Volatilidad Bursátil, Crisis, Econometría y Machine Learning

Análisis predictivo sobre el rol de la volatilidad en mercados de capitales sobre variaciones en el PBI real de Estados Unidos entre 1962-2024 combinando herramental econométrico y de aprendizaje automático.

Tesis doctoral - Universidad del CEMA

Candidato: Agustin A. Shehadi C.

Director: Guillermo Lopez Dumrauf, PhD.

2025

Índice general

0.1. Introducción

¿Es posible predecir crisis económicas a partir de la volatilidad en los mercados de capitales?

La relación entre las fluctuaciones bursátiles y la estabilidad macroeconómica ha sido un tema central en la literatura económica, y eventos históricos como el *Lunes Negro* de 1987, la *Burbuja Punto Com* de principios del siglo XXI y la *Crisis Financiera de 2008* han contribuido de manera decisiva a poner en foco la necesidad de integrar el análisis de la volatilidad en las políticas económicas y de supervisión.

El *Lunes Negro*, ocurrido el 19 de octubre de 1987, evidenció cómo la interconexión de los mercados y la ausencia de mecanismos efectivos para contener la volatilidad pueden desencadenar caídas abruptas en los mercados de capitales que, aun sin generar de inmediato una recesión profunda, obligaron a los *policy makers* a replantear los métodos de valuación de riesgos bursátiles y a desarrollar estrategias de mitigación ante shocks financieros (Roll, 1988; Eichengreen, 1989).

De igual forma, la *Burbuja Punto Com* se caracterizó por un entusiasmo especulativo desmedido y una sobrevaloración de activos vinculados a las nuevas tecnologías, lo que culminó en una corrección abrupta a comienzos de los 2000. Este fenómeno, ampliamente analizado en estudios como los de Shiller (2000) y Schularick and Taylor (2012), resaltó la necesidad de desarrollar herramientas rigurosas que midan volatilidad y detecten, a partir de indicadores tempranos, la formación de burbujas especulativas que podrían desembocar en crisis económicas de mayor envergadura.

La experiencia acumulada con estos episodios fue ampliada por la *Crisis Financiera de 2008*, donde la acumulación de riesgos en el sector hipotecario estadounidense, combinada con la complejidad y la interconexión de los productos financieros, evidenció de forma contundente la vulnerabilidad del sistema financiero global (Acharya and Richardson, 2011; Baron and Xiong, 2016; Stiglitz, 2010). Dicho acontecimiento impulsó una revisión profunda de los marcos regulatorios, orientada hacia la implementación de modelos predictivos y sistemas de monitoreo continuo que integren la volatilidad del mercado como una variable clave para la detección de desequilibrios subyacentes. En este sentido, los *policy makers* e institutos regulatorios han dirigido esfuerzos hacia la actualización de normativas y el fortalecimiento de la coordinación internacional, aspectos esenciales para prevenir contagios sistémicos y mitigar el impacto de crisis financieras. La integración de indicadores de volatilidad

bursátil ha contribuido a dotar a las autoridades de herramientas más precisas para la anticipación y gestión de riesgos, lo que representa un cambio paradigmático en la política económica contemporánea. En definitiva, estos eventos han servido no solo para evidenciar las fragilidades inherentes al sistema financiero, sino también para estimular una evolución en las estrategias de regulación, orientadas a la prevención de crisis y a la consolidación de la estabilidad macroeconómica.

Así, el concepto de volatilidad bursátil juega un rol central en este contexto. Autores como Bhowmik and Wang (2020) destacan el rol de la volatilidad bursátil como consecuencia directa de la falta de información o incertidumbre de los agentes económicos, influyendo directamente en las decisiones de ahorro e inversión de empresas e individuos. Tradicionalmente, la volatilidad se mide a través de la varianza de rendimientos, modelos de volatilidad subyacente en derivados, u enfoques econométricos que incluyen, entre otros, la regresión lineal, métodos de máxima verosimilitud y modelos autoregresivos de algún tipo (Bollerslev et al., 2003; Baker et al., 2016; Gulen and Ion, 2016; Engle et al., 2013). Estos modelos tienen la ventaja analítica de identificar de manera precisa los parámetros involucrados en la estimación, lo que permite la realización de testeos de inferencia estadística con propiedades bien definidas en términos de la teoría de la probabilidad [1]. En este sentido, al aplicar modelos econométricos tradicionales se puede determinar la dependencia temporal de la volatilidad y establecer relaciones funcionales entre las variables, lo que facilita la interpretación de los resultados y la identificación de efectos específicos en los mercados financieros.

Sin embargo, una limitación importante de estos enfoques tradicionales reside en que cada modelo está condicionado por la estructura o forma funcional-correlacional entre variables que el analista determina *ex-ante*. Esta especificación previa puede restringir la capacidad del modelo para capturar la complejidad y la naturaleza dinámica de los mercados, especialmente en entornos donde las relaciones entre variables pueden evolucionar o presentar comportamientos no lineales. Estudios como el de Lee (2005) han propuesto modelos que incorporan la naturaleza dinámica de la volatilidad implícita, permitiendo una mayor flexibilidad en la estimación y ofreciendo métricas estandarizadas en la industria financiera para evaluar y comparar precios de activos derivados. En consecuencia, aunque los modelos econométricos tradicionales ofrecen un claro beneficio en términos de análisis estructurado y validación estadística, su dependencia de supuestos funcionales predefinidos puede

[1] La precisión en la estimación de parámetros y la robustez de los test de hipótesis son aspectos fundamentales que facilitan la validación empírica de los modelos, como se discute en Engle et al. (2013).

limitar su capacidad para adaptarse a escenarios de alta incertidumbre o a cambios abruptos en las condiciones del mercado.

En este campo, el avance de nuevas técnicas como la inteligencia artificial está cobrando particular relevancia en los últimos tiempos. El concepto de inteligencia artificial (IA) se define como el conjunto de técnicas y métodos computacionales que permiten a las máquinas simular capacidades cognitivas humanas, tales como el aprendizaje, el razonamiento y la adaptación [2]. En el contexto de la predicción de series temporales, y más precisamente en el análisis de fluctuaciones macroeconómicas relacionadas con los mercados de capitales, la IA se ha posicionado como una herramienta innovadora que posibilita la identificación de patrones complejos y no lineales en grandes volúmenes de datos. Este enfoque ha permitido superar algunas de las limitaciones de los modelos econométricos tradicionales, ya que algoritmos como las redes neuronales y las máquinas de soporte vectorial han demostrado una alta capacidad para anticipar comportamientos futuros en entornos de incertidumbre (El-Aal and Mohamed, 2024; Bluwstein et al., 2023; Yu et al., 2019; Chatzis et al., 2018).

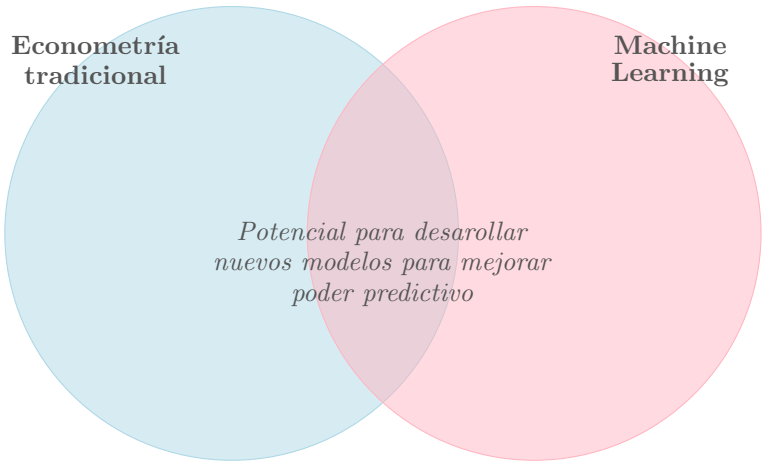
Una característica central de los modelos basados en inteligencia artificial es que priorizan la calidad predictiva por sobre la interpretabilidad de los parámetros o la rigidez en la estructura correlacional. Mientras que en los modelos econométricos tradicionales se busca una explicación detallada de los vínculos causales y se pueden estimar de forma precisa los parámetros involucrados, en los modelos de IA el objetivo principal es lograr la mayor precisión en la predicción, aun cuando ello implique que el funcionamiento interno del modelo se convierta en una 'caja negra'. Esta estrategia resulta especialmente valiosa en contextos donde la capacidad de prever crisis o fluctuaciones abruptas tiene una importancia crítica, pero también requiere de especial precaución para evitar el sobreajuste (overfitting). El sobreajuste se presenta cuando el modelo se adapta excesivamente a las peculiaridades de los datos históricos, lo que puede introducir sesgos y limitar su capacidad de generalización a nuevos escenarios.

En cuanto a la evaluación de la calidad de ajuste en las predicciones de series temporales, la literatura especializada ha señalado diversas métricas, siendo el RMSE (Raíz del Error Cuadrático Medio) una de las más destacadas por su interpretabilidad y su eficacia para penalizar errores de gran magnitud (Hyndman and Koehler, 2006). El RMSE se ha convertido en un estándar para la comparación de

[2] Ver informe online de IBM en tópicos generales de modelos de aprendizaje automático en https://www.ibm.com/mx-es/topics/machine-learning

modelos predictivos, ya que proporciona una medida clara y cuantitativa del desempeño del modelo. En este trabajo se abordará de forma rigurosa el desarrollo de estas medidas de calidad de ajuste, analizando en detalle sus fortalezas y limitaciones.

El objetivo de esta investigación es desarrollar una herramienta predictiva de alta calidad que combine técnicas econométricas tradicionales y aprendizaje automático para modelar la relación entre volatilidad bursátil y fluctuaciones del PBI real. Utilizando datos de los más de 500 constituyentes del S&P 500 entre enero de 1962 y diciembre de 2024, se busca construir un puente entre la práctica econométrica y el creciente campo de la inteligencia artificial. Diagramalmente:



Las preguntas que surgen naturalmente son ¿se pueden unir ambos mundos? En tal caso, ¿se genera algún valor predictivo por sobre las distintas corrientes de modelización individuales?¿es posible automatizar una rutina de modelización para ser aplicable en distintos casos de estudios, distintos mercados, o incluso distintos países? Estas preguntas busca responder esta investigación.

El resto de este trabajo se organiza de la siguiente manera: el Capítulo 1 introduce la motivación para el estudio de la volatilidad bursátil para anticipar fluctuaciones de PBI real. El Capítulo 2 presenta un resumen de los modelos econométricos tradicionales más usados actualmente en la literatura especializada. El Capítulo 3 describe los métodos de inteligencia artificial que tanto practitioners como académicos están aplicando en el campo. El Capítulo 4 se introducen los métodos más comunes a la hora de medir calidad predictiva entre modelos. En el Capítulo 5 se presenta el desarrollo de una

rutina algorítmica propia que incorpora tanto herramental econométrico tradicional como modelos de inteligencia artificial. Finalmente, el Capítulo 6 presenta las conclusiones.

Capítulo 1

Volatilidad y Crisis

1.1. ¿Qué es la volatilidad? ¿Cómo se mide?

Intuitivamente, volatilidad refiere a una medida de cuán disperso está un conjunto de datos; o, visto de otra forma, al grado de incertidumbre que tenemos sobre el objeto de estudio. A mayor dispersión, más difícil será para nosotros predecir futuros resultados de la variable en cuestión. A modo de ejemplo, la Figura 1.1 debajo muestra dos conjuntos de datos simulados en $\mathbb{R}^2$. Ambos comparten el mismo valor promedio, pero el conjunto de la izquierda reporta una mayor volatilidad que el de la derecha. En este ejemplo es fácil notar que predecir nuevos datos del conjunto de la izquierda es mucho más desafiante que con el conjunto de la derecha. Es por cosas como estas que decimos que el conjunto de la izquierda presenta una mayor *incertidumbre*.

En el campo de la economía financiera se aplica esta misma lógica. Decimos que un activo financiero es más volátil cuando su retorno es muy disperso respecto a su promedio histórico. Haciendo el paralelismo con el concepto de incertidumbre, podemos afirmar además que en el ámbito de las finanzas mayor volatilidad implica mayor incertidumbre, y mayor incertidumbre implica a su vez mayor *riesgo*.

Llevando este concepto al mundo de la teoría de la probabilidad, Evans and Rosenthal (2004) explica

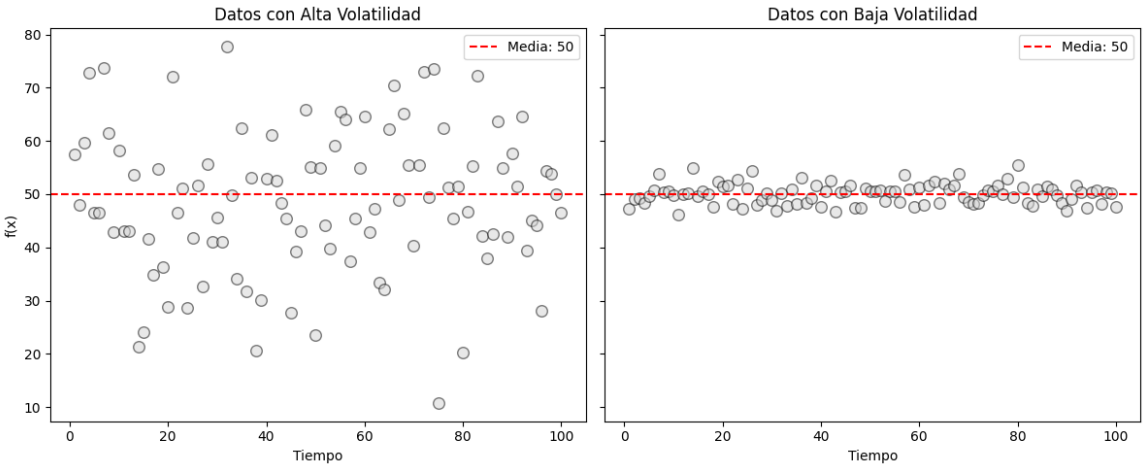


Figura 1.1: Simulación de datos con alta y baja volatilidad

que muchas veces no basta con conocer el valor promedio de una variable aleatoria, sino que es necesario conocer además alguna medida de qué tan *dispersa* es la probabilidad de ocurrencia de cierto resultado. Con esta motivación en mente, decimos que si estudiamos un activo i con retorno esperado $\mathbf{E}(R_i)$ sería absurdo esperar que *siempre y en todo momento* el retorno de este activo sea $\mathbf{E}(R_i)$. Dicho de otro modo, a veces el retorno en un periodo determinado $R_{i,t}$ puede ser mayor o menor que su valor promedio $\mathbf{E}(R_i)$. Gráficamente, asociamos esta dispersión de los datos respecto a su valor promedio con el grado de *pesadez* (i.e., qué tan ancho) de las colas de la distribución de probabilidad. La Figura 1.2 debajo muestra un ejemplo de esto: El primer diagrama muestra cómo el retorno de un activo $R_{i,t}$ puede oscilar en torno a su valor medio $\mathbf{E}(R_i)$ y cómo esta volatilidad se traduce en el grado de dispersión que presenta su distribución de probabilidad; esto último es precisamente lo que ilustra el gráfico de la derecha, en donde se muestra la distribución de probabilidad del retorno del activo alrededor del valor promedio $\mathbf{E}(R_i)$ y con el ancho de las colas ilustrando el grado de dispersión en los resultados potenciales.

