



Facultad de Ciencias Económicas y Empresariales
ICADE

Predicción del Déficit Fiscal de Estados Unidos con Machine Learning: XGBoost, LSTM-CNN y Combinaciones Dinámicas

Autor: María Teresa Piergili de la Escalada
Director: Peter Guenther Antoon Claeys

MADRID | Diciembre 2024

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

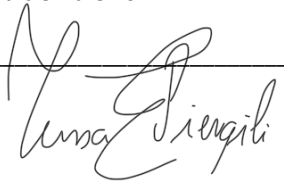
Por la presente, yo, **MARÍA TERESA PIERGILI DE LA ESCALADA**, estudiante de **MASTER DE INGENIERIA INDUSTRIAL Y ADE** de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado **PREDICCIÓN DEL DÉFICIT FISCAL DE ESTADOS UNIDOS MEDIANTE MACHINE LEARNING: XGBOOST, LSTM-CNN Y COMBINACIONES DINÁMICAS**, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 1 de diciembre 2024

Firma: _____





Facultad de Ciencias Económicas y Empresariales
ICADE

Predicción del Déficit Fiscal de Estados Unidos con Machine Learning: XGBoost, LSTM-CNN y Combinaciones Dinámicas

Autor: María Teresa Piergili de la Escalada
Director: Peter Guenther Antoon Claeys

MADRID | Diciembre 2024

PREDICCIÓN DEL DÉFICIT FISCAL DE ESTADOS UNIDOS CON MACHINE LEARNING: XGBOOST, LSTM-CNN Y COMBINACIONES DINÁMICAS

Autor: Piergili de la Escalada, María Teresa

Director: Claeys, Peter Antoon Guenther

RESUMEN DEL PROYECTO

Este estudio presenta un marco metodológico innovador para la predicción del déficit fiscal anual de Estados Unidos, abordando las limitaciones y sesgos inherentes a las previsiones realizadas por instituciones financieras. A través de la aplicación de modelos de aprendizaje automático y técnicas avanzadas de combinación, se busca superar los enfoques tradicionales, proporcionando predicciones más precisas y dinámicas. En particular, se proponen métodos de combinación basados en criterios como AIC, BIC y BMA, destacando especialmente una ponderación dinámica automatizada mediante el estadístico Theil's U, diseñado específicamente en este trabajo como una métrica clave para optimizar la precisión.

La metodología integra dos modelos avanzados, XGBoost y LSTM-CNN, cuyos *forecast errors* son evaluados mediante métricas y pruebas estadísticas para validar su robustez. El análisis destaca que el modelo combinado ponderado por pesos dinámicos, impulsado por Theil's U, emerge como el enfoque más sólido, logrando un valor mínimo de Theil's U del 9%. Este resultado supera tanto a los modelos individuales como a las instituciones financieras, especialmente en periodos caracterizados por alta volatilidad económica.

Este trabajo no solo demuestra la efectividad de los enfoques combinados, sino que establece un marco replicable y adaptable para futuros estudios en previsión económica. Los resultados subrayan la capacidad del modelo basado en Theil's U para capturar patrones dinámicos y complejos, ofreciendo una solución robusta frente a desafíos económicos. Así, se consolida la relevancia de adoptar metodologías avanzadas en contextos de incertidumbre económica, posicionando esta investigación como un aporte significativo en el campo de la modelización fiscal.

Palabras clave: Déficit Fiscal, XGBoost, LSTM-CNN, Theil's U, Pesos Dinámicos

PREDICTING THE US FISCAL DEFICIT WITH MACHINE LEARNING: XGBOOST, LSTM-CNN AND DYNAMIC COMBINATIONS

Autor: Piergili de la Escalada, María Teresa

Director: Claeys, Peter Antoon Guenther

ABSTRACT

This study presents an innovative methodological framework for the US annual fiscal deficit prediction, addressing the limitations and inherent biases of forecasts produced by financial institutions. Through the application of machine learning models and advanced combination techniques, this research aims to surpass traditional approaches by providing more accurate and adaptable predictions. Specifically, it proposes combination methods based on criteria such as AIC, BIC, and BMA, with a particular focus on a dynamic weighting mechanism automated through Theil's U statistic, specifically designed in this study as a key metric to optimize precision.

The methodology integrates two advanced models, XGBoost and LSTM-CNN, whose residuals are evaluated through metrics and statistical tests to validate their robustness. The analysis highlights that the combined model weighted by dynamic factors, driven by Theil's U, emerges as the most reliable approach, achieving a minimum Theil's U value of 9%. This performance surpasses both individual models and financial institutions, particularly during periods of high economic volatility.

This work not only demonstrates the effectiveness of combined approaches but also establishes a replicable and adaptable framework for future economic forecasting studies. The results underline the capacity of the Theil's U-based model to capture dynamic and complex patterns, offering a robust solution to economic challenges. Thus, this research reinforces the importance of adopting advanced methodologies in contexts of economic uncertainty, positioning itself as a significant contribution to the field of fiscal modelling.

Keywords: Fiscal Deficit, XGBoost, LSTM-CNN, Theil's U, Dynamic Weights

INDICE

1	INTRODUCCIÓN.....	1
1.1	Objetivos.....	3
1.2	Metodología.....	5
1.3	Fuentes de Información.....	6
1.4	Revisión de la Literatura	7
1.4.1	Superando Sesgos en las Previsiones Fiscales	7
1.4.2	Innovación en la Predicción del Déficit Fiscal	8
1.5	Impacto en la Sociedad.....	11
1.5.1	En línea con los Objetivos de Desarrollo Sostenible.....	12
2	DESCRIPCIÓN DE LOS DATOS.....	15
2.1	El Déficit Fiscal Histórico.....	15
2.1.1	Análisis Detallado del Déficit Fiscal	20
2.2	El Déficit Fiscal Predicho por Instituciones.....	23
3	DESCRIPCIÓN LAS METODOLOGÍAS.....	31
3.1	Modelos y Combinaciones.....	31
3.1.1	Modelo XGBoost	31
3.1.1.1	Función Objetivo: XGBoost	32
3.1.1.2	Método de Optimización	32
3.1.1.3	Estructura del Árbol	33
3.1.1.4	Regularización y Poda.....	33
3.1.1.5	Personalización.....	33
3.1.2	Modelo LSTM-CNN	34
3.1.2.1	Función Objetivo: LSTM-CNN	34
3.1.2.2	Método de Optimización	34
3.1.2.3	Estructura de las redes neuronales	35
3.1.2.4	Regularización	36
3.1.2.5	Personalización.....	37
3.1.3	Modelo Combinado: Ponderación por AIC	37
3.1.4	Modelo Combinado: Ponderación por BIC	37
3.1.5	Modelo Combinado: Ponderación por BMA	38
3.1.6	Modelo Combinado: Ponderación Dinámica por Theil's U	38
3.2	Métricas	39

3.2.1	Error Cuadrático Medio (MSE).....	39
3.2.2	Error Absoluto Medio (MAE).....	39
3.2.3	Error Raíz del Error Cuadrático Medio (RMSE).....	40
3.2.4	Coefficiente de Determinación (R^2).....	40
3.2.5	Autocorrelación (ACF y PACF).....	40
3.3	<i>Pruebas</i>	41
3.3.1	Prueba de Sesgo (<i>Bias Test</i>).....	41
3.3.2	Prueba de <i>Stickiness</i>	41
3.3.3	Prueba de Dickey-Fuller Aumentado (ADF).....	42
3.3.4	Prueba de HEGY.....	43
4	RESULTADOS.....	45
4.1	<i>Resultados de cada modelo</i>	46
4.1.1	XGBoost.....	47
4.1.1.1	Parámetros del XGBoost.....	47
4.1.1.2	Resultados del XGBoost.....	49
4.1.2	LSTM-CNN.....	52
4.1.2.1	Parámetros del LSTM-CNN.....	52
4.1.2.2	Resultados del LSTM-CNN.....	54
4.1.3	Análisis de los <i>forecast errors</i> : XGBoost y LSTM – CNN.....	56
4.1.4	Modelo Combinado.....	61
4.1.4.1	Parámetros del modelo combinado.....	61
4.1.4.2	Resultados del modelo combinado.....	62
4.1.5	<i>Forecast errors</i> de los modelos combinados.....	63
4.2	<i>Comparación con instituciones</i>	65
4.2.1	Instituciones frente al déficit.....	65
4.2.2	Modelos combinados frente instituciones y déficit real.....	68
5	CONCLUSIÓN.....	75
5.1	<i>Machine learning vs Instituciones Privadas</i>	75
6	BIBLIOGRAFÍA.....	79

INDICE DE FIGURAS

Figura 1. Déficit Fiscal de EE. UU. como porcentaje del PIB (anual, sin ajuste estacional)	16
Figura 2. Distribución del Déficit Fiscal sobre el PIB.....	19
Figura 3. Graficas de las pruebas ACF Y PACF según el retardo	23
Figura 4. Grafica de las predicciones de las tres instituciones seleccionadas (JPMorgan, First Trust y Econ).....	27
Figura 5. Diagramas de Cajas y Bigotes para JPMorgan, First Trust, Econ y el déficit real desde 1993.....	27
Figura 6. Predicciones anuales del modelo XGBoost (split 50%-50%).....	51
Figura 7. Comparación de las predicciones el déficit real (Train 60% - Test 40%).....	56
Figura 8. Representación gráfica de los forecast errors de XGBoost y LSTM-CNN desde 1993-2022.....	57
Figura 9. Graficas de la prueba KPSS: XGBoost y LSTM – CNN.....	58
Figura 10. Test de ARCH: XGBoost	59
Figura 11. Histograma de forecast errors: XGBoost y LSTM-CNN.....	59
Figura 12. Diagramas QQ de forecast errors: XGBoost y LSTM – CNN	60
Figura 13. Comparación de los modelos combinados frente a los valores reales.....	62
Figura 14. Grafica de los forecast errors de los modelos combinados.....	64
Figura 15. Comparación de predicciones: déficit real, instituciones y el mejor modelo combinado.....	72

INDICE DE TABLAS

Tabla 1. Estadísticas descripticas del déficit fiscal anual de EE. UU.	18
Tabla 2. Resultados de la Prueba ADF: Estadísticos y Visualización	20
Tabla 3. Prueba HEGY: Resultados de la regresión lineal.....	22
Tabla 4. Distribución de Predicciones Trimestrales del Déficit Fiscal por Institución en Estados Unidos (1993-2022)	24
Tabla 5. Hiperparámetros utilizados en XGBoost para la automatización	48
Tabla 6. Mejores resultados del XGBoost para cada split.....	49
Tabla 7. Hiperparámetros utilizados en LSTM-CNN para la automatización	52
Tabla 8. Mejores resultados del LSTM-CNN para cada split.....	54
Tabla 9. Métricas principales de los forecast errors de los modelos: XGBoost y LSTM – CNN.....	57
Tabla 10. Resultados de las predicciones anuales de las instituciones frente al déficit real	66
Tabla 11. Resultados de los modelos combinados para cada institución.....	68

1 Introducción

El aumento continuo del déficit fiscal y, en consecuencia, de la deuda en las economías modernas plantea la cuestión de por qué, año tras año, tanto los organismos públicos como privados han normalizado la presencia de grandes errores en sus predicciones. Esto abre una línea de investigación orientada a determinar si, gracias a las nuevas técnicas de análisis de datos, es posible mejorar la precisión de dichas predicciones.

En los últimos años, ha surgido un creciente interés por esclarecer esta variable, que durante mucho tiempo ha sido dejada de lado. Este interés, catalizado por los cambios recientes en las políticas fiscales, el incremento de la inflación y la evolución de la situación económica global tiene lugar en el contexto de un entorno enormemente dinámico.

Las teorías de Keynes, que han dominado el pensamiento económico del occidente industrial desde la posguerra, instauraron una visión positiva del gasto público, marcando un punto de inflexión en la economía moderna (Keynes, 1936). Keynes fue un ferviente defensor del uso del gasto público durante periodos de recesión, argumentando que este puede estimular el crecimiento económico, mantener la demanda agregada y estabilizar el empleo.

Según su pensamiento, respaldado por la idea de que en contextos de recesión la población tiende a ahorrar y tiene menos incentivos para contraer préstamos e invertir, la economía rara vez opera a su máximo potencial. Ante esta insuficiencia, Keynes propuso al Estado como la autoridad necesaria para llenar ese vacío mediante la inversión pública, lo que denominó la necesidad de “socializar la inversión” (Dwyer, 2012).

No obstante, estas políticas también dieron origen a un patrón recurrente de ciclos económicos marcados por déficits presupuestarios persistentes, reflejando los desafíos de mantener el equilibrio fiscal en un modelo keynesiano (Cooley & Ohanian, 1997).

Según la escuela de Chicago, las políticas fiscales anticíclicas, fallan debido a las racionalizaciones de los agentes económicos.

Jesús Huerta de Soto en su libro, *Dinero, Crédito Bancario y Ciclos Económicos*, contradice a Keynes haciendo resaltar la escuela austriaca de economía mediante su opinión sobre el gasto público, asegurando que el aumento del gasto público no resuelve los problemas económicos, sino que tiende a exacerbar los ciclos económicos (Huerta de Soto, 1998).

David Ricardo en el siglo XIX, mediante sus distintas hipótesis aclaró que el individuo racional es consciente de que la deuda del gobierno ha de ser pagada de forma directa o indirecta y por tanto este tiende al ahorro anticipado, ralentizando así el crecimiento subyacente del país.

La evolución del pensamiento económico, desde las teorías keynesianas hasta las críticas de la Escuela de Chicago y la perspectiva austriaca, resalta la complejidad inherente a la gestión del déficit fiscal y sus implicaciones económicas. Las ideas de Keynes marcaron un punto de inflexión al promover el gasto público como herramienta de estabilización económica, pero, como han señalado Huerta de Soto (1998) y otros economistas críticos, estas políticas pueden acentuar los ciclos económicos si no se gestionan adecuadamente. A esto se suma la racionalidad de los agentes, descrita por Ricardo, quien anticipó cómo el comportamiento de los individuos frente a la deuda gubernamental puede limitar el impacto de las políticas fiscales.

Por estas razones, los errores entre las proyecciones presupuestarias y los resultados reales tienden a persistir con el tiempo. Las técnicas modernas de análisis de datos, como las basadas en *machine learning*, ofrecen un enfoque innovador para integrar las deficiencias económicas, las cuales están intrínsecamente reflejadas en los datos históricos, ya sea de forma parcial o en su totalidad. Estas herramientas permiten considerar el comportamiento anticipado de los agentes económicos e intentan generar proyecciones más robustas y precisas.

El uso de modelos predictivos avanzados permite abordar las limitaciones inherentes a los enfoques tradicionales, proporcionando a las instituciones una herramienta capaz de predecir de forma objetiva y con mayor precisión el déficit para que los individuos puedan diseñar políticas fiscales efectivas. Este trabajo pretende otorgar una aplicación práctica tanto para organismos públicos como para el sector privado, con el fin de promover una planificación económica más informada y eficiente en un entorno económico global cada vez más complejo.

1.1 Objetivos

El trabajo de investigación se centra en el análisis de las dinámicas del déficit fiscal de Estados Unidos, con el objetivo principal de desarrollar un modelo univariante predictivo que permite anticipar el déficit fiscal de manera más precisa mediante modelos de *machine learning* empleando tanto arboles de decisión como redes neuronales tratando de reducir la componente humana. En concreto, este trabajo implementará dos modelos de *machine learning* autorregresivos univariantes: XGBoost y LSTM-CNN, los cuales han mostrado ser herramientas efectivas en el pronóstico económico (Ba et al., 2020) (Shejul et al., 2024), combinando técnicas tradicionales con metodologías modernas.

Posteriormente, estos modelos serán combinados en cuatro formas distintas (ponderaciones basadas en AIC y BIC, ponderación dinámica mediante Theil's U, y Bayesian Model Averaging (BMA)) buscando aprovechar las fortalezas de cada uno y mitigar sus deficiencias individuales (Carabotta & Claeys, 2024). Este enfoque combinado busca superar las limitaciones tradicionales del modelado económico como los sesgos inherentes (sesgo optimista o político) y las restricciones metodológicas (rigideces informativas o previsiones manipuladas) de las proyecciones gubernamentales y privadas.

Finalmente, los resultados obtenidos serán comparados con las proyecciones de las tres instituciones privadas seleccionadas (JPMorgan, First Trust Advisors y Econ Intelligence Unit), evaluando la precisión de las predicciones combinadas realizadas frente a las instituciones.

Este análisis permitirá adoptar una perspectiva más integral y objetiva en el estudio del déficit fiscal, y de su impacto en la economía estadounidense, garantizando un enfoque centrado en los datos y dejando que los números hablen por sí mismos. Se busca demostrar cómo los modelos numéricos pueden convertirse en una herramienta esencial para predecir con mejor precisión el déficit fiscal y mediante la combinación con la “humanización” del código, conseguir una mejora sustancial en el resultado final.

Al integrar estas nuevas tecnologías, también se pretende anticipar ciclos económicos y, en consecuencia, prever y mitigar crisis económicas, una necesidad subrayada por la literatura. Como lo destaca Celik (2024), el uso de técnicas de aprendizaje automático proporciona soluciones innovadoras para predecir crisis macroeconómicas y gestionar las posibles cargas fiscales, permitiendo así un enfoque más proactivo y eficiente frente a los retos financieros actuales.

1.2 Metodología

Para alcanzar el objetivo planteado, se han empleado técnicas de modelización predictiva implementadas mediante programación en Python, utilizando dos modelos: XGBoost y LSTM-CNN. Estos modelos fueron seleccionados por sus ventajas frente a modelos más simples: XGBoost para capturar relaciones no lineales, por ejemplo, este modelo se utilizó para predecir con precisión el Producto Interno Bruto (PIB) y los precios al consumidor en los países del G7 (Kurihara & Fukushima, 2019), mientras que LSTM-CNN es útil a la hora de modelar patrones temporales complejos y con presencia de estacionalidades. En el ámbito de la economía, LSTM-CNN se ha aplicado con éxito para predecir el PIB regional y para analizar tendencias macroeconómicas, como en el caso del PIB de la provincia de Zhejiang, donde logró identificar patrones de crecimiento a largo plazo y fluctuaciones estacionales con alta precisión (Wan et al., 2024).

Con el propósito principal de predecir el déficit fiscal de Estados Unidos y compararlo con las proyecciones realizadas por instituciones privadas, se implementó una combinación de XGBoost y LSTM-CNN utilizando cuatro métodos distintos: ponderación basada en el parámetro AIC, ponderación basada en el parámetro BIC, Bayesian Model Averaging (BMA) y una ponderación dinámica fundamentada en el estadístico Theil's U.

Estas combinaciones buscan generar una predicción combinada que aproveche las fortalezas de cada modelo individual, minimizando los sesgos inherentes que podrían introducirse al depender de un solo modelo. La integración de estos métodos garantiza un enfoque robusto y versátil para abordar las complejidades de la predicción fiscal y convertir la variable del déficit fiscal en dinámica, más alineada con la naturaleza del entorno actual.

En cuanto al flujo del modelo, se implementó un proceso automatizado de ajuste de hiperparámetros mediante el método de Grid Search aplicado a cada uno de los modelos. La selección del modelo óptimo se realiza utilizando el estadístico Theil's U como criterio principal, eligiendo el conjunto de parámetros que minimiza este valor. Además, se llevó a cabo una selección automatizada de splits de datos con un decalaje progresivo del 10%, variando desde el 50% hasta el 90% del conjunto de datos. Este enfoque asegura que los modelos se ajusten y evalúen en diferentes particiones, optimizando su rendimiento y evitando sesgos.

El objetivo de esta automatización es garantizar resultados precisos y robustos, minimizando el error potencial que podría surgir al imponer manualmente parámetros o configuraciones de splits que, aunque factibles, podrían no ser las más adecuadas para el conjunto de datos específico.

