el IVAE que usó como índice de normalización o escalado, los resultados del modelo se muestran a continuación, ver figura 29.

ARDL Model Results						
=====						
Dep. Variable:	itn_iv12	No. Observations:	175			
Model:	ARDL(3, 3, 3, 3)	Log Likelihood	377.518			
Method:	Conditional MLE	S.D. of innovations	0.027			
Date:	Sun, 07 Sep 2025	AIC	-721.036			
Time:	18:56:32	BIC	-667.529			
Sample:	3	HQIC	-699.327			
	175					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.0619	0.049	1.274	0.204	-0.034	0.158
itn_iv12.L1	0.1501	0.076	1.984	0.049	0.001	0.300
itn_iv12.L2	0.1623	0.072	2.269	0.025	0.021	0.304
itn_iv12.L3	0.3651	0.073	5.017	0.000	0.221	0.509
rem_iv12.L0	0.1791	0.061	2.955	0.004	0.059	0.299
rem_iv12.L1	0.4531	0.065	6.931	0.000	0.324	0.582
rem_iv12.L2	-0.1903	0.070	-2.704	0.008	-0.329	-0.051
rem_iv12.L3	-0.1921	0.077	-2.495	0.014	-0.344	-0.040
iva_itn.L0	-0.9259	0.036	-25.719	0.000	-0.997	-0.855
iva_itn.L1	0.1831	0.081	2.253	0.026	0.023	0.344
iva_itn.L2	0.2622	0.079	3.322	0.001	0.106	0.418
iva_itn.L3	0.3743	0.080	4.692	0.000	0.217	0.532
inflacion_mes.L0	-0.0064	0.006	-1.147	0.253	-0.017	0.005
inflacion_mes.L1	0.0040	0.006	0.693	0.490	-0.007	0.016
inflacion_mes.L2	-0.0017	0.006	-0.292	0.770	-0.013	0.010
inflacion_mes.L3	-0.0025	0.006	-0.440	0.660	-0.014	0.009
=====						

Ilustración 29: Resultados modelo ARDL para los ingresos tributarios

Lo primero que se destaca son los signos fuertes en los lags (rezagos) de remesas: L0 (+0.179, $p < 0.01$) y L1 (+0.453, $p < 0.001$), con lo que se puede concluir que las remesas incrementan los ingresos tributarios netos en el corto plazo. Pero como L2 y L3 son negativos, se puede decir que hay un efecto de ajuste, es decir, que lo que entra primero como consumo e impuestos se desvanece o incluso revierte en los meses siguientes.

Para el caso del IVA el coeficiente contemporáneo (L0) es negativo muy fuerte (-0.92, $p < 0.001$), lo cual confirma que, la recaudación depende mucho del IVA y el ingreso neto total es más frágil. Lags positivos sugieren que el efecto se

compensa en meses posteriores. Y para el caso de la Inflación, la conclusión es que su impacto no es significativo.

4.2.3 Segundo Modelo ARDL

Para este segundo modelo la variable Y es gasto corriente, y el proceso es totalmente análogo al modelo 1 con las mismas variables predictoras, ver figura 30 con los resultados del modelo.

ARDL Model Results						
=====						
Dep. Variable:	gc_iv12	No. Observations:	175			
Model:	ARDL(3, 3, 3, 3)	Log Likelihood	270.649			
Method:	Conditional MLE	S.D. of innovations	0.050			
Date:	Sun, 07 Sep 2025	AIC	-499.298			
Time:	18:56:32	BIC	-433.201			
Sample:	3	HQIC	-472.481			
	175					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.0618	0.066	0.943	0.347	-0.068	0.191
gc_iv12.L1	0.1249	0.079	1.572	0.118	-0.032	0.282
gc_iv12.L2	-0.0406	0.083	-0.487	0.627	-0.205	0.124
gc_iv12.L3	0.1648	0.090	1.827	0.070	-0.013	0.343
rem_iv12.L0	0.3331	0.129	2.584	0.011	0.078	0.588
rem_iv12.L1	-0.1191	0.173	-0.687	0.493	-0.462	0.223
rem_iv12.L2	0.4426	0.178	2.492	0.014	0.092	0.793
rem_iv12.L3	-0.4344	0.173	-2.509	0.013	-0.776	-0.092
iva_iv12.L0	1.7395	0.580	2.998	0.003	0.593	2.886
iva_iv12.L1	-1.5398	0.618	-2.491	0.014	-2.761	-0.319
iva_iv12.L2	-0.2639	0.565	-0.467	0.641	-1.381	0.853
iva_iv12.L3	1.2044	0.468	2.575	0.011	0.280	2.129
inflacion_mes.L0	-0.0081	0.011	-0.767	0.444	-0.029	0.013
inflacion_mes.L1	-0.0059	0.011	-0.539	0.591	-0.028	0.016
inflacion_mes.L2	-0.0008	0.011	-0.073	0.942	-0.023	0.021
inflacion_mes.L3	0.0049	0.011	0.450	0.654	-0.017	0.026