Teniendo todo eso en mente, surge la pregunta ¿cómo podemos medir la volatilidad? Una de las maneras más usadas acorde a Evans and Rosenthal (2004) son las medidas de dispersión basadas en distancia respecto a la media. La intuición es sencilla. En la Figura 1.3 marcamos algunos puntos al azar (puntos en verde) y su distancia respecto al retorno promedio (líneas azules). Así, gráficamente,

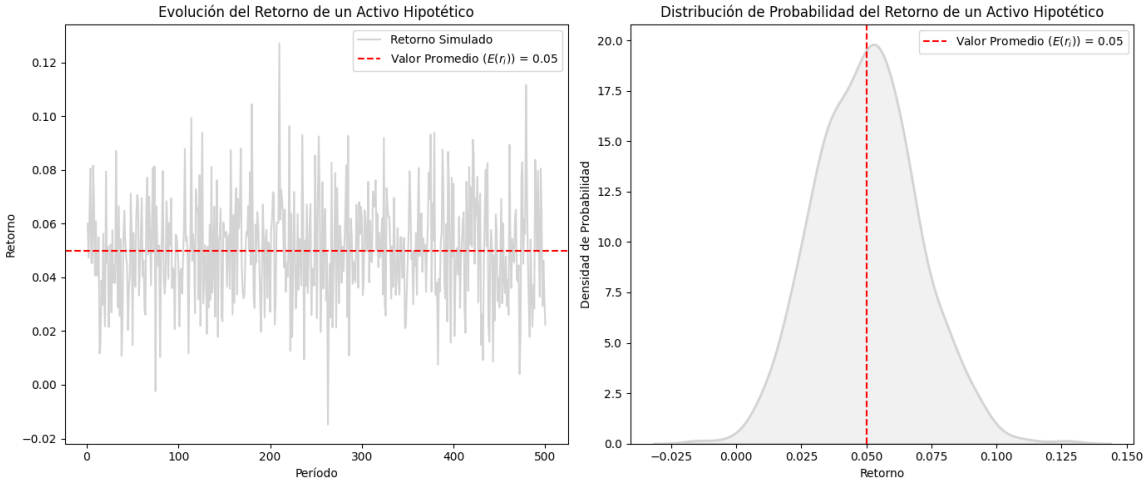


Figura 1.2: Simulación de datos con alta y baja volatilidad

las medidas de dispersión basadas en distancia respecto a la media no son más que funciones que evalúan la magnitud de las distancias entre todos los puntos del conjunto de datos (i.e., todas las líneas azules que se pueden trazar) y el valor promedio $\mathbf{E}(R_i)$. Un poco más formalmente, lo que estamos buscando con este tipo de medidas es una función $f(x) : \mathbb{R}^n \to \mathbb{R}$ que evalúe las distancias entre cada valor de la variable aleatoria x respecto a su promedio $\mathbf{E}(x)$. Es decir, definiendo la familia de medidas de dispersión basadas en distancia respecto a la media como $V(x)$ definimos de manera general como:

$$V(x) = f(x - \mathbf{E}(x)) \tag{1.1}$$

Esta función $V(x)$ a su vez puede tomar varias formas. Las más usuales son las siguientes:

- Desvío standard: $V(x) \equiv \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[x_i - \mathbf{E}(x)]^2}$,
- Varianza: $V(x) \equiv \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}[x_i - \mathbf{E}(x)]^2$,
- Skewness: $V(x) \equiv \tilde{\mu}^3 = \frac{\sum_{i=1}^{n}[x_i - \mathbf{E}(x)]^3}{n\sigma^3}$,

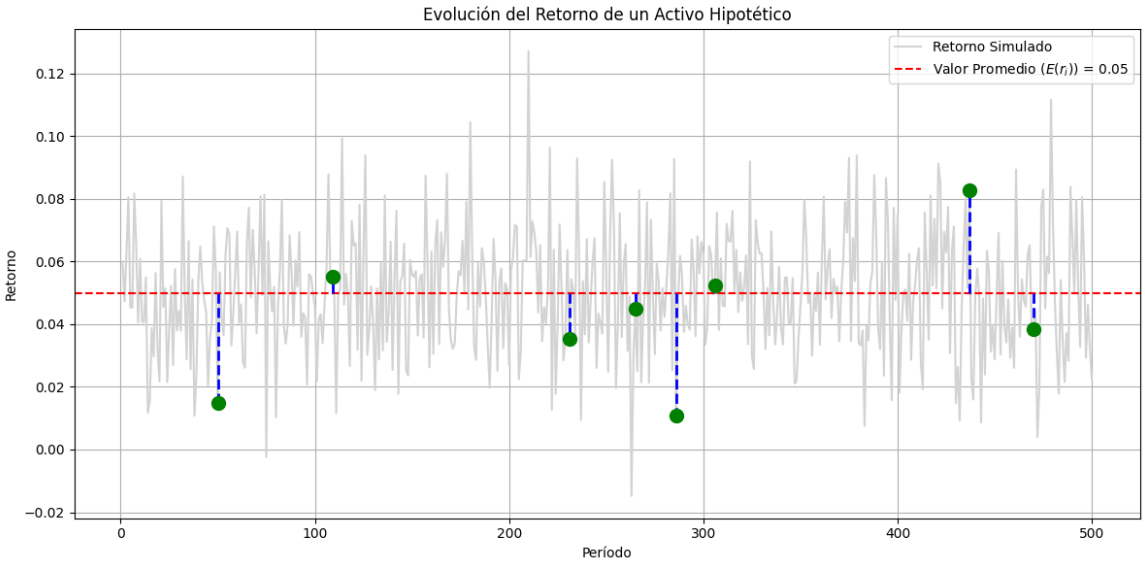


Figura 1.3: Simulación de datos con alta y baja volatilidad

- Kurtosis: $V(x) \equiv \tilde{\mu}^4 = \frac{\sum_{i=1}^{n}[x_i - \mathbf{E}(x)]^4}{n\sigma^4}$,

donde n denota el tamaño del conjunto de datos.

A simple vista se puede apreciar que estas medidas de dispersión tienen una forma similar entre sí. Esto es porque, en teoría de la probabilidad, estas medidas de dispersión pertenecen a la familia de *momentos estandarizados*. Según Ramsey et al. (2002), estas medidas están directamente asociadas con los momentos de la distribución y son funciones que responden a la siguiente forma

$$\sigma^k = (\sqrt{\mathbf{E}[x - \mathbf{E}(x)]^2})^k. \tag{1.2}$$

En donde el valor de k determina el orden de la medida de dispersión.

La motivación del uso de este tipo de medidas de dispersión se destaca en el campo de las finanzas empíricas. Como se menciona en Perez et al. (2003), en la literatura financiera se suele utilizar la kurtosis y/o skewness para probar la normalidad de la distribución de datos financieros (mediante

la evaluación de potenciales 'colas pesadas' en su función de densidad). Este tipo de tests ayuda a académicos y practitioners a validar supuestos y procedimientos utilizados para determinar el valor de un activo financiero.

Siguiendo la corriente estadística aplicada, otra forma usual de medir volatilidad en series de tiempo es mediante la llamada *volatilidad estocástica*, producto de estimar un modelo GARCH. Los modelos GARCH (por sus siglas en inglés, *Generalized Autoregressive Conditional Heteroskedasticity*) son ampliamente utilizados en econometría financiera para modelar y predecir la volatilidad de series temporales, particularmente en el análisis de retornos de activos financieros. Estos modelos fueron introducidos por Bollerslev (1986a), como una extensión del modelo ARCH desarrollado por Engle (1982).

En un modelo GARCH(p, q), la varianza condicional de los retornos (σ_t^2) se modela como una función autorregresiva de los términos de varianza pasados y de los errores al cuadrado de períodos previos. El modelo puede expresarse de la siguiente manera:

$$R_t = \mu + \epsilon_t, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim \mathcal{N}(0,1), \tag{1.3}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{1.4}$$

donde R_t representa el retorno del activo financiero en el tiempo t; μ es el término constante o el retorno esperado; ϵ_t es el término de error o *innovación* del modelo en t; σ_t^2 es la varianza condicional, que captura la volatilidad estocástica en t; z_t es un término de ruido blanco estandarizado, normalmente distribuido con media cero y varianza uno;$\omega > 0$, $\alpha_i \geq 0$ (para $i = 1, \ldots, q$), y $\beta_j \geq 0$ (para $j = 1, \ldots, p$) son los parámetros del modelo, que deben cumplir ciertas restricciones para garantizar la positividad de σ_t^2.

Así se desprenden dos propiedades del modelo GARCH particularmente útiles en el contexto de estudiar mercados bursátiles. La primera es la capacidad de capturar la llamada *volatilidad agrupada*; i.e., períodos de alta volatilidad que son seguidos por períodos similares, indicando dependencia temporal en la varianza condicional. La segunda propiedad es la capacidad de capturar *leptocurtosis*,

i.e., distribuciones de retornos con 'colas pesadas' comparadas con una modelización estándar de distribución normal.

La estimación de un modelo GARCH se realiza típicamente mediante métodos de máxima verosimilitud, maximizando la función de log-verosimilitud condicional:

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{t=1}^{T} \left[\ln(2\pi) + \ln(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2} \right], \tag{1.5}$$

donde T es el número total de observaciones. Los parámetros estimados ω, α_i y β_j permiten computar dinámicamente la volatilidad estocástica σ_t^2, lo que proporciona una medida clave para el análisis del riesgo y la toma de decisiones financieras.

Por otro lado, siguiendo la práctica meramente financiera sobre la medición de volatilidad bursátil, en la práctica también se suele usar lo que se conoce como *índice VIX*. El VIX es el símbolo bursátil del *Índice de Volatilidad de la Bolsa de Opciones de Chicago (CBOE)*, una medida ampliamente utilizada para evaluar las expectativas de volatilidad del mercado basadas en las opciones del índice S& P 500.

El origen del VIX se remonta a la investigación en economía financiera de Brenner and Galai (1989). En una serie de estudios iniciados en 1989, estos autores propusieron la creación de una serie de índices de volatilidad, comenzando por uno centrado en la volatilidad del mercado de acciones y ampliándose hacia índices de volatilidad de tasas de interés y tipos de cambio. No fue hasta Brenner (1992) que se llegó al índice VIX que conocemos hoy en día, con el fin de que se actualizara frecuentemente y sirviera como activo subyacente para futuros y opciones.

El VIX mide la volatilidad esperada del mercado, calculando el cambio anualizado esperado en el índice S&P 500 durante los 30 días posteriores a la fecha de medición, a través de sus derivados financieros[1] . Este cálculo se basa en la teoría de opciones y los datos actuales del mercado de

[1]La metodología de cálculo del VIX se fundamenta en la valoración de la volatilidad implícita a partir de los precios de opciones del S&P 500, considerando tanto *calls* como *puts* . Específicamente, se utiliza una fórmula que integra, de forma continua, los precios de opciones a lo largo de una variedad de precios de ejercicio para estimar la varianza esperada del índice en un horizonte de 30 días.

opciones. Así, el VIX busca reflejar las expectativas de volatilidad futura del mercado en el corto plazo.

Finalmente, hay un conjunto de medidas de dispersión que no son tan usadas en el mundo de las finanzas tradicionales, pero son relevantes de poner en foco. Entre ellas se destaca la medida de *entropía*. La entropía es un concepto de mecánica estadística que está comúnmente asociado al grado de desorden, aleatoriedad o incertidumbre de un sistema. Según Wehrl (1978), el concepto tradicional de entropía se deriva de consideraciones termodinámicas fenomenológicas basadas en la segunda ley de la termodinámica, y puede ser considerada como una medida del *caos* o de falta de información sobre un sistema.

Siguiendo a Karaca et al. (2022), una de las formas usuales de calcular entropía es mediante la llamada *Entropía de Shannon*. Esta medida es capaz de detectar aspectos no lineales en series de datos, contribuyendo a una explicación más confiable respecto a la dinámica no lineal de diferentes puntos de análisis, lo que a su vez mejora la comprensión de la naturaleza de los sistemas complejos. Formalmente, para una variable x_i con M_x posibles estados, cada cual con una probabilidad $p(x_i)$, la entropía de Shannon $H_s(x)$ se calcula como

$$H_s(x) = -\sum_{i=1}^{M_x} p(x_i) \log p(x_i). \tag{1.6}$$

1.2. ¿Qué relación hay entre volatilidad y estabilidad económica?

En esta sección proponemos un modelo económico-financiero de carácter exclusivamente teórico, cuya finalidad es ofrecer un marco lógico bajo el cual la volatilidad bursátil puede considerarse un motor relevante para explicar fluctuaciones del PBI. La validación empírica y/o parametrización

Para obtener una estimación de la volatilidad esperada a 30 días, se realiza una interpolación lineal entre las estimaciones de varianza derivadas de dos vencimientos contiguos, ponderando de acuerdo con el número de días restantes hasta cada vencimiento. Esta metodología, desarrollada y refinada por la CBOE, permite capturar de manera robusta la expectativa de volatilidad implícita en el mercado, evitando sesgos que podrían derivarse de la dependencia de modelos paramétricos y considerando la información completa contenida en la estructura de precios de las opciones (véase Chicago Board Options Exchange (2009)).

explícita de este esquema exceden el foco de este trabajo [2]. Más bien, recurriendo a nociones de microeconomía clásica, finanzas y teoría del crecimiento, buscamos ilustrar de manera intuitiva los canales a través de los cuales la incertidumbre de los mercados de capitales afecta al ritmo de expansión económica. De este modo, motivamos la razonabilidad de emplear la volatilidad como variable clave a la hora de diseñar modelos predictivos del crecimiento del PBI[3].

1.2.1. Producto nacional

Para comenzar, vamos a asumir que en esta economía hay una sola firma, por lo que la producción nacional en un periodo dado Y_t es igual al producto de esta única firma. A su vez, asumimos que esta firma utiliza dos insumos de producción: capital físico k_t y la ganancia de capital financiero γ_t[4]. Formalmente:

$$Y_t = f(k_t, \gamma_t), \tag{1.7}$$

Por simplicidad asumiremos que Y_t es continua, diferenciable y homogénea de grado uno. En este modelo es fácil notar que la tasa de crecimiento del producto nacional $\dot{Y}_t$ vendrá dada básicamente por dos componentes, la forma funcional de $f(k_t, \gamma_t)$ y de la tasa de crecimiento de los distintos

[2]Este punto se deja pendiente para futuras investigaciones.

[3]A continuación, en los capítulos posteriores, no regresaremos a este modelo teórico como objeto de prueba o contraste empírico. En su lugar, se desarrollará un número de especificaciones predictivas -algunas basadas en técnicas de econometría tradicional y otras en algoritmos de machine learning- cuyo propósito es cuantificar la capacidad explicativa de la volatilidad bursátil sobre las variaciones reales de PBI. El andamiaje teórico aquí expuesto sirve exclusivamente para justificar la elección de hipótesis y variables en dichos ejercicios predictivos, garantizando así una coherencia intelectual que atraviesa todo este trabajo.

[4]Véase esta especificación como referencia en (Bencivenga and Smith, 1991; Greenwood and Jovanovic, 1990; King and Levine, 1993). La incorporación de capital físico y *ganancia* del capital físico como insumos de producción no es sutil. Cabe destacar que el uso del *stock* de capital financiero como insumo productivo puede resultar en una doble contabilización del capital total, dado que los activos financieros representan derechos sobre activos reales ya existentes (Tobin (1969), Gurley and Shaw (1955)). En consecuencia, en este modelo se propone como proxy del capital financiero no su stock contable, sino su ganancia de capital, a los fines de capturar el canal de transmisión financiera sin incurrir en duplicación conceptual.

tipos de capital. Es decir:

$$\dot{Y}_t \equiv \Delta Y_t = \Delta f(k_t, \gamma_t) = \frac{\partial f(k_t, \gamma_t)}{\partial k_t} \Delta k_t + \frac{\partial f(k_t, \gamma_t)}{\partial \gamma_t} \Delta \gamma_t, \tag{1.8}$$

De la ecuación 1.8 se desprende que la tasa de crecimiento del producto guarda una estrecha relación con la tasa de crecimiento de capital físico Δk_t y con la variación de la ganancia de capital financiero $\Delta \gamma_t$. Lo que resulta interesante notar aquí es que, incluso si asumimos un estado invariante en capital físico (e.g., en caso de llegar a un máximo de capacidad instalada), el producto nacional puede variar a causa de fluctuaciones en los mercados financieros. Es decir, si asumimos una economía que dejó de acumular capital físico, el crecimiento del producto vendrá dado por:

$$\dot{Y}_t = \frac{\partial f(k_t, \gamma_t)}{\partial \gamma_t} \Delta \gamma_t, \tag{1.9}$$

Cabe preguntarse entonces ¿qué es la ganancia de capital financiero? En su forma más sencilla podemos asumir que este insumo de la función de producción viene dado por el retorno del activo financiero subyacente R_t y la cantidad de este activo disponibilizado por la compañía w_t (i.e., variación de la capitalización bursátil, o *market cap*, de la firma). Así, la ganancia de capital financiero se puede expresar como:

$$\gamma_t = w_t R_t, \tag{1.10}$$

Ahora bien, si asumimos que la empresa no emite nuevas acciones y no recompra acciones ya emitidas, el stock de activo bursátil queda fijo ($w_t = w \ \forall t$), por lo que la variación de la ganancia de capital financiero se simplifica a una función de la tasa de crecimiento del retorno del activo. Por tanto, la tasa de crecimiento del producto resulta en:

$$\dot{Y}_t = w \frac{\partial f(k_t, \gamma_t)}{\partial \gamma_t} \Delta R_t, \tag{1.11}$$

1.2.2. Mercados de capitales

Respecto al funcionamiento de los mercados de capitales en esta economía, asumimos que el precio de las acciones que emite la firma cotiza en t a un precio P_t, y que la firma representativa distribuye dividendos D_t a los tenedores de este activo. Así, podemos definir que el retorno del activo R_t vendrá dado por la suma entre dividendos recibidos en t y la ganancia (pérdida) de capital por la tenencia del instrumento $(P_t - P_{t-1})$:

$$R_t = \frac{D_t + P_t - P_{t-1}}{P_{t-1}}, \tag{1.12}$$

Esta relación puede ser reexpresada como

$$R_t = \frac{D_t}{P_{t-1}} + \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{1.13}$$

En donde el primer componente de la suma representa la rentabilidad por dividendos (i.e., la parte del rendimiento procedente de dividendos relativo al precio actual), y el segundo componente representa la rentabilidad por ganancias de capital (i.e., la parte del rendimiento proveniente del cambio en el precio del activo).

Asimismo, como se elabora en Williams (2014), es de común acuerdo en la literatura clásica de economía financiera (introducido originalmente por Gordon and Shapiro (1956)), que el precio de cualquier activo en equilibrio competitivo se lo puede definir como el valor presente del flujo de dividendos futuros. Formalmente

$$P_t = \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1 + r_t)^i}, \tag{1.14}$$

En donde r_t denota el costo de capital (o tasa de descuento) por la cual se trae a valor presente la capitalización de flujos futuros.