Finalmente, se destacan las ventajas de la metodología de combinación de modelos, como su flexibilidad para ajustar parámetros y su capacidad para realizar predicciones dinámicas utilizando distintos métodos de ponderación. Sin embargo, también se reconocen sus limitaciones, entre ellas la dependencia de datos históricos, el carácter univariable de los datos empleados y el propio acceso de información gubernamental por parte de las instituciones privadas (Tovar Jalles et al., 2015), entre otros. A pesar de estas restricciones, la combinación de múltiples métodos y modelos tiene como objetivo dinamizar los resultados finales, mejorando la precisión y ofreciendo herramientas valiosas tanto para organismos públicos como privados en la toma de decisiones estratégicas.

1.3 Fuentes de Información

Para llevar a cabo el análisis, se ha recurrido a varias fuentes de datos confiables, en concreto:

Base de datos FRED (*Federal Reserve Economic Data*): Los datos del *Federal Surplus or Deficit [-] as Percent of Gross Domestic Product* fueron obtenidos para el periodo comprendido entre 1929 y 2023. Los datos son anuales y no están ajustados por estacionalidad. Esta fuente garantiza la calidad y consistencia de las series temporales utilizadas, fundamentales para la modelización del déficit fiscal.

Instituciones no públicas (*data_90_US*): Los datos de las proyecciones realizadas por instituciones no públicas, como JPMorgan, First Trust Advisors y Econ Intelligence Unit fueron obtenidas utilizando Stata, para el periodo comprendido entre 1993 y 2022. Las predicciones se realizan de manera trimestral, ya divididas en relación con el PIB y sin ajustes estacionales. Para convertir estos datos a valores anuales, se ha calculado la media de los cuatro trimestres correspondientes.

Las fuentes de información seleccionadas destacan por su fiabilidad y por la disponibilidad de los datos a las fechas requeridas.

1.4 Revisión de la Literatura

1.4.1 Superando Sesgos en las Previsiones Fiscales

Este trabajo busca superar las limitaciones tradicionales del modelado económico, abordando los sesgos y errores identificados en diversos estudios. Frankel (2011) destaca el sesgo optimista en las previsiones fiscales oficiales, caracterizado por la sobreestimación del crecimiento económico y la subestimación de los déficits presupuestarios, a menudo motivado por presiones políticas y reglas fiscales más restrictivas de las estimadas. Coibion y Gorodnichenko (2011) resaltaron la existencia de rigideces informativas, como la actualización incompleta de

información (información pegajosa) y la recepción de señales imperfectas (información ruidosa), que generan errores persistentes en las predicciones. Carabotta y Claeys (2015) remarcaron cómo las previsiones fiscales individuales, influenciadas por sesgos políticos y discrepancias metodológicas, pueden mejorar significativamente cuando se combinan mediante métodos ponderados, como el Weighted Forecast Combination (WFC), para reducir errores y aumentar la robustez. Jonung y Larch (2006) señalaron que los sesgos políticos en la Unión Europea son comunes, ya que los gobiernos ajustan sus previsiones para cumplir con reglas fiscales, lo que afecta la credibilidad y precisión de las predicciones. Asimismo, Beetsma, Giuliadori y Wiertz (2009) enfatizaron que estos ajustes estratégicos llevan a previsiones manipuladas, diseñadas para aparentar cumplimiento de las metas, pero que resultan en errores sistemáticos. Por último, Merola y Pérez (2013) compararon las previsiones de gobiernos con las de agencias independientes, mostrando que las instituciones independientes tienen menor sesgo y mayor precisión debido a su menor influencia política.

En conjunto, estos estudios resaltan la necesidad de adoptar métodos combinados y enfoques independientes que reduzcan los sesgos, mejoren la precisión y fortalezcan la credibilidad de las previsiones fiscales.

1.4.2 Innovación en la Predicción del Déficit Fiscal

El avance en técnicas de *machine learning* como el XGBoost, introducido en 2014, o técnicas híbridas como LSTM-CNN han cambiado radicalmente la forma en la que los analistas abordan problemas complejos en la economía, especialmente en escenarios caracterizados por la incertidumbre y la alta volatilidad. Estas nuevas tecnologías han sido incorporadas a los análisis gracias a su eficiencia y rapidez de procesamiento de grandes cantidades de datos, de distintitos escenarios no

lineales y de identificar patrones ocultos, con el fin de adaptarse dinámicamente a cambios estructurales en las economías.

De acuerdo con Elbeltagi et al. (2024), la combinación de algoritmos de aprendizaje profundo y métodos de clasificación, como redes neuronales convolucionales (CNN), permite modelar simultáneamente patrones espaciales y temporales, sobrepasando ciertas limitaciones de los modelos más tradicionales, que a menudo asumen relaciones lineales.

Rahman y Elbeltagi (2022) refuerzan esta afirmación al demostrar que modelos híbridos como XGBoost, en combinación con LSTM-CNN, no solo mejoran la precisión de la predicción, sino que también aumentan la robustez frente a eventos inesperados o *black swans*. Este enfoque aprovecha las ventajas de ambos modelos mediante el uso de la capacidad de LSTM para capturar estacionalidades y secuencias temporales complejas, y el potencial que XGBoost ofrece para identificar interacciones no lineales entre múltiples variables económicas.

Además, Jo et al. (2022) destacan la importancia de los modelos donde las combinaciones pueden extraer información relevante incluso de datos ruidosos. Por ejemplo, Biswas et al. (2023) presentan un enfoque innovador al integrar factores macroeconómicos con redes neuronales y XGBoost, logrando predicciones precisas en mercados de alta volatilidad.

Finalmente, Kotsiantis et al. (2006), en su revisión, argumentan que la combinación de técnicas de *machine learning* no solo mejora la precisión, sino que también aborda temas como la incertidumbre y la identificación de relaciones escondidas. Este enfoque permite recalcar la robustez y la adaptabilidad de los modelos en contextos complejos.

Estos modelos permiten a los analistas prever déficits fiscales con mayor precisión y diseñar políticas económicas más efectivas. En conjunto, la integración de modelos avanzados como XGBoost y LSTM-CNN, ofrece un enfoque transformador para la predicción fiscal moderna, alineándose con la necesidad de herramientas fiables y dinámicas en un entorno económico global cada vez más complejo.

Por otro lado, la implementación de técnicas combinadas como el Criterio de Información de Akaike (AIC), el Criterio de Información Bayesiano (BIC) y el Bayesian Model Averaging (BMA) son cruciales para mejorar la precisión en predicciones económicas. Según Diks y Vrugt (2010), los métodos basados en AIC y BIC ofrecen una metodología precisa que facilita la selección de un modelo de predicción óptimo, particularmente con relaciones intravariantes complejas y no lineales. Además, resalta que la ponderación por estos criterios permite integrar múltiples modelos en una única proyección, reduciendo los errores residuales.

Steel (2020) destaca que BMA no solo mejora las predicciones al integrar modelos con diferentes supuestos estructurales, sino que también aborda problemas de incertidumbre al considerar las probabilidades de cada modelo explicativo, siendo especialmente útil cuando los datos económicos presentan ruido y volatilidad. Por su parte, Tsangari (2007) argumenta que el uso combinado de AIC, BIC y métodos promedios ponderados incrementa la precisión de las proyecciones a través de una mejor identificación de relaciones ocultas en los datos.

Feldkircher (2012) refuerza esta perspectiva al aplicar BMA en la predicción de la producción industrial, demostrando que la combinación de modelos supera significativamente a las predicciones basadas en un único modelo. Finalmente, Sheikh y Coulibaly (2024) señalan, a través de modelos hidrológicos, que la

integración de AIC y BIC junto con BMA proporciona no solo mayor precisión, sino también mayor eficacia ante escenarios dinámicos, como es el caso del análisis económico.

Los enfoques dinámicos en la ponderación de modelos han demostrado ser esenciales para mejorar la precisión y adaptabilidad de las predicciones económicas. Sineglazov y Chumachenko (2014) proponen un sistema que ajusta los pesos de los modelos en tiempo real según su desempeño, evaluado mediante métricas como el estadístico MSE.

Esta capacidad de ajuste dinámico no solo incrementa la precisión de las predicciones, sino que también aporta robustez frente a cambios estructurales en los datos, lo que resulta ideal en contextos de volatilidad y relaciones no lineales. Integrar este enfoque con técnicas consolidadas crea un marco avanzado que optimiza la precisión y la adaptabilidad de los modelos.

En este trabajo, se ha desarrollado un modelo propio dinámico que ajusta los pesos de los modelos individuales según el estadístico Theil's U para cada una de las predicciones, permitiendo que el modelo más eficiente tenga mayor influencia en cada periodo predicho y ajustando los pesos automáticamente conforme cambian las condiciones económicas. Este diseño representa una innovación al combinar herramientas tradicionales con un enfoque más adaptados a la problemática de hoy en día, y que pretende mejorar la precisión, la robustez y la eficacia en las predicciones del déficit fiscal.

1.5 Impacto en la Sociedad

El déficit fiscal no es solo una cifra económica; es un reflejo de las decisiones políticas, la distribución de recursos y las prioridades de un país. Su análisis y predicción afectan directamente a las personas, ya que de ello depende en parte

de la capacidad del gobierno para financiar servicios públicos esenciales, implementar programas sociales y sostener inversiones clave en infraestructura y desarrollo. No solo es de vital importancia para las instituciones públicas, cuyo objetivo principal es promover el bienestar económico del país y actuar como catalizador clave en la generación de empleo, sino también para la inversión privada, tanto nacional como extranjera. Esto permite fomentar la atracción de capital hacia el país, impulsando la llegada de capital humano y tecnología, lo que, en última instancia, contribuye al crecimiento económico nacional.

Una gestión fiscal efectiva, apoyada por modelos matemáticos, puede marcar la diferencia en cómo se distribuyen los recursos, asegurando que lleguen a los más desfavorecidos y promoviendo la estabilidad económica a largo plazo. Además, un déficit controlado fortalece la confianza en la economía nacional, tanto a nivel interno como internacional, mejorando la percepción de los mercados nacionales y reduciendo los costes de financiación. De este modo, el déficit fiscal tiene un impacto directo en el bienestar de las personas y en la calidad de vida de toda la sociedad y las herramientas de predicción de este deben evolucionar junto con los cambios en la coyuntura económica.

1.5.1 En línea con los Objetivos de Desarrollo Sostenible

La sostenibilidad fiscal no es solo un objetivo económico, sino también un pilar esencial para construir una sociedad más equitativa y justa. Este trabajo busca contribuir positivamente a varios de los Objetivos de Desarrollo Sostenible (ODS) al aplicar soluciones tecnológicas basadas en análisis de datos, mejorando la toma de decisiones y contribuyendo al fortalecimiento de la economía tanto nacional como global.

ODS 1: Fin de la pobreza. Un déficit fiscal bien gestionado, acompañado de una clara visibilidad sobre cómo podría variar en el futuro, permite hacer realidad la financiación de programas destinados a reducir la pobreza y a generar valor para los sectores más vulnerables de la sociedad (Ozili, 2024).

ODS 8: Trabajo decente y crecimiento económico . Una economía estable, respaldada por una gestión fiscal responsable, fomenta el crecimiento sostenible y la creación de empleos.

ODS 10: Reducción de las desigualdades . Una buena predicción del déficit crea un entorno donde hay una distribución equitativa y más justa de las ayudas y subvenciones, cerrando brechas entre diferentes grupos de la sociedad (Benito et al., 2023).

ODS 16: Paz, justicia e instituciones sólidas . La transparencia y precisión en las proyecciones fiscales refuerzan la confianza en las instituciones públicas, otorgando efectividad a la gobernanza (Bisogno et al., 2024).

Por ello, el desarrollo de herramientas avanzadas para predecir el déficit fiscal no solo busca mejorar la planificación económica, sino también contribuir a un futuro más sostenible y equitativo. Estas proyecciones no solo benefician a instituciones públicas, sino también al sector privado, al asegurar que las decisiones económicas estén alineadas con el bienestar de las personas y los compromisos globales hacia el desarrollo sostenible.

2 Descripción de los Datos

El objetivo de este trabajo es analizar y modelar el déficit fiscal de Estados Unidos como porcentaje del Producto Interno Bruto (PIB), utilizando técnicas avanzadas de *machine learning* para identificar patrones en los datos históricos y generar predicciones precisas, estables y significativas que permitan mejorar la planificación fiscal y la toma de decisiones estratégicas.

Las bases de datos del Budget Balance como porcentaje del PIB anual sin ajuste estacional han sido obtenidas del Federal Reserve Bank of St. Louis (FRED)¹.

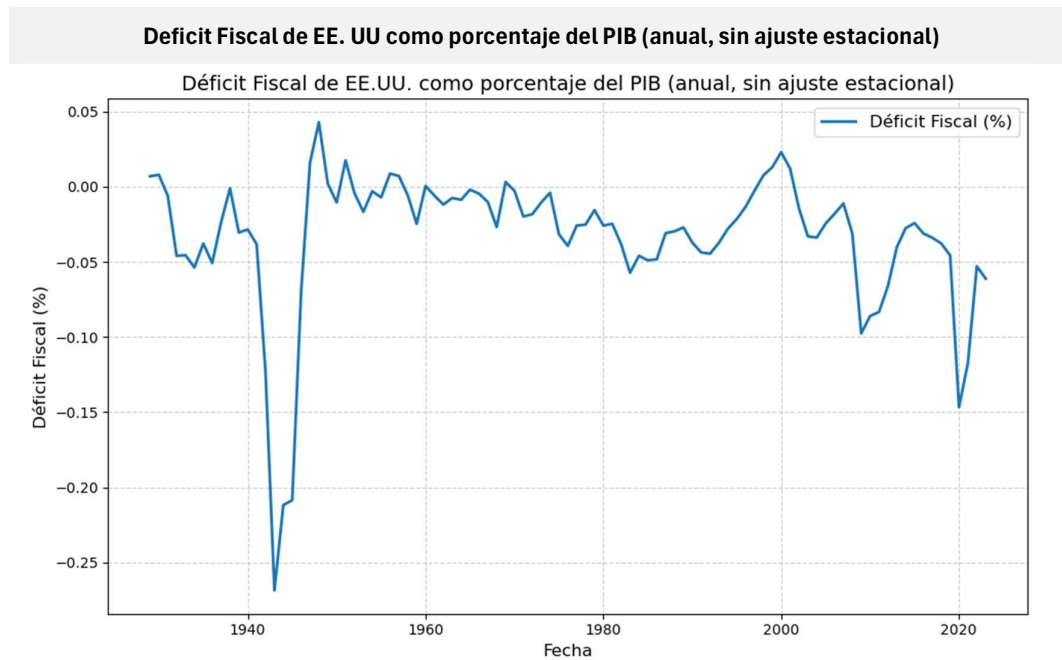
Todos los modelos presentados en el apartado 3 Descripción las Metodologías son univariantes y anuales. Los modelos de series temporales trabajan bajo la suposición que los valores pasados influyen en el futuro.

En este apartado se presentan, por un lado, los datos históricos del déficit fiscal utilizados en el análisis y, por otro, las predicciones del déficit fiscal realizadas por diversas instituciones. El objetivo es comparar las predicciones obtenidas en este trabajo con aquellas generadas por las diferentes instituciones, permitiendo evaluar la precisión y consistencia de los resultados.

2.1 El Déficit Fiscal Histórico

A continuación (Figura 1) se muestra el Déficit Fiscal sobre PIB anual desde 1929 hasta 2023.

¹ Hoy en día, la FRED desempeña un papel crucial dentro del sistema de la Reserva Federal. Se encarga de supervisar instituciones financieras, implementar políticas monetarias y proporcionar información económica fomentando la transparencia y el análisis económico de los Estados Unidos.



Fuente: Elaboración Propia

Figura 1. Déficit Fiscal de EE. UU. como porcentaje del PIB (anual, sin ajuste estacional)

En la Figura 1, se pueden analizar varios patrones importantes en la evolución del déficit fiscal de Estados Unidos como porcentaje del PIB, que reflejan tanto cambios en las políticas fiscales como en el contexto económico global.

La Gran Depresión (1929-1933) marcó un punto de inflexión en la política fiscal de Estados Unidos. Durante este período, el colapso económico llevó a un aumento significativo del desempleo y una caída drástica de la producción industrial. Para hacer frente a esta crisis, el gobierno comenzó a incurrir en déficits presupuestarios más amplios, utilizando el gasto público como herramienta de estabilización económica. Esto incluyó la implementación del New Deal por Franklin D. Roosevelt, que expandió los programas sociales y las inversiones en infraestructura, aumentando sustancialmente el gasto público y dejando atrás la política de presupuestos equilibrados que había predominado anteriormente (De Long, 1998).

Entre 1939 y 1945, durante la Segunda Guerra Mundial, los déficits presupuestarios alcanzaron niveles sin precedentes, ya que el gobierno financió masivamente el conflicto bélico. Este período consolidó el déficit fiscal como una herramienta clave de la política económica. Según Renshaw (1999), marcó un cambio estructural en el papel del gobierno en la economía.

En las décadas posteriores a la guerra, los déficits fiscales comenzaron a reducirse. Durante los años 50 y 60, las políticas fiscales buscaron estabilizar la economía, alternando entre superávits y déficits moderados en función del ciclo económico.

Entre 1960 y 1980, los movimientos del presupuesto fiscal en Estados Unidos se caracterizaron por una estabilidad relativa, con pequeñas fluctuaciones que reflejan políticas fiscales cíclicas y moderadas. Este período muestra un presupuesto enfocado en mantener el equilibrio (Auerbach, 2003).

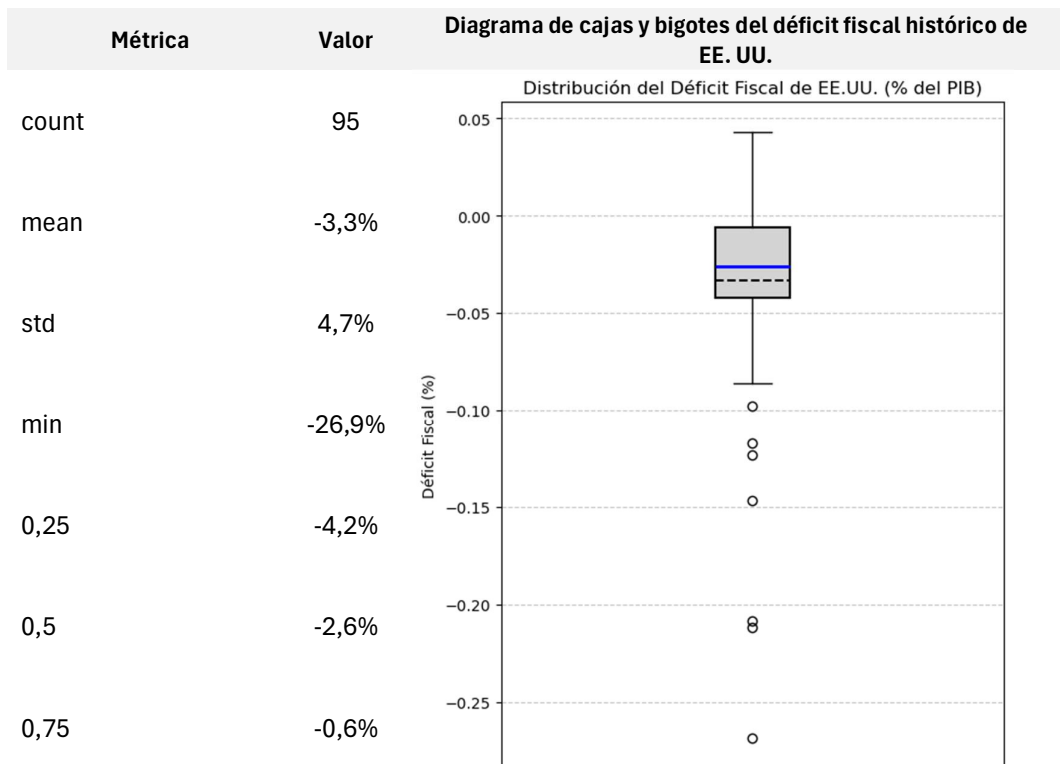
En el período de 1980 a 2000, se observa un cambio significativo. Durante la presidencia de Ronald Reagan, se implementaron grandes recortes de impuestos bajo la premisa de la economía de oferta, junto con un aumento sustancial del gasto en defensa, lo que resultó en una ampliación notable del déficit fiscal. Sin embargo, hacia finales de los años 90, bajo la administración de Bill Clinton, las políticas enfocadas en la reducción del déficit y el crecimiento económico derivado del auge tecnológico llevaron a un superávit fiscal, marcando un contraste con las décadas anteriores (Blanchard & Leigh, 2013).