Ilustración 30: resultados del modelo ARDL que predice el gasto corriente

Las remesas muestran un efecto contemporáneo positivo (+0.33, $p < 0.01$), es decir, más remesas coinciden con más gasto, pero los lags tienen signos cambiantes (positivo en L2, negativo en L3), lo que podría indicar un efecto procíclico o inconsistente, se podría decir que el gasto no responde linealmente a

las remesas. Para el caso del IVA, este es significativo en varios lags, por lo que parece influir en el gasto, aunque con signos mixtos, lo que podría indicar que no hay una relación lineal.

4.2.3 Tercer modelo ARDL

Hasta este punto se puede decir que hay evidencia de un impacto claro de las remesas y el IVA en los ingresos, y un impacto importante sobre todo del IVA en el gasto, por tanto, lo que sigue es comprobar el nivel del impacto de las remesas en el IVA, para poder sacar conclusiones robustas sobre el impacto de las remesas directa o indirectamente en los ingresos y los gastos.

ARDL Model Results						
=====						
Dep. Variable:	iva_itn	No. Observations:	175			
Model:	ARDL(3, 3)	Log Likelihood	231.108			
Method:	Conditional MLE	S.D. of innovations	0.063			
Date:	Sun, 07 Sep 2025	AIC	-444.217			
Time:	18:56:33	BIC	-415.890			
Sample:	3	HQIC	-432.724			
	175					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.4570	0.061	7.443	0.000	0.336	0.578
iva_itn.L1	-0.0714	0.073	-0.985	0.326	-0.215	0.072
iva_itn.L2	-0.0695	0.066	-1.050	0.295	-0.200	0.061
iva_itn.L3	0.1387	0.067	2.068	0.040	0.006	0.271
rem_iv12.L0	-0.3103	0.130	-2.395	0.018	-0.566	-0.054
rem_iv12.L1	-0.8573	0.127	-6.737	0.000	-1.109	-0.606
rem_iv12.L2	0.4542	0.135	3.374	0.001	0.188	0.720
rem_iv12.L3	0.7755	0.142	5.481	0.000	0.496	1.055
=====						

Figura 31: Modelo que relaciona remesas con Iva

```

Granger Causality
number of lags (no zero) 1
ssr based F test:      F=1.9340 , p=0.1661 , df_denom=171, df_num=1
ssr based chi2 test:  chi2=1.9680 , p=0.1607 , df=1
likelihood ratio test: chi2=1.9569 , p=0.1618 , df=1
parameter F test:     F=1.9340 , p=0.1661 , df_denom=171, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:      F=17.0835 , p=0.0000 , df_denom=168, df_num=2
ssr based chi2 test:  chi2=35.1840 , p=0.0000 , df=2
likelihood ratio test: chi2=32.0276 , p=0.0000 , df=2
parameter F test:     F=17.0835 , p=0.0000 , df_denom=168, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:      F=22.3789 , p=0.0000 , df_denom=165, df_num=3
ssr based chi2 test:  chi2=69.9849 , p=0.0000 , df=3
likelihood ratio test: chi2=58.7175 , p=0.0000 , df=3
parameter F test:     F=22.3789 , p=0.0000 , df_denom=165, df_num=3

```

Ilustración 32: resumen de prueba de causalidad de Granger de las remesas sobre el Iva

Con esto se vuelve evidente que para los Lags 2 y 3, las remesas son muy significativas para el IVA, con lo que queda muy claro el impacto directo e indirecto de las remesas en los ingresos tributarios y por tanto, en los ingresos corrientes, y de igual manera en el gasto corriente. Para visualizar de forma más clara el impacto de las remesas en los aspectos claves de las finanzas públicas que se han definido en este trabajo (ingresos tributarios netos y gastos corrientes) se pueden ver las figuras 33 y 34.

Los gráficos que se presentan en la última parte de esta sección son de tipo impulso-respuesta, primero de las remesas sobre los ingresos y en el segundo de las remesas sobre los gastos, en los cuales se puede ver que las remesas impactan con mayor fuerza los ingresos sobre todo en la parte inicial, para los gastos el impacto es menor, pero es consistente con el impacto en los ingresos.