En la práctica financiera básica, el concepto de dividendos está estrechamente relacionado con los

flujos de fondos de las compañías. Acorde al Corporate Finance Institute[5], la distribución de dividendos de cualquier compañía depende de los beneficios (netos de impuestos) que las firmas tienen para libre distribución entre sus accionistas. Dicho de otro modo, los dividendos dependen de qué tan rentable sea la compañía. En términos económicos, decimos que los dividendos son función directa de la productividad marginal de capital físico ($\frac{\partial Y_t}{\partial k_t} := PMg_{k_t}$) de la empresa en cuestión.

$$D_{t+i} = f(PMg_{k,t+i}), \tag{1.15}$$

Por simplicidad, asumimos que la empresa representativa re-invierte una porción de sus beneficios (κ) para mantener su capacidad instalada o para hacer crecer su capital físico. Siendo este el caso, podemos asumir entonces que los dividendos a repartir tienen la siguiente dinámica

$$D_{t+i} = \kappa PMg_{k,t+i}, \tag{1.16}$$

Ahora bien, sustituyendo la ecuación (1.16) en (1.14) obtenemos

$$P_t = \kappa \sum_{i=1}^{\infty} \frac{PMg_{k,t+i}}{(1+r_t)^i}, \tag{1.17}$$

Así, la ecuación (1.17) muestra que el precio actual del activo (P_t) viene dado por el valor presente de los flujos futuros de la productividad marginal del capital, ajustado por la constante de reinversión de capital (κ). Como corolario, de la ecuación (1.17) podemos notar que, aun manteniendo constante la productividad marginal del capital (i.e., aun asumiendo que la firma es igualmente productiva y/o rentable); shocks exógenos en r_t (e.g., shocks de política monetaria, restricciones de financiación, etc.), van a afectar el precio actual del activo P_t generando volatilidad exógena fuera del control del management de la compañía.

Sustituyendo las ecuaciones (1.17) y (1.16) en 1.13 obtenemos la siguiente relación entre el retorno

[5]Ver website oficial disponible en este enlace

del activo y la productividad marginal del capital

$$R_t = \frac{PMg_{k,t}}{\sum_{i=1}^{\infty} \frac{PMg_{k,t-1+i}}{(1+r_{t-1})^i}} + \frac{\sum_{i=1}^{\infty} \frac{PMg_{k,t+i}}{(1+r_t)^i} - \sum_{i=1}^{\infty} \frac{PMg_{k,t-1+i}}{(1+r_{t-1})^i}}{\sum_{i=1}^{\infty} \frac{PMg_{k,t-1+i}}{(1+r_{t-1})^i}}, \tag{1.18}$$

Si la esperanza de supervivencia de la compañía es lo suficientemente larga (i.e., en caso de que podamos efectivamente aproximar el valor presente como una suma infinita), y si asumimos que el costo de capital es estable (i.e., $r_t = r_{t-1} = r$); entonces

$$\sum_{i=1}^{\infty} \frac{PMg_{k,t+i}}{(1+r)^i} \approx \sum_{i=1}^{\infty} \frac{PMg_{k,t-1+i}}{(1+r)^i}, \tag{1.19}$$

Por lo que la ecuación (1.18) se simplifica a

$$R_t = \frac{PMg_{k,t}}{\sum_{i=1}^{\infty} \frac{PMg_{k,t+i}}{(1+r)^i}}, \tag{1.20}$$

La ecuación (1.20) tiene implicancias muy importantes. La primera es notar que el retorno corriente del activo viene dado por cuánto del valor presente de la compañía representa la productividad marginal actual. Dado que un supuesto usual en teoría de la firma es asumir *rendimientos decrecientes a escala* (i.e., productividad marginal positiva pero decreciente a tasa creciente), entonces la ecuación (1.20) nos indicaría que firmas en etapa de crecimiento (a.k.a., todavía en proceso de acumulación de capital respecto a su estado estacionario) tenderán a reportar retornos más altos que aquellas firmas que ya estén consolidadas [6].

Un segundo punto a destacar es que una simple derivación de la ecuación (1.20) nos indica que la *volatilidad* del retorno del activo σ_{R_t} viene dado por las fluctuaciones (tanto transitorias como

[6] Un punto interesante a analizar es que para aquellas firmas que exhiban rendimientos crecientes a escala (a.k.a., monopolios naturales) tenderían a reportar, asimismo, retornos bajos en su etapa de crecimiento y retornos más elevados en medida que acumulan capital. Estudios sobre este tipo de dinámicas se deja abierto para futuras investigaciones.

permanentes)[7] en la productividad de la compañía respecto a su valor inherente de *largo plazo*. Formalmente, definiendo la volatilidad de R_t como

$$\sigma_{R_t} \equiv \Delta R_t, \tag{1.21}$$

Y el valor presente de la productividad marginal del capital como

$$P\bar{M}g_k \equiv \sum_{i=1}^{\infty} \frac{PMg_{k,t+i}}{(1+r)^i}, \tag{1.22}$$

La volatilidad del retorno de un activo viene dada por fluctuaciones de su productividad marginal respecto de su valor presente de largo plazo. Dicho de otro modo, incluso en ausencia de imperfecciones de mercado, la volatilidad en los mercados de capitales proporciona información valiosa sobre cómo está fluctuando la productividad marginal del capital actual respecto a su valor de equilibrio.

$$\sigma_{R_t} = \Delta(PMg_{k,t} - P\bar{M}g_k), \tag{1.23}$$

1.2.3. Interés para la toma de decisiones en políticas públicas

Si consideramos las ecuaciones 1.11, 1.21, y 1.23 derivamos que la tasa de crecimiento del producto nacional está estrechamente relacionada con las fluctuaciones de productividad marginal del capital físico respecto a su valor de equilibrio.

$$\dot{Y}_t = w\frac{\partial f(k_t, \gamma_t)}{\partial \gamma_t}\Delta(PMg_{k,t} - P\bar{M}g_k), \tag{1.24}$$

De aquí la relevancia de policy makers por monitorizar las fluctuaciones de los mercados bursátiles. La

[7] No se explora en este trabajo, pero una fluctuación transitoria puede ser definida como un shock espontaneo dado, por ejemplo, por regulaciones favorables (e.g., como adjutidaciones de licitaciones por períodos de tiempo prefijados); mientras que una fluctuación permanente puede ejemplificarse como la incorporación de una nueva tecnología que aumente de forma sostenida la productividad de la compañía.

volatilidad financiera trae implícita información importante sobre la productividad de las empresas subyacentes y, en última instancia, sobre el crecimiento del producto nacional.

Pragmáticamente, las fluctuaciones de PBI son reportadas con frecuencia lenta (usualmente de manera trimestral), mientras que la volatilidad de los mercados de capitales es de publicación de alta frecuencia (incluso de manera horaria o por minuto). Es por este tema que el seguimiento minucioso de la volatilidad bursátil puede resultar relevante como factor predictivo de la tasa de crecimiento del PBI y, en última instancia, de anticipación de potenciales crisis económicas.

Otro factor interesante a notar es cuando la firma representativa presenta rendimientos decrecientes a escala[8]. Siendo este el caso, entonces la relación en el tiempo entre la tasa de crecimiento del producto y la volatilidad pasa a ser negativa.

$$\frac{\partial \dot{Y}_t}{\partial \sigma_{R_t}} < 0, \tag{1.25}$$

Lo que implicaría que la presencia de volatilidad en los mercados de capitales afecta negativamente a la tasa de crecimiento del producto nacional. Intuitivamente, esto refiere a que la presencia de volatilidad en los mercados bursátiles son un reflejo de la incertidumbre sobre la capacidad productiva de las firmas. A mayor incertidumbre sobre el devenir de la productividad marginal, tanto inversores como empresarios se verán cautelosos a la hora de derivar recursos escasos para fines productivos. Esto, a fin de cuentas enfría la tasa de crecimiento del producto nacional. Gráficamente, decimos que la pendiente de la curva *ceteris paribus* entre tasa de crecimiento del producto $\dot{Y}_t$ y la volatilidad bursátil σ_{R_t} tiene pendiente negativa.

En suma, el marco teórico aquí planteado busca capturar una intuición pragmática. La incertidumbre *paraliza*: frena decisiones de inversión, detiene patrones de consumo y afecta decisiones de ahorro de los agentes. La volatilidad bursátil es una medida tangible que los hacedores de política tienen a mano para medir incertidumbre. De allí que resulta relevante para *policymakers* poder hacer uso de herramientas predictivas que tomen como input la volatilidad de los mercados de capitales para poder predecir potenciales devenires en producto bruto.

[8]Por ejemplo, tras asumir una función de producción del tipo CES (a.k.a., elasticidad de sustitución constante entre insumos productivos), y rendimientos decrecientes a escala respecto a k_t y γ_t.

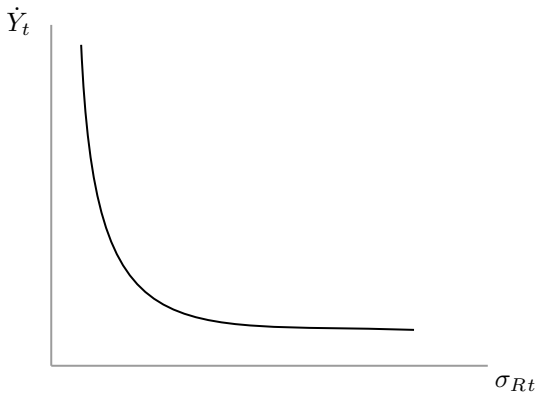


Figura 1.4: Relación entre crecimiento del producto nacional y volatilidad bursátil

1.3. Descripción, análisis y procesamiento de datos de interés

Dentro del marco teórico expuesto en la sección anterior, procedemos a describir la naturaleza y procedimiento de los datos utilizados en este trabajo. Para ello, hacemos uso de la serie histórica más larga posible dada la naturaleza de los datos, tomando como punto de llegada diciembre de 2024 para contemplar el año calendario cerrado más reciente al momento de redactar este trabajo.

1.3.1. Datos Macro

Respecto a la serie histórica macroeconómica consideramos información de acceso público provista por la Reserva Federal de St. Louis (FRED)[9]. Como principal variable macroeconómica consideramos la tasa de crecimiento del PBI real de los Estados Unidos ajustada por estacionalidad [10], con información trimestral reportada desde abril de 1947 en adelante. La Figura 1.5 debajo ilustra la información disponible.

Esta información es producida por el Bureau of Economic Analysis, y reporta el valor de los bienes y servicios producidos por la economía americana, menos el valor de los bienes y servicios utilizados en la producción. Esta métrica refleja la suma de los gastos de consumo personal, la inversión privada interna bruta, las exportaciones netas de bienes y servicios, y el gasto e inversión bruta del gobierno. Los valores reportados son estimaciones ajustadas por inflación.

[9] Ver sitio web oficial disponible en el siguiente enlace: https://fred.stlouisfed.org/

[10] Disponible bajo el código de la FRED A191RL1Q225SBEA, disponible en el siguiente enlace: https://fred.stlouisfed.org/series/A191RL1Q225SBEA

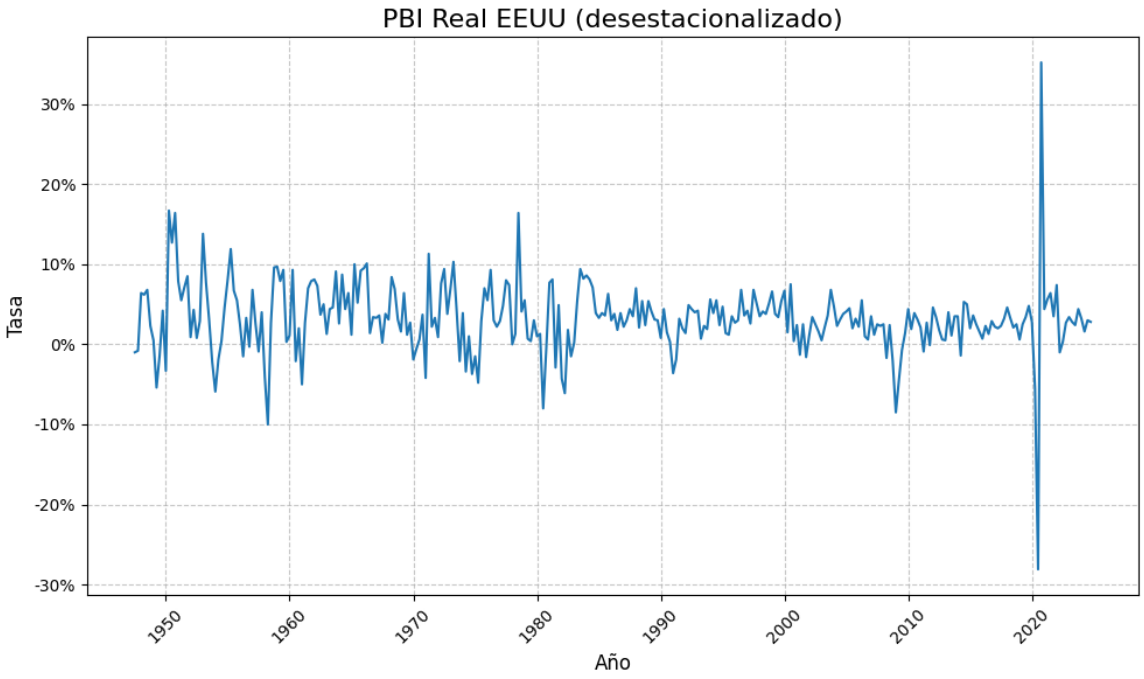


Figura 1.5: Evolución histórica de la tasa de crecimiento del PBI real de los Estados Unidos

De esta fuente también se dispone de otras variables macroeconómicas que se usarán en este trabajo para incorporar robustez al conjunto de variables independientes. Una de ellas es la serie histórica de la tasa de desempleo mensual desestacionalizada de los Estados Unidos[11]. Esta información se reporta desde enero de 1948 y se presenta en la Figura 1.6 debajo.

La FRED define la tasa de desempleo como el porcentaje de personas desempleadas sobre la población habilitada para pertenecer a la fuerza laboral. Los datos sobre la fuerza laboral se limitan a personas de 16 años o más, que residen actualmente en uno de los 50 estados o el Distrito de Columbia, que no viven en instituciones (por ejemplo, cárceles, centros psiquiátricos, hogares para personas mayores) y que no están en servicio activo en las Fuerzas Armadas.

[11] Disponible bajo el código de la FRED UNRATE, disponible en el siguiente enlace: https://fred.stlouisfed.org/series/UNRATE

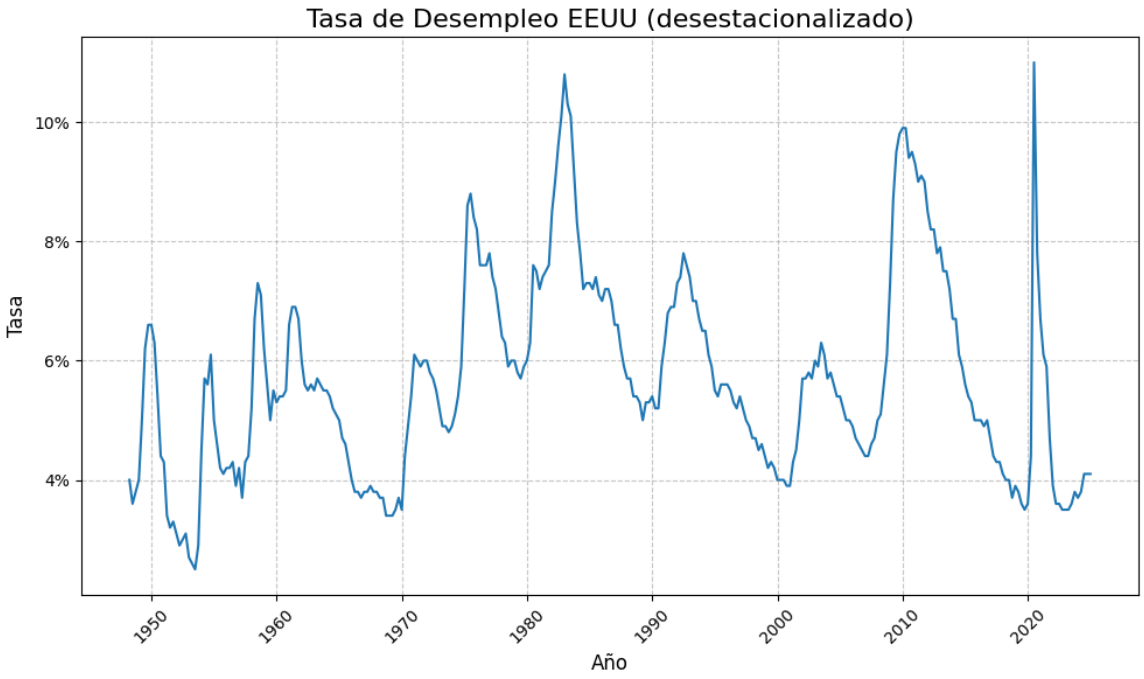


Figura 1.6: Evolución histórica de la tasa de desempleo de los Estados Unidos

Adicionalmente, evaluamos la tasa de inflación histórica de los Estados Unidos, medido mediante la variación del índice de precios al consumidor (CPI)[12]. Como se ilustra en la Figura 1.7, contamos con información reportada desde enero de 1957.

Esta medida de inflación la produce el Bureau of Labor Statistics, y busca reflejar el precio promedio pagado por consumidores urbanos por una canasta típica de bienes y servicios, excluyendo alimentos y energía. Esta medición, conocida también como 'CPI Core', es utilizada comúnmente por economistas y practitioners debido a que los precios de los alimentos y la energía tienden estar fuertemente influenciados por fluctuaciones estacionales.