A partir del año 2000, los déficits fiscales comenzaron a ampliarse nuevamente. Los recortes impositivos realizados durante la presidencia de George W. Bush, junto con los costos asociados a las guerras en Afganistán e Irak, financiados principalmente con deuda, tuvieron un impacto significativo en el presupuesto. Entre 2008 y 2009, la Gran Recesión agravó esta tendencia, con un aumento

drástico del déficit fiscal debido a los estímulos económicos masivos diseñados. La administración de Barack Obama implementó políticas expansivas que buscaron mitigar los efectos de la recesión y estabilizar la economía, lo que se reflejó en déficits persistentes a lo largo de la década.

En 2020, la pandemia de COVID-19 llevó a una caída drástica del déficit fiscal, como resultado de medidas como el CARES Act y otros paquetes de estímulos destinados a enfrentar la emergencia sanitaria y sus consecuencias económicas. Estas políticas expansivas, aunque necesarias, marcaron un punto crítico en la acumulación de deuda pública (Blinder, 2023).

A lo largo de estas décadas, el manejo del presupuesto fiscal en Estados Unidos ha sido moldeado por decisiones políticas, eventos económicos globales y crisis extraordinarias.

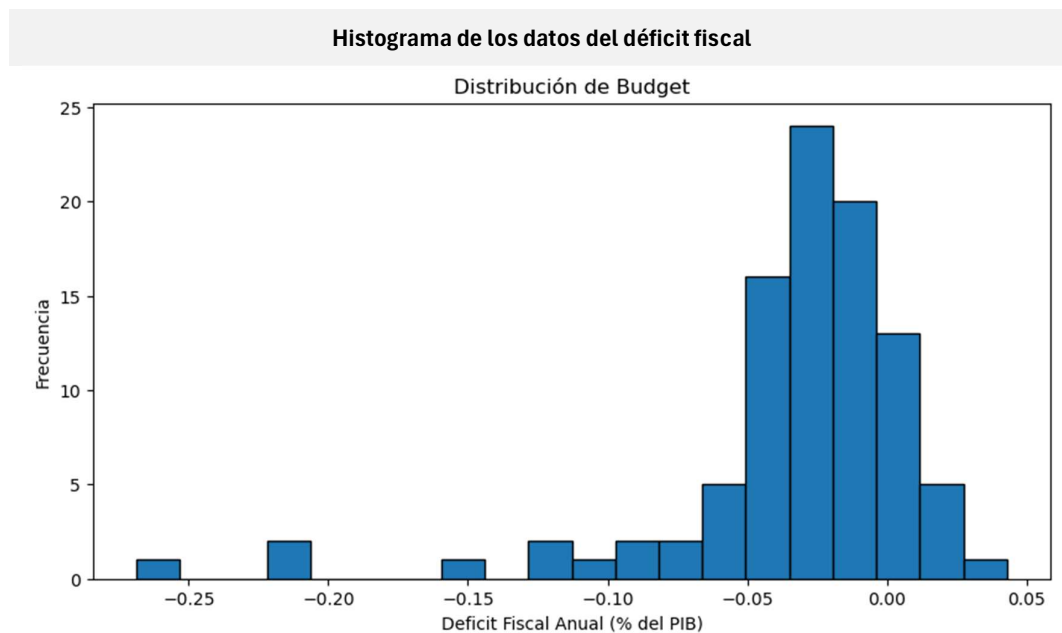


Fuente: Elaboración Propia

Tabla 1. Estadísticas descriptivas del déficit fiscal anual de EE. UU.

La Tabla 1, muestra las principales estadísticas descriptivas junto a una descripción grafica de los datos. Se puede observar la segunda guerra mundial (-26,9%) o el superávit máximo de 4,3 % del PIB ocurrido en 1948, bajo la presidencia de Harry S. Truman, como parte de los esfuerzos para reducir la deuda pública acumulada durante la Segunda Guerra Mundial. Este período fue marcado por un fuerte crecimiento económico en la posguerra y políticas fiscales enfocadas en estabilizar las finanzas nacionales públicas (Calda, 2009).

El data set total cuenta con un total de 95 datos anuales registrados entre 1929 a 2023 con un promedio de -3,3% y una variación estándar del 4,7%.



Fuente: Elaboración Propia

Figura 2. Distribución del Déficit Fiscal sobre el PIB

La distribución representada en la Figura 2 muestra que los valores se concentran ligeramente hacia el lado negativo del cero, lo que sugiere que, en general, el presupuesto fiscal de Estados Unidos ha tendido a estar en déficit en la mayoría de los períodos analizados. Este comportamiento es característico de economías

que enfrentan altos niveles de gasto en sectores clave como defensa e infraestructura, financiados principalmente a través del endeudamiento público.

2.1.1 Análisis Detallado del Déficit Fiscal

En este apartado se realiza un análisis exhaustivo del déficit fiscal de Estados Unidos utilizando herramientas estadísticas avanzadas para identificar patrones de comportamiento, estacionalidades y dinámicas subyacentes. Se aplicaron diferentes pruebas (ADF, HEGY, ACF y PACF) para evaluar la naturaleza estacionaria de la serie, la presencia de raíces unitarias estacionales y las correlaciones temporales.

La prueba de Dickey-Fuller aumentado (ADF) se utilizó para verificar si la serie del déficit fiscal es estacionaria, evaluando la existencia de raíces unitarias globales en los datos. El modelo consideró tanto una constante como una tendencia lineal, dado el comportamiento histórico del déficit. En la Tabla 2 se presentan los resultados obtenidos.

Métrica	Valor	Prueba ADF según nivel de significancia
Estadístico de Prueba ADF	-4,671	
Valor P	0,001	
Número de Retardos	3	
Número de Observaciones	100	
Valor Crítico (1%)	-4,060	
Valor Crítico (5%)	-3.459	
Valor Crítico (10%)	-3.155	

Fuente: Elaboración Propia

Tabla 2. Resultados de la Prueba ADF: Estadísticos y Visualización

La prueba ADF realizada muestra que la serie del déficit fiscal es estacionaria al nivel de significancia del 1%. Esto se evidencia al comparar el estadístico de prueba ($-4,671$) con los valores críticos ($-4,060$ para el 1%). Como el estadístico es significativamente menor que el valor crítico, se rechaza la hipótesis nula de no estacionariedad. Este resultado implica que las propiedades estadísticas de la serie, como la media y la varianza, son constantes en el tiempo, lo que permite aplicar modelos predictivos sin necesidad de diferenciación adicional para estabilizar la serie.

Por otro lado, se aplicó la prueba HEGY para evaluar la existencia de raíces unitarias estacionales en la serie del déficit fiscal, considerando componentes como las fluctuaciones trimestrales y anuales, características de este tipo de datos. Los resultados indicaron que las raíces unitarias asociadas a las frecuencias estacionales son significativas, lo que revela que una parte importante de las fluctuaciones del déficit fiscal está influida por patrones estacionales. Este hallazgo refuerza la necesidad de utilizar modelos avanzados como LSTM-CNN, diseñados para capturar de manera eficiente las estacionalidades y las dependencias temporales complejas presentes en los datos.

Prueba HEGY: Resultados de la regresión lineal

Estadístico HEGY: -1.454781642288853
P-valor HEGY: 0.14945648851196613

OLS Regression Results

```
=====
Dep. Variable:          y_diff      R-squared:                0.235
Model:                  OLS         Adj. R-squared:           0.189
Method:                 Least Squares  F-statistic:              5.159
Date:                   Sun, 01 Dec 2024  Prob (F-statistic):       0.000355
Time:                   20:49:49      Log-Likelihood:          193.22
No. Observations:      90            AIC:                     -374.4
Df Residuals:          84            BIC:                     -359.4
Df Model:               5
Covariance Type:       nonrobust
```

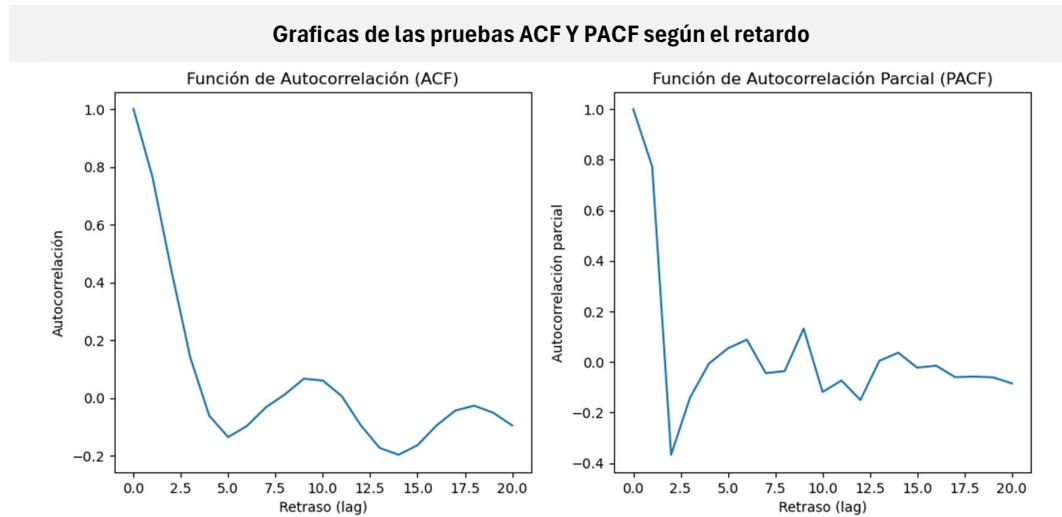
	coef	std err	t	P> t	[0.025	0.975]
const	-0.0116	0.004	-2.642	0.010	-0.020	-0.003
y_lag1	-0.0739	0.051	-1.455	0.149	-0.175	0.027
y_lag2	-0.1400	0.046	-3.017	0.003	-0.232	-0.048
y_lag3	-0.1285	0.053	-2.419	0.018	-0.234	-0.023
lag_diff_1	0.0660	0.068	0.976	0.332	-0.069	0.201
lag_diff_2	-0.0114	0.068	-0.167	0.868	-0.147	0.125
lag_diff_3	0.0213	0.114	0.187	0.852	-0.205	0.247
lag_diff_4	-0.0451	0.120	-0.377	0.707	-0.283	0.193
Omnibus:		65.405	Durbin-Watson:		1.985	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		389.017	
Skew:		-2.263	Prob(JB):		3.36e-85	
Kurtosis:		12.125	Cond. No.		2.82e+17	

Fuente: Elaboración Propia

Tabla 3. Prueba HEGY: Resultados de la regresión lineal

La prueba HEGY realizada indica que no se puede rechazar la hipótesis nula de presencia de raíces unitarias estacionales al nivel de significancia del 5%. El estadístico de prueba obtenido (-1,455) no es menor al valor crítico correspondiente (-3,458 para el 5%), lo que sugiere que la serie presenta raíces unitarias estacionales. Este resultado implica que las fluctuaciones en el déficit fiscal podrían estar influenciadas por patrones estacionales. Por lo tanto, es necesario incorporar modelos capaces de capturar estas estacionalidades, como LSTM-CNN, para mejorar la precisión en las predicciones.

Por último, el análisis de autocorrelación (ACF) y autocorrelación parcial (PACF) revela información sobre los patrones de dependencia temporal en la serie del déficit fiscal. En la Figura 3, se pueden observar ambos gráficos para cada retardo.



Fuente: Elaboración Propia

Figura 3. Graficas de las pruebas ACF Y PACF según el retardo

Los resultados indican una autocorrelación significativa en el lag 1 (0,765), lo que demuestra una fuerte dependencia anual. Sin embargo, a partir del lag 2, los valores de ACF disminuyen rápidamente, sugiriendo que los efectos de los valores previos se desvanecen con el tiempo. Por otro lado, los valores de PACF también destacan en el lag 1, confirmando una relación directa entre los datos consecutivos. En retardos superiores, los valores fluctúan, reflejando posibles patrones más complejos o efectos indirectos en la serie.

Estos hallazgos refuerzan el uso de modelos avanzados como LSTM-CNN, que pueden abordar tanto dependencias inmediatas como más complejas, para mejorar la precisión en la predicción del déficit fiscal.

2.2 El Déficit Fiscal Predicho por Instituciones

El déficit fiscal ha sido un tema de interés tanto para instituciones públicas como privadas, dado su impacto en la estabilidad económica. En este trabajo, se consideran las predicciones realizadas por tres instituciones reconocidas: Econ Intelligence Unit, First Trust Advisors y JPMorgan, cuyas estimaciones se

recopilaron de bases de datos económicas para ser analizadas comparativamente con los modelos combinados propuestos.

La presente tabla (Tabla 4) demuestra la inconsistencia del trabajo de las predicciones realizadas por instituciones sobre estados Unidos.

Distribución de predicciones trimestrales del déficit fiscal por institución en estados unidos (1993-2022)																											
Años	Instituciones																										
	3	9	13	14	25	31	37	41	42	48	51	53	55	56	58	59	67	76	93	100	106	111	124	129	134	135	
1993	3	2	3				3	3	3	3																	
1994	3		4				4	1	4	4	4							1									
1995							4		4	4	3																
1996							3		3	3	4																
1997							4			4	4							1									
1998						3	4			4																	
1999						4	4			3																	
2000						4	3			4																	
2001						4	2		1	3		1					3	1									
2002									4			4					4	1									
2003				1					4			4						2									
2004				4					4			4						4									
2005				4					4			4						4									
2006				4					4			4						4									
2007				4					4			4						4									
2008				4					3			4						4									
2009	3			4									2					4									
2010	4			4										1				4									
2011	1			1		2	2		2	2	3	1	3			2	2	2	2	2			3	1	1		3
2012			1	2	2	3	4	2		1	2	1	2	2	1	5		2	2	2			2				
2013	3		3	2		3	1	1	1	3	1	1			2	1	2	2	2	1		4	2	1			2
2014	1		1	2	1	2	4	2		5	2		1	2	1	3	2	2	1	1	2		2	3	1		2
2015	1		3	1	2	1	1		2	3	1	2	2	3	1	2	1	1	3	2	1	3	1	3	2		2
2016	4		2	2	1				1	1	1	2	2	1	1	3	2		1	3	2	4		1			3
2017	1		2		3			1	1	1	1	1	3	2		2		3			4	1			3		
2018			2	1	1				4	1	3	1	1	2	1				1	2	3						1
2019	1	2	3		3				1	1	2		2		1		1		2	2		2	2				1
2020		1	1		2	1			2	2			2				1		1			1					1
2021		1	2						2		1		1					1			1	1	1				1
2022		1	1	2					2	1		1	2	1				2			1	1	1	1			2
Grand Total	19	11	23	19	45	25	43	8	48	31	43	34	43	20	15	10	22	46	12	15	18	12	19	8	9	18	

Fuente: Elaboración Propia

Tabla 4. Distribución de Predicciones Trimestrales del Déficit Fiscal por Institución en Estados Unidos (1993-2022)

La Tabla 4 muestra la inconsistencia en las predicciones trimestrales del déficit fiscal realizadas por las instituciones analizadas en Estados Unidos durante el período 1993-2022. Esta irregularidad es notable en varios aspectos que afectan la calidad y la continuidad de las series temporales necesarias para evaluar el comportamiento fiscal del país.

En el análisis de la tabla (Tabla 4), se puede observar que solo algunas instituciones, marcadas en rojo, presentan predicciones más frecuentes y completas, justificando su selección para análisis comparativos. Sin embargo,

incluso estas muestran lagunas temporales, limitando su capacidad para capturar tendencias continuas. En contraste, otras instituciones presentan vacíos significativos en sus predicciones, con años completos sin datos registrados o patrones interrumpidos, lo que dificulta su uso como *benchmarks* fiables.

La distribución desigual de predicciones refleja esfuerzos concentrados en ciertos periodos, dejando años relevantes sin cobertura, lo que reduce la utilidad de estas instituciones para análisis consistentes. Además, la falta de estandarización en metodologías y frecuencia de publicaciones refuerza la necesidad de modelos avanzados, como los desarrollados en este trabajo, que ofrecen predicciones más precisas, completas y útiles para la toma de decisiones fiscales.

En resumen, la inconsistencia de los datos institucionales subraya la importancia de enfoques metodológicos robustos y avanzados para superar estas limitaciones y mejorar las previsiones del déficit fiscal.

Cabe destacar que cada institución emplea metodologías propias para generar sus proyecciones fiscales. No obstante, estas predicciones presentan limitaciones significativas. Entre ellas, destacan los errores residuales que, en muchos casos, se manifiestan como sesgos optimistas o conservadores, influenciados por los datos proporcionados por los agentes económicos y los supuestos empleados en sus modelos. Además, el uso de metodologías tradicionales, muchas veces lineales, ha demostrado ser insuficiente para capturar la complejidad inherente al déficit fiscal. Estas limitaciones, entre otras, han sido ya detalladas en el apartado 1 donde se analizan las debilidades de los enfoques convencionales y la necesidad de adoptar modelos avanzados para abordar esta problemática.

Estas entidades representan sectores clave: Econ Intelligence Unit como líder en análisis macroeconómico, First Trust Advisors como gestora de inversiones

destacada, y JPMorgan como institución financiera influyente en los mercados globales.

La Figura 4 presenta las predicciones anuales realizadas por las tres instituciones (Econ Intelligence Unit, First Trust Advisors y JPMorgan) para el período correspondiente al rango de datos utilizados en el conjunto de prueba (1993-2022). Es importante destacar que el split entre datos de entrenamiento y prueba se realiza de manera dinámica, resultando en el split que más minimice el estadístico de Theil's U. Se ha pretendido evitar fugas de información entre los datos históricos y las proyecciones. Las fugas de información ocurren cuando un modelo de datos utiliza información que no debería tener al momento de hacer predicciones, como pueden ser datos futuros o características relacionadas directamente con la respuesta o salida del modelo. Evitarlas es crucial, ya que podrían generar una falsa percepción de precisión en el modelo y conducir a resultados engañosos. Este enfoque asegura que las predicciones de las instituciones puedan compararse con los valores predichos por los modelos combinados de forma fiable.