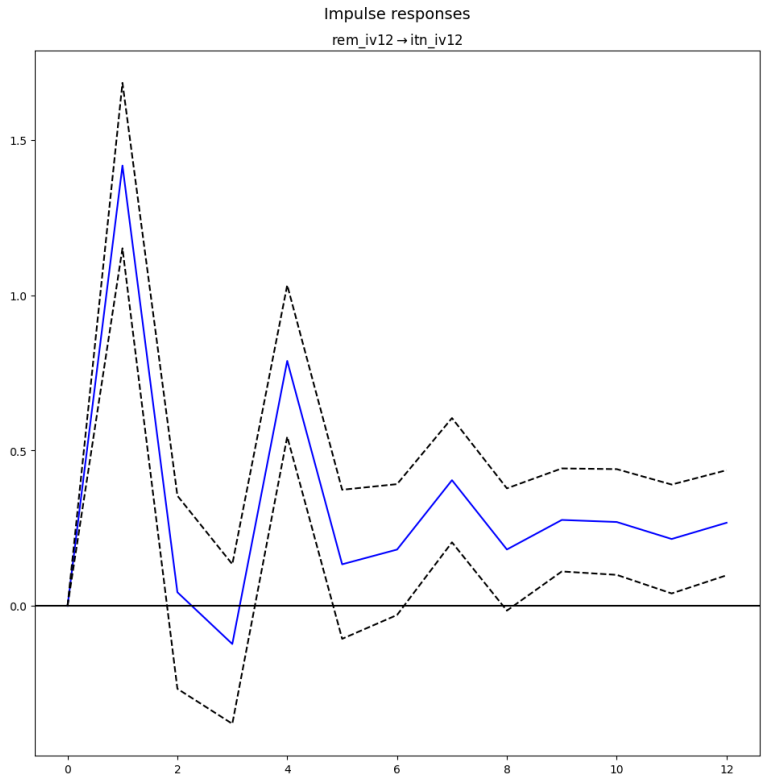


Ilustración 33: Gráfico de impulso-respuesta de las remesas sobre los ingresos tributarios netos

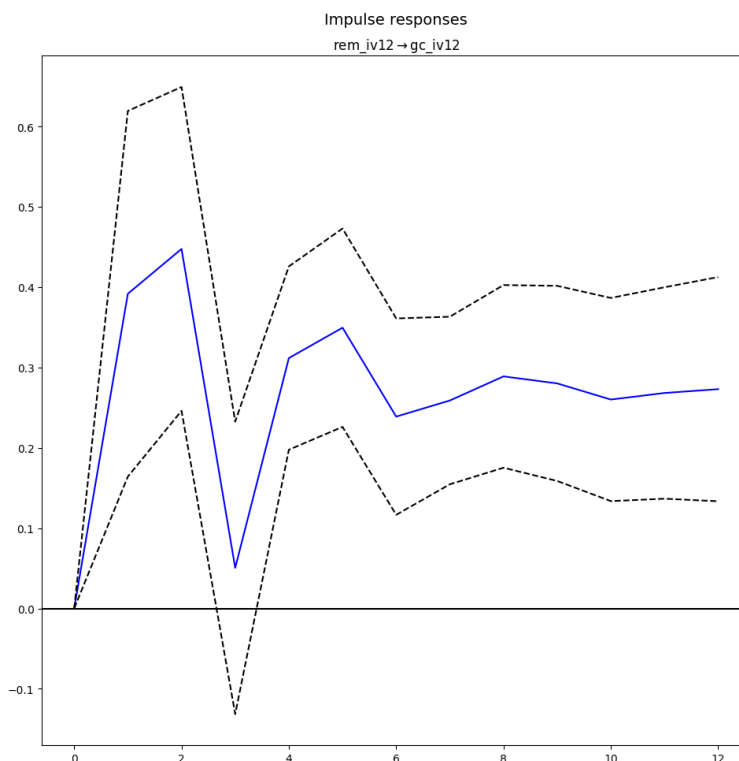


Ilustración 34: Gráfico de impulso-respuesta de las remesas sobre el gasto corriente

4.2.4 Análisis de impacto y proyecciones

Hasta este punto está claro que las remesas tienen un impacto muy significativo tanto en los ingresos como en los gastos, por ello, se siguió con un análisis de impacto usando proyecciones y escenarios, para esto, se afinaron los modelos ARDL para que este análisis fuese lo más realista posible.