[12]Disponible bajo el código de la FRED CPILFESL, disponible en el siguiente enlace: https://fred.stlouisfed.org/series/CPILFESL

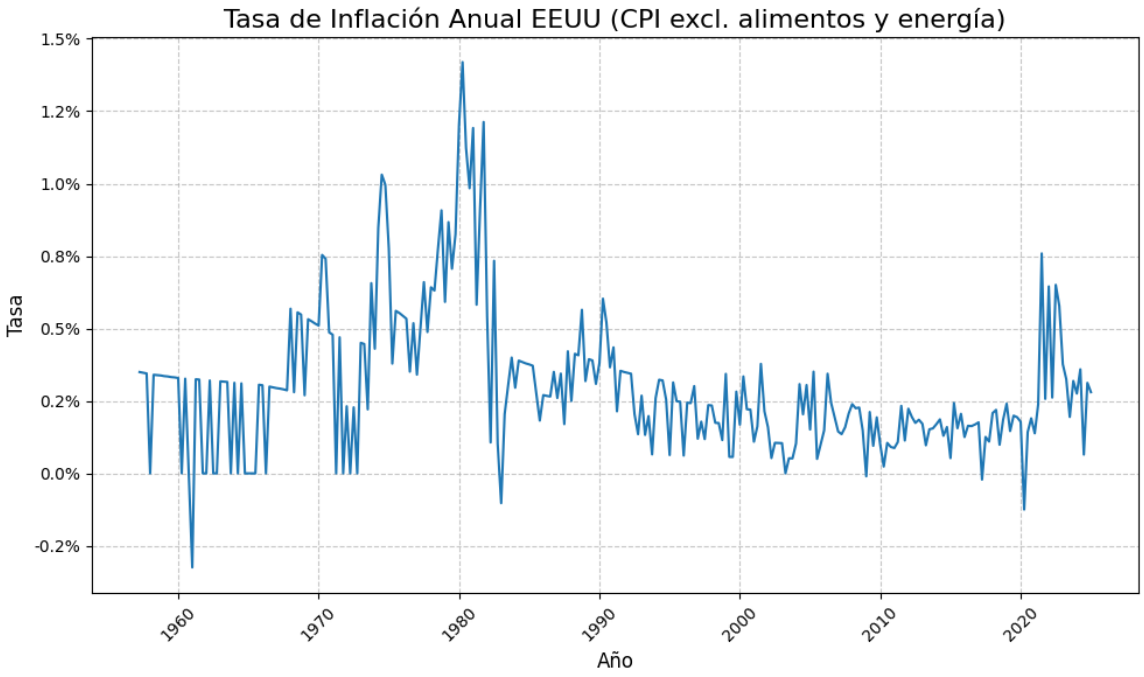


Figura 1.7: Evolución histórica de la tasa de inflación de los Estados Unidos

A modo de resumen, la Tabla 1.3.1 presenta una serie de estadísticas descriptivas de las variables macroeconómicas a utilizar.

	Crecimiento del PBI	**Tasa de Desempleo**	**Tasa de Inflación**
count	310	308	272
mean	3.21 %	5.67 %	0.30 %
std	4.55 %	1.68 %	0.25 %
min	-28.10 %	2.50 %	-0.32 %
25 %	1.30 %	4.40 %	0.15 %
50 %	3.10 %	5.50 %	0.26 %
75 %	5.07 %	6.70 %	0.37 %
max	35.20 %	11.00 %	1.42 %

Cuadro 1.1: Estadísticas descriptivas de variables macro

1.3.2. Datos Financieros

La información bursátil que se toma como punto de partida es un conjunto de acceso público, provisto por la API de la plataforma Yahoo Finance[13]. Esta provee series históricas de frecuencia diaria para los 506 componentes del S&P 500, sin límite de descarga. La serie bursátil disponible cubre desde enero 1962 en adelante, dependiendo de la longevidad de las firmas disponibles. De esta base de datos se puede acceder libremente a información sobre precio, volumen, capitalización de mercado, etcétera. Dada la complejidad del procedimiento propuesto en este trabajo, mientras mayor sea la cantidad de datos, más robustos serán los resultados obtenidos.

Se eligieron estas 506 compañías por capturar aproximadamente el 80 % de toda la capitalización de mercado en Estados Unidos, y por corresponder a uno de los índices más seguidos por la práctica profesional. Según el Reporte Metodológico Oficial del S&P 500[14], las empresas que componen este índice son elegidas por un comité que selecciona compañías de forma que sean representativas de las industrias que operan en la economía de Estados Unidos. Para ser agregada al S& P 500, una empresa debe satisfacer los siguientes requerimientos de liquidez y tamaño: (i) la capitalización bursátil debe ser igual o mayor a 4000 millones de dólares norteamericanos; (ii) la relación entre el monto anual en dólares negociado a la capitalización bursátil ajustada debe ser superior a 1,0; y (iii), el volumen de acciones negociado mensualmente debe por lo menos ser de 250.000 acciones en cada uno de los seis meses previos a la fecha de evaluación. Finalmente, las acciones deben ser ofrecidas de manera pública bien en el NYSE (NYSE Arca o NYSE MKT) o en el NASDAQ (NASDAQ Global Select Market, NASDAQ Select Market o el NASDAQ Capital Market).

Habiendo descargado la serie de retornos diarios R_t para todas las compañías del S&P500, desde el momento en que cada firma fue listada en el índice hasta el dato más reciente disponible, resta preguntarse ¿qué medida de volatilidad usamos?. En la Sección 1.1 ilustramos que existen muchas medidas de volatilidad para elegir. Las mismas pueden ser calculadas para cada compañía a través del tiempo (a.k.a., volatilidad *within*); o también pueden ser calculadas para cada momento del tiempo, pero midiendo *entre compañías* (a.k.a., volatilidad *between*). En el presente trabajo vamos a adoptar el conjunto de medidas discutidas en la Sección 1.1, tomando todas las medidas de momentos estandarizados para medir volatilidad within; y tomando la medida de entropía para medir volatilidad

[13] Disponible de manera gratuita en el sitio web de Yahoo Finance: https://finance.yahoo.com/.

[14] Ver reporte metodológico de S& P 500 disponible en el sitio web oficial: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf

tanto within como between.

Como el lector habrá advertido, teniendo i compañías, n periodos de tiempo, j medidas de volatilidad within, y k medidas de volatilidad between; el dataset de trabajo contiene un volumen de datos considerable (de hecho, contiene $i \times n \times j \times k$ observaciones). Para ilustrar, de tener 506 compañías, 5 medidas de volatilidad within, 1 medida de volatilidad between e información histórica diaria desde 1962 hasta 2024; fácilmente llegamos a una base de datos de más de 50 millones de observaciones.

1.3.3. Reducción de dimensionalidad mediante Análisis de Componentes Principales

Según Max Garzon (2022), el manejo de un gran volumen de información plantea desafíos prácticos especialmente en el contexto del entrenamiento de modelos de *machine learning*, donde es común calcular una gran cantidad de parámetros para un mismo conjunto de datos. En este campo, el concepto de dimensionalidad refiere a las características estructurales del conjunto regresores, entendida como la cantidad de columnas que compone la matriz de variables explicativas. En nuestro caso, el dataset de trabajo se representa mediante una matriz en la que cada columna corresponde a una variable (o covariable), que luego se utilizarán para el entrenamiento de los modelos predictivos.

En este trabajo, se adoptan técnicas de reducción de dimensionalidad para sintetizar la mayor cantidad posible de información de las $i \times j \times k$ covariables (i.e., todas las medidas de volatilidad, tanto within como between, de cada uno de los componentes del S&P500); en pos de identificar y eliminar redundancias en la información, al mismo tiempo que preservamos la factibilidad computacional de los algoritmos empleados. Una de las principales herramientas empleadas es el Análisis de Componentes Principales (*Principal Component Analysis*, PCA), una técnica ampliamente utilizada en estadística y aprendizaje automático [15].

[15] Como menciona Max Garzon (2022), existen diversos métodos de reducción de dimensionalidad (por ejemplo, Escalamiento Multi-Dimensional, Entropía Condicional y Mapeo Isométrico, entre otros). No obstante, el método de Componentes Principales (PCA) resulta especialmente adecuado para nuestro caso, ya que permite obtener un conjunto reducido de j componentes, todos ortogonales entre sí. Esta propiedad es fundamental para sintetizar de manera eficaz la información subyacente de un amplio conjunto de medidas de volatilidad que, en esencia, buscan capturar el mismo concepto fundamental. En efecto, mediante el PCA se extrae la estructura subyacente en los datos, garantizando que cada componente aporte una fracción significativa de la varianza explicada del sistema y, por ende, una contribución clara y libre de redundancias. Así, se mitiga el problema de la multicolinealidad y se reduce el ruido, lo cual resulta crucial para la robustez y la interpretabilidad de los modelos econométricos y de aprendizaje automático

Acorde a Johnson et al. (2002), el PCA se utiliza para explicar la estructura de covarianzas de un conjunto de datos con ρ variables mediante combinaciones lineales de un conjunto reducido de componentes. Matemáticamente, se parte de un vector aleatorio $\mathbf{X} = [X_1, X_2, \ldots, X_\rho]$, cuya matriz de covarianzas es Σ, y se buscan combinaciones lineales específicas de las variables aleatorias $X_1, X_2, \ldots, X_\rho$, tal que las nuevas variables, llamadas componentes principales, capturen la mayor cantidad de varianza posible. Formalmente, las combinaciones lineales están definidas como:

$$
\begin{aligned}
Y_1 &= \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1\rho}X_\rho, \\
Y_2 &= \mathbf{a}_2'\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2\rho}X_\rho, \\
&\vdots \\
Y_\rho &= \mathbf{a}_\rho'\mathbf{X} = a_{\rho 1}X_1 + a_{\rho 2}X_2 + \cdots + a_{\rho\rho}X_\rho,
\end{aligned}
$$

donde $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_\rho$ son vectores de coeficientes. Las varianzas y covarianzas de Y_i están dadas por:

$$
\mathrm{Var}(Y_i) = \mathbf{a}_i'\Sigma\mathbf{a}_i, \quad i = 1, 2, \ldots, \rho, \tag{1.26}
$$

$$
\mathrm{Cov}(Y_i, Y_k) = \mathbf{a}_i'\Sigma\mathbf{a}_k, \quad i, k = 1, 2, \ldots, \rho. \tag{1.27}
$$

Los componentes principales son aquellas combinaciones lineales ortogonales (i.e., no correlacionadas) que maximizan la varianza. Específicamente:

- El primer componente principal, Y_1, es la combinación lineal $\mathbf{a}_1'\mathbf{X}$ que maximiza $\mathrm{Var}(\mathbf{a}_1'\mathbf{X})$ sujeto a que $\mathbf{a}_1'\mathbf{a}_1 = 1$. - El segundo componente principal, Y_2, maximiza $\mathrm{Var}(\mathbf{a}_2'\mathbf{X})$ sujeto a que $\mathbf{a}_2'\mathbf{a}_2 = 1$ y $\mathrm{Cov}(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0$. - El i-ésimo componente principal maximiza $\mathrm{Var}(\mathbf{a}_i'\mathbf{X})$ sujeto a $\mathbf{a}_i'\mathbf{a}_i = 1$ y $\mathrm{Cov}(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0$ para $k < i$.

Este proceso puede expresarse de manera más compacta utilizando autovalores y autovectores de la matriz de covarianzas Σ. Si Σ tiene pares autovalor-autovector $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_\rho, \mathbf{e}_\rho)$, donde $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\rho \geq 0$, entonces:

empleados en este estudio.

$$Y_i = \mathbf{e}_i'\mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{i\rho}X_\rho, \quad i = 1, 2, \ldots, \rho, \tag{1.28}$$

$$\text{Var}(Y_i) = \mathbf{e}_i'\Sigma\mathbf{e}_i = \lambda_i, \tag{1.29}$$

$$\text{Cov}(Y_i, Y_k) = \mathbf{e}_i'\Sigma\mathbf{e}_k = 0, \quad i \neq k. \tag{1.30}$$

Por tanto, el primer componente principal corresponde al autovector asociado al mayor autovalor λ_1, el segundo al segundo mayor autovalor λ_2, y así sucesivamente. Una propiedad importante del PCA es que permite sintetizar una proporción ω de la varianza total utilizando únicamente $j < \rho$ componentes principales, tal que:

$$\frac{\sum_{i=1}^{j} \lambda_i}{\sum_{i=1}^{\rho} \lambda_i} = \omega. \tag{1.31}$$

Esto permite reducir la dimensionalidad del conjunto de datos, conservando la mayor parte de la información.

En términos computacionales, si se normalizan las variables originales para tener media cero y varianza unitaria, el PCA puede realizarse calculando los autovalores y autovectores de la matriz de correlaciones en lugar de la matriz de covarianzas. Esta normalización es crucial cuando las variables tienen diferentes escalas [16].

En el caso que nos concierne, en este trabajo utilizamos el procedimiento de componentes principales para sintetizar la información de volatilidad bursátil (dataset de más de 50 millones de observaciones, incluyendo tanto medidas de volatilidad *between* como *within*), a unas pocas variables de manera tal que se cubra al menos un 85 % de la variabilidad total del sistema [17]. En el Gráfico 1.8 debajo se

[16] Para más detalles, véase Hastie et al. (2009) y Jolliffe (2002).

[17] La elección de fijar el umbral de varianza explicada en el 85 % no es arbitraria. Se observó que, al aumentar dicho umbral hasta el 90 %, el tiempo de cómputo del algoritmo se incrementaría en torno a un 20 %, lo que correspondería únicamente a la inclusión de aproximadamente 5 componentes adicionales. Con el objetivo de garantizar la factibilidad computacional de los procedimientos presentados en este trabajo, se optó por conservar el umbral en el 85 %. Esta práctica, que busca equilibrar la precisión del modelo y los costos computacionales, es consistente con criterios presentados en la literatura, como los expuestos por Jolliffe (2002).

puede ver cuánto porcentaje de la varianza global queda cubierta a medida que incorporamos más componentes principales. De allí se desprende que la cantidad de componentes que cubre al menos 85 % de la varianza global es de 35 componentes.

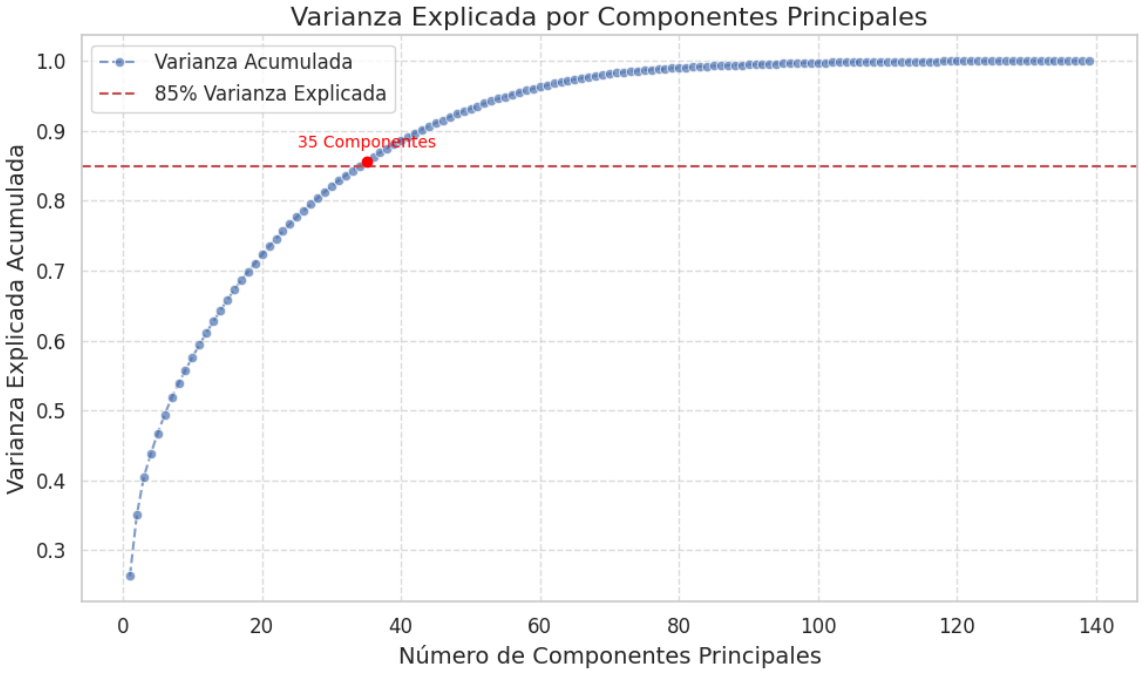


Figura 1.8: Porcentaje de Varianza Acumulada por Componentes Principales

Capítulo 2

Tradición econométrica: Modelos macroeconométricos de series de tiempo

2.1. ¿Cómo tratan los practitioners económicos la predicción de fluctuaciones macroeconómicas?

La capacidad para anticipar las fluctuaciones macroeconómicas resulta fundamental tanto para los responsables de las políticas públicas como para los actores del sector privado. En este contexto, los *practitioners* económicos y financieros han confiado tradicionalmente en modelos econométricos que permiten analizar y predecir la evolución de variables críticas, tales como el Producto Bruto Interno (PBI) o la volatilidad en los mercados financieros. Por ejemplo, según Stock and Watson (2002), dichos modelos ofrecen un marco formal que facilita la identificación de relaciones estadísticas entre variables y la evaluación del impacto de shocks exógenos, lo que resulta especialmente útil para la toma de decisiones en entornos complejos.

En la práctica, entidades como bancos centrales, organismos internacionales y firmas financieras aplican estos modelos para guiar decisiones estratégicas. El análisis de series temporales a través de modelos autorregresivos (AR) y de promedios móviles (MA) se ha establecido como una herramienta estándar para generar pronósticos sobre el crecimiento del PBI Box and Jenkins (1970), mientras que modelos como el GARCH permiten capturar la volatilidad condicional en los mercados financieros Bollerslev (1986b), aportando información clave para la gestión de riesgos y la valoración de activos. En resumen, estos modelos combinan un sólido rigor estadístico con la capacidad de incorporar información histórica y estructural, permitiendo evaluar escenarios alternativos bajo diversos supuestos económicos, aunque también presentan limitaciones ante episodios de alta incertidumbre o cambios estructurales abruptos.