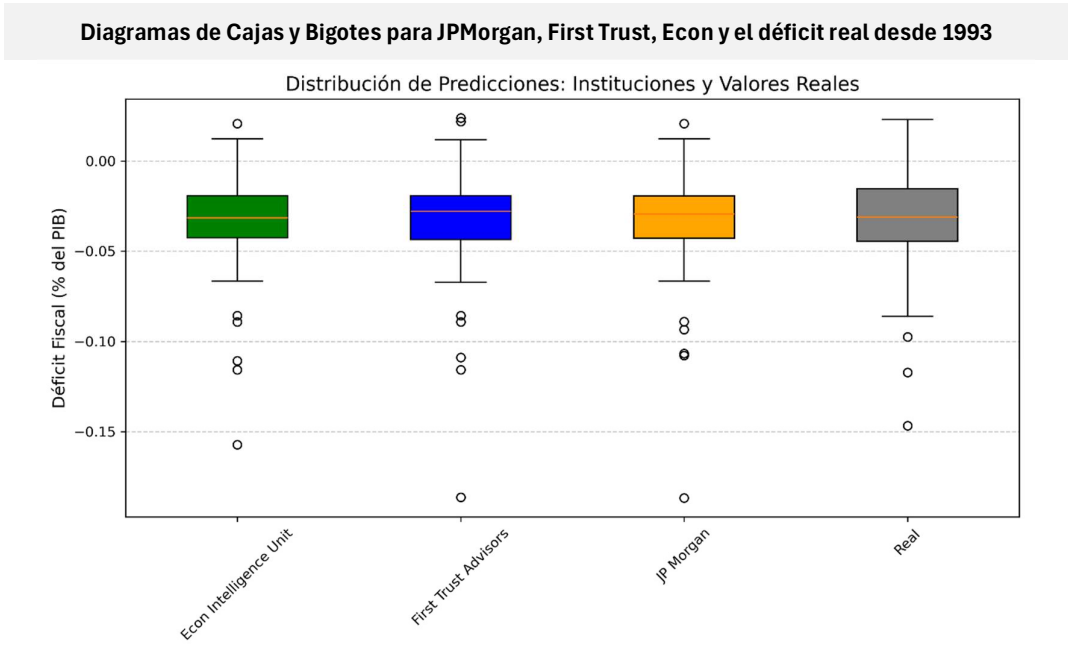
Por ello, en la Figura 4, se representan solo los datos desde 1993, ya que son los que se van a comparar en el apartado 4.2 Comparación con instituciones.



Fuente: Elaboración Propia

Figura 4. Grafica de las predicciones de las tres instituciones seleccionadas (JPMorgan, First Trust y Econ)

En los diagramas de caja y bigotes representados en la Figura 5, se puede observar, por un lado, para cada institución, la estructura de las predicciones del déficit fiscal de Estados Unidos a lo largo de la historia, por otro lado, un diagrama adicional de caja y bigotes representando la estructura de los datos reales del presupuesto fiscal desde 1993.



Fuente: Elaboración Propia

Figura 5. Diagramas de Cajas y Bigotes para JPMorgan, First Trust, Econ y el déficit real desde 1993

Analizando la Figura 5, se puede observar que la dispersión de las predicciones de las tres instituciones es relativamente similar con rangos IQR compactos, lo que indica que las instituciones tienden a predecir valores dentro de un rango moderado, indicando una posible subestimación de los valores reales por parte de las instituciones. Por otro lado, las instituciones parecen estar modelando adecuadamente el comportamiento central del déficit fiscal, pero fallan a la hora de modelar las fluctuaciones extremas observadas en los valores reales. Esto refuerza la necesidad de métodos más avanzados, como combinaciones de modelos *machine learning*, para mejorar la precisión en la predicción.

El análisis presentado en la sección 1 resalta la complejidad del comportamiento histórico del déficit fiscal de Estados Unidos. Por un lado, las pruebas estadísticas, como ADF y HEGY, permitieron validar la estacionariedad de la serie, pero con existencia de patrones estacionales, lo que refuerza la implementación de modelos avanzados (LSTM-CNN). Asimismo, las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) resaltaron en las predicciones de las instituciones la presencia de dependencias significativas en el primer año, justificando el uso de modelos que aprovechen tanto relaciones lineales como no lineales en los datos.

Por otro lado, el análisis de las predicciones institucionales mostró que, aunque estas representan una base sólida para anticipar el déficit fiscal, tienen ciertas limitaciones en términos de precisión y cobertura de valores extremos. Las instituciones, al emplear modelos tradicionales, tienden a subestimar la variabilidad del déficit fiscal.

En este contexto, los resultados refuerzan la importancia de los modelos avanzados de *machine learning* y las combinaciones metodológicas descritas en este trabajo. La comparación entre las proyecciones institucionales y los valores

reales subraya la relevancia de enfoques que combinen precisión, adaptabilidad y capacidad para capturar las dinámicas.

En la sección 3 Descripción las Metodologías se detallarán la metodología y se dará luz a los *black box* que rodean los modelos de *machine learning* para justificar el uso de cada uno.

3 Descripción las Metodologías

El análisis preliminar de los datos históricos del déficit fiscal de Estados Unidos, descrito en el apartado 0, fue clave para seleccionar la metodología de este trabajo. La alta volatilidad y las dinámicas no lineales observadas justificaron el uso de modelos avanzados como XGBoost y LSTM-CNN, cuya eficacia en entornos complejos está respaldada por la literatura (Chen & Guestrin, 2016; Hochreiter & Schmidhuber, 1997; Zheng et al., 2024). La combinación de estos enfoques permitió aprovechar sus fortalezas individuales, generando un modelo robusto.

Además, se implementó un sistema automatizado que optimiza hiperparámetros y splits entre porcentajes de datos de entrenamiento y de prueba dinámicos para maximizar el rendimiento de los modelos, validando los hallazgos de Carabotta y Claeys (2024) sobre la eficacia de las combinaciones. Las predicciones obtenidas se integraron mediante AIC, BIC, BMA y ponderaciones dinámicas basadas en Theil's U, permitiendo comparar su desempeño con las proyecciones institucionales y los valores reales del déficit fiscal.

3.1 Modelos y Combinaciones

3.1.1 Modelo XGBoost

XGBoost (*Extreme Gradient Boosting*) es un algoritmo de aprendizaje supervisado basado en árboles de decisión que se utiliza para tareas de regresión y clasificación. Su principal fortaleza radica en la capacidad para manejar relaciones no lineales y datos de alta dimensionalidad de manera eficiente. Este modelo optimiza las predicciones mediante una técnica de *boosting*, en la que los errores de los modelos anteriores son corregidos iterativamente en cada etapa (Chen & Guestrin, 2016).

3.1.1.1 Función Objetivo: XGBoost

Sea \hat{y}_i la predicción para una instancia i , entonces el modelo aprende aditivamente mediante la optimización del siguiente objetivo en la t -ésima iteración (1).

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (1)$$

donde:

$l(y_i, \hat{y}_i)$ es la función de pérdida que mide la diferencia entre los valores reales y_i y las predicciones \hat{y}_i .

$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$ es el término de regularización, que controla la complejidad del modelo penalizando el número de hojas T y los pesos w_j en cada árbol.

Cada árbol de decisión f_k se define como una función mapeada del espacio de características al espacio de predicción $f_k: \mathbb{R}^d \rightarrow \mathbb{R}$

3.1.1.2 Método de Optimización

En cada iteración, se optimiza el objetivo mediante la expansión del modelo previo $\hat{y}_i^{(t-1)}$ añadiendo un nuevo árbol f_t :

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (2)$$

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{1}{2} f''(a)(x - a)^2 \quad (3)$$

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i} \quad (4)$$

$$h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad (5)$$

El algoritmo utiliza una aproximación de segundo orden para actualizar los árboles, calculando los gradientes² de pérdida con respecto a la predicción (4) y los hessianos³ (segunda derivada) de la pérdida con respecto a la predicción (5). En cada iteración, los nuevos pesos del árbol se actualizan para minimizar la función objetivo basada en estos gradientes.

3.1.1.3 Estructura del Árbol

Cada árbol $f_k(x)$ se construye para minimizar los errores residuales de los árboles anteriores. El proceso de construcción incluye la ganancia representada en la ecuación (6).

$$Ganancia = \frac{1}{2} \left[\frac{(\sum g_i)^2}{\sum h_i + \lambda} - \frac{(\sum g_L)^2}{\sum h_L + \lambda} - \frac{(\sum g_R)^2}{\sum h_R + \lambda} \right] - \gamma \quad (6)$$

Donde:

g_i y h_i : Gradiente y Hessiano en el nodo izquierdo

g_R y h_R : Gradiente y Hessiana en el nodo derecho

γ : Penalización por dividir un nodo

3.1.1.4 Regularización y Poda

XGBoost emplea regularización explícita con un control del tamaño de los árboles mediante λ y γ y poda para reducir el sobreajuste, eliminando ramas irrelevantes al final de cada iteración para mejorar la simplicidad del modelo.

3.1.1.5 Personalización

² Gradientes que representan la tasa de cambio de la pérdida

³ Hessianos que evalúan la curvatura de la pérdida

El modelo selecciona automáticamente parámetros como la profundidad máxima de los árboles (*max – depth*), la ratio de aprendizaje (η) y el número de árboles (*n – estimators*) mediante una búsqueda optimizada.

3.1.2 Modelo LSTM-CNN

El modelo LSTM-CNN combina la capacidad de aprendizaje secuencial de las redes neuronales de memoria a largo plazo (LSTM) con la capacidad de extracción de características espaciales de las redes neuronales convolucionales (CNN), creando una arquitectura híbrida adecuada para el análisis de series temporales complejas y altamente correlacionadas.

3.1.2.1 Función Objetivo: LSTM-CNN

El modelo busca minimizar el error entre las predicciones \hat{y}_i y entre los valores reales y_i (se muestra en la ecuación (7))

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

Donde:

y_i Valor real del déficit fiscal

\hat{y}_i Predicción generada por el modelo

θ Conjunto de parámetros del modelo que se optimizan durante el entrenamiento

3.1.2.2 Método de Optimización

El modelo utiliza el algoritmo Adam (*Adaptive Moment Estimation*) para ajustar los parámetros. Adam combina los gradientes de primer y segundo orden, proporcionando una convergencia más rápida y estable.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(\theta_t) v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla L(\theta_t))^2 \quad (8) \quad (9)$$

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \widehat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad \theta_{t+1} = \theta_t - \eta \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \quad (10)(11)(12)$$

Donde:

η Tasa de aprendizaje

ϵ Factor de estabilización numérica

β_1 y β_2 Parámetros de decaimiento

3.1.2.3 Estructura de las redes neuronales

3.1.2.3.1 Componente CNN (Convolutivo)

La convolución se define como:

$$Z_{i,j,k} = \sigma \left(\sum_{m,n} X_{i+m,j+n} W_{m,n,k} + b_k \right) \quad (13)$$

Donde:

$Z_{i,j,k}$ Activación del mapa de características

X Entrada de datos normalizada (déficit fiscal)

$W_{m,n,k}$ Pesos del filtro convolutivo

b_k Sesgo del filtro

σ Función de activación (ReLU)

3.1.2.3.2 Componente LSTM

Cada celda LSTM sigue las siguientes ecuaciones.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (14)(15)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (16)(17)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad h_t = o_t \odot \tanh(C_t) \quad (18)(19)$$

Donde:

f_t, i_t, o_t Puertas de olvido, entrada y salida

\tilde{C}_t, h_t Estados de celda y estados ocultos

\odot Producto elemento a elemento

3.1.2.3.3 Integración CNN-LSTM

Las características extraídas por las capas convolucionales son procesadas como secuencias por las capas LSTM, lo que permite capturar patrones espaciales y temporales combinados.

3.1.2.3.4 Capa Densa (Fully Connected)

Genera la predicción final del déficit fiscal.

3.1.2.4 Regularización

Para evitar el sobreajuste, el modelo emplea

Dropout: Apaga neuronas de manera aleatoria durante el entrenamiento:

$$h_i = \begin{cases} \frac{h_i}{p}, & \text{con probabilidad } p \\ 0, & \text{con probabilidad } 1 - p \end{cases} \quad (20)$$

Donde:

p es la tasa de retención

Penalización L2: Regularización aplicada a los pesos de la red:

$$L_{\text{reg}} = L(\theta) + \frac{\lambda}{2} \sum \theta^2 \quad (21)$$

Donde:

λ Controla la penalización

3.1.2.5 Personalización

El modelo se ajusta dinámicamente mediante la selección de hiperparámetros clave, como el tamaño de los filtros en la CNN, el número de capas y unidades en la LSTM, y la tasa de aprendizaje η .

3.1.3 Modelo Combinado: Ponderación por AIC

El AIC (22) evalúa el ajuste de un modelo penalizando la complejidad para evitar sobreajuste.

$$AIC = n \cdot \log(MSE) + 2 \cdot k \quad (22)$$

La asignación de los pesos (23) se basa en la fórmula de Akaike.

$$w_{AIC} = \frac{e^{-\frac{1}{2}\Delta_i}}{\sum_{j=1}^m e^{-\frac{1}{2}\Delta_j}} \quad (23)$$

Donde $\Delta_i = AIC_i - AIC_{min}$ es la diferencia entre el AIC del modelo i y el menor AIC entre todos los modelos por lo que los modelos con menor AIC obtienen mayor peso.

3.1.4 Modelo Combinado: Ponderación por BIC

El BIC (24) penaliza más severamente la complejidad del modelo que el AIC, favoreciendo modelos más simples en conjuntos de datos grandes.

$$BIC = n \cdot \log(MSE) + k \cdot \log(n) \quad (24)$$

La asignación de los pesos se basa en la fórmula de Bayesiano (25)

$$w_{BIC} = \frac{e^{-\frac{1}{2}\Delta_i}}{\sum_{j=1}^m e^{-\frac{1}{2}\Delta_j}} \quad (25)$$

Donde $\Delta_i = BIC_i - BIC_{min}$ es la diferencia entre el BIC del modelo i y el menor BIC entre todos los modelos por lo que los modelos con menor BIC obtienen mayor peso.

3.1.5 Modelo Combinado: Ponderación por BMA

El BMA (Bayesian Model Averaging) integra múltiples modelos basándose en su probabilidad posterior, considerando la incertidumbre en la selección del modelo.

La probabilidad posterior para un modelo i $P(M_i|D)$ se calcula como en la ecuación (26).

$$P(M_i|D) = \frac{P(D|M_i) \cdot P(M_i)}{\sum_{j=1}^m P(D|M_j) \cdot P(M_j)} \quad (26)$$

Los pesos (w_{BMA}) del BMA (27) son equivalentes a las probabilidades posteriores.

$$w_{BMA} = P(M_i|D) \quad (27)$$

3.1.6 Modelo Combinado: Ponderación Dinámica por Theil's U

El estadístico Theil's U (28) compara la precisión del modelo con un *benchmark* (normalmente un modelo *naïve*).

$$Theil's U = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n \hat{y}_i^2}} \quad (28)$$

Un valor de Theil's U menor que 1 indica que el modelo es mejor que el *benchmark*; valores mayores a 1 indican un peor rendimiento.

Los pesos dinámicos $w_{Theil U,t}$ se asignan en cada fecha t en función del valor relativo de Theil's U para los modelos (29).

$$w_{Theil's U, t} = \frac{1}{\frac{Theil's U_t}{\sum_{j=1}^m \frac{1}{Theil's U_{j,t}}}} \quad (29)$$

A diferencia de los métodos estáticos (AIC, BIC, BMA), este enfoque adapta los pesos para cada fecha según el rendimiento actual de los modelos. Por lo que modelos con menor Theil's U obtienen mayor peso.

3.2 Métricas

La evaluación de los modelos se realiza mediante un conjunto de métricas que permiten medir la precisión, robustez y capacidad explicativa del modelo frente a los valores reales y las predicciones de las instituciones. A continuación, se describen las métricas utilizadas con sus fórmulas y explicaciones.

3.2.1 Error Cuadrático Medio (MSE)

El Error Cuadrático Medio mide la magnitud promedio de los errores al cuadrado entre los valores predichos \hat{y}_i entre los valores reales y_i .

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (30)$$

Valores más bajos de MSE indican un modelo con mejores predicciones. Penaliza fuertemente los errores grandes.

3.2.2 Error Absoluto Medio (MAE)

El Error Absoluto Medio mide el promedio de las diferencias absolutas entre los valores reales y las predicciones:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (31)$$

Proporciona una métrica menos sensible a errores grandes en comparación con MSE, reflejando el error promedio en las predicciones.

3.2.3 Error Raíz del Error Cuadrático Medio (RMSE)

La Raíz del Error Cuadrático Medio es la raíz cuadrada del MSE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (32)$$

Similar al MSE, pero se expresa en las mismas unidades que los datos originales, facilitando su interpretación.

3.2.4 Coeficiente de Determinación (R^2)

El R^2 mide qué tan bien las predicciones explican la variabilidad de los datos reales:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (33)$$

Un R^2 cercano a 1 indica que el modelo explica casi toda la variabilidad; valores negativos indican un mal ajuste.

3.2.5 Autocorrelación (ACF y PACF)

La función de autocorrelación mide la relación entre los valores de una serie temporal y sus propios retardos pasados. Es útil para identificar patrones repetitivos, como estacionalidad, y determinar la dependencia temporal en los datos.

$$ACF(k) = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (34)$$

Una alta autocorrelación en ciertos retardos puede indicar patrones estacionales.

La función de autocorrelación parcial mide la correlación directa entre un valor y sus retardos, eliminando los efectos intermedios.

$$\text{PACF}(k) = \text{corr}(y_t, y_{t-k} \mid y_{t-1}, \dots, y_{t-k+1}) \quad (35)$$

Un gráfico de PACF ayuda a identificar cuántos retardos son relevantes para explicar la dinámica de la serie.

3.3 Pruebas

3.3.1 Prueba de Sesgo (*Bias Test*)

Evalúa si las predicciones están sistemáticamente sesgadas. Esto se hace ajustando una regresión lineal sobre los residuos ($e_i = y_i - \hat{y}_i$). Supone que las predicciones son imparciales y no están influenciadas por errores sistemáticos.

Hipótesis Nula (H_0): No hay sesgo sistemático en las predicciones ($\beta_0 = 0$)

$$e_i = \beta_0 - \epsilon_i \quad (36)$$

Si el coeficiente β_0 no es significativamente diferente de 0, no hay sesgo sistemático en las predicciones.

3.3.2 Prueba de *Stickiness*

La prueba de *Stickiness* se utiliza para evaluar la persistencia de los errores en las predicciones a lo largo del tiempo, detectando si los errores tienden a propagarse o son independientes entre períodos consecutivos (calculado mediante las revisiones ($r_t = \hat{y}_t - y_t$)). Este concepto es crucial en modelos económicos, ya que errores persistentes pueden indicar problemas sistemáticos en la metodología predictiva.

Hipótesis Nula (H_0): Los errores de predicción no son persistentes (independencia temporal de los errores) con $\phi = 0$.

El modelo utiliza las revisiones de predicciones r_t como variable dependiente y sus valores en períodos anteriores r_{t-1} como regresores. La especificación del modelo es la siguiente se presenta en la ecuación (37).

$$r_t = \alpha + \phi \cdot r_{t-1} + \epsilon_t \quad (37)$$

Si ϕ (coeficiente de autocorrelación) es cercano a 0, los errores no son persistentes; valores altos indican que los errores de predicción tienden a propagarse.

3.3.3 Prueba de Dickey-Fuller Aumentado (ADF)

La prueba ADF (Dickey & Fuller, 1979) se utiliza para verificar si una serie temporal es estacionaria de carácter general. Es decir, evalúa si la serie tiene una raíz unitaria, lo que indicaría no estacionariedad.

Hipótesis Nula (H_0): La serie tiene una raíz unitaria; por lo tanto, no es estacionaria.

La ecuación del modelo con constante y tendencia (nivel y tendencia) se muestra en la ecuación (38).

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t \quad (38)$$

El estadístico de la prueba (39) es la estimación $\hat{\phi}$ del coeficiente de y_{t-1} en el modelo ajustado, comparado con valores críticos para determinar si $\hat{\phi} < 0$

Rechazo de H_0 .

$$\tau = \frac{\hat{\phi}}{SE(\hat{\phi})} \quad (39)$$

3.3.4 Prueba de HEGY

La prueba de Hylleberg, Engle, Granger y Yoo (HEGY) evalúa la presencia de raíces unitarias estacionales en series temporales con frecuencia definida (por ejemplo, trimestral o anual) con el fin de determinar si los patrones en una serie temporal son estables o cambian con el tiempo. La prueba supone que las series pueden descomponerse en componentes determinísticas (tendencias, estacionalidades) y aleatorias.

Hipótesis Nula (H_0): la serie tiene raíces unitarias estacionales, lo que indica no estacionariedad estacional.

$$\Delta y_t = \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \sum_{i=1}^{s-1} \phi_i \Delta y_{t-i} + \epsilon_t \quad (40)$$

Donde π_1, π_2, ϕ_i son los coeficientes que se pretenden estimar.