Se hizo un diagnóstico de los modelos y luego se ajustaron sus parámetros para mejorar su poder predictivo y simplificarlos, quitando variables poco significativas, dando como resultado a los tres modelos siguientes:

```
🏆 MEJOR MODELO PARA itn_iv12:  
Variables: ['rem_iv12']  
Orders: [2]  
AIC: -315.89  
Error último: 0.1%  
Buscando mejor modelo para gc_iv12...
```

```
🏆 MEJOR MODELO PARA gc_iv12:  
Variables: ['rem_iv12']  
Orders: [2]  
AIC: -478.21  
Error último: 13.4%  
Buscando mejor modelo para ratio_itn_gc...  
🏆 MEJOR MODELO PARA ratio_itn_gc:  
Variables: ['rem_iv12']  
Orders: [1]  
AIC: 70.55  
Error último: 8.0%
```

Como se puede observar, los modelos finales muestran errores en sus predicciones por debajo del 14%, siendo el primero con un mejor resultado y un error del 0.1%, estos modelos mostraron que la variable principal predictora es el flujo de remesas, por tanto, para este apartado se descartó el uso de la inflación y el IVA.

Luego, se realizó un análisis del impacto de las remesas tanto en el corto como en el largo plazo, para esto, se toma el factor de largo plazo de un modelo ARDL (LR), el efecto acumulado de corto plazo q y su beta.

```
LR ITN: {'beta_sum_short_run': 0.4658049200619329, 'phi_sum': 0.036981542984068964, 'LR': 0.4836926194595533}  
LR GC: {'beta_sum_short_run': 0.5582658367423294, 'phi_sum': 0.15734334125554184, 'LR': 0.6625068833777858}  
LR RATIO: {'beta_sum_short_run': 0.0033076023191185855, 'phi_sum': 0.004201498541202197, 'LR': 0.0033215578395359145}
```

Impacto de las remesas sobre el ingreso

- El impacto acumulado de corto plazo (0.466) indica que un incremento de 1% en las remesas está asociado a un aumento de 0.47 % en el ratio de ingresos tributarios sobre actividad económica en los siguientes 1 o 2 meses.
- El multiplicador de largo plazo (0.484) es casi idéntico, lo que implica que la mayor parte del efecto se mantiene (no se disipa con el tiempo).
- El coeficiente $\phi_sum \approx 0.04$ sugiere una dinámica bastante estable, la recaudación responde rápido y no muestra excesiva persistencia.

Se puede confirmar que las remesas tienen un efecto positivo y persistente sobre la recaudación tributaria, consolidando la hipótesis de que el flujo de remesas aumenta la base imponible y mejora el esfuerzo fiscal en el corto y mediano plazo.

Impacto de las remesas sobre el gasto

- El impacto de corto plazo (0.56) es aún mayor que en ingresos. Cada 1 % adicional en remesas está asociado a un aumento de ~ 0.56 % en el gasto corriente relativo a la actividad económica.
- El multiplicador de largo plazo (0.66) también es positivo y más alto que el de los ingresos. Es decir, las remesas inducen un aumento persistente del gasto público.
- La persistencia dinámica ($\phi_sum \approx 0.16$) es moderada, indicando ajuste relativamente rápido.

Los modelos capturan de forma muy positiva el efecto de las remesas sobre los ingresos y los gastos, para cerrar con esa idea, a continuación, se presentan

gráficos de las series y los valores generados por los modelos.

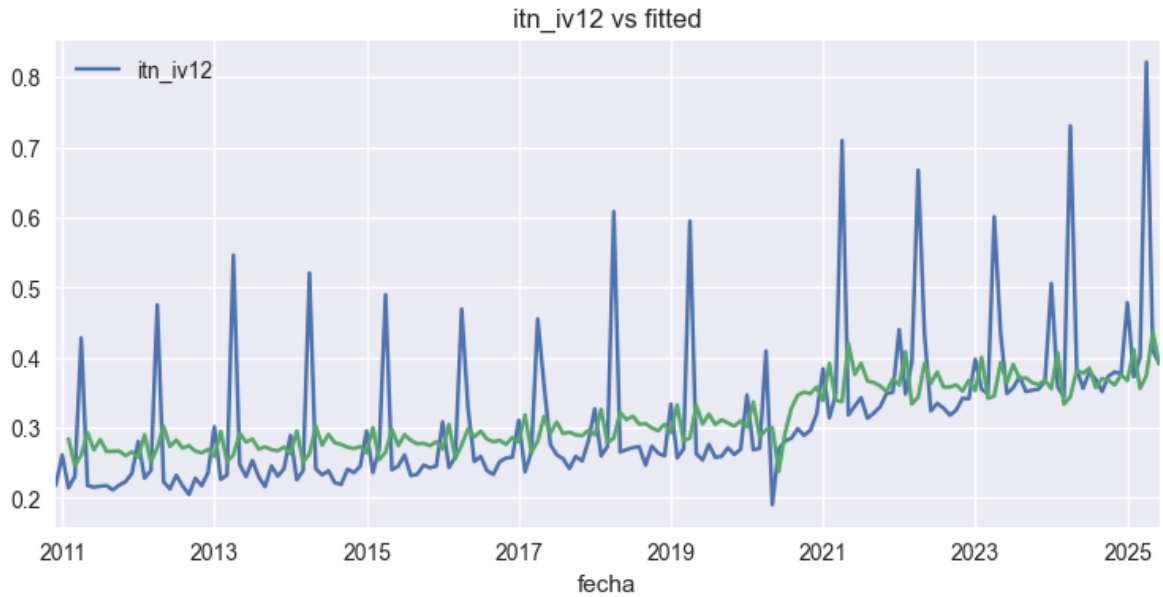


Figura 35: gráfico de las series de tiempo del ingreso neto sobre el IVAE anualizado real vs predicción