2.2. Mínimos Cuadrados Ordinarios (MCO)

El método de Mínimos Cuadrados Ordinarios (MCO) fue introducido inicialmente por Legendre (1805) y formalizado posteriormente por Gauss (1809). Esta técnica resulta esencial en econometría y estadística para estimar los parámetros de un modelo de regresión lineal, ya que se basa en la minimización de la suma de los cuadrados de los errores (denotados por ϵ_i) entre los valores observados y los predichos. En un modelo de regresión lineal simple, la relación se expresa mediante la ecuación

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

donde y_i representa la variable dependiente, x_i la variable independiente, β_0 y β_1 los coeficientes a estimar, y ϵ_i el error asociado, asumido con media cero y varianza constante ($\epsilon_i \sim \mathcal{N}(0, \sigma^2)$). La representación matricial del modelo es

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

en la que $\mathbf{y}$ es un vector columna de n observaciones, $\mathbf{X}$ es la matriz de diseño (que incluye una columna de unos para representar el intercepto), $\boldsymbol{\beta}$ es el vector de parámetros y $\boldsymbol{\epsilon}$ es el vector de errores. El objetivo del MCO es encontrar el vector $\boldsymbol{\beta}$ que minimiza la suma de los cuadrados de los errores, lo que se traduce en la función

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2.$$

La minimización de esta función conduce a una solución cerrada, obtenida al derivar $Q(\boldsymbol{\beta})$ respecto a $\boldsymbol{\beta}$ e igualar a cero, resultando en la fórmula

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

Bajo los supuestos clásicos (linealidad en los parámetros, exogeneidad, homocedasticidad y ausencia de autocorrelación en los errores), los estimadores $\hat{\boldsymbol{\beta}}$ son insesgados y poseen varianza mínima entre los estimadores lineales insesgados (lo cual se enuncia en el teorema de Gauss-Markov, ver Greene (2003) y Wooldridge (2010)), además de seguir una distribución normal asintótica cuando los errores son normales.

La Figura 2.1 ilustra el proceso de estimación mediante MCO en un espacio en $\mathbb{R}^2$. En ella se muestran los puntos correspondientes a las observaciones simuladas $\{(x_i, y_i)\}$ y la línea recta ajustada (en color rojo) obtenida al estimar los parámetros β_0 y β_1 mediante el método de MCO. Esta línea recta minimiza la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos, es decir, minimiza la función

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2.$$

La solución obtenida garantiza que la línea de regresión pase por la 'centroide' de los datos, lo que implica que, en promedio, la distancia entre los datos y la línea ajustada es mínima. Este resultado visual ilustra cómo el MCO implementa el principio de "minimización del error cuadrático medio" para obtener el modelo de regresión que mejor se ajusta a la información contenida en la muestra.

2.3. Redes Elásticas (Elastic Nets)

El método de Redes Elásticas, desarrollado por Zou and Hastie (2005), es un enfoque que combina la penalización Ridge (L_2) y la penalización Lasso (L_1) para regularizar y realizar selección de variables en modelos de regresión. La penalización Ridge, basada en la norma L_2 (es decir, $\|\boldsymbol{\beta}\|_2^2 = \sum_{j=1}^{p} \beta_j^2$), tiende a reducir la magnitud de los coeficientes sin llevarlos a cero, lo que es útil en presencia de multicolinealidad. Por su parte, la penalización Lasso, basada en la norma L_1 ($\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$), puede reducir algunos coeficientes a cero, promoviendo así la selección de variables relevantes. En este

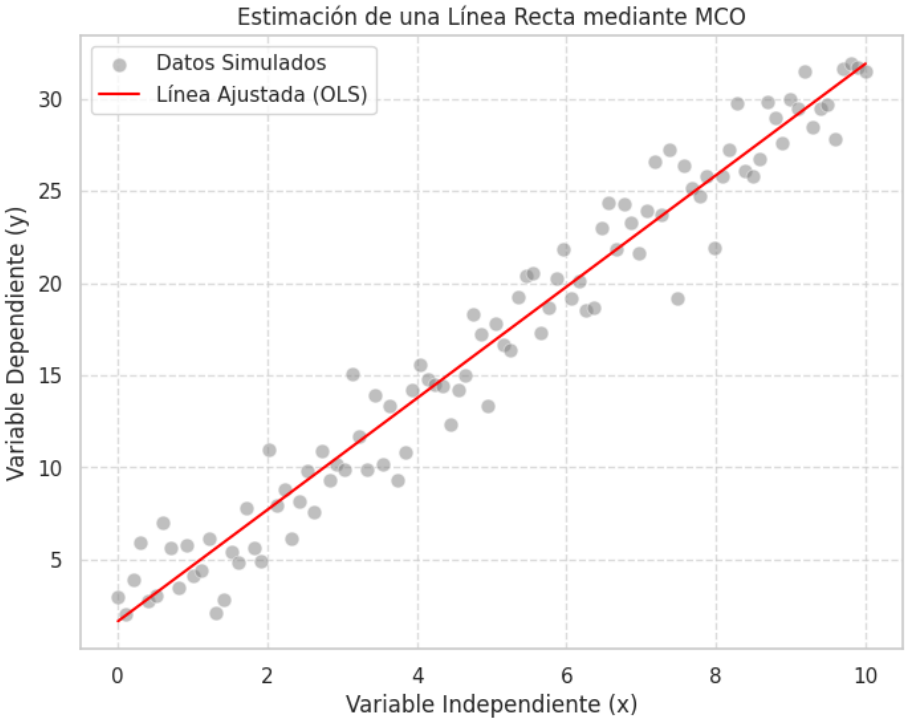


Figura 2.1: Gráfico de dispersión de datos simulados y línea de regresión ajustada mediante MCO.

sentido, la combinación de ambas penalizaciones en el Elastic Net permite aprovechar las ventajas de cada uno, sobre todo en situaciones donde las variables predictoras están altamente correlacionadas o cuando el número de predictores excede el número de observaciones.

Si se considera una variable dependiente $\mathbf{y} \in \mathbb{R}^n$ y una matriz de predictores $\mathbf{X} \in \mathbb{R}^{n \times p}$, el modelo de regresión lineal tradicional se formula como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

donde $\boldsymbol{\beta}$ es el vector de coeficientes y $\boldsymbol{\epsilon}$ es el vector de errores, asumidos con media cero y varianza constante. La función objetivo del Elastic Net incorpora la penalización en la pérdida del MCO de la siguiente manera:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{2n}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \left(\alpha \|\boldsymbol{\beta}\|_1 + \frac{1-\alpha}{2}\|\boldsymbol{\beta}\|_2^2 \right).$$

En esta expresión, el parámetro $\lambda > 0$ controla la magnitud de la penalización, mientras que el parámetro $\alpha \in [0, 1]$ determina el peso relativo entre la penalización L_1 y la L_2: un valor de $\alpha = 1$

equivale al Lasso, $\alpha = 0$ al Ridge, y valores intermedios proporcionan un equilibrio entre ambos. La aplicación de esta metodología resulta particularmente valiosa en escenarios de alta dimensionalidad o cuando se requiere la selección automática de variables, ya que permite obtener estimaciones robustas y estables.

Para comprender de forma intuitiva el efecto de las penalizaciones Ridge y Lasso, es útil visualizar las regiones de penalización en un plano cartesiano donde el vector de parámetros se reduce a dos componentes, $\boldsymbol{\beta} = (\beta_1, \beta_2)$.

Para el caso de la penalización Ridge, la restricción se impone sobre la norma L_2, de manera que se considera el conjunto de puntos que satisfacen:

$$\beta_1^2 + \beta_2^2 \leq t,$$

para algún $t > 0$. Esta región corresponde a un círculo centrado en el origen. La forma circular implica que, a medida que se minimiza la función de pérdida con la restricción impuesta, las soluciones tienden a reducir la magnitud de ambos coeficientes de forma equitativa, sin favorecer que alguno se haga exactamente cero.

En contraste, la penalización Lasso impone una restricción sobre la norma L_1:

$$|\beta_1| + |\beta_2| \leq t.$$

Geométricamente, esta región es un rombo (o diamante) centrado en el origen. La característica distintiva de esta forma es la presencia de 'esquinas' en los ejes. Durante el proceso de minimización, es más probable que la solución óptima se ubique en una de estas esquinas, lo que induce que uno de los coeficientes sea exactamente cero y, por tanto, promueve la selección de variables.

La Figura 2.2 debajo ilustra estas dos regiones de penalización para $t = 1$: el círculo en rojo representa la región L_2 (Ridge) y el rombo en azul la región L_1 (Lasso).

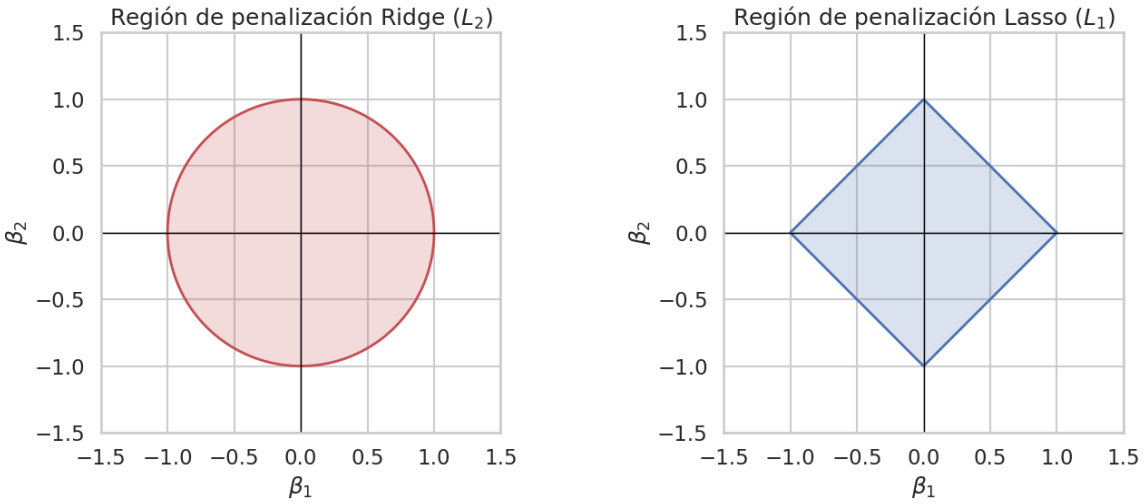


Figura 2.2: Representación geométrica de las regiones de penalización Ridge (L_2) y Lasso (L_1) en el espacio de parámetros (β_1, β_2).

2.4. Regresión Polinómica

La regresión polinómica amplía el modelo lineal clásico al incluir términos polinómicos de la variable independiente, lo que posibilita la modelación de relaciones no lineales. Este enfoque es muy empleado en situaciones donde la relación entre la variable dependiente y la independiente no se ajusta adecuadamente a una simple recta, tal como se discute en la literatura especializada (Draper and Smith (1998), Montgomery et al. (2012)). En un modelo de regresión polinómica de grado d, la relación se expresa de la forma

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d + \epsilon,$$

donde $\beta_0, \beta_1, \ldots, \beta_d$ son los coeficientes del modelo y ϵ representa el error, asumido con media cero y varianza constante. La representación matricial consiste en construir una matriz de diseño en la que cada columna corresponde a x elevado a una potencia desde 0 hasta d. Los coeficientes se estiman minimizando la suma de los cuadrados de los residuos, lo que lleva a la solución analítica familiar

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

siempre que la matriz $\mathbf{X}^\top\mathbf{X}$ sea invertible. Es importante destacar que la elección del grado d debe ser cuidadosamente balanceada para evitar tanto el sobreajuste como el subajuste del modelo, generalmente mediante técnicas de validación cruzada.

Para obtener una intuición geométrica sobre la regresión polinómica, consideremos un caso en el que se generan datos en $\mathbb{R}^2$ a partir de una función polinómica de grado 3 con una tendencia creciente, y se ajusta una curva polinómica a dichos datos. Geométricamente, mientras que en la regresión lineal la estimación corresponde a una recta, en la regresión polinómica la solución es una curva que puede capturar la curvatura en los datos.

La Figura 2.3 ilustra, mediante datos simulados, cómo la curva estimada se ajusta a la dispersión de puntos generados por una función cúbica, mostrando visualmente la capacidad del modelo para adaptarse a la no linealidad. Esta representación ayuda a comprender cómo la incorporación de términos de mayor orden permite que la función de ajuste capture patrones complejos en la relación entre x y y.

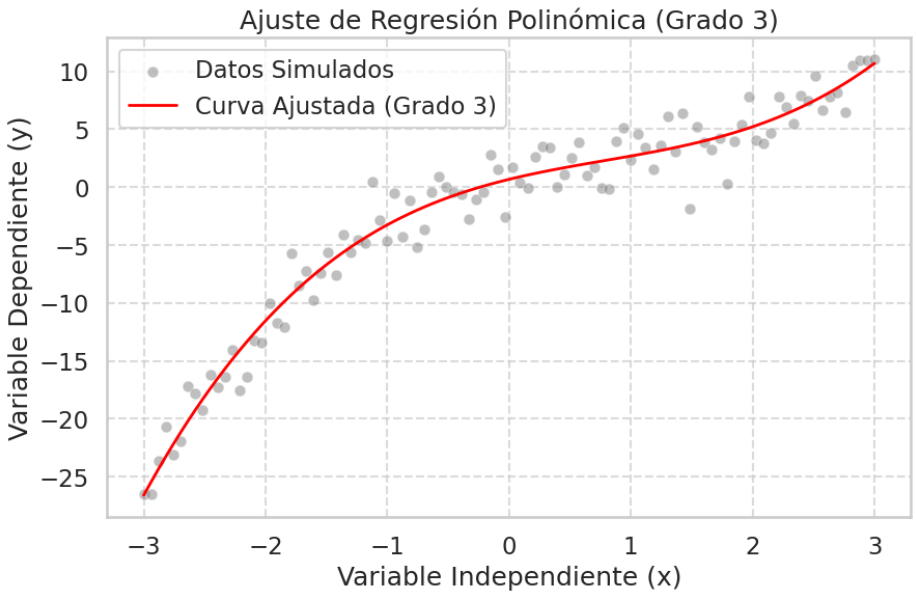


Figura 2.3: Ajuste de regresión polinómica de grado 3 a datos simulados. La curva roja representa la función estimada, que se ajusta a la tendencia y la curvatura presentes en los datos.

2.5. Modelos Autorregresivos (AR)

Los modelos autorregresivos (AR por sus siglas en inglés) se utilizan para describir series temporales basándose en la idea de que el valor actual de la serie depende de sus propios valores pasados y de un término de error. Formalmente, un modelo AR de orden p se expresa como

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t,$$

donde y_t es el valor en el tiempo t, c es una constante, y $\phi_1, \phi_2, \ldots, \phi_p$ son los coeficientes que determinan la influencia de los rezagos, mientras que ϵ_t es el error, asumido con media cero y varianza constante. La condición de estacionariedad, fundamental para la predicción, exige que las raíces del polinomio característico

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

se encuentren fuera del círculo unitario (es decir, que $|z| > 1$). Los parámetros del modelo pueden estimarse mediante métodos de máxima verosimilitud o, alternativamente, mediante mínimos cuadrados aplicados a los datos disponibles. Además, la determinación del orden p se realiza a través de criterios de información, ya que resulta esencial encontrar un equilibrio adecuado entre la complejidad del modelo y su capacidad explicativa.

Una herramienta fundamental para analizar series temporales es la función de autocorrelación (ACF), la cual mide la correlación entre los valores de la serie en distintos rezagos. En el contexto de modelos autoregresivos, el gráfico de la ACF permite identificar la dependencia temporal en los datos y es crucial para determinar el orden p del modelo. Por ejemplo, en un proceso AR(2), se espera que la ACF decaiga gradualmente y que exhiba patrones característicos que reflejen la influencia de los dos rezagos anteriores en el valor actual. La visualización de la ACF facilita la detección de la estacionalidad y la persistencia de la dependencia temporal, ayudando así a ajustar y validar el modelo.

La Figura 2.4 muestra, a modo de ejemplo, la simulación de un proceso AR(2) y su correspondiente gráfico de autocorrelación. En este caso, los datos se generan a partir del proceso

$$y_t = c + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \epsilon_t, \tag{2.1}$$

donde se han seleccionado parámetros de ejemplo que garantizan la estacionariedad [1]. La función de autocorrelación resultante permite observar cómo los rezagos 1 y 2 tienen una correlación significativa con el valor actual, lo que respalda la estructura del modelo AR(2).