La hipótesis nula evalúa si hay raíces unitarias estacionales $\pi = 0$ y, por tanto, la serie no es estacionaria en sus componentes estacionales. La prueba de HEGY proporciona información sobre qué componentes estacionales presentan raíces unitarias, lo que ayuda a identificar patrones estacionales.

En el apartado 4, se presentan los resultados obtenidos para cada uno de los modelos y las combinaciones descritas en este apartado. Asimismo, se detallan los valores calculados en los diferentes tests y métricas de evaluación, lo que permite identificar tanto las ventajas como las limitaciones de cada enfoque implementado.

4 Resultados

En este apartado se presentan los resultados derivados de la implementación de las metodologías descritas previamente (sección 3), con el objetivo de evaluar la capacidad predictiva y el rendimiento de los modelos desarrollados frente a las tres instituciones seleccionadas. La importancia de este análisis radica en demostrar cómo las técnicas avanzadas de *machine learning* (XGBoost y LSTM-CNN) combinadas mediante métodos de ponderación y ensamblaje, superan las limitaciones (especificadas en la sección 1) de las predicciones tradicionales utilizadas por las instituciones analizadas.

El análisis de resultados está estructurado para proporcionar una visión integral del desempeño de los modelos. En primer lugar, se detallan los outputs individuales de XGBoost y LSTM-CNN, argumentando y comprobando las fortalezas y características específicas de cada uno mediante los resultados obtenidos. En este contexto, se evalúan métricas clave, como el Error Cuadrático Medio (MSE), el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE), el Coeficiente de Determinación (R^2), y el estadístico Theil's U, para medir la precisión, robustez y generalización de las predicciones.

Asimismo, el apartado incluye una sección dedicada al análisis de las combinaciones de modelos mediante técnicas de ensamblaje como AIC, BIC, BMA y ponderaciones dinámicas basadas en Theil's U. Estas combinaciones permiten aprovechar las fortalezas individuales de cada modelo y mitigar sus debilidades, resultando en predicciones más precisas y confiables. Este enfoque innovador no solo valida la literatura existente, como los hallazgos de Carabotta y Claeys (2024), sino que también introduce una metodología nueva y adaptativa

específicamente desarrollada en este trabajo (ponderación dinámica mediante el estadístico Theil's U).

Posteriormente, se realiza una comparación exhaustiva entre los modelos desarrollados mediante la combinación de los modelos (XGBoost y LSTM-CNN) por cuatro métodos distintos y las proyecciones generadas por las instituciones seleccionadas (Econ Intelligence Unit, First Trust Advisors y JPMorgan). Esta comparación no solo busca identificar diferencias en el rendimiento, sino también analizar cómo los enfoques basados en datos y aprendizaje automático pueden abordar los sesgos y limitaciones inherentes a las metodologías tradicionales. Las instituciones, a menudo influenciadas por decisiones macroeconómicas y datos ruidosos, tienden a mostrar predicciones con errores sistemáticos, lo que refuerza la necesidad de explorar soluciones alternativas más adaptativas.

Finalmente, los resultados son interpretados detalladamente para identificar las fortalezas, limitaciones y áreas de mejora de los modelos presentados. Además, se reflexiona sobre la aplicabilidad y escalabilidad de estas metodologías en otros contextos económicos, destacando el potencial de estos enfoques para transformar la manera en que se predice y gestiona el déficit fiscal. Este análisis integral busca no solo validar la viabilidad de las metodologías propuestas, sino también sentar las bases para futuras investigaciones en este campo.

4.1 Resultados de cada modelo

La metodología implementada en este trabajo ha sido diseñada para maximizar la precisión y adaptabilidad de los modelos mediante la automatización de varios procesos. En primer lugar, la selección de parámetros se realizó mediante una búsqueda en cuadrícula, siguiendo las recomendaciones de Bergstra y Bengio (2012). Este enfoque permite explorar sistemáticamente combinaciones de

hiperparámetros, buscando que los modelos sean configurados de manera óptima. Además, se automatizó la selección de los porcentajes para los splits de entrenamiento, evaluando rangos dinámicos del 50% al 90% de los datos disponibles, según las mejores prácticas descritas por Kohavi (1995). Esto garantiza que los modelos sean entrenados con configuraciones que maximizan su capacidad predictiva, optimizando automáticamente en función del estadístico Theil's U.

Uno de los principales desafíos abordados fue encontrar el equilibrio adecuado entre precisión y eficiencia, un aspecto crítico en la automatización de modelos avanzados, como señalan Domingos (2012). Este equilibrio ha sido central en el diseño de la metodología, dado que la búsqueda de una alta precisión no debe comprometer la viabilidad computacional ni la capacidad de generalización de los modelos.

Finalmente, los modelos han sido evaluados utilizando métricas que incluyen medidas de errores absolutos y relativos, ofreciendo una visión detallada de la precisión, robustez y capacidad predictiva de los modelos. Esta combinación de automatización y evaluación rigurosa constituye la base para los resultados y análisis presentados en los apartados siguientes.

4.1.1 XGBoost

4.1.1.1 Parámetros del XGBoost

El modelo desarrollado mediante XGBoost (*Extreme Gradient Boosting*) se ha realizado con los siguientes hiperparámetros (Tabla 5):

Hiperparámetros utilizados en XGBoost para la automatización			
Parámetros del GridSearchCV			
Numero de estimadores	50	100	150
Tasa de aprendizaje	0,01	0,1	0,2
Profundidad del árbol	3	5	7
Características del XGBoost (Formula)			
Objetivo	reg: squarederror		
CV	3		
Scoring	theil_scorer		

Fuente: Elaboración Propia

Tabla 5. Hiperparámetros utilizados en XGBoost para la automatización

Mediante el uso de *GridSearchCV*, se exploraron diversas combinaciones de hiperparámetros con el objetivo de identificar la configuración que maximiza el rendimiento del modelo. A continuación, se detallan los aspectos clave de este proceso:

El número de estimadores define el número de árboles de decisión que el modelo construye durante cada iteración de entrenamiento.

La Tasa de Aprendizaje controla la contribución de cada árbol al modelo final y se entiende como la tasa de ajustes que realiza el modelo (cuanto más pequeño, más preciso, pero menos eficiente). Se seleccionaron para probar tasas de aprendizaje conservadoras (0,01), intermedias (0,10), y más agresivas (0,20).

La Profundidad del Árbol define la profundidad máxima de cada árbol al modelo final.

Se configuró el modelo *XGBoost* para automatizar la búsqueda de hiperparámetros utilizando *GridSearchCV*, optimizando el error cuadrático medio (MSE) mediante validación cruzada (CV). La validación cruzada consiste en dividir los datos de entrenamiento en varios subconjuntos, en este caso 3, para entrenar y validar el modelo de forma alternada, evaluando así su capacidad de

generalización a nuevos datos. Esto asegura que los resultados sean estables y robustos, minimizando el riesgo de sobreajuste.

Para evaluar el rendimiento relativo de las predicciones frente a los valores reales, se utilizó *Theil's U* como métrica personalizada de puntuación. Esta métrica es particularmente adecuada para los objetivos del proyecto, ya que permite analizar la precisión relativa del modelo en comparación con una línea base o *benchmark*.

En conclusión, la configuración implementada permitió una exploración exhaustiva de los hiperparámetros del modelo, optimizando su precisión y eficiencia. La validación cruzada, combinada con el uso de métricas especializadas, asegura que los resultados obtenidos sean tanto precisos como generalizables.

4.1.1.2 Resultados del XGBoost

Los mejores resultados de cada uno de los splits se presentan a continuación (Tabla 6):

Resultados del modelo XGBoost por Split													
Train	Test	MSE	RMSE	MAE	R2	AIC	BIC	Theil's U	Bias p-value	Stickiness p-value	E	LR	MD
90%	9%	0,0004	0,020	0,008	0,748	-73	-72	0,127	0,230	0,778	100	0,1	7
80%	19%	0,0004	0,020	0,010	0,685	-143	-140	0,136	0,022	0,090	100	0,1	3
70%	30%	0,0003	0,016	0,008	0,821	-233	-229	0,134	0,010	0,030	100	0,1	5
60%	40%	0,0002	0,014	0,006	0,822	-317	-312	0,127	0,007	0,011	50	0,2	5
50%	50%	0,0002	0,013	0,006	0,823	-412	-406	0,122	0,002	0,004	50	0,2	3

Fuente: Elaboración Propia

Tabla 6. Mejores resultados del XGBoost para cada split

En la Tabla 6, se destaca el split 50%-50%, sombreado en verde, como la configuración óptima seleccionada automáticamente por su menor estadístico de Theil's U. Este resultado subraya la capacidad del modelo para lograr predicciones precisas, superando significativamente al benchmark evaluado (detallado en la sección 4.2.1). Además, los parámetros óptimos seleccionados incluyen 50 árboles,

una tasa de aprendizaje de 0,2 y una profundidad máxima de 3, equilibrando precisión y eficiencia.

El estadístico Theil's U refleja la precisión relativa del modelo en comparación con un benchmark, con un valor mínimo de 0,122 en el split 50%-50%, destacando la efectividad de este enfoque para mejorar las predicciones. Este split también presenta los mejores resultados en las métricas de error: el Error Cuadrático Medio (MSE) y su raíz cuadrada (RMSE) alcanzan valores mínimos de 0,0002 y 0,013, respectivamente, indicando un ajuste preciso del modelo. Por su parte, el Error Absoluto Medio (MAE), con un valor de 0,006, refuerza la estabilidad del modelo en esta configuración.

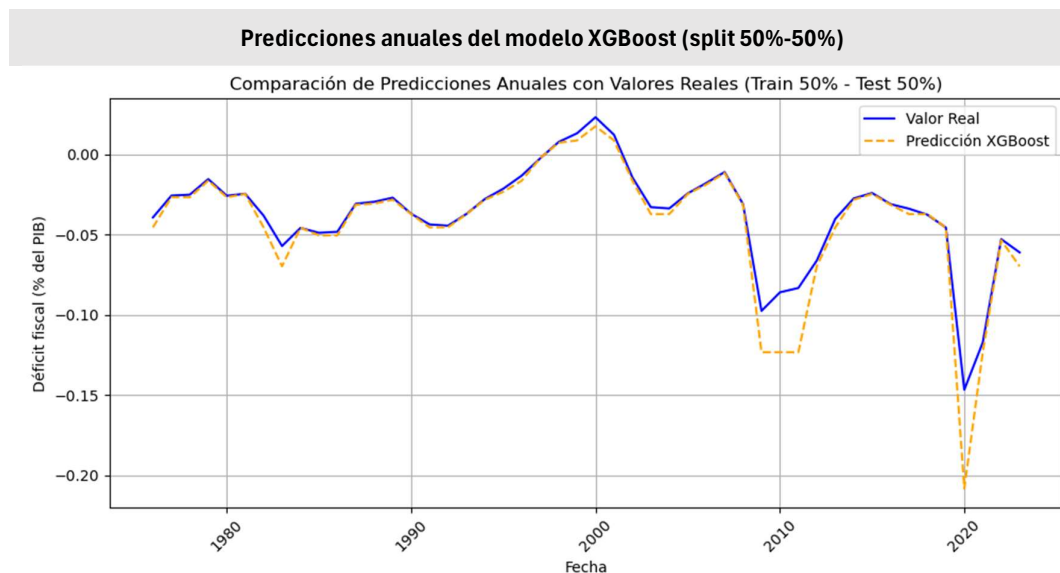
El Coeficiente de Determinación (R^2), que mide qué tan bien las predicciones explican la variabilidad de los datos reales, alcanza un valor máximo de 82,3% en el split 50%-50%, evidenciando que esta configuración logra capturar de manera óptima las variaciones en los datos.

Los criterios de selección del modelo, Akaike (AIC) y Bayesiano (BIC), también confirman la superioridad del split 50%-50%, con valores mínimos de -412 y -406, respectivamente. Estos resultados indican que el modelo no solo es preciso, sino también eficiente en términos de complejidad.

En cuanto a robustez, las pruebas de sesgo y stickiness refuerzan la confiabilidad del modelo. Los p-valores bajos en la prueba de sesgo para los splits 50%-50% (0,002) y 60%-40% (0,007) confirman la ausencia de sesgos sistemáticos en las predicciones. Por otro lado, los p-valores altos en la prueba de stickiness en todos los splits indican que los errores no son persistentes, validando la capacidad del modelo para adaptarse a las fluctuaciones temporales sin comprometer su desempeño.

En resumen, los resultados obtenidos demuestran que el split 50%-50% no solo es la configuración más equilibrada y precisa, sino también la más robusta, lo que subraya la eficacia de los modelos desarrollados y la metodología empleada.

La representación gráfica del mejor resultado del modelo XGBoost se muestra en la Figura 6 donde se puede ver los valores reales frente a las predicciones realizadas anualmente entre 1981 y 2022.



Fuente: Elaboración Propia

Figura 6. Predicciones anuales del modelo XGBoost (split 50%-50%)

El modelo XGBoost logra capturar patrones tanto de tendencias generales como las variaciones específicas en los datos históricos. La gráfica muestra evidencias que el modelo XGBoost es capaz de identificar patrones significativos y predecir con precisión los cambios en el déficit fiscal. Particularmente, el modelo refleja con éxito las caídas abruptas, como las registradas durante la crisis financiera de 2008 y el impacto de la pandemia de COVID-19 en 2020.

En los períodos de menor volatilidad, como las décadas de 1980 y 1990, las predicciones son casi indistinguibles de los valores reales, lo que destaca la robustez del modelo en contextos estables. Sin embargo, en los puntos de mayor

variabilidad, como en 2008 y 2020, se observan pequeñas discrepancias en la magnitud de los cambios, lo cual es esperable en cualquier modelo predictivo cuando enfrenta eventos extremos e inesperados.

En conjunto, la gráfica valida el enfoque metodológico adoptado y resalta el potencial del modelo para aplicaciones prácticas en escenarios futuros.

4.1.2 LSTM-CNN

4.1.2.1 Parámetros del LSTM-CNN

El modelo desarrollado mediante LSTM-CNN se ha realizado mediante las siguientes especificaciones (Tabla 7):

Hiperparámetros utilizados en LSTM-CNN para la automatización	
Parámetros del Grid Search	
Capas de Convolución (Conv1D)	1 capa: 64 filtros, kernel size 2, activación ReLU
Capa de MaxPooling (MaxPooling1D)	Pool size 2
Capas Recurrentes (LSTM)	2 capas de 50 unidades cada una
Dropout (Dropout)	20% en ambas capas
Capa de Salida (Dense)	1 neurona para regresión

Fuente: Elaboración Propia

Tabla 7. Hiperparámetros utilizados en LSTM-CNN para la automatización

En este modelo también se ha empleado un buscador de parámetros para las distintas combinaciones de hiperparámetros, como el número de filtros en la capa convolucional, unidades en las capas LSTM, tasas de *dropout* y tipos de optimizadores, con el fin de encontrar la configuración que optimiza el rendimiento del modelo.

En la capa inicial, una red convolucional (Conv1D) con 64 filtros y un tamaño de kernel de 2 se encarga de extraer patrones locales significativos de las secuencias temporales. Esta capa, al identificar fluctuaciones y tendencias en los datos, proporciona una base sólida para que las capas posteriores procesen las

características más relevantes. A esto se añade una capa de MaxPooling1D, que reduce la dimensionalidad de los datos, mejorando la eficiencia del modelo y ayudando a prevenir el sobreajuste.

El núcleo de la arquitectura lo forman dos capas LSTM consecutivas, diseñadas para manejar dependencias temporales complejas. La primera capa LSTM, con 50 unidades, procesa las secuencias generadas por la capa convolucional y captura patrones temporales a largo plazo. A su vez, esta capa devuelve una salida en forma de secuencia, que es refinada por la segunda capa LSTM, también con 50 unidades, encargada de generar una representación más compacta y enriquecida de la serie temporal. Este enfoque en cascada permite modelar tanto las relaciones inmediatas como las de largo plazo en los datos.

Para mitigar el riesgo de sobreajuste y mejorar la generalización del modelo, se incluyen dos capas de Dropout, que desactivan aleatoriamente el 20% de las conexiones durante el entrenamiento. Este mecanismo resulta particularmente útil para trabajar con conjuntos de datos ruidosos, asegurando que el modelo mantenga su robustez frente a fluctuaciones en los datos.

Finalmente, la predicción del déficit fiscal se realiza a través de una capa densa (Dense) de salida, que consta de una única neurona encargada de traducir las características aprendidas en un valor predictivo. Este diseño simplificado garantiza una interpretación clara y precisa de las predicciones generadas por el modelo.

Esta configuración permite explorar distintas combinaciones de filtros, unidades LSTM y distintos *dropouts*, optimizando la precisión y eficiencia del modelo.

4.1.2.2 Resultados del LSTM-CNN

Los mejores resultados de cada uno de los splits se presentan a continuación (Tabla 8):

Resultados del modelo LSTM-CNN por Split										
Train	Test	MSE	RMSE	MAE	R2	AIC	BIC	Theil's U	Bias p-value	Stickiness p-value
90%	9%	0,0029	0,054	0,036	-0,824	-25	-26	0,373	0,183	0,457
80%	19%	0,0019	0,043	0,030	-0,529	-84	-83	0,383	0,092	0,019
70%	30%	0,0013	0,036	0,026	0,128	-156	-154	0,352	0,552	0,004
60%	40%	0,0011	0,034	0,024	0,140	-220	-217	0,378	0,585	0,000
50%	50%	0,0010	0,031	0,023	0,039	-293	-290	0,334	0,344	0,000

Fuente: Elaboración Propia

Tabla 8. Mejores resultados del LSTM-CNN para cada split

En la Tabla 8, se destaca el split 50%-50%, sombreado en verde, como la configuración óptima seleccionada automáticamente por su menor estadístico de Theil's U.

El estadístico Theil's U refleja la precisión relativa del modelo en comparación con el modelo *naïve*, logrando un valor mínimo de 0,334 en el split 50%-50%, lo que destaca la efectividad de este enfoque para mejorar las predicciones. Este split también presenta los mejores resultados en las métricas de error: el Error Cuadrático Medio (MSE) y su raíz cuadrada (RMSE) alcanzan valores mínimos de 0,0010 y 0,031, respectivamente, indicando un ajuste preciso del modelo. Por su parte, el Error Absoluto Medio (MAE), con un valor de 0,023, refuerza la estabilidad del modelo en esta configuración. El menor MSE y RMSE en el split 50%-50% indican que este ajuste es el más preciso entre las configuraciones analizadas.

El Coeficiente de Determinación (R^2), que mide qué tan bien las predicciones explican la variabilidad de los datos reales, resulta ser un valor bajo de 3,9% en el split 50%-50%, reflejando que, aunque las predicciones son precisas en términos

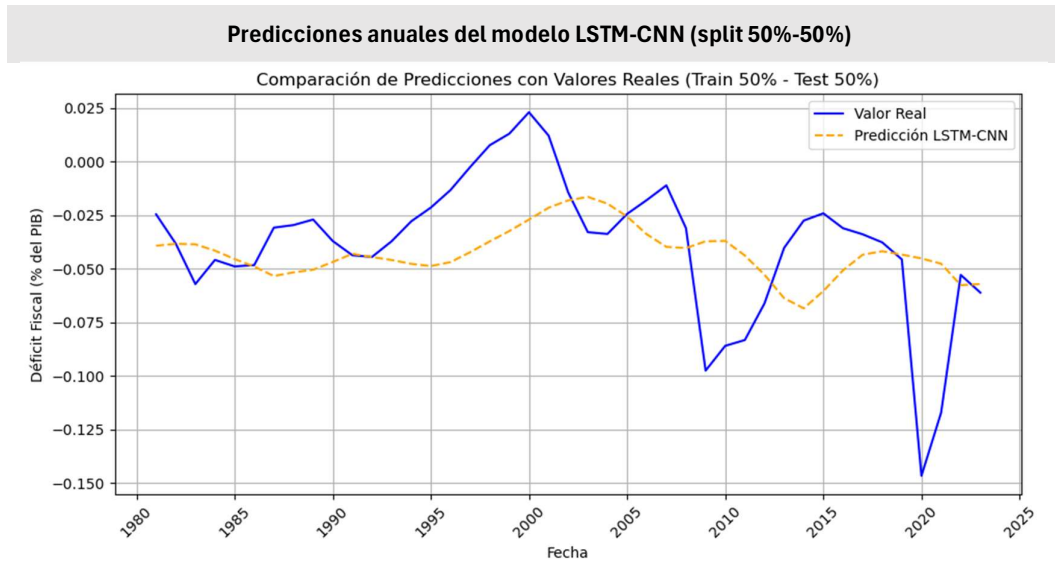
de error absoluto, el modelo aún no captura toda la estructura subyacente de la serie temporal. Este valor es bajo en comparación con otros modelos (como XGBoost), lo que indica que el modelo captura menos de las variaciones en los datos reales. Sin embargo, puede deberse a la complejidad intrínseca de los datos fiscales.