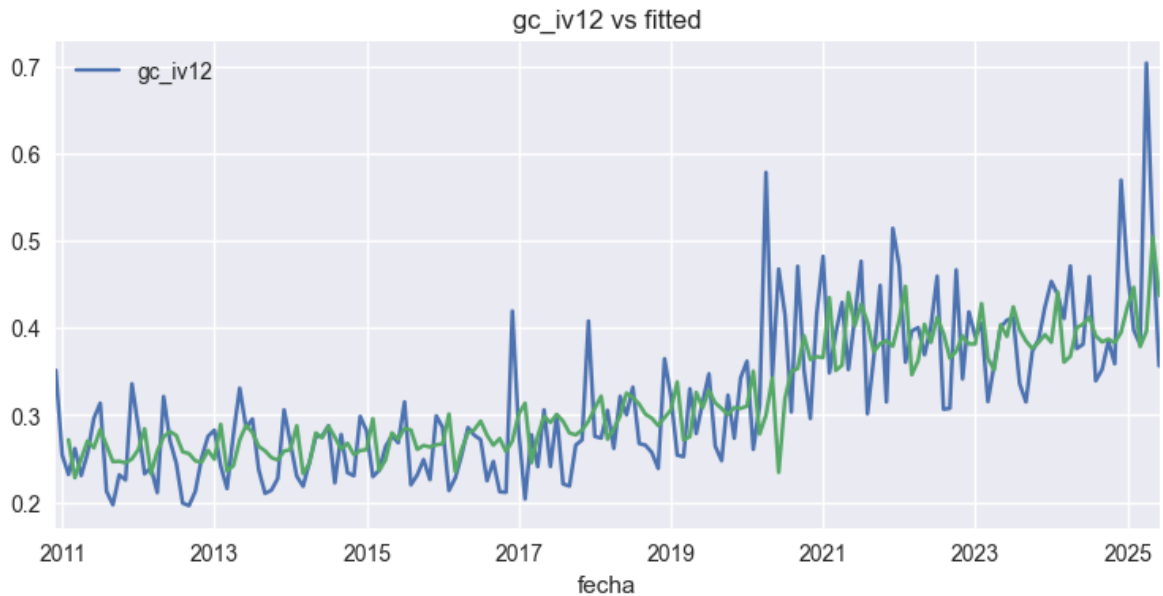


Figura 36: gráfico de las series de tiempo del gasto corriente sobre el IVAE anualizado real vs predicción

Se observa que los modelos generan curvas más suavizadas, pero logran capturar las tendencias, filtrando el ruido estacional, modelando la relación estructural y no los saltos mensuales puntuales.

Proyecciones a 6 meses

Para este apartado se tomó un horizonte de 6 meses a partir del último mes (para el caso de este trabajo fue julio de 2025), asumiendo un escenario donde las remesas aumentan un 10%, mostrando los siguientes resultados:

```
ITN_IV12
Base: mean=0.4152, path=['0.4245', '0.4140', '0.4131', '0.4132', '0.4132', '0.4132']
Shock PERM: mean=0.4334, Δmean=0.0183
Shock TEMP: mean=0.4290, Δmean=0.0138

GC_IV12
Base: mean=0.4538, path=['0.4502', '0.4558', '0.4546', '0.4542', '0.4542', '0.4542']
Shock PERM: mean=0.4780, Δmean=0.0242
Shock TEMP: mean=0.4727, Δmean=0.0189
```

Para este análisis se tomó como shock permanente un incremento del 10% en los flujos de remesas y como escenario temporal el mismo incremento, pero sólo para 3 meses, teniendo como resultado que un aumento del 10% en remesas eleva el ITN/IVAE12 en torno a un **4.4%** y el GC/IVAE12 en un **5.3%** en promedio durante los 6 meses posteriores; bajo un shock transitorio de 3 meses los efectos son **3.3%** y **4.2%**, respectivamente. Lo que indica que las remesas **expanden simultáneamente ingresos y gasto**, sin mejorar de forma apreciable la

sostenibilidad fiscal. A 6 meses se materializan dos terceras partes **del impacto de largo plazo**, coherente con la dinámica ARDL estimada.