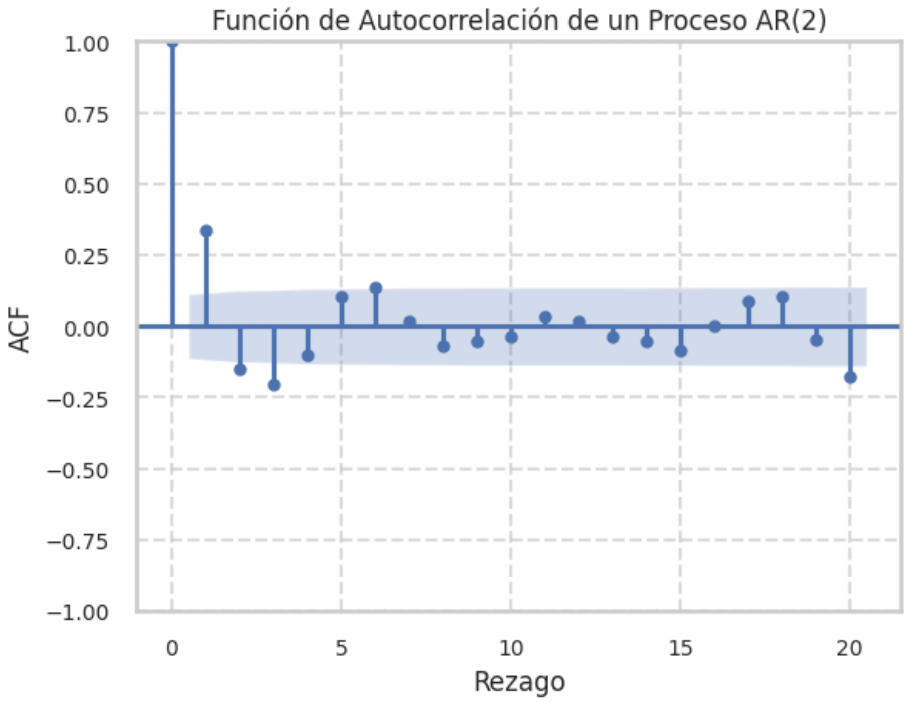


Figura 2.4: Gráfico de autocorrelación (ACF) para el proceso AR(2) simulado, mostrando la dependencia significativa en los rezagos 1 y 2.

2.6. Modelos de Medias Móviles (MA)

En contraste con los modelos AR, los modelos de medias móviles (MA) se basan en la hipótesis de que el valor actual de la serie depende únicamente de los errores pasados. Un modelo MA de orden q se define mediante la ecuación

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{2.2}$$

[1] Se simula un proceso AR(2) con parámetros $\phi_1 = 0{,}5$ y $\phi_2 = -0{,}3$, y generaron 300 observaciones

donde μ es la media de la serie, $\theta_1, \theta_2, \ldots, \theta_q$ son los coeficientes que cuantifican la influencia de los shocks pasados, y ϵ_t es el término de error con distribución normal, media cero y varianza constante. La característica esencial de los modelos MA es su inherente estacionariedad, ya que dependen de errores aleatorios que se asumen independientes. La estimación de los parámetros y la determinación del orden q se realizan mediante técnicas como la máxima verosimilitud y el análisis de la función de autocorrelación, complementadas por criterios de información que permiten equilibrar el ajuste del modelo con su complejidad.

Para ilustrar de forma gráfica el funcionamiento de estos modelos, consideremos el caso de un proceso MA(2):

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}. \tag{2.3}$$

En este modelo, un shock unitario aplicado en el tiempo t produce un efecto de 1 en y_t, un efecto de θ_1 en y_{t+1} y un efecto de θ_2 en y_{t+2}, quedando para $t > t+2$ la respuesta igual a 0. Geométricamente, la función de impulso-respuesta (IRF por sus siglas en inglés) para un MA(2) es:

$$\text{IRF}_{\text{MA(2)}} = \{1,\, \theta_1,\, \theta_2,\, 0,\, 0, \ldots\}.$$

Por ejemplo, si se toma $\theta_1 = 0{,}8$ y $\theta_2 = 0{,}3$, la IRF será $\{1,\, 0{,}8,\, 0{,}3,\, 0,\, \ldots\}$.

En contraste, consideremos un modelo AR(2) de forma acorde a la Ecuación 2.1 donde los parámetros, por ejemplo, son $\phi_1 = 0{,}5$ y $\phi_2 = -0{,}3$. En este modelo, el efecto de un shock se propaga de forma recursiva. La IRF se calcula de la siguiente forma:

$$\text{IRF}_0 = 1,$$

$$\text{IRF}_1 = \phi_1 = 0{,}5,$$

$$\text{IRF}_2 = \phi_1 \times \text{IRF}_1 + \phi_2 \times \text{IRF}_0 = 0{,}5 \times 0{,}5 - 0{,}3 = -0{,}05,$$

$$\text{IRF}_3 = \phi_1 \times \text{IRF}_2 + \phi_2 \times \text{IRF}_1 = 0{,}5 \times (-0{,}05) - 0{,}3 \times 0{,}5 = -0{,}175,$$

$$\text{IRF}_4 = \phi_1 \times \text{IRF}_3 + \phi_2 \times \text{IRF}_2 \approx 0{,}5 \times (-0{,}175) - 0{,}3 \times (-0{,}05) \approx -0{,}0725.$$

Así, la IRF para el modelo AR(2) es:

$$\text{IRF}_{AR(2)} = \{1,\ 0{,}5,\ -0{,}05,\ -0{,}175,\ -0{,}0725,\ \dots\}.$$

La Figura 2.5 muestra, en un panel a la izquierda, la función de impulso-respuesta para el proceso MA(2) (extendida hasta y_{t+4}, donde se observa que la respuesta es 0 a partir de y_{t+3}) y, en un panel a la derecha, la IRF para el proceso AR(2), donde se evidencia que el efecto del shock se propaga recursivamente más allá de y_{t+3}.

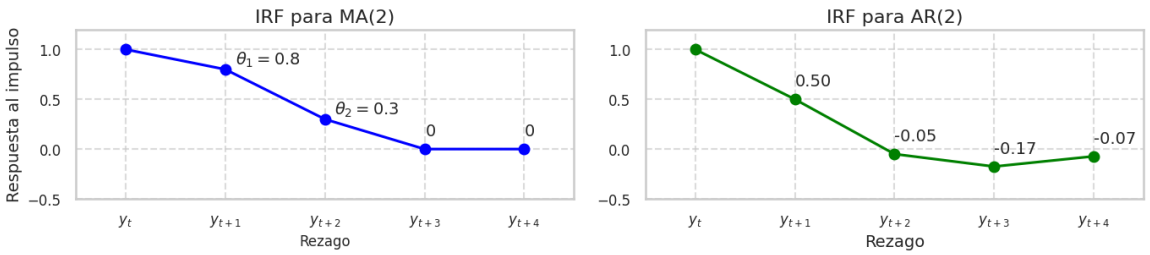


Figura 2.5: A la izquierda: Función de impulso-respuesta de un proceso MA(2) con $\theta_1 = 0{,}8$ y $\theta_2 = 0{,}3$. A la derecha: IRF de un proceso AR(2) con $\phi_1 = 0{,}5$ y $\phi_2 = -0{,}3$.

2.7. Modelos Autorregresivos Integrados con Medias Móviles (ARIMA)

Los modelos ARIMA combinan las ideas de los modelos autorregresivos y de medias móviles, además de incorporar un componente de integración para abordar la no estacionariedad de las series temporales. Un modelo ARIMA se denota como $ARIMA(p, d, q)$, donde p indica el orden del componente autorregresivo, d el número de diferenciaciones requeridas para lograr la estacionariedad, y q el orden del componente de medias móviles. La formulación general del modelo es

$$\phi(L)(1 - L)^d y_t = \theta(L)\epsilon_t,$$

en la que L es el operador de rezago, $\phi(L)$ es el polinomio autorregresivo (definido como $1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$), y $\theta(L)$ es el polinomio de medias móviles (definido como $1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$). El proceso de análisis de un modelo ARIMA implica determinar primero el número de diferenciaciones d necesarias (usualmente a través de pruebas de raíz unitaria, como la prueba de Dickey-Fuller aumentada, Dickey and Fuller (1979)), y posteriormente identificar los órdenes p y q mediante el análisis de las funciones de autocorrelación y autocorrelación parcial. La estimación de los parámetros se efectúa por métodos de máxima verosimilitud o de mínimos cuadrados, y la validez del modelo se verifica a través del análisis de los residuales y la comparación de criterios de información.

2.8. Ventajas y desventajas de los modelos tradicionales

Los modelos econométricos tradicionales ofrecen un sólido fundamento teórico y una gran interpretabilidad, lo que resulta ventajoso para la comunicación de resultados a audiencias tanto técnicas como no técnicas. Su formulación relativamente sencilla permite que cada coeficiente tenga un significado económico directo, y la consistencia con teorías económicas bien establecidas facilita la verificación de hipótesis específicas. Además, al requerir menos recursos computacionales, estos modelos pueden ser aplicados de manera eficiente incluso en contextos en los que la cantidad de datos es moderada. Por otro lado, la existencia de una amplia variedad de herramientas de diagnóstico para evaluar la validez de los supuestos del modelo (como pruebas de heterocedasticidad, autocorrelación y normalidad de los residuales) permite detectar problemas potenciales en la especificación.

Sin embargo, estas ventajas se contraponen a ciertas limitaciones. La dependencia de supuestos

estrictos —tales como la linealidad, la homocedasticidad y la ausencia de multicolinealidad— puede afectar la validez de los estimadores en situaciones donde estos supuestos no se cumplen. Asimismo, en escenarios de alta dimensionalidad, cuando el número de predictores es elevado en relación con el tamaño muestral, la aplicación del MCO puede volverse inestable o ineficaz. La incapacidad para capturar relaciones no lineales y la sensibilidad ante problemas de multicolinealidad son otros inconvenientes importantes. Finalmente, aunque estos modelos permiten realizar inferencias sobre las relaciones entre variables, su capacidad predictiva a menudo resulta inferior cuando se comparan con métodos modernos, especialmente en contextos en los que la dinámica de las series temporales varía a lo largo del tiempo.

La comparación de estos modelos tradicionales con métodos contemporáneos, como los modelos de machine learning, resulta especialmente relevante en estudios que buscan mejorar la precisión predictiva sin sacrificar la claridad y el rigor interpretativo.

Capítulo 3

Nuevas metodologías: Modelos de inteligencia artificial

3.1. Disrupción de los modelos de machine learning y su uso en problemas predictivos

En los últimos años, los avances en *machine learning* (ML) han transformado significativamente el panorama de los métodos predictivos en ciencias económicas y financieras. A diferencia de los modelos econométricos tradicionales, que se basan en supuestos estructurales explícitos acerca de los datos y su generación, los algoritmos de ML enfatizan la flexibilidad y la capacidad de aprender patrones directamente de los datos sin requerir fuertes hipótesis paramétricas. Esta característica, como señala Bishop (2006), ha permitido que los modelos de ML sobresalgan en contextos en los que las relaciones entre variables resultan altamente no lineales o en situaciones con una elevada dimensionalidad de los datos.

En el ámbito de las series temporales, estos modelos han demostrado un potencial disruptivo al abordar desafíos como la incorporación de variables exógenas, el manejo de datos heterogéneos y la detección de patrones complejos en series altamente volátiles, características comunes en variables

macroeconómicas y financieras. Por ejemplo, algoritmos basados en redes neuronales recurrentes (RNNs) y sus variantes han sido aplicados exitosamente para predecir series no estacionarias y con alta dependencia temporal Hochreiter and Schmidhuber (1997). Asimismo, según Goodfellow et al. (2016), los enfoques de ML han modificado la manera en que se evalúan y optimizan los modelos predictivos; en lugar de depender de soluciones analíticas cerradas y de criterios tradicionales como el Mínimo Cuadrado Ordinario, estos algoritmos emplean técnicas basadas en validación cruzada y en optimización por gradiente para minimizar errores en conjuntos de prueba independientes. Tal enfoque iterativo y empírico ha abierto nuevas oportunidades para la predicción del Producto Bruto Interno (PBI) y de la volatilidad en los mercados financieros.

A lo largo de este capítulo se exploran diversos modelos de ML aplicados a la predicción de series temporales, desde técnicas clásicas como la Suavización Exponencial hasta métodos más avanzados basados en aprendizaje profundo. En cada caso se analizan las ventajas y limitaciones de cada enfoque, haciendo énfasis en su aplicabilidad práctica y en su capacidad para superar algunas de las restricciones inherentes a los métodos econométricos tradicionales.

3.2. Suavización Exponencial

Basándose en los trabajos de Hyndman and Athanasopoulos (2018) y Holt (1957), el modelo de suavización exponencial es una técnica estadística diseñada para pronosticar series temporales, en especial aquellas en las que las observaciones más recientes poseen una mayor relevancia respecto a las pasadas. La idea central es asignar un peso mayor a las observaciones recientes, lo que permite que el pronóstico se adapte rápidamente a cambios en la dinámica de la serie. En este modelo, el parámetro de suavización, denotado por α y restringido al intervalo $[0, 1]$, determina la tasa de decaimiento de los pesos; valores cercanos a 1 implican que se otorga un peso predominante a la última observación, mientras que valores próximos a 0 implican una mayor influencia de datos anteriores.

El modelo de suavización exponencial simple se define de forma recursiva mediante la expresión

$$\hat{y}_{t+1} = \alpha\, y_t + (1 - \alpha)\, \hat{y}_t \tag{3.1}$$

donde $\hat{y}_{t+1}$ representa el pronóstico para el periodo $t + 1$, y_t es el valor observado en el periodo t, y

$\hat{y}_t$ es el pronóstico realizado en t. La estructura del modelo implica que el pronóstico futuro es un promedio ponderado entre el valor actual observado y el pronóstico previo.

Existen también extensiones para incorporar tendencias y estacionalidades. Cuando la serie presenta una tendencia lineal, se utiliza el modelo de suavización exponencial doble. En este caso, el pronóstico se obtiene sumando la estimación del nivel $\hat{y}_t$ y una componente de tendencia $\hat{b}_t$, según la relación

$$\hat{y}_{t+1} = \hat{y}_t + \hat{b}_t \tag{3.2}$$

donde la tendencia se actualiza recursivamente mediante

$$\hat{b}_t = \beta\,(\hat{y}_t - \hat{y}_{t-1}) + (1-\beta)\,\hat{b}_{t-1}, \tag{3.3}$$

con β siendo el parámetro de suavización para la tendencia. Para series que muestran tanto tendencia como patrones estacionales, el modelo de suavización exponencial triple, o modelo de Holt-Winters, descompone la serie en componentes de nivel, tendencia y estacionalidad. En este esquema, el nivel se actualiza ajustando la relación entre el valor observado y la estacionalidad del ciclo anterior, la tendencia se estima a partir de la diferencia entre niveles consecutivos, y la estacionalidad se actualiza comparando el valor observado con el nivel estimado. El pronóstico futuro se obtiene combinando estos componentes mediante la expresión

$$\hat{y}_{t+h} = (l_t + h\,b_t)\,s_{t-m+h}, \tag{3.4}$$

donde l_t es el nivel, b_t la tendencia, s_t la componente estacional, m es la longitud del ciclo estacional y h es el horizonte de pronóstico. La elección entre estos modelos se determina habitualmente mediante técnicas de validación cruzada, considerando la estabilidad y la complejidad de la serie en estudio.

3.3. Modelos de Vectores Soporte (SVM)

Los Modelos de Vectores de Soporte (SVM), descritos en detalle por Vapnik (1995) y Smola and Scholkopf (2004), constituyen un conjunto de métodos de aprendizaje supervisado aplicables tanto a problemas de clasificación como de regresión. En el caso de la regresión, denominada SVR (Support Vector Regression), el objetivo es encontrar una función $f(x)$ que se aproxime a los valores observados y de forma que el error no exceda una tolerancia predefinida ϵ. La función de predicción lineal se expresa de la forma

$$f(x) = w^T x + b, \tag{3.5}$$

donde $x \in \mathbb{R}^n$ es el vector de características, $w \in \mathbb{R}^n$ es el vector de pesos y b es el sesgo. El modelo se entrena minimizando la norma de w para obtener una solución lo más plana posible, siempre que la diferencia entre y y $f(x)$ se mantenga dentro de la tolerancia ϵ. Esta idea se implementa a través de la función de pérdida ϵ-insensible, definida de manera que errores menores o iguales a ϵ no son penalizados, mientras que aquellos que la exceden se castigan en proporción a su magnitud. Formalmente, la función de pérdida se define como

$$\mathcal{L}(y, f(x)) = \begin{cases} 0, & \text{si } |y - f(x)| \leq \epsilon, \\ |y - f(x)| - \epsilon, & \text{si } |y - f(x)| > \epsilon. \end{cases} \tag{3.6}$$

La formulación del problema de optimización consiste en minimizar el término de complejidad, medido por $\frac{1}{2}\|w\|^2$, junto con una penalización proporcional a la suma de las variables de holgura que permiten que algunos errores excedan la tolerancia, lo que se expresa de la siguiente manera:

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*), \tag{3.7}$$

sujeto a las restricciones

$$y_i - (w^T x_i + b) \leq \epsilon + \xi_i, \tag{3.8}$$

$$(w^T x_i + b) - y_i \leq \epsilon + \xi_i^*, \tag{3.9}$$

$$\xi_i, \xi_i^* \geq 0, \tag{3.10}$$

$$i = 1, \ldots, m. \tag{3.11}$$

Aquí, C es un parámetro que equilibra el compromiso entre la complejidad del modelo y la tolerancia al error. Cuando la relación entre las variables de entrada no es lineal, se aplica una transformación mediante una función de kernel $\phi(x)$ que permite trabajar en un espacio de características de mayor dimensión, donde la función de predicción se expresa como

$$f(x) = \sum_{i=1}^{m} \alpha_i \langle \phi(x_i), \phi(x) \rangle + b. \tag{3.12}$$

El producto interno $\langle \phi(x_i), \phi(x) \rangle$ se calcula mediante funciones de kernel, tales como el kernel lineal, polinómico o el RBF (Radial Basis Function). Este enfoque confiere a los modelos SVR la capacidad de capturar relaciones complejas y, a la vez, controlar el sobreajuste mediante el uso del parámetro de regularización C.