Los criterios de selección del modelo, Akaike (AIC) y Bayesiano (BIC), también confirman la superioridad del split 50%-50%, con valores mínimos de -293 y -290, respectivamente, lo que indica que esta configuración es la más eficiente al balancear precisión y complejidad del modelo.

En cuanto a robustez, las pruebas de sesgo y stickiness aportan información clave sobre el desempeño del modelo. Los p-valores de la prueba de sesgo para el split 50%-50% (0,348) sugieren que no hay evidencia estadísticamente significativa de sesgos sistemáticos en las predicciones, lo que refuerza la confiabilidad del modelo en este aspecto. Sin embargo, los p-valores bajos en la prueba de stickiness (0,000) indican una persistencia en los errores a lo largo del tiempo, lo que señala una posible área de mejora. Esto sugiere que el modelo podría beneficiarse de ajustes adicionales, como la incorporación de técnicas que mitiguen la dependencia temporal de los errores.

El split 50%-50% se mantiene como la configuración más equilibrada y precisa para el modelo LSTM-CNN, logrando los mejores resultados en métricas como MSE, RMSE, y criterios de información (AIC y BIC). Sin embargo, el bajo R^2 y la persistencia de errores reflejan que el modelo aún enfrenta desafíos para capturar completamente la dinámica del déficit fiscal. Complementar el modelo con variables exógenas podría mejorar su desempeño global.

La representación gráfica del mejor resultado del modelo LSTM-CNN se muestra en la Figura 7, donde se puede ver los valores reales frente a las predicciones anuales realizadas entre 1981 hasta 2022.



Fuente: Elaboración Propia

Figura 7. Comparación de las predicciones el déficit real (Train 60% - Test 40%)

El modelo LSTM-CNN logra capturar patrones cíclicos y tendencias generales del déficit fiscal, pero presenta desviaciones significativas en varios periodos clave. Su principal fortaleza es la capacidad de combinar la detección de patrones locales mediante convoluciones con el aprendizaje de dependencias temporales a largo plazo a través de LSTM. Sin embargo, su precisión disminuye en periodos de alta variabilidad del déficit fiscal, lo que indica que el modelo podría beneficiarse de la incorporación de datos adicionales que reflejen mejor las fluctuaciones extremas.

4.1.3 Análisis de los *forecast errors*: XGBoost y LSTM – CNN

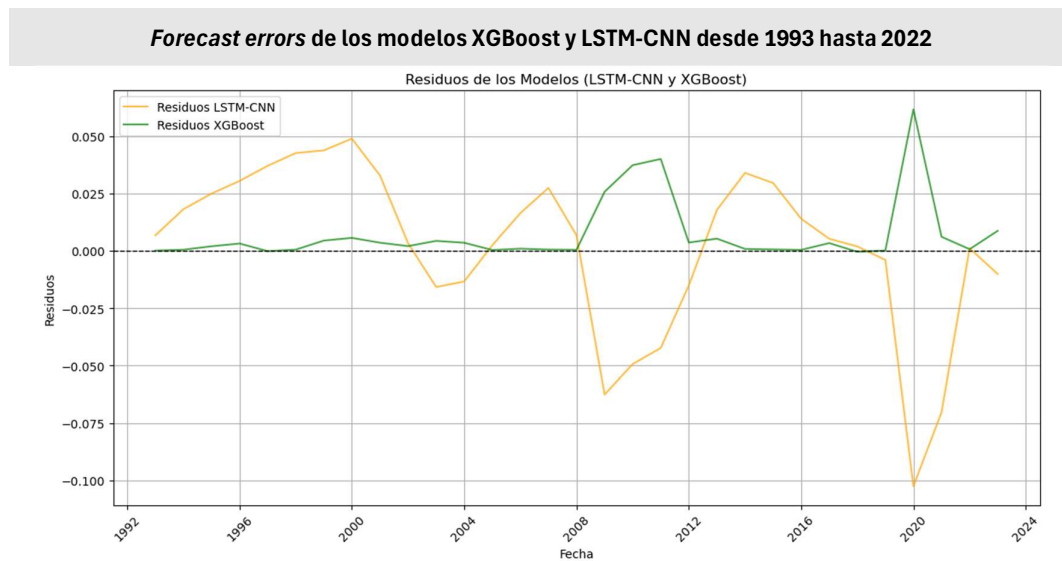
Un análisis preliminar de las métricas para valorar los *forecast errors* obtenidos en el train set (50%-50%) de cada modelo se muestra en la Tabla 9 .

Métricas de los <i>forecast errors</i>		
Test	LSTM-CNN	XGBoost
ADF Stat	-3,079	-1,311
ADF p-value	0,028	0,624
KPSS Stat	0,337	0,226
KPSS p-value	0,100	0,100
ARCH Stat	5,111	1,266
ARCH p-value	0,530	0,974
Shapiro-Wilk Stat	0,898	0,546
Shapiro-Wilk p-value	0,007	0,000
Jarque-Bera Stat	8,099	82,738
Jarque-Bera p-value	0,017	0,000
Ljung-Box Stat	20,606	12,204
Ljung-Box p-value	0,024	0,272

Fuente: *Elaboración Propia*

Tabla 9. Métricas principales de los *forecast errors* de los modelos: XGBoost y LSTM – CNN

La representación gráfica de los *forecast errors* de los dos modelos predichos se muestra en la Figura 8.



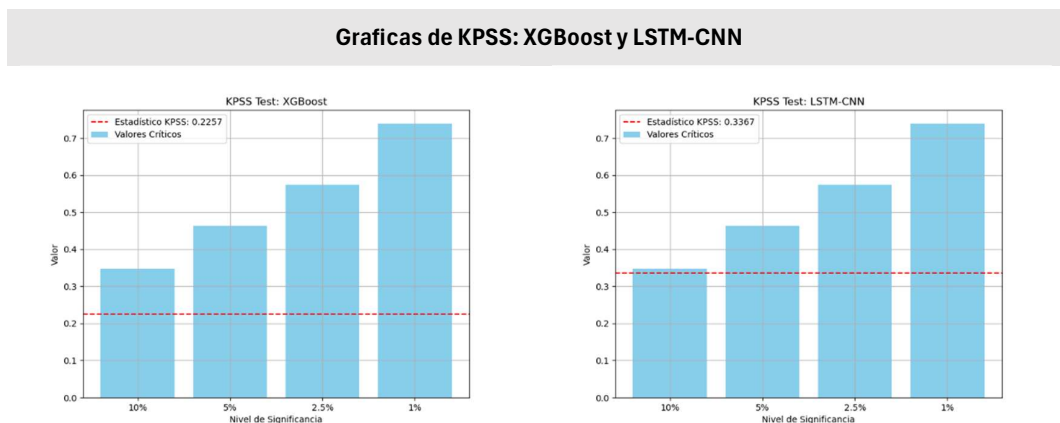
Fuente: *Elaboración Propia*

Figura 8. Representación gráfica de los *forecast errors* de XGBoost y LSTM-CNN desde 1993-2022

El análisis de los *forecast errors* para los modelos LSTM-CNN y XGBoost revela información importante sobre su desempeño y limitaciones. En primer lugar, en la gráfica de los *forecast errors* a lo largo del tiempo (Figura 8), el modelo LSTM-CNN muestra fluctuaciones moderadas, con valores que permanecen relativamente cercanos a la línea base de cero. Esto refleja su capacidad para modelar dependencias temporales, aunque se observan ligeros desajustes en periodos de alta variabilidad. Por otro lado, los *forecast errors* del modelo XGBoost muestran mayor estabilidad y cercanía a la línea base, destacando por su precisión en períodos normales, aunque presentan picos en eventos extremos, lo que indica que este modelo puede ser más sensible a cambios abruptos en los datos.

Por un lado, el análisis de autocorrelación (ACF) y autocorrelación parcial (PACF) muestra que ambos modelos presentan una correlación significativa en el primer año. Sin embargo, estas correlaciones disminuyen rápidamente, indicando que las dependencias más largas han sido adecuadamente modeladas.

El estadístico KPSS⁴ evalúa la estacionariedad de la serie (Figura 9).



Fuente: Elaboración Propia

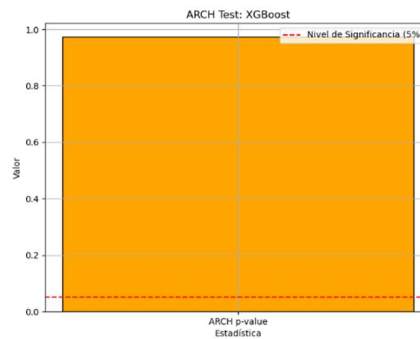
Figura 9. Graficas de la prueba KPSS: XGBoost y LSTM – CNN

⁴ Kwiatkowski-Phillips-Schmidt-Shin

Como el estadístico KPSS supera los valores críticos para cada uno de los niveles de significancia, se puede rechazar la hipótesis nula de estacionariedad. El tests sugiere estacionariedad en los *forecast errors*.

La prueba de ARCH⁵ para XGBoost evalúa la heteroscedasticidad en los *forecast errors*.

Test de ARCH: XGBoost

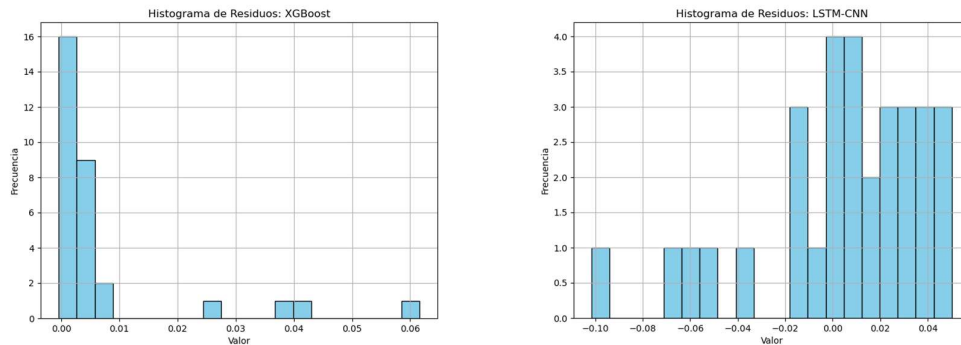


Fuente: Elaboración Propia

Figura 10. Test de ARCH: XGBoost

Para XGBoost, el p-valor es alto, lo que confirma que no hay heterocedasticidad en los *forecast errors*, reforzando la estabilidad del modelo.

Histograma de Forecast Errors: XGBoost y LSTM-CNN



Fuente: Elaboración Propia

Figura 11. Histograma de forecast errors: XGBoost y LSTM-CNN

⁵ Autoregressive Conditional Heteroskedasticity

El histograma de los *forecast errors* (Figura 11) revela que los *forecast errors* de ambos modelos están centrados en torno a 0, pero no presentan una distribución normal. Esto se confirma en el gráfico QQ plot (Figura 12), donde los *forecast errors* de ambos modelos se desvían de la línea teórica, evidenciando que las predicciones no se ajustan completamente a una distribución gaussiana. Según la prueba de Shapiro, los *forecast errors* no son normales y la prueba de Jarque-Bera afirma la misma hipótesis. Este comportamiento puede estar influido por la naturaleza compleja y no lineal del déficit fiscal, que introduce características difíciles de modelar con precisión.

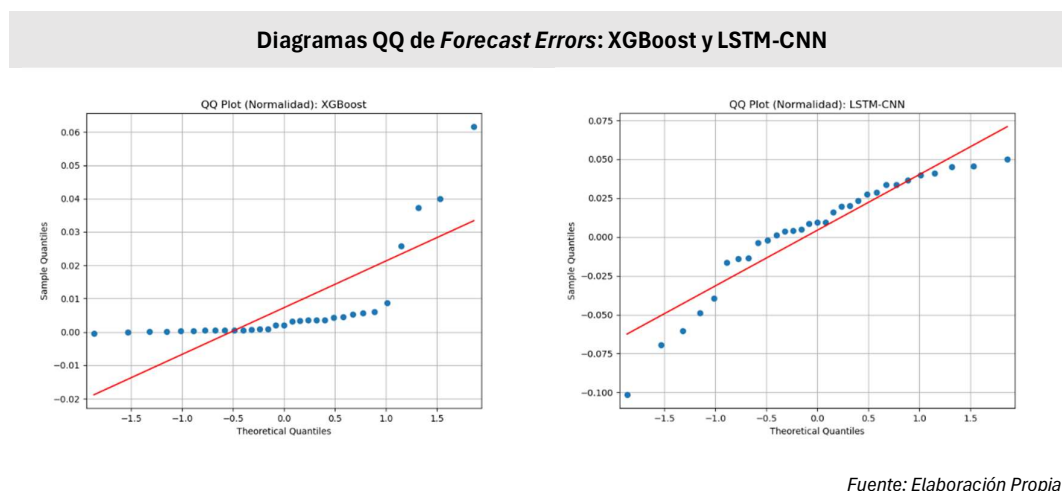


Figura 12. Diagramas QQ de forecast errors: XGBoost y LSTM – CNN

Por último, el análisis del estadístico Ljung-Box confirma que no hay autocorrelaciones significativas en los *forecast errors* a niveles más altos de lag, especialmente en el modelo XGBoost. Esto refuerza la idea de que los modelos han logrado capturar correctamente las dependencias temporales dentro del rango de lag analizado, aunque las fluctuaciones residuales observadas en el modelo LSTM-CNN podrían mejorarse mediante ajustes adicionales o la incorporación de más datos de entrenamiento.

En conjunto, estos resultados destacan las fortalezas de ambos modelos: LSTM-CNN por su capacidad de capturar patrones temporales complejos y XGBoost por su precisión general y manejo eficiente de los *forecast errors*. Sin embargo, ambos presentan margen de mejora en cuanto a garantizar una distribución más normal de los *forecast errors*. Esto subraya la importancia de utilizar técnicas complementarias para abordar la complejidad del problema del déficit fiscal.

4.1.4 Modelo Combinado

La intención del modelo combinado es mejorar la precisión, robustez y eficiencia de las predicciones, integrando los resultados de los dos modelos mediante cuatro métodos de combinación distintos. El objetivo es evaluar y determinar el más adecuado para predecir el déficit fiscal. Es importante mencionar que se utilizará el mejor split de cada modelo, y el modelo combinado adoptará el split más restrictivo para garantizar que se disponga de datos consistentes.

4.1.4.1 Parámetros del modelo combinado

El modelo combinado se implementó utilizando cuatro métodos distintos de combinación para evaluar y seleccionar el más adecuado para predecir el déficit fiscal. Los cuatro métodos que se detallarán a continuación son: ponderación por el parámetro AIC, ponderación por el parámetro BIC, ponderación por *Bayesian Model Averaging* (BMA) y por último una ponderación según pesos dinámicos según el estadístico Theil's U en cada fecha.

El primer método se basa en el *Akaike Information Criterion* (AIC), asignando probabilidades relativas a cada modelo según su calidad del ajuste y simplicidad, mientras que el segundo utiliza el *Bayesian Information Criterion* (BIC), que favorece modelos más simples al aplicar una penalización más estricta por complejidad. El tercer método, *Bayesian Model Averaging* (BMA), combina las

predicciones de múltiples modelos ponderándolas según su evidencia bayesiana, integrando las fortalezas de cada modelo para mejorar la precisión y robustez.

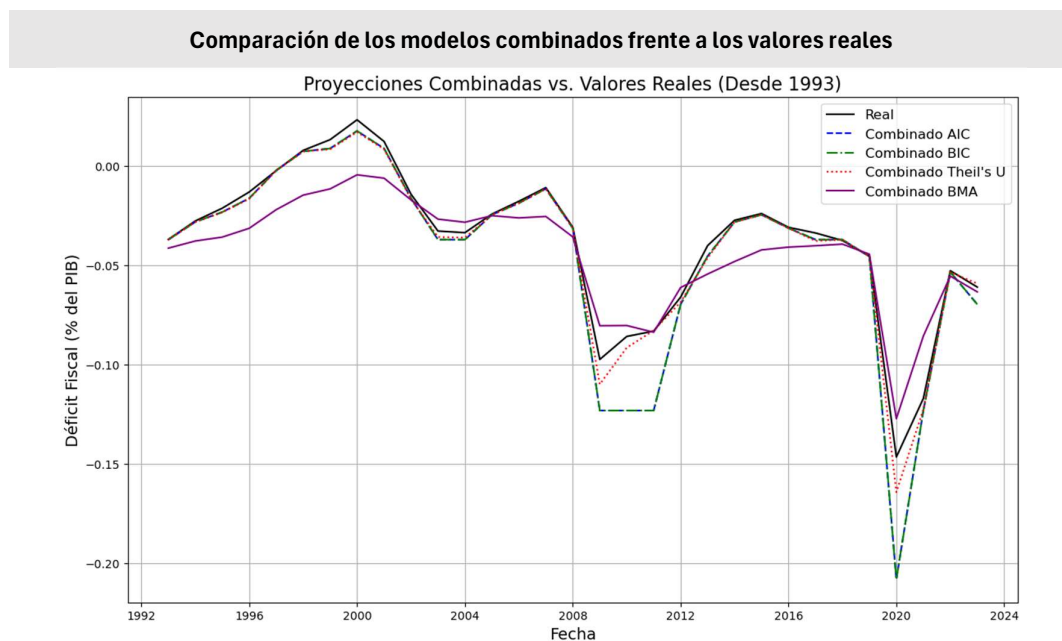
Por último, el cuarto método emplea ponderaciones dinámicas basadas en Theil's U, asignando pesos inversamente proporcionales a este valor para cada fecha, dando más peso a los modelos con mejor desempeño en cada momento específico.

De este modo, se permite que las predicciones finales se ajusten dinámicamente según la precisión relativa de cada modelo en distintos periodos, favoreciendo consistentemente a al menor valor de Theil's U.

Estos enfoques permiten evaluar las fortalezas de los modelos individuales y combinarlas de manera efectiva para obtener predicciones más precisas y consistentes.

4.1.4.2 Resultados del modelo combinado

En la Figura 13, se puede observar los valores reales del déficit fiscal frente a los resultados de los modelos combinados desde 1993 hasta el 2022 (50%-50% split).



Fuente: Elaboración Propia

Figura 13. Comparación de los modelos combinados frente a los valores reales

Gracias a la combinación de los mejores modelos de cada split según los cuatro métodos explicados en el apartado 3, se puede concluir que las predicciones combinadas logran suavizar los errores individuales de cada modelo, integrando las fortalezas de cada uno y reduciendo las debilidades individuales.