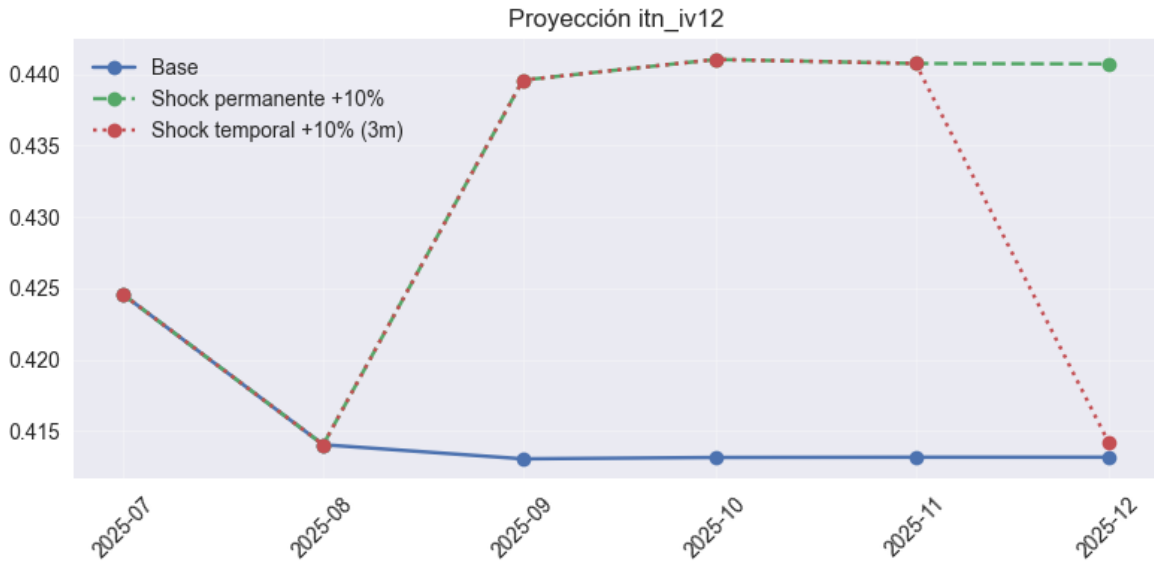


Figura 37: proyecciones del indicador de ingresos neto con un shock de +10% de remesas

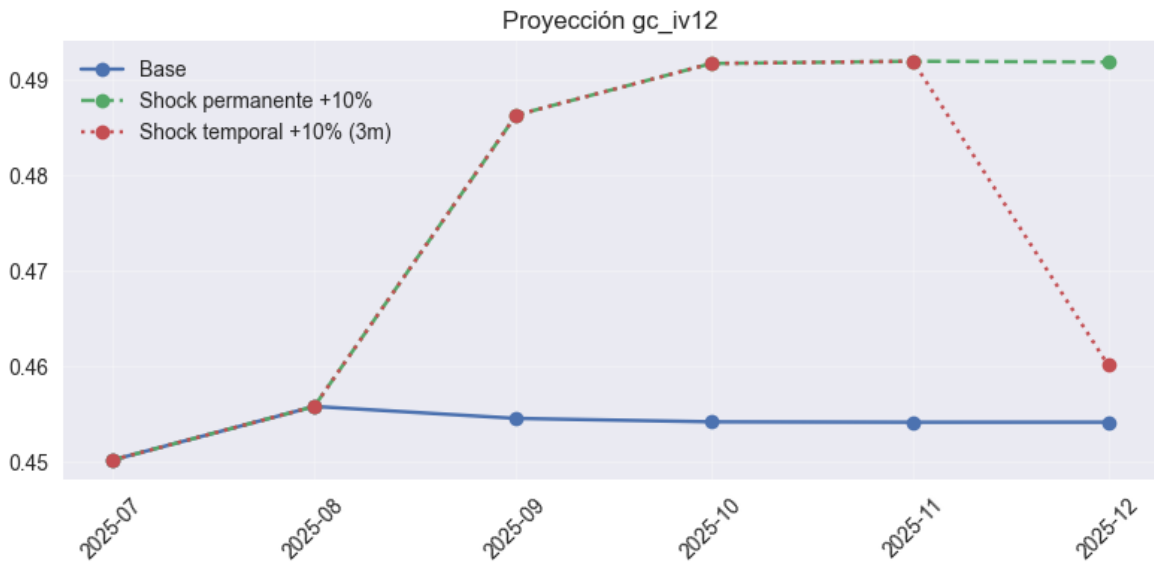


Figura 38: proyecciones del indicador de gastos corrientes con un shock de +10% de remesas

Conclusiones

Con el desarrollo de este trabajo se ha logrado determinar el impacto de los flujos de remesas en las finanzas públicas, en específico en los ingresos tributarios netos, pero también se puede concluir que las remesas impactan de forma directa a los gastos corrientes del gobierno, con los cuales se financian la gran mayoría de instituciones del estado.