3.4. Redes Neuronales Recurrentes (RNN)

Las redes neuronales recurrentes (RNN) son una clase de modelos diseñados específicamente para procesar datos secuenciales, como las series temporales, aprovechando conexiones recurrentes que permiten mantener un estado interno y, de este modo, capturar dependencias temporales. A diferencia de las redes neuronales tradicionales en las que la información fluye de manera unidireccional desde la capa de entrada hasta la de salida, en una RNN la salida en cada instante t depende tanto de la entrada x_t como del estado oculto h_{t-1} proveniente del instante anterior. Esta relación se puede expresar de forma compacta mediante la ecuación

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b), \tag{3.13}$$

donde h_t representa el estado oculto en el tiempo t, W_h y W_x son matrices de pesos que conectan, respectivamente, el estado anterior y la entrada actual, b es el sesgo y σ es una función de activación. La capacidad de estas redes para aprender dependencias temporales se optimiza mediante el algoritmo de retropropagación a través del tiempo (BPTT) (véase, por ejemplo, Rumelhart et al. (1986)). No obstante, las RNN tradicionales pueden sufrir problemas de desvanecimiento del gradiente, lo que dificulta el aprendizaje de dependencias a largo plazo.

Para solventar esta limitación, se han desarrollado arquitecturas más sofisticadas como las Redes Neuronales de Memoria a Largo Plazo (LSTM) y las Gated Recurrent Units (GRU). En el modelo LSTM, se incorporan mecanismos de puertas que regulan el flujo de información a través de la celda de memoria, permitiendo que la red decida qué información conservar y cuál olvidar. De manera resumida, en una celda LSTM se definen las puertas de olvido f_t, de entrada i_t y de salida o_t, junto con la memoria candidata $\tilde{C}_t$, mediante las siguientes ecuaciones:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \tag{3.14}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \tag{3.15}$$

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C), \tag{3.16}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{3.17}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{3.18}$$

$$h_t = o_t * \tanh(C_t). \tag{3.19}$$

Estas ecuaciones permiten que la celda LSTM aprenda de forma más efectiva dependencias a largo plazo, controlando la información a través de mecanismos internos. Las GRU, por su parte, simplifican el modelo al fusionar algunas de las puertas, manteniendo resultados competitivos con menor complejidad computacional. Pese a los elevados costos computacionales que puede implicar el entrenamiento de estas redes, su capacidad para modelar dependencias temporales las convierte en herramientas poderosas en la predicción de series temporales, procesamiento de lenguaje natural, entre otros campos.

3.5. Redes Neuronales Profundas (DNN)

Las redes neuronales profundas (DNN) representan una extensión de las redes neuronales tradicionales, caracterizadas por disponer de múltiples capas ocultas que permiten modelar representaciones cada vez más abstractas y complejas de los datos. En estos modelos, la salida de cada capa se convierte en la entrada de la siguiente, lo que se formaliza con la expresión

$$a^{(l)} = \sigma(W^{(l)}\, a^{(l-1)} + b^{(l)}), \tag{3.20}$$

donde $a^{(l)}$ denota el vector de activaciones en la capa l, $W^{(l)}$ y $b^{(l)}$ son, respectivamente, la matriz de pesos y el vector de sesgos asociados a dicha capa, y σ es una función de activación no lineal, siendo la función ReLU una elección popular por su eficiencia y capacidad para mitigar el desvanecimiento del gradiente. La capa de salida transforma la última representación oculta en la predicción final mediante

$$\hat{y} = f_{\text{output}}(W_{\text{output}}\, a^{(L)} + b_{\text{output}}), \tag{3.21}$$

donde L es la última capa de la red y f_{output} es la función de activación apropiada para la tarea, como softmax en clasificación o una función lineal en regresión. El entrenamiento de estas redes se lleva a cabo utilizando la retropropagación del error y algoritmos de optimización basados en el descenso de gradiente, con la función de pérdida, por ejemplo, definida para regresión como el error cuadrático medio

$$\mathcal{L}(y, \hat{y}) = \frac{1}{2} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2, \tag{3.22}$$

y los parámetros se actualizan iterativamente mediante expresiones del tipo

$$W^{(l)} \leftarrow W^{(l)} - \eta\, \frac{\partial \mathcal{L}}{\partial W^{(l)}}, \tag{3.23}$$

donde η representa la tasa de aprendizaje. Aunque el entrenamiento de DNN puede verse afectado por el desvanecimiento del gradiente, el uso de funciones de activación apropiadas, técnicas de normalización y arquitecturas especiales ha permitido superar en gran medida este problema, habilitando aplicaciones exitosas en tareas de alta complejidad.

3.6. Splines

Los modelos Spline son herramientas de regresión no paramétrica ampliamente utilizadas para capturar relaciones no lineales en los datos. Su desarrollo se basa en trabajos como los de Green and Silverman (1993) y Friedman (1991), posteriormente enriquecido por aportes de autores más contemporáneos como Chen (2007). La contribución de Chen (2007), en particular, enriqueció el marco teórico existente al elaborar sobre aspectos cruciales como la selección óptima de nodos y la implementación de penalizaciones de suavidad, entre otros.

La idea central de este tipo de modelos es aproximar una función $g(x)$ mediante segmentos polinomiales que se unen en puntos denominados 'nodos' (o *knot points*) de forma que se garantice la continuidad de la función y, en el caso de splines suaves, la continuidad de sus derivadas hasta cierto orden. Sea $\{y_t\}_{t=1}^{T}$ una serie temporal y x_t una variable asociada al tiempo, se modela la relación mediante

$$y_t = g(x_t) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{3.24}$$

donde $g(x_t)$ es la función a estimar y ϵ_t es un error gaussiano. Entre las variantes de splines, los splines lineales ajustan funciones polinomiales de grado uno en cada intervalo definido por los nodos, utilizando expresiones de la forma

$$g(x_t) = \beta_0 + \beta_1 x_t + \sum_{j=1}^{m} \beta_{j+1} \left(x_t - k_j\right)_+, \tag{3.25}$$

donde $(x_t - k_j)_+ = \text{máx}(0, x_t - k_j)$ asegura que sólo se considere la diferencia cuando x_t supera el nodo k_j. Si bien esta formulación es sencilla, puede resultar insuficiente para capturar dinámicas complejas.

Por ello, los splines cúbicos, que utilizan polinomios de grado tres y garantizan la continuidad de las primeras y segundas derivadas, son comúnmente empleados. Su formulación es

$$g(x_t) = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3 + \sum_{j=1}^{m} \gamma_j \left(x_t - k_j\right)_+^3. \tag{3.26}$$

La continuidad en las derivadas evita cambios bruscos en la pendiente y la curvatura de la función estimada. Dado que los splines pueden sobreajustarse en presencia de ruido, se incorpora una penalización de suavidad, de modo que el ajuste se obtiene minimizando la función

$$\sum_{t=1}^{T} (y_t - g(x_t))^2 + \lambda \int \left[g''(x)\right]^2 dx, \tag{3.27}$$

donde el parámetro λ controla el compromiso entre el ajuste a los datos y la suavidad de la función estimada. En el análisis de series temporales, los splines resultan particularmente útiles para modelar tendencias no lineales y efectos estacionales, pudiendo incluso integrarse en modelos autoregresivos para capturar la dependencia temporal junto a una estructura flexible de tendencia.

3.7. Redes Temporales Convolucionales (TCN)

Las Redes Temporales Convolucionales (TCN) constituyen una extensión de las Redes Neuronales Convolucionales (CNN) adaptadas para trabajar con datos secuenciales o series temporales. En el contexto de la predicción de series temporales económicas y financieras, las TCN se han posicionado como una alternativa eficaz frente a modelos tradicionales como las Redes Neuronales Recurrentes (RNN) o las redes de Memoria a Largo Plazo (LSTM). La principal ventaja de las TCN reside en su capacidad para aplicar operaciones de convolución de forma paralela sobre toda la secuencia de entrada, lo cual permite capturar de manera eficiente relaciones temporales de largo alcance sin depender de cálculos secuenciales, como ocurre en las RNN.

El funcionamiento de una TCN se basa en la aplicación de filtros convolucionales sobre la serie temporal. Formalmente, si denotamos por x_t el valor de la serie en el instante t y por w_k los

coeficientes de un filtro de tamaño K, la operación de convolución causal se define mediante

$$y_t = \sum_{k=0}^{K-1} w_k \cdot x_{t-k}. \tag{3.28}$$

Esta ecuación garantiza que la salida en el tiempo t dependa únicamente de valores anteriores o iguales a t, respetando la causalidad, lo cual es esencial en aplicaciones económicas y financieras. Una característica crucial de las TCN es la incorporación de *convoluciones dilatadas*. La dilatación consiste en espaciar los elementos del filtro mediante un factor d (denominado factor de dilatación), de modo que el índice correspondiente a cada coeficiente se reescala a $r_k = k \cdot d$. La operación dilatada queda entonces expresada como

$$y_t = \sum_{k=0}^{K-1} w_k \cdot x_{t-k \cdot d}. \tag{3.29}$$

Este mecanismo permite aumentar de forma exponencial el campo receptivo de la red (es decir, la cantidad de pasos temporales que influyen en la salida) sin incrementar significativamente el número de parámetros, lo cual es especialmente útil para modelar dependencias a largo plazo en series donde la información histórica extendida resulta determinante.

Además de la capacidad para capturar relaciones a largo plazo, otro aspecto destacable es que, al emplear operaciones convolucionales, las TCN pueden beneficiarse de la paralelización en el cómputo. A diferencia de las RNN, donde cada paso depende del cálculo anterior, las operaciones de convolución se pueden realizar en paralelo sobre diferentes segmentos de la secuencia, reduciendo considerablemente el tiempo de entrenamiento y aprovechando eficientemente hardware especializado como GPUs o TPUs. Este enfoque también mitiga problemas numéricos comunes en las RNN, como el desvanecimiento o la explosión de gradientes, ya que la estructura convolucional permite un flujo de información más estable a través de la red.

En aplicaciones económicas y financieras, las TCN han sido empleadas para la predicción de la volatilidad de los mercados, la estimación de tasas de interés y la predicción de precios de activos, entre otras tareas. Por ejemplo, para la predicción del Producto Bruto Interno (PBI) o de indicadores

de inflación, es fundamental poder integrar información histórica de manera robusta, y la capacidad de las TCN para modelar dependencias a largo plazo resulta muy valiosa en este contexto.

No obstante, la implementación exitosa de una TCN requiere de un cuidadoso ajuste de hiperparámetros, como el tamaño de los filtros, el factor de dilatación y la profundidad de la red, ya que estos determinan el equilibrio entre la complejidad del modelo y su capacidad de generalización. Un diseño inadecuado puede derivar en sobreajuste o en una representación insuficiente de las dinámicas temporales presentes en los datos.

3.8. Ventajas y desventajas de modelos de machine learning

Los modelos de machine learning ofrecen, en el análisis y la predicción de series temporales, una serie de ventajas que los diferencian de los métodos tradicionales, aunque también presentan ciertos desafíos que deben ser abordados con cuidado. Una de las principales ventajas es la flexibilidad para modelar relaciones no lineales y complejas, lo que permite que algoritmos como las redes neuronales profundas, los SVM o las TCN capturen patrones que se escapan a los supuestos lineales y estructurados de los modelos econométricos. Esta capacidad resulta especialmente útil en series donde la dinámica subyacente varía de forma compleja, como en el caso de mercados financieros o indicadores económicos en períodos de alta volatilidad.

Otra fortaleza es la habilidad para trabajar con grandes volúmenes de datos y alta dimensionalidad. Los métodos de machine learning están diseñados para escalar y, en muchos casos, integran técnicas de reducción de dimensionalidad o aprendizaje jerárquico, lo cual mejora tanto la precisión predictiva como la eficiencia computacional en contextos de datos extensos. Asimismo, modelos avanzados como las RNN, TCN o las redes profundas han demostrado una alta capacidad para capturar dependencias temporales a largo plazo, lo que es crucial para la predicción en series con estructuras de dependencia complejas.

Sin embargo, estas ventajas se ven acompañadas de desafíos significativos. La complejidad computacional es un aspecto importante, ya que el entrenamiento de modelos de machine learning, en especial de redes neuronales profundas, puede requerir recursos considerables y largos tiempos de cómputo. Además, existe un riesgo inherente de sobreajuste, donde el modelo se adapta demasiado a los datos de entrenamiento y pierde capacidad de generalización; este problema debe ser mitigado

mediante técnicas de regularización, validación cruzada y la adecuada selección de hiperparámetros. Por otro lado, a pesar de los avances en interpretabilidad, muchos de estos métodos aún son percibidos como "cajas negras", lo que dificulta extraer conclusiones claras sobre las relaciones causales o estructurales presentes en los datos. Finalmente, la efectividad de estos modelos depende en gran medida de la calidad y la cantidad de los datos disponibles, lo cual puede limitar su aplicabilidad en escenarios donde la información es escasa o ruidosa.

En resumen, aunque los modelos de machine learning presentan un gran potencial para mejorar la precisión predictiva en series temporales, su implementación exitosa requiere un manejo cuidadoso de la complejidad computacional, la prevención del sobreajuste y la integración de estrategias que permitan interpretar los resultados de forma significativa. Estos aspectos deben ser considerados en conjunto para aprovechar al máximo las capacidades de estos métodos en aplicaciones económicas y financieras.

Capítulo 4

Lo mejor de dos mundos. Desarrollo de un algoritmo global entre técnicas macroeconométricas y rutinas de machine learning

4.1. Rutina de modelización de hipermodelo

Además de correr tanto el conjunto de modelos econométricos tradicionales (expuesto en el Capítulo 2), como el conjunto de modelos de machine learning (expuesto en el Capítulo 3), el presente trabajo busca desarrollar un hiper-modelo (i.e., un modelo compuesto por las predicciones de los demás conjuntos de modelos) con el fin último de ofrecer una alternativa superadora en términos de calidad predictiva.

Como primer paso para correr nuestros modelos de predicción de fluctuaciones de PBI en función de volatilidad de mercados bursátiles, necesitamos delimitar con mucho cuidado (i) cómo entrenamos

nuestros modelos (i.e., con qué dataset computamos los parámetros relevantes) y (ii) cómo evaluamos eficientemente la calidad predictiva de los mismos. Para ello, vamos a dividir nuestro dataset total en tres conjuntos:

- **Conjunto 1**: Este será el conjunto de entrenamiento (i.e., cómputo de parámetros) de los modelos econométricos tradicionales y de machine learning por separado. Aquí el objetivo es estimar el conjunto de parámetros $\{\hat{\beta}_{1,n}\}_{n=1}^{N}$ que surge de entrenar nuestros modelos con el conjunto de variables dependientes $\{y_{1,t}\}_{t=1}^{T}$ y el conjunto de variables explicativas $\{\mathbf{X}_{1,t}\}_{t=1}^{T}$.

- **Conjunto 2**: Este será el conjunto de test en los que los modelos entrenados en el Conjunto 1 serán puestos a prueba. Los modelos deberán tratar de predecir el conjunto $\{y_{2,t}\}_{t=1}^{T}$ utilizando los parámetros calculados en el Conjunto 1 ($\{\hat{\beta}_{1,n}\}_{n=1}^{N}$), y la matriz de variables explicativas del Conjunto 2 ($\{\mathbf{X}_{2,t}\}_{t=1}^{T}$). Denotaremos a estas predicciones como $\{\hat{y}_{2,t}\}_{t=1}^{T}|\{\mathbf{X}_{2,t}\}_{t=1}^{T}, \{\hat{\beta}_{1,n}\}_{n=1}^{N}$. Dado que el conjunto $\{y_{2,t}\}_{t=1}^{T}$ es conocido en el Conjunto 2, podemos calcular el RMSE de cada modelo individual (como se definió en el Capítulo **??**). Denotaremos a este conjunto de RMSE's como $\{\text{RMSE}_{2,n}\}_{n=1}^{N}$. Así podemos determinar un ranking de la calidad predictiva de los modelos de manera individual (i.e., se puede listar los RMSE de menor a mayor, refiriendo al menor RMSE a aquel modelo con mejor calidad predictiva).

  Asimismo, las predicciones del Conjunto 2 ($\{\hat{y}_{2,t}\}_{t=1}^{T}|\{\mathbf{X}_{2,t}\}_{t=1}^{T}, \{\hat{\beta}_{1,n}\}_{n=1}^{N}$) nos sirven para entrenar nuestro hiper-modelo. Esto es, calcularemos un nuevo conjunto de parámetros $\{\hat{\beta}_2\}$ que surja de entrenar un modelo que tome como input las predicciones del Conjunto 2 $\{\hat{y}_{2,t}\}_{t=1}^{T}|\{\mathbf{X}_{2,t}\}_{t=1}^{T}, \{\hat{\beta}_{1,n}\}_{n=1}^{N}$ para tratar de predecir el conjunto conocido $\{y_{2,t}\}_{t=1}^{T}$.