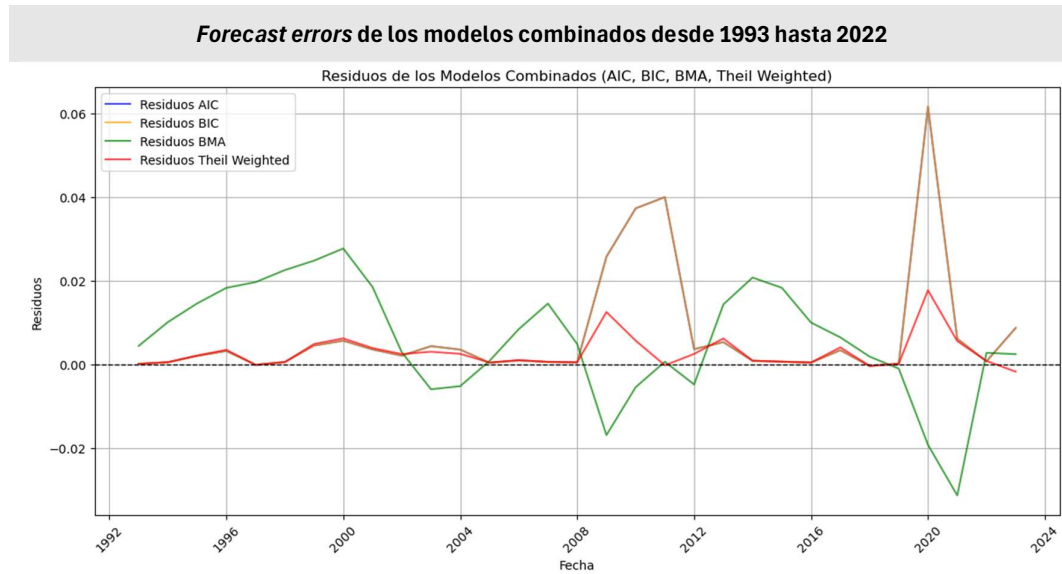
Los modelos combinados según los parámetros AIC y BIC muestran resultados similares, donde se prioriza la simplicidad. El modelo BMA ofrece una integración balanceada de las predicciones de los modelos. Esto permite predicciones robustas, pero carece de sensibilidad a errores específicos de cada fecha en comparación con el enfoque dinámico de Theil's U.

El método de ponderación dinámica por Theil's U parece ser el enfoque más robusto debido a su capacidad de adaptarse a las condiciones de cada fecha. Sin embargo, los métodos basados en AIC, BIC y BMA ofrecen soluciones robustas para escenarios donde se prioriza simplicidad y balance en las predicciones. La combinación ponderada permite mitigar errores individuales y ofrecer un resultado más fiable y preciso, aunque siendo un modelo más complejo.

4.1.5 *Forecast errors* de los modelos combinados

El análisis de los errores de pronóstico (*forecast errors*) en modelos combinados es fundamental para evaluar la precisión, la estabilidad y la adaptabilidad de diferentes enfoques predictivos. En este caso, se examinan los errores generados por cuatro combinaciones clave: AIC, BIC, BMA y Theil Weighted. Estos modelos han sido diseñados para capturar patrones temporales en datos económicos y superar las limitaciones inherentes a las estrategias individuales.

Por último, la Figura 14 muestra los *forecast errors* de los modelos combinados.



Fuente: Elaboración Propia

Figura 14. Gráfica de los forecast errors de los modelos combinados

La representación gráfica de los *forecast errors* muestra diferencias significativas en cómo cada modelo maneja las fluctuaciones de los datos reales. Los modelos basados en AIC y BIC, presentan un comportamiento más estable en períodos de baja volatilidad, pero muestran desviaciones marcadas en eventos extremos, como en 2008 y 2020. Este comportamiento sugiere que, aunque ambos modelos son efectivos en capturar tendencias generales, tienen dificultades para adaptarse a cambios abruptos en los datos. Por su parte, el modelo BMA exhibe errores sustancialmente más grandes en escenarios de alta incertidumbre, destacando limitaciones en su capacidad de generalización y adaptación a la complejidad de las fluctuaciones económicas.

En contraste, el modelo Theil Weighted, desarrollado específicamente para este análisis y representado por la línea roja, demuestra un comportamiento más equilibrado. Este modelo mantiene los *forecast errors* más cercanos a cero, incluso durante períodos de alta volatilidad como en 2008 y 2020, destacándose como el enfoque más robusto y adaptativo entre los combinados. Su diseño optimizado le

permite manejar tanto las dinámicas generales como las fluctuaciones específicas del déficit fiscal, lo que lo convierte en una herramienta confiable para escenarios económicos dinámicos.

El análisis estadístico destaca al modelo Theil Weighted como el más robusto, mostrando estacionariedad (ADF p-value: 0,004) y estabilidad (ARCH p-value: 0,838). Sin embargo, todos los modelos, incluido Theil Weighted, presentan desafíos en la normalidad de los *forecast errors* según la prueba de Jarque-Bera, con el modelo BMA mostrando una distribución más cercana a la normalidad (p-value: 0,278).

4.2 Comparación con instituciones

4.2.1 Instituciones frente al déficit

El análisis de las predicciones anuales realizadas por las instituciones financieras EconIntelligenceUnit, FirstTrustAdvisors y JPMorgan frente a los datos reales del déficit fiscal proporciona una evaluación del desempeño predictivo de las instituciones y establece el benchmark inicial para la comparación con modelos avanzados. Estos resultados destacan tanto fortalezas como limitaciones en sus enfoques, proporcionando un marco claro para evaluar la efectividad de alternativas más sofisticadas.

En la Tabla 10, se presentan los resultados de las predicciones realizadas por las tres instituciones financieras (EconIntelligenceUnit, FirstTrustAdvisors y JPMorgan) en comparación con los datos reales del déficit fiscal. Este análisis tiene como objetivo establecer benchmarks claros que permitan evaluar la efectividad de modelos avanzados, como el combinado ponderado por Theil's U,

mediante la evaluación de métricas clave como el MSE, MAE, R², Theil's U y las pruebas de sesgo y stickiness.

Resultados de las predicciones de cada institución frente al déficit real									
Institution	MSE	MAE	RMSE	R2	Theil_U	Bias_Test	Bias Test p-value	Stickiness_Test	Stickiness p-value
EconIntelligenceUnit	3,0E-05	0,004	0,006	98%	10%	0,002	0,076	-0,041	0,847
FirstTrustAdvisors	8,7E-05	0,006	0,009	94%	18%	0,002	0,271	-0,115	0,564
JPMorgan	8,8E-05	0,005	0,009	94%	18%	0,002	0,204	-0,216	0,315

Fuente: Elaboración Propia

Tabla 10. Resultados de las predicciones anuales de las instituciones frente al déficit real

Se puede observar en la Tabla 10 que EconIntelligenceUnit se posiciona como la institución con el mejor desempeño entre las evaluadas, mostrando un MSE y un MAE menor, lo que resalta su capacidad para generar predicciones más precisas en comparación con las demás. Además, su RMSE también menor refuerza su ventaja en la captura de los patrones generales del déficit fiscal. En términos del Theil's U, con un valor de 10,49%, Econ supera a las demás instituciones y establece una referencia sólida frente a la que comparar otros modelos, aunque aún queda margen para mejorar su capacidad de capturar fluctuaciones más complejas.

Por otro lado, FirstTrustAdvisors presenta un desempeño inferior, con un Theil's U de 17,75% reflejando un modelo más limitado, especialmente durante períodos de alta volatilidad económica, como 2020. Este resultado sugiere que FirstTrust adopta un enfoque más conservador, lo que, si bien puede ser beneficioso en períodos estables, limita su capacidad para capturar dinámicas más complejas. De manera similar, JPMorgan muestra resultados comparables a los de FirstTrust, con un Theil's U de 17,79% reflejando la incapacidad para ajustarse con precisión en períodos de alta variabilidad.

En términos del Coeficiente de Determinación (R²), las predicciones de Econ alcanzan un valor notable de 97,85%, lo que confirma su capacidad para explicar

una proporción significativa de la variabilidad de los datos reales. Sin embargo, tanto FirstTrust (93,85%) como JPMorgan (93,81%) muestran un ajuste más limitado, lo que indica que sus predicciones son menos efectivas para capturar patrones más complejos o dinámicas subyacentes del déficit.

Las pruebas de sesgo y *stickiness* proporcionan información adicional sobre la robustez de las predicciones. Tanto las pruebas de sesgo como las pruebas de *stickiness* sugieren que no hay evidencia estadísticamente significativa de sesgos sistemáticos en sus predicciones.

En términos generales, las instituciones financieras analizadas (EconIntelligenceUnit, FirstTrustAdvisors y JPMorgan) establecen un benchmark sólido para la evaluación de modelos avanzados, con una media de Theil's U de aproximadamente 15,3%. Este valor refleja la capacidad de las instituciones para capturar patrones generales del déficit fiscal, proporcionando una base útil para comparar el desempeño de modelos más sofisticados. Sin embargo, este benchmark presenta limitaciones, especialmente en la capacidad de adaptarse a contextos de alta volatilidad económica, como los eventos críticos observados en períodos recientes.

Las instituciones tienden a subestimar las desviaciones en períodos extremos, reflejando enfoques conservadores que, aunque efectivos en escenarios estables, demuestran una menor adaptabilidad en contextos más dinámicos. Esta limitación es evidente en sus métricas generales, que, si bien alcanzan niveles aceptables de precisión en términos de MSE y R^2 , no logran capturar completamente las fluctuaciones más complejas y dinámicas del déficit fiscal.

En conclusión, las predicciones institucionales, representadas por su desempeño promedio, ofrecen un benchmark inicial útil, pero con restricciones significativas

frente a la complejidad de las series temporales del déficit fiscal. Esto subraya la necesidad de desarrollar enfoques más avanzados y adaptativos, como el modelo combinado ponderado por Theil's U, para superar las limitaciones inherentes de estas estrategias tradicionales y proporcionar predicciones más robustas y precisas en escenarios económicos complejos.

4.2.2 Modelos combinados frente instituciones y déficit real

El objetivo principal de este trabajo es comparar los modelos combinados frente a las tres instituciones seleccionadas mencionadas en el apartado 2.2.

Para cada modelo combinado se pretende comparar los valores predichos mediante los modelos combinados frente a las instituciones seleccionadas y que el propio código devuelva la selección del mejor modelo combinado según el mínimo estadístico de Theil's U.

La siguiente tabla (Tabla 11) muestra cada modelo combinado frente a cada una de las instituciones con las métricas obtenidas.

Resultados de los modelos combinados para cada institución										
Model	Institution	MSE	MAE	RMSE	R2	Theil_U	Bias_Test	Bias Test p-value	Stickiness_Test	Stickiness p-value
Pred_AIC	Econ	2.0E-04	0.008	0.014	92%	21%	0.007	0.011	0.347	0.064
Pred_AIC	FirstTrust	1.4E-04	0.007	0.012	94%	18%	0.007	0.011	0.523	0.003
Pred_AIC	JPMorgan	1.3E-04	0.007	0.011	95%	17%	0.007	0.011	0.713	0.000
Pred_AIC	Real	2.6E-04	0.007	0.016	82%	30%	0.007	0.011	0.246	0.197
Pred_BIC	Econ	2.0E-04	0.008	0.014	92%	21%	0.007	0.011	0.347	0.064
Pred_BIC	FirstTrust	1.4E-04	0.007	0.012	94%	18%	0.007	0.011	0.523	0.003
Pred_BIC	JPMorgan	1.3E-04	0.007	0.011	95%	17%	0.007	0.011	0.713	0.000
Pred_BIC	Real	2.6E-04	0.007	0.016	82%	30%	0.007	0.011	0.246	0.197
Pred_BMA	Econ	2.4E-04	0.013	0.016	66%	31%	0.006	0.026	0.521	0.004
Pred_BMA	FirstTrust	3.5E-04	0.014	0.019	51%	37%	0.006	0.026	0.475	0.010
Pred_BMA	JPMorgan	3.2E-04	0.014	0.018	55%	36%	0.006	0.026	0.443	0.018
Pred_BMA	Real	2.2E-04	0.012	0.015	85%	28%	0.006	0.026	0.673	0.000
Pred_Theil_Weighted	Econ	2.6E-05	0.004	0.005	98%	9%	0.003	0.000	0.393	0.047
Pred_Theil_Weighted	FirstTrust	4.9E-05	0.005	0.007	97%	12%	0.003	0.000	0.048	0.809
Pred_Theil_Weighted	JPMorgan	5.4E-05	0.005	0.007	97%	13%	0.003	0.000	-0.031	0.887
Pred_Theil_Weighted	Real	2.4E-05	0.003	0.005	98%	9%	0.003	0.000	0.059	0.760

Fuente: Elaboración Propia

Tabla 11. Resultados de los modelos combinados para cada institución

Los análisis representados en la Tabla 11, proporciona información clave sobre la precisión y el rendimiento de los modelos combinados, distribuidos en cuatro enfoques principales: Pred_AIC, Pred_BIC, Pred_BMA, y Pred_Theil_Weighted. Estas categorías destacan diferencias en métricas como MSE, MAE, R^2 , Theil's U, y las pruebas de sesgo y stickiness. A continuación, se presenta un análisis detallado consolidado de los resultados, con énfasis en el modelo Pred_Theil_Weighted, una creación original de este Trabajo de Fin de Grado (TFG).

El modelo combinado basado en el criterio AIC muestra un desempeño razonablemente consistente entre las instituciones financieras (Econ, FirstTrust y JPMorgan). Los valores de Theil's U para estas instituciones oscilan entre el 17% y el 21%, lo que indica que, aunque el modelo es relativamente preciso, no alcanza los niveles de precisión obtenidos por el modelo combinado Pred_Theil_Weighted (valores de Theil's U de 9%-13%). Al comparar las predicciones realizadas entre el modelo combinado ponderado por AIC y los datos reales, el Theil's U alcanza el 30%, destacando una mayor imprecisión en este caso. Esto puede indicar que el modelo combinado por AIC no consigue escapar de los sesgos y limitaciones identificados en análisis previos. En cuanto a la capacidad para explicar la variabilidad de los datos reales, indicada por los valores de R^2 , el modelo Pred_AIC logra un rango de 92%-95% para las instituciones, lo cual es alto. Sin embargo, este desempeño no es comparable con el 98% alcanzado por el modelo Pred_Theil_Weighted, lo que evidencia que Pred_AIC es limitado al enfrentarse a fluctuaciones complejas. Además, las pruebas de sesgo y stickiness para Pred_AIC arrojan valores p-value menores a 0,05, lo que indica que el modelo podría no ser completamente robusto frente a cambios estructurales en los datos, especialmente en períodos de alta volatilidad económica (Figura 13).

El modelo ponderado por el criterio BIC presenta resultados similares a Pred_AIC, lo cual es esperado dado que ambos criterios tienden a seleccionar modelos simples y efectivos (Figura 13). Esto se debe, en parte, al reducido tamaño del conjunto de datos utilizado en este análisis, lo que provoca que AIC y BIC seleccionen modelos muy similares.

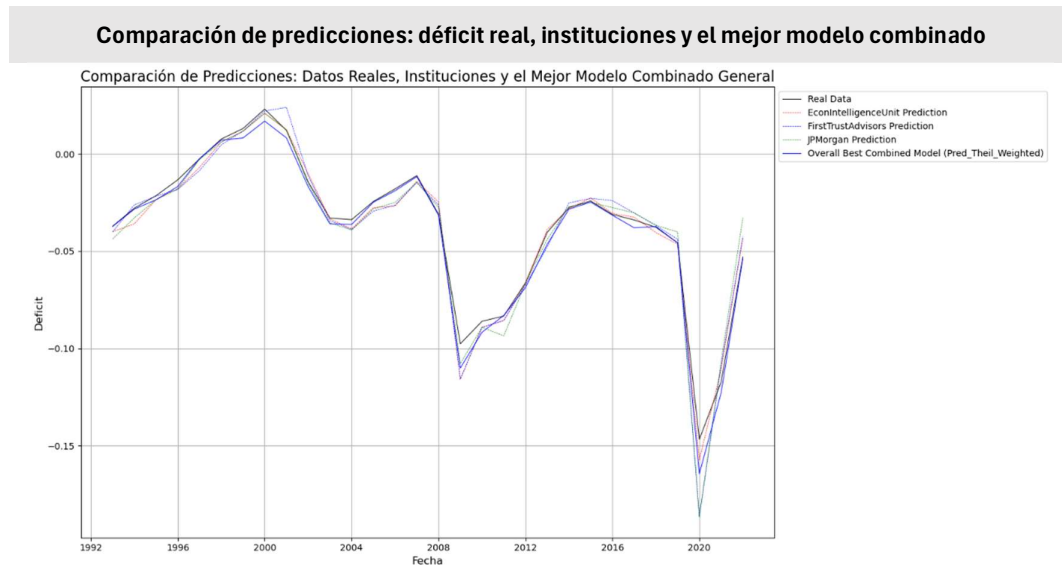
El modelo Pred_BMA, basado en un promedio bayesiano, exhibe un desempeño inferior en comparación con Pred_AIC y Pred_BIC. Los valores de Theil's U para las instituciones varían entre 31% y 37%, destacando una mayor imprecisión en comparación con los modelos combinados anteriores. Sin embargo, el Theil's U para los datos reales alcanza el 28%, acercándose más a las predicciones reales en comparación con Pred_AIC y Pred_BIC. En cuanto al R^2 , Pred_BMA logra valores entre 51%-66%, lo que indica una menor capacidad para explicar la variabilidad de los datos. Las pruebas de sesgo y stickiness también confirman este desempeño, con p-values menores a 0,05, lo que refuerza que el modelo tiene dificultades para adaptarse a fluctuaciones dinámicas en los datos.

En contraste, el modelo Pred_Theil_Weighted, desarrollado específicamente en este TFG, supera significativamente a los modelos Pred_AIC, Pred_BIC y Pred_BMA en todas las métricas clave. Con un Theil's U de 9%-13% para las instituciones y 9% al compararse con los datos reales, este modelo demuestra su capacidad para superar los sesgos y limitaciones observados en los modelos anteriores. Los valores excepcionalmente bajos de MSE ($2,4E-05$) y MAE (0,003), junto con un R^2 del 98%, posicionan a Pred_Theil_Weighted como el modelo más robusto y preciso del análisis. Además, las pruebas de sesgo y stickiness refuerzan su solidez: p-values altos en stickiness (por ejemplo, 0,809 para FirstTrust) indican que los errores no son persistentes, mientras que los p-values cercanos a 0 en las pruebas de sesgo confirman la ausencia de sesgos sistemáticos.

Las fortalezas de cada modelo se alían mediante el `Pred_Theil_Weighted` para mejorar la precisión del modelo, destacada por valores mínimos en MSE, MAE y Theil's U. Esto indica que la metodología combinada no solo captura patrones generales, sino que también responde adecuadamente a fluctuaciones específicas. Su generalización es notable, reflejada en los altos valores de R^2 y la consistencia en todas las instituciones evaluadas. Además, el modelo supera claramente las predicciones institucionales, que presentan ligeros sesgos conservadores y dificultades para capturar desviaciones en períodos críticos, como en 2020.

En conclusión, el modelo `Pred_Theil_Weighted`, creado específicamente en este TFG, se posiciona como una herramienta excepcionalmente precisa y robusta para predecir déficits fiscales. Su desempeño en métricas clave, junto con su capacidad para superar las predicciones de las instituciones financieras, demuestra el valor de las metodologías de combinación avanzadas. Este enfoque no solo mejora la precisión, sino que también ofrece una mayor adaptabilidad a contextos económicos dinámicos. Las instituciones podrían beneficiarse de adoptar estrategias similares, integrando modelos combinados avanzados para mejorar la confiabilidad y robustez de sus predicciones.

Por último, la representación gráfica de la comparación se muestra en la Figura 15.



Fuente: *Elaboración Propia*

Figura 15. *Comparación de predicciones: déficit real, instituciones y el mejor modelo combinado*

La gráfica ilustra la comparación entre las predicciones realizadas por varias instituciones financieras, los datos reales del déficit fiscal, y el modelo combinado Theil Weighted, representado por la línea azul. Este modelo destaca por su capacidad para aproximarse con consistencia y precisión a los datos reales, superando en general las predicciones de las instituciones. Sin embargo, en periodos de alta volatilidad, como la crisis económica de 2020, el modelo no logra reflejar completamente los valores extremos del déficit fiscal, aunque sigue siendo el más robusto entre las opciones presentadas y frente a las instituciones.