Se logró analizar y describir la forma en que se da el impacto de las remesas sobre los ingresos y los gastos, siendo este en el corto plazo, logrando el desarrollo de modelos con un comportamiento aceptable, que permiten la realización de escenarios de shock, y con ello se determinó que la variable remesa se impone en importancia sobre el IVA y la inflación en importancia sobre los ingresos y gastos.

Para el caso de las predicciones sobre los flujos de remesas se logró el desarrollo de un modelo con un funcionamiento muy aceptable, con una leve sobreestimación para el caso de El Salvador, pero en términos generales con un buen rendimiento y destacable poder predictivo.

Se identificaron cuales variables o indicadores son los que más afectan la predicción de los flujos de remesas, siendo las más importantes relacionadas con la economía de Estados Unidos.

Se identifico cuales son las variables no económicas más importantes para la predicción de los flujos de remesas, dentro de ellas, las principales están relacionadas con la capacidad de producción agrícola, con la inmigración de Estados Unidos (específicamente los inmigrantes expulsados) y finalmente la tasa

de homicidios por cada cien mil habitantes, aunque esta última, mostró una importancia muy baja en comparación con otras variables predictoras.

El área de las finanzas públicas que más se ve impactada por los flujos de remesas es el ingreso corriente del estado, el cual particularmente depende de los ingresos tributarios y, por tanto, del IVA, lo que pone en riesgo el funcionamiento de casi todas las instituciones del estado en caso de fluctuaciones grandes y negativas de los flujos de remesas.

Recomendaciones

Se recomienda una **estrategia administrativa** (que coincide con sugerencias de organismos internacionales) de disminución del déficit, es decir, una disminución sistemática en el gasto, a la vez de una mejor canalización del gasto privado hacia la recaudación tributaria, junto con políticas que permitan y mejoren el envío y recepción de remesas de forma que estas impliquen menos costos para las familias, finalmente se podrían implementar planes que ayuden a generar inversión privada o emprendimientos para las familias receptoras de remesas con apoyo burocrático, planes de financiamiento o incluso con excepciones fiscales.

Para el caso de modelos de Machine Learning el principal desafío consiste en el tamaño limitado de los datos para cualquier selección de variables, por lo que se podría implementar procesos para homologar los datos de múltiples países, aunque no tengan características tan similares, con lo que se podría extender el conjunto de datos para otros países de América, o incluso otros continentes, siempre que tengan flujos de remesas significativos desde EEUU.

Los modelos ARDL en este trabajo se usaron con el único objetivo de analizar el impacto de los flujos de remesas en las finanzas públicas, sin embargo, hay aspectos sobre cómo se da el impacto, qué rezagos son los más importantes y sobre el tipo de relación que describen las variables de entrada, que podrían ser estudiados a profundidad y de una forma más completa en futuros trabajos.

Referencias

- Acosta, P., Calderón, C., Fajnzylber, P., & López, H. (2008). What is the impact of international remittances on poverty and inequality in Latin America? *World Development*, 36(1), 89–114.
<https://doi.org/10.1016/j.worlddev.2007.02.016>
- Amuedo-Dorantes, C., & Pozo, S. (2004). Workers' remittances and the real exchange rate: A paradox of gifts. *World Development*, 32(8), 1407–1417.
<https://doi.org/10.1016/j.worlddev.2004.02.003>
- Banco Central de Reserva de El Salvador. (2023). Informe anual 2023.
<https://www.bcr.gob.sv>
- Banco Central de Reserva de El Salvador. (2023). *Índice de Volumen de Actividad Económica (IVAE): Metodología y resultados*. San Salvador: BCR.
Recuperado de
<https://www.bcr.gob.sv/>
- Banco Central de Reserva de El Salvador (BCR). (2023). Informe de remesas familiares 2023. San Salvador: BCR.
- Banco Central de Reserva de El Salvador (BCR). (2024). Indicadores macroeconómicos y series estadísticas. Recuperado de
<https://www.bcr.gob.sv>
- Banco Mundial. (2024). Migration and Development Brief 38: Resilience through diversification. World Bank.
- Fondo Monetario Internacional (FMI). (2022). Manual de estadísticas de finanzas públicas. Washington, D.C.: FMI.

- Ministerio de Hacienda de El Salvador. (2024). Informe de ingresos y gastos del sector público no financiero. San Salvador: MH.
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2023). Revenue Statistics in Latin America and the Caribbean 2023. Paris: OECD Publishing.
- Banco Central de Reserva de El Salvador. (2022). *Informe de política económica y social*. San Salvador: BCR.
- Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). *A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities*. Journal of Internet Services and Applications, 9(1), 16.
<https://doi.org/10.1186/s13174-018-0087-2>
- Becerra, G. (2023). Árboles de decisión y Random Forest. En Ensambladores: Random Forest - Parte I. Recuperado el 4 de octubre de 2025, de <https://bookdown.org/content/2031/ensambladores-random-forest-parte-i.html>
- Banco Central de Honduras. (2025). Informe Encuesta de Remesas Familiares Enero 2025. <https://www.bch.hn/estadisticas-y-publicaciones-economicas/sector-externo/informes-y-publicaciones/resultado-de-encuesta-semestral-de-remesas-familiares>
- Banco Central República Dominicana. (2024). Banco Central informa que los flujos de remesas alcanzaron los US\$2,635.6 millones en el primer trimestre de 2024. <https://www.bancentral.gov.do/a/d/5925-banco-central->

[informa-que-los-flujos-de-remesas-alcanzaron-los-us26356-millones-en-el-primer-trimestre-de-2024](#)

- Banco Mundial. (2022). Migration and Development Brief 36.
<https://www.worldbank.org/en/topic/migrationremittancesdiasporaissues>
- Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Chami, R., Fullenkamp, C., & Jahjah, S. (2003). Are immigrant remittance flows a source of capital for development? (IMF Working Paper No. 03/189). International Monetary Fund. <https://doi.org/10.5089/9781451859372.001>
- Chollet, F. (2018). *Deep learning with Python* (2nd ed.). Manning Publications.
- Durand, J., Massey, D. S., & Parrado, E. A. (1996). Migradollars and development: A reconsideration of the Mexican case. *International Migration Review*, 30(2), 423–444. <https://doi.org/10.1177/019791839603000201>
- FUSADES. (2021). *Impacto económico de las remesas familiares en El Salvador*. Fundación Salvadoreña para el Desarrollo Económico y Social.
- Gujarati, D. N., & Porter, D. C. (2010). *Econometría*. McGraw-Hill.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Econometrics*, 100(2), 289-326. *Journal of Econometrics*.

- Ratha, D. (2013). The impact of remittances on economic growth and poverty reduction. Migration Policy Institute. <https://www.migrationpolicy.org>
- Peña, D. (2005). Análisis de Series Temporales. Alianza.
- Ezequiel Uriel Jiménez. (1985). Análisis de Series Temporales modelos ARIMA. Paraninfo SA

Anexos

A continuación, se presentan algunos de los casos de uso más conocidos en finanzas, de tecnologías de tipo Machine Learning e Inteligencia Artificial y que empresas son las que los implementan.

SCORING CREDITICIO Y EVALUACIÓN DE RIESGOS



Modelos predictivos basados en ML analizan múltiples variables para estimar la probabilidad de impago, mejorando la precisión frente a métodos tradicionales.

 **FICO FICO BBVA**

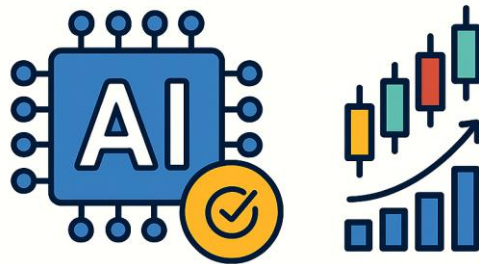


DETECCIÓN DE FRAUDES Y ANOMALÍAS

Algoritmos de ML analizan transacciones en tiempo real para identificar patrones sospechosos.

PayPal, Wells Fargo, Capital One

TRADING ALGORÍTMICO Y GESTIÓN DE PORTAFOLIOS



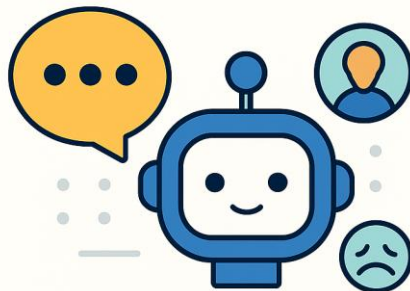
Algoritmos de alta frecuencia y modelos predictivos optimizan la ejecución de órdenes y la composición de carteras.

JPMorgan Chase

Goldman Sachs



CITADEL



ATENCIÓN AL CLIENTE Y ASESORÍA PERSONALIZADA

Chatbots y asistentes virtuales ofrecen soporte 24/7 y recomendaciones personalizadas.

Instituciones: Bank of America (Erica), BBVA, ING