En cuanto a las predicciones de las instituciones financieras (líneas de colores), se observa una tendencia a subestimar el déficit fiscal en momentos de crisis. Las instituciones presentan patrones diferenciados: FirstTrustAdvisors y Econ Intelligence Unit adoptan un enfoque conservador, que les permite ajustarse bien en periodos estables, pero no logra capturar correctamente las fluctuaciones extremas en los déficits. Por otro lado, JPMorgan muestra aun mayor sensibilidad a los cambios bruscos en los datos, reflejando desviaciones más pronunciadas en momentos de crisis, aunque con un ajuste más razonable en periodos estables.

En conjunto, estas instituciones tienden a converger en sus predicciones en periodos de estabilidad, pero el modelo Theil Weighted supera sus resultados al adaptarse mejor a cambios dinámicos sin comprometer la consistencia. Este enfoque destaca como una herramienta más equilibrada y precisa, y podría servir como referencia para mejorar las proyecciones institucionales en escenarios de alta incertidumbre económica.

5 Conclusión

Este trabajo analizó el déficit fiscal anual de Estados Unidos mediante técnicas avanzadas de *machine learning*, explorando su capacidad para generar predicciones más precisas y robustas, tanto frente a los valores reales como a las proyecciones de instituciones reconocidas. Los resultados obtenidos subrayan la eficacia de las combinaciones de modelos, particularmente cuando se emplean metodologías innovadoras como la ponderación dinámica basada en el estadístico Theil's U.

5.1 *Machine learning* vs Instituciones Privadas

Uno de los objetivos de este trabajo era el de corroborar hallazgos previamente estudiados en la literatura. Por un lado, se examinó si las técnicas de *machine learning*, especialmente la combinación de modelos basados en redes neuronales y árboles de decisión, podían mejorar las predicciones del déficit fiscal de Estados Unidos. Por otro lado, se quiso replicar y extender el trabajo de Carabotta y Claeys (2024), quienes demostraron que las combinaciones (en su caso, de instituciones) generan predicciones significativamente más precisas que los resultados individuales.

Una contribución clave de este trabajo es la creación de un modelo de ponderaciones dinámicas basado en el estadístico Theil's U. Este enfoque, desarrollado específicamente en esta investigación, asigna pesos dinámicamente a cada modelo combinado según su desempeño en cada fecha específica, adaptándose a las fluctuaciones en los datos y optimizando la precisión de las proyecciones. Este modelo es una extensión y adaptación original del marco

teórico propuesto por Sineglazov y Chumachenko (2014), aplicado aquí por primera vez al análisis del déficit fiscal.

Los resultados obtenidos respaldan la eficacia de este enfoque innovador. La media del estadístico Theil's U de las predicciones realizadas por las tres instituciones seleccionadas frente al déficit real fue de aproximadamente 15,34%, lo que estableció el benchmark inicial. El modelo de ponderaciones dinámicas creado en este trabajo fue seleccionado automáticamente por el código al presentar los mejores resultados ya que se logró reducir esta media a un 11,38% frente a las instituciones y a un 9,43% frente al déficit fiscal real. Estas reducciones, de aproximadamente 396 y 591 puntos básicos respectivamente, evidencian una mejora sustancial en la precisión de las proyecciones y demuestran la capacidad del modelo para superar tanto las predicciones institucionales como los enfoques estáticos tradicionales.

Los resultados confirman que la combinación del modelo XGBoost y LSTM-CNN, utilizando técnicas de ensamblaje como AIC, BIC, BMA y ponderaciones dinámicas, mejora significativamente la precisión al capturar patrones complejos. Este enfoque es escalable y aplicable a contextos fiscales internacionales. Sin embargo, su eficacia depende de la calidad de los datos históricos y requiere capacidades computacionales y analíticas avanzadas para su implementación.

Los resultados contribuirán significativamente al bienestar social y económico. Su capacidad para mejorar la planificación presupuestaria, fomentar una distribución equitativa de los recursos y fortalecer la confianza en las instituciones refuerza la estabilidad económica y social alineándose con los principios y objetivos delineados en el apartado 1.5 Impacto en la Sociedad, demostrando su compromiso

con el desarrollo sostenible y con la promoción de una sociedad más justa y equilibrada mediante la inclusión de las nuevas tecnologías.

Como posibles vías de investigación, destacan direcciones clave para ampliar los avances de este estudio. Las normativas fiscales de la Unión Europea, como el Pacto de Estabilidad y Crecimiento, resaltan la necesidad de controles más rigurosos sobre los déficits presupuestarios, y este trabajo presenta una solución adaptable a este y futuros cambios regulatorios, que podría fortalecerse con actualizaciones e investigaciones adicionales.

Una línea prometedora es la integración de factores externos, como fluctuaciones de mercado y choques macroeconómicos, en los modelos de predicción fiscal. Herramientas avanzadas como LSTM pueden capturar dinámicas complejas y mejorar la precisión de las proyecciones (Zheng et al., 2024). Asimismo, convertir los modelos en multivariantes permitiría incorporar las variables inherentes al déficit fiscal, aumentando su capacidad explicativa y alcance predictivo.

Este trabajo demuestra el potencial de las técnicas de *machine learning* para transformar la predicción y gestión del déficit fiscal, mejorando significativamente la precisión de las proyecciones. Estas líneas de investigación subrayan la importancia de seguir avanzando con nuevas tecnologías y arquitecturas híbridas, lo que requiere colaboración interdisciplinaria y compromiso continuo para enfrentar los desafíos económicos globales.

6 Bibliografía

- Ademmer, M., & Boysen-Hogrefe, J. (2022). The Impact of Forecast Errors on Fiscal Planning and Debt Accumulation. *Jahrbücher für Nationalökonomie und Statistik*, *242*(2), 171-190. doi:10.1515/jbnst-2020-0054
- AIRef. (2024). *Opinión sobre la reforma de la gobernanza fiscal de la Unión Europea*. Madrid: AIRef.
- Akaike, H. (Diciembre de 1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705
- Alonso, A. M., Peña, D., & Romo, J. (Enero de 2002). Forecasting time series with sieve bootstrap. *100*(1), págs. 1-11.
- Armstrong, J. (2001). Principles of forecasting: A handbook for researchers and practitioners. Springer.
- Auerbach, A. J. (2003). Fiscal Policy, Past and Present. *Brookings Papers on Economic Activity*, *1*, 75-122. doi:10.3386/w10023
- Ba, C. M., Fanny, S. S., & Qureshi, S. (2020). Forecasting Canadian GDP growth using XGBoost. *Carleton Economic Papers*, *20-14*. Obtenido de <https://www.carleton.ca/economics/wp-content/uploads/cewp20-14.pdf>
- Barro, R. J. (1981). Output Effects of Government Purchases. *Journal of Political Economy*, *89*(6), 1086-1121. doi:10.3386/w0432
- Beetsma, R., Giuliodori, M., & Wierdsma, P. (1 de Octubre de 2009). Planning to cheat: EU fiscal policy in real time. *Economic Policy*, *24*(60), 753-804. doi:10.1111/j.1468-0327.2009.00230.x
- Benito, B., Guillamón, M.-D., & Ríos, A.-M. (2023). The sustainable development goals: How does their implementation affect the financial sustainability of the largest Spanish municipalities. *Sustainable Development*, *31*(4), 2836-2850. doi:10.1002/sd.2551C
- Bentéjac, C., Anna Csörgö, A., & Martínez-Muñoz, G. (2021). A Comparative Analysis of XGBoost. *Artificial Intelligence Review*, *54*, 1937-1967. doi:10.48550/arXiv.1911.01914
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(10), págs. 281-305. doi:10.5555/1643031.1643047
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *24th International Conference on Neural Information Processing Systems*, (págs. 2546-2554). Granada. Obtenido de

https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf

- Bergstra, J., Yamins, D., & Cox, D. D. (2013). Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. *30th International Conference on International Conference on Machine Learning*. 28, págs. I-115 - I-123. Atlanta: JMLR.
- Bisogno, M., Cuadrado-Ballesteros, B., Manes-Rossi, F., & Peña-Miguel, N. (2024). Financial Sustainability and Sustainable Development in Local Governments: Empirical Insights. *Public Performance & Management Review*, 784-811. doi:10.1080/15309576.2024.2340116
- Biswas, A., Ahmed Uday, I., Rahat, K. M., Akter, M. S., & Mahdy, M. R. (2023). Forecasting the United States Dollar (USD)/Bangladeshi Taka (BDT) exchange rate with deep learning models. *PLoS One*, 18(2). doi:10.1371/journal.pone.0279602
- Blanchard, O. J., & Leigh, D. (2013). Growth Forecast Errors and Fiscal Multipliers. *American Economic Review*, 103(3), 117-120. doi:10.3386/w18779
- Blinder, A. S. (2023). A Monetary and Fiscal History of the United States, 1961-2022. *Milken Institute Review*. Obtenido de <https://www.milkenreview.org/articles/a-monetary-and-fiscal-history-of-the-united-states-1961-2022>
- Bordo, M. D., & Levy, M. D. (2021). Do enlarged fiscal deficits cause inflation? The historical record. *Economic Affairs*, 41, 59-83. doi:10.1111/ecaf.12446
- Calda, M. (2009). War Time to Peace Time: the Truman Administration during the Early Phase of Post-WW2 Reconversion. *Acta Universitatis Carolinae*, 9(4), 47-63. Obtenido de <https://biblio.hiu.cas.cz/records/0d3b294d-48e1-4664-bd31-875d12134b1d>
- Carabotta, L., & Claeys, P. (6 de Diciembre de 2024). Combine to compete: Improving fiscal forecast accuracy. *Journal of Forecasting*, 43, 948-982. doi:10.1002/for.3058
- Celik, S. S. (2024). *Predicting External Macroeconomic Crises: Machine Learning Perspective*. Middle East Technical University.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, págs. 785-794. doi:doi.org/10.1145/2939672.2939785
- Coibion, O., & Gorodnichenko, Y. (10 de June de 2011). What Can Survey Forecasts Tell Us about Information Rigidities? *Journal of Political Economy*, 120(1). doi:10.1086/665662

- Cooley, T. F., & Ohanian, L. (1997). Postwar British Economic Growth and the Legacy of Keynes. *The Journal of Political Economy*, 105(3), 439-472. doi:10.1086/262079
- De Long, J. B. (1998). Fiscal Policy in the Shadow of the Great Depression. En M. D. Bordo, C. Goldin, E. N. White, & N. B. Research (Ed.), *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century* (págs. 67-86). University of Chicago Press. doi:10.1515/9781400873326-011
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *74*(366), 427-431. doi:10.2307/2286348
- Diks, C. G., & Vrugt, J. A. (2010). Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment*, 24, 809-820. doi:10.1007/s00477-010-0378-z
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87. doi:10.1145/2347736.2347755
- Dwyer, J. (2012). Keynes's economics and the question of public debt. The University of Sydney.
- Efe, L. D. (9 de Septiembre de 2020). EEUU logra el mayor superávit fiscal presupuestario desde 1948. *Libre Mercado*. Obtenido de <https://www.libremercado.com/2000-09-27/eeuu-logra-el-mayor-superavit-fiscal-presupuestario-desde-1948-2172/>
- Elbeltagi, A., Srivastava, A., Ehsan, M., Sharma, G., Yu, J., Khadke, L., . . . Jinsong, D. (2024). Advanced stacked integration method for forecasting long-term drought severity: CNN with machine learning models. *Journal of Hydrology: Regional Studies*, 53. doi:10.1016/j.ejrh.2024.101759
- European Union. (03 de 10 de 2024). Obtenido de EUR-Lex: <https://eur-lex.europa.eu/ES/legal-content/glossary/stability-and-growth-pact.html>
- Feldkircher, M. (2010). Forecast combination and Bayesian model averaging: A prior sensitivity analysis. *Journal of Forecasting*, 31(4), 361-376. doi:doi.org/10.1002/for.1228
- Frankel, J. (2011). Over-optimism in Forecasts by Official Budget Agencies and Its Implications. *Oxford Review of Economic Policy*, 27(4), 536-562. doi:10.1093/oxrep/grr025
- FRED. (s.f.). *Federal Reserve Bank of St. Louis*. Obtenido de History of the St. Louis Fed's Headquarters: <https://www.stlouisfed.org/about-us/history>
- Gentiloni, P. (15 de Mayo de 2024). Spring 2024 Economic Forecast: A gradual expansion amid high geopolitical risks. *European Economic Forecast* .

- Gorbatiuk, V., Sineglazov, V., & Chumachenko, O. (2014). A method for building a forecasting model with dynamic weights. *Eastern-European Journal of Enterprise Technologies*, 2(4(68)). doi:10.15587/1729-4061.2014.21189
- Hallett, A. H., & Weymark, D. (2005). Independence Before Conservatism: Transparency, Politics, and Central Bank Design. *German Economic Review*, 6(1), 1-21. doi:10.1111/j.1465-6485.2005.00119.x
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), págs. 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (Noviembre de 1999). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statist. Sci.*, 14(4), 382-417. doi:10.1214/ss/1009212519
- Huerta de Soto, J. (1998). *Dinero, crédito bancario y ciclos económicos* (8 ed.). Madrid: Union Editorial.
- Hylleberg, S., Engle, R., Granger, C. W., & Sam Yoo, B. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44(1-2), 215-238. doi:10.1016/0304-4076(90)90080-D
- Jenkins, G. E. (1976). *Time series analysis: Forecasting and control*. Holden-Day.
- Jo, T., Nho, K., Bice, P., & Saykin, J. A. (2022). Deep learning-based identification of genetic variants: Deep learning-based identification of genetic variants:. *Briefings in Bioinformatics*, 23(2), 1-11. doi:10.1093/bib/bbac022
- Jonung, L., & Larch, M. (1 de Julio de 2006). Improving fiscal policy in the EU: The case for independent forecasts. *Economic Policy*, 21(47), 492-534. doi:10.1111/j.1468-0327.2006.00162.x
- Kahneman, D. S. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Keynes, J. M. (1936). *Teoría General del Empleo, el Interés y el Dinero* (2006 ed.). Sussex: Fondo de Cultura Economica de España.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 2, págs. 1137–1143.
- Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26(3), 159-190. doi:10.1007/s10462-007-9052-3
- Kurihara, Y., & Fukushima, A. (Mayo de 2019). AR Model or Machine Learning for Forecasting GDP and Consumer Price for G7 Countries. *Applied Economics and Finance*, 6(3). doi:10.11114/aef.v6i3.4126
- Leal, T., Pérez, J. J., Tujula, M., & Vidal, J.-P. (2008). Fiscal forecasting: Lessons from the literature and challenges. *Fiscal Studies*, 29(3), 347-386. Obtenido de <http://www.jstor.org/stable/24440159>

- Merola, R., & Pérez, J. J. (2013). Fiscal forecast errors: Governments versus independent agencies. *European Journal of Political Economy*, *32*, 285-299. doi:10.1016/j.ejpoleco.2013.09.002
- Merola, R., & Pérez, J. J. (Diciembre de 2013). Fiscal forecast errors: governments versus independent agencies? *European Journal of Political Economy*, *32*, 285-299. doi:10.1016/j.ejpoleco.2013.09.002
- Mia, M. U., Rahman, M., Elbeltagi, A., Al-Mahbub, M. A., Sharma, G., Touhidul Islam, H. M., . . . Ali Washakh, R. u. (2022). Sustainable flood risk assessment using deep learning-based algorithms with a blockchain technology. *Geocarto International*. doi:10.1080/10106049.2022.2112982
- Ozili, P. K. (2024). Economic Policy for Sustainable Development: Role of Monetary Policy, Fiscal Policy and Regulatory Policy. *Circular Economy and Sustainability*, 1-32. Obtenido de <https://mpira.ub.uni-muenchen.de/121523/>
- Pina, Á. M., & Venes, M. N. (Septiembre de 2011). The political economy of EDP fiscal forecasts: An empirical assessment. *European Journal of Political Economy*, *27*(3), 534-546. doi:10.1016/j.ejpoleco.2011.01.005
- Renshaw, P. (1999). Was there a Keynesian Economy in the USA between 1933 and 1945? *Journal of Contemporary History*, *34*(3), 337-364. doi:10.1177/002200949903400302
- Sahib, M. R., Elkina, H., & Zaki, T. (2024). From Technical Indicators to Trading Decisions: A Deep Learning Model Combining CNN and LSTM. *International Journal of Advanced Computer Science and Applications*, *15*(6), 847-855. doi:10.14569/IJACSA.2024.0150685
- Schwarz, G. (Marzo de 1978). Estimating the Dimension of a Model. *Ann. Statist.*, *6*(2), 461-464. doi:10.1214/aos/1176344136
- Shah, D., & Thaker, M. (2024). A Review of Time Series Forecasting Methods. *International Journal of Research and Analytical Reviews (IJRAR)*, *11*(2), 749-755. doi:10.1729/Journal.38816
- Sheikh, M., & Coulibaly, P. (2024). Review of Recent Developments in Hydrologic Forecast Merging Techniques. *Water*, *16*(2), 301. doi:10.3390/w16020301
- Shejul, K., Harikrishnan, R., & Gupta, H. (2024). The improved integrated Exponential Smoothing based CNN-LSTM algorithm to forecast the day ahead electricity price. *MethodsX*, *13*. doi:10.1016/j.mex.2024.102923
- Spilimbergo, A., Symansky, S., Blanchard, O., & Cottarelli, C. (29 de Diciembre de 2008). Fiscal Policy for the Crisis. *INTERNATIONAL MONETARY FUND*. Obtenido de <https://ssrn.com/abstract=1339442> or <http://dx.doi.org/10.2139/ssrn.1339442>

- Steel, M. F. (2020). Model Averaging and its Use in Economics. *Journal of Economic Literature*, 58(3), 644-719. doi:10.1257/jel.20191385
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37-45. doi:10.1080/00031305.2017.1380080
- Taylor, S., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), págs. 37-45. doi:https://doi.org/10.1080/00031305.2017.1380080
- Theil, H. (1 de Junio de 1967). Applied Economic Forecasting. Studies in Mathematical and Managerial Economics Get access Arrow. *The Economic Journal*, 77(306), 368-370. doi:10.2307/2229320
- Tovar Jalles, J., Karibzhanov, I., & Loungani, P. (Septiembre de 2015). Cross-country evidence on the quality of private sector fiscal. *Journal of Macroeconomics*, 45, 186-201. doi:10.1016/j.jmacro.2015.04.009
- Tsangari, H. (2007). An Alternative Methodology for Combining Different Forecasting Models. *Journal of Applied Statistics*, 34(4), 403-421. doi:10.1080/02664760701231633
- Tversky, A., & Kahneman, D. (27 de Septiembre de 1974). Judgment under Uncertainty:Heuristics and Biases. *Science*, 185, 1124-1131. Obtenido de <https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf>
- U.S. Bureau of Economic Analysis. (2024). *Federal government budget surplus or deficit (-) [M318501Q027NBEA]*. Obtenido de FRED, Federal Reserve Bank of St. Louis.: <https://fred.stlouisfed.org/series/M318501Q027NBEA>.
- U.S. Bureau of Economic Analysis. (2024). *Gross Domestic Product [GDP]*. Obtenido de FRED, Federal Reserve Bank of St. Louis: <https://fred.stlouisfed.org/series/GDP>
- von Hagen, J. (2010). Sticking to fiscal plans: the role of institutions. *Public Choice*, 144(3/4), 487-503. Obtenido de <http://www.jstor.org/stable/40835625>
- Wan, Y., Tao, H., & Ma, L. (2024). Forecasting Zhejiang Province's GDP Using a CNN-LSTM. *Frontiers in Business, Economics and Management*, 13(3). doi:10.54097/bmq2dy63
- Yueh, L. (2019). *The Great Economists*. Penguin.