- **Conjunto 3**: Finalmente, este conjunto es aquel con el que se evaluará la calidad predictiva de nuestro hiper-modelo [1]. Aquí se tratará de predecir el conjunto $\{y_{3,t}\}_{t=1}^{T}$ utilizando los parámetros de nuestro hiper-modelo $\{\hat{\beta}_2\}$ y el conjunto de variables explicativas $\{\mathbf{X}_{3,t}\}_{t=1}^{T}$. Dado que este nuevo modelo sintetiza la calidad predictiva de todos los modelos entrenados en el Conjunto 1, es esperable que el RMSE (RMSE_3) que surge de comparar $\{\hat{y}_{3,t}\}_{t=1}^{T}|\{\mathbf{X}_{3,t}\}_{t=1}^{T}, \{\hat{\beta}_2\}$

[1] Se busca que este conjunto sea convexo y estrictamente ortogonal a los anteriores para así evitar un potencial sobreajuste sobre los parámetros del hipermodelo. Esto eso, se busca evaluar la performance de los parámetros del hipermodelo poniéndolos a prueba en la predicción de un conjunto que sea totalmente independiente (ortogonal) de cualquier conjunto involucrado en su entrenamiento.

con el conjunto conocido $\{y_{3,t}\}_{t=1}^{T}$, sea menor que cualquiera de los RMSE del conjunto $\{\text{RMSE}_{n,2}\}_{n=1}^{N}$.

donde $t = 1, 2, ..., T$ denota el espacio temporal que cubren nuestras variables, y $n = 1, 2, ..., N$ refiere a los distintos modelos estimados. Para asegurar la robustez del procedimiento, este esquema requiere que el número de observaciones en el Conjunto 1 sea mayor al del Conjunto 2, y que el número de observaciones del Conjunto 2 sea mayor al del Conjunto 3. Respetando la estructura temporal de los datos, definimos que el Conjunto 1 abarcará el 80 % del total de observaciones disponibles; el Conjunto 2 abarcará el 16 %; y el Conjunto 3 (que se usa exclusivamente para la validación del hiper-modelo) abarcará el restante 4 %. La Figura 4.1 ilustra la partición usada

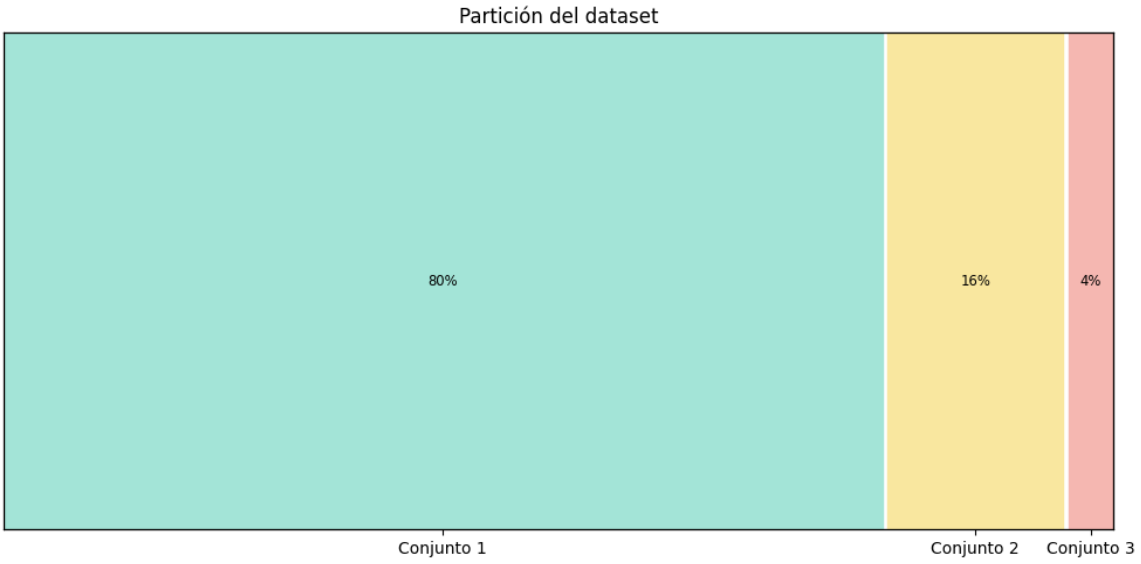


Figura 4.1: Partición del dataeset total para entrenamiento/testeo de modelos

Para evitar el problema de sobreajuste en los procesos de entrenamiento, para cada modelo vamos a incorporar regularizaciones del tipo de redes elásticas, como las descritas en la Sección 2.3 (tanto para los modelos econométricos tradicionales como para los modelos de machine learning)[2]. Para toda rutina de estimación individual se utilizó validación cruzada, para determinar los parámetros

[2]Se conserva el modelo de MCO, sin regularización de redes elásticas, para mantener un modelo de benchmark entre las estimaciones

óptimos de cada modelo. Finalmente, el modelo elegido para estimar el conjunto de parámetros $\hat{\beta}_2$ es el de redes neuronales recurrentes (RNN). Se optó por este modelo debido a la flexibilidad que tiene frente a potenciales no-linealidades en la predicción de series temporales[3], y por su floreciente uso en el campo de la literatura especializada en inteligencia artificial.

4.2. Optimización de parámetros e hiperparámetros

Con las especificaciones descritas en la sección anterior, procederemos a la evaluación de nuestros modelos. Comenzando con el entrenamiento de los modelos econométricos tradicionales, la Tabla 4.2 reporta los errores cuadráticos medios (RMSE)[4] de cada uno de los seis modelos desarrollados en el Capítulo 2. De allí se observa que el modelo con menor RMSE es la *Regresión Polinómica*, con un valor de 0.071, seguido por el modelo de *Promedio Móvil (MA)* y *Elastic Net*, con valores de 0.072 y 0.073, respectivamente. Estos resultados indican que los modelos no lineales, como la *Regresión Polinómica*, parecen capturar mejor las fluctuaciones en el PIB en comparación con los modelos tradicionales como *MCO* o *ARIMA*.

[3] Ver desarrollo de la Sección 3.4

[4] Aunque el RMSE es ampliamente utilizado como métrica para evaluar la precisión predictiva de modelos econométricos y de machine learning, es importante reconocer sus limitaciones desde el punto de vista inferencial. Como indican Escanciano and Parra (2024), la dificultad para establecer tests estadísticos de significancia sobre el RMSE impide, en muchos casos, determinar de forma concluyente qué modelo posee un desempeño predictivo superior. Esta limitación radica en la naturaleza del RMSE, que, al resumir la dispersión de los errores, puede ocultar diferencias estadísticas relevantes entre modelos, especialmente en presencia de heterocedasticidad o muestras de tamaño reducido.

En este trabajo se utiliza el RMSE como medida de comparación, dado que ofrece una aproximación directa y fácilmente interpretable del error de predicción. Sin embargo, reconocemos que dicha medida puede no capturar completamente las sutilezas en el desempeño predictivo de distintos modelos. Por ello, el desarrollo de metodologías alternativas o la aplicación de pruebas de inferencia más robustas en torno a esta métrica queda pendiente para futuras investigaciones.

Modelo	**RMSE**
Regresión Polinómica	0.071
Promedio Móvil (MA)	0.073
Elastic Net	0.073
ARIMA	0.074
Modelo Autorregresivo (AR)	0.074
MCO	0.076

Cuadro 4.1: RMSE de modelos econométricos tradicionales

Complementariamente, la Tabla 4.2 reporta los RMSE resultantes de los seis modelos de *machine learning* desarrollados en el Capítulo 3. Se destaca que el modelo con menor RMSE es la *Máquina de Soporte Vectorial (SVM)*, con un valor de 0.069, seguido de las *Redes Neuronales Recurrentes (RNN)* y los *Redes Temporales Convolucionales (TCN)*, con valores de 0.071 y 0.073, respectivamente. En comparación con los modelos econométricos tradicionales, los modelos de *machine learning* presentan menores valores de RMSE en promedio, lo que sugiere una mejor capacidad de ajuste y predicción frente a posibles no linealidades en las series temporales.

Modelo	**RMSE**
Máquina de Soporte Vectorial (SVM)	0.069
Redes Neuronales Recurrentes (RNN)	0.071
Redes Temporales Convolucionales (TCN)	0.073
Splines	0.073
Redes Neuronales Profundas	0.074
Suavizamiento Exponencial	0.076

Cuadro 4.2: RMSE de modelos de *machine learning*

Finalmente, tras entrenar nuestro hiper-modelo utilizando como predictor al conjunto de predicciones individuales (de la unión entre modelos econométricos tradicionales y de *machine learning*), llegamos al RMSE reportado en la Tabla 4.2. Este hiper-modelo logra un RMSE de 0.029, significativamente menor que cualquiera de los modelos evaluados previamente. Esto refuerza la utilidad de combinar las predicciones de múltiples enfoques en un modelo híbrido para mejorar la calidad predictiva.

Modelo	**RMSE**
Hipermodelo	0.029

Cuadro 4.3: RMSE del hipermodelo

En conclusión, se entrenaron un total de 13 modelos (seis econométricos tradicionales, seis de *machine learning* y un hiper-modelo). Entre los modelos individuales, la *Máquina de Soporte Vectorial (SVM)* y la *Regresión Polinómica* fueron los más efectivos dentro de sus respectivos grupos. Sin embargo, el hiper-modelo demostró una notable capacidad de síntesis predictiva al lograr reducir el error cuadrático medio en más del 50 % respecto al mejor modelo individual. Esto pone de manifiesto la relevancia de utilizar enfoques híbridos para la predicción de series temporales complejas, como las fluctuaciones del PIB.

Capítulo 5

Conclusiones

El presente trabajo se propuso explorar la relación entre la volatilidad de los mercados de capitales y las fluctuaciones del Producto Bruto Interno (PBI) real. Esta conexión es particularmente relevante, ya que la volatilidad bursátil refleja la incertidumbre de los agentes económicos, un factor clave que puede influir significativamente en las decisiones de inversión y, en última instancia, en el crecimiento económico. El objetivo principal fue desarrollar herramientas predictivas robustas que permitan anticipar crisis económicas mediante el análisis de la dinámica entre estos dos fenómenos.

La volatilidad se definió como una medida de dispersión en las series temporales de retornos financieros. Para cuantificarla, se utilizaron métricas clásicas como la varianza y el desvío estándar, además de enfoques más avanzados como la entropía de Shannon y los modelos GARCH de volatilidad estocástica. Dado el alto volumen de datos provenientes de 506 empresas del S&P 500, fue necesario implementar un análisis de componentes principales (PCA) para reducir la dimensionalidad del problema. Este enfoque permitió sintetizar la información en 35 componentes principales que explican al menos el 85 % de la varianza total del sistema, facilitando el manejo computacional y preservando la mayor parte de la información relevante.

En cuanto a los modelos econométricos tradicionales, se evaluaron metodologías como el MCO, los modelos autoregresivos (AR) y los modelos GARCH. Estos enfoques, respaldados por una sólida

base teórica, destacan por su simplicidad e interpretabilidad, lo que los hace ideales para realizar inferencias y contrastar hipótesis económicas. Sin embargo, sus limitaciones se hicieron evidentes en escenarios complejos, donde no logran capturar relaciones no lineales ni dinámicas altamente no estacionarias. Además, la dependencia de supuestos estrictos y su limitada capacidad predictiva en comparación con métodos modernos resaltan la necesidad de complementarlos con enfoques más flexibles.

Por otro lado, se desarrollaron y evaluaron modelos de *machine learning*, incluyendo redes neuronales recurrentes (RNN), redes temporales convolucionales (TCN), modelos de vectores soporte (SVM) y splines. Estos métodos demostraron una capacidad notable para modelar relaciones no lineales complejas y dependencias temporales a largo plazo, así como para manejar grandes volúmenes de datos y detectar patrones sutiles en series temporales. Su desempeño predictivo, medido mediante el RMSE, fue significativamente superior al de los modelos econométricos tradicionales. No obstante, también enfrentan desafíos importantes, como su alta complejidad computacional, el riesgo de sobreajuste y la falta de interpretabilidad en comparación con enfoques más simples.

Uno de los aportes más destacados de este trabajo fue la construcción de un hipermodelo que combina las predicciones de los modelos econométricos tradicionales y de *machine learning*. Este enfoque híbrido mostró un desempeño superior en términos de RMSE, lo que sugiere que ambos paradigmas pueden complementarse eficazmente. El hipermodelo logró capturar tanto las relaciones lineales como no lineales presentes en los datos, aprovechando las fortalezas de cada tipo de modelo.

En resumen, este trabajo contribuye al campo de la predicción macroeconómica al proponer un enfoque novedoso que integra técnicas tradicionales y modernas. Se demostró que la volatilidad bursátil contiene información valiosa para predecir fluctuaciones del PBI real, especialmente cuando se combina con herramientas avanzadas de reducción de dimensionalidad y aprendizaje automático. Si bien los modelos econométricos tradicionales siguen siendo relevantes por su simplicidad e interpretabilidad, su capacidad predictiva es limitada en entornos complejos. Por el contrario, los modelos de *machine learning* ofrecen un gran potencial predictivo, aunque requieren un manejo cuidadoso de sus limitaciones inherentes. La combinación de ambos enfoques en un hipermodelo representa una alternativa prometedora para mejorar la capacidad predictiva y abordar problemas económicos y financieros de manera más integral.

Finalmente, este enfoque puede extenderse a otros contextos, como la predicción de crisis económicas en mercados emergentes o la modelización de dinámicas sectoriales específicas. La implementación práctica de estas herramientas constituye un área fértil para futuras investigaciones y aplicaciones, con el potencial de transformar la forma en que se analiza la relación entre los mercados financieros y la economía real.

Bibliografía

Acharya, V. V. and Richardson, M. P. (2011). Restoring financial stability: How to repair a failed system. *Journal of Financial Stability*, 7(3):123–142.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Baron, M. and Xiong, W. (2016). Credit Expansion and Neglected Crash Risk. NBER Working Papers 22695, National Bureau of Economic Research, Inc.

Bencivenga, V. R. and Smith, B. D. (1991). Financial intermediation and endogenous growth. *Journal of Political Economy*, 99(5):607–631.

Bhowmik, R. and Wang, S. (2020). Stock market volatility and return analysis: A systematic literature review. *Entropy*, 22(5):522.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York, 1 edition. Capítulo 5 aborda el uso de redes neuronales en problemas predictivos.

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., and Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, 145:103773.

Bollerslev, T. (1986a). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1986b). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T., Diebold, F., Andersen, T., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, Inc., San Francisco, CA.

Brenner, Menachem, G. D. (1992). Hedging volatility in foreign currencies.

Brenner, M. and Galai, D. (1989). New financial instruments for hedge changes in volatility. *Financial Analysts Journal*, 45(4):61–65.

Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., and Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert systems with applications*, 112:353–371.

Chen, X. (2007). Chapter 76 large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 6:5549–5632.

Chicago Board Options Exchange (2009). Cboe volatility index methodology. Documento técnico. Disponible en: `http://www.cboe.com/products/vix-volatility-index-volatility-statistics/`.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. John Wiley & Sons, New York, 3rd edition.

Eichengreen, B. (1989). *Globalizing Capital: A History of the International Monetary System*. Norton, New York.

El-Aal, A. and Mohamed, F. (2024). Economic crisis treatment based on artificial intelligence. *Institutions & Economies*, 16(3).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4):987–1007.

Engle, R. F., Ghysels, E., and Sohn, B. (2013). Stock Market Volatility and Macroeconomic Fundamentals. *The Review of Economics and Statistics*, 95(3):776–797.

Escanciano, J. C. and Parra, R. (2024). Extending the scope of inference about predictive ability to machine learning methods. *arXiv preprint arXiv:2402.12838*.

Evans, M. and Rosenthal, J. (2004). *Probability and Statistics: The Science of Uncertainty*. W. H. Freeman.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Perthes et Besser.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA. Referencia exhaustiva sobre aprendizaje profundo, incluyendo redes recurrentes y su aplicación.

Gordon, M. J. and Shapiro, E. (1956). Capital equipment analysis: The required rate of profit. *Management Science*, 3(1):102–110. Reprinted in *Management of Corporate Capital*, Free Press of Glencoe, 1959.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall/CRC, London.

Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall, 5th edition.

Greenwood, J. and Jovanovic, B. (1990). Financial development, growth, and the distribution of income. *Journal of Political Economy*, 98(5):1076–1107.

Gulen, H. and Ion, M. (2016). Policy uncertainty and corporate investment. 29:523–564.

Gurley, J. G. and Shaw, E. S. (1955). Financial aspects of economic development. *The American Economic Review*, 45(4):515–538.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2nd edition.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780. Trabajo seminal que introduce las LSTM, ampliamente utilizadas en series temporales.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. In *Proceedings of the 1st International Symposium on Forecasting*, pages 97–110.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, 2nd edition.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Johnson, R. A., Wichern, D. W., et al. (2002). Applied multivariate statistical analysis.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, New York, NY, 2nd edition.

Karaca, Y., Baleanu, D., Zhang, Y.-D., Gervasi, O., and Moonis, M. (2022). *Multi-chaos, fractal and multi-fractional artificial intelligence of different complex systems*. Academic Press.

King, R. G. and Levine, R. (1993). Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics*, 108(3):717–737.

Lee, R. W. (2005). *Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation*, pages 241–268. Springer US, Boston, MA.

Legendre, A.-M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot.

Max Garzon, Ching-Chi Yang, D. V. N. K. K. J. L.-Y. D. (2022). *Dimensionality Reduction in Data Science*. Springer.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons, Hoboken, NJ, 5th edition.

Perez, N. H., Menendez, S. C., and Seco, L. (2003). A theoretical comparison between moments and l-moments. *Unpublished paper retrieved from http://risklab. erin. utoronto. ca/members. htm.*

Ramsey, J., Newton, H., and Harvill, J. (2002). *The Elements of Statistics: With Applications to Economics and the Social Sciences*. Duxbury/Thomson Learning.

Roll, R. (1988). The crash of 1987: A reassessment. *Journal of Finance*, 43(4):1063–1080.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). *Learning Representations by Back-Propagating Errors*, volume 323.

Schularick, M. and Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, 102(2):1029–61.

Shiller, R. J. (2000). Measuring bubble expectations and investor confidence. *Journal of Psychology and Financial Markets*, 1(1):49–60.

Smola, A. J. and Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

Stiglitz, J. E. (2010). Risk, uncertainty, and the global economy. *Economic Policy*, 25(63):149–168.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking*, 1(1):15–29.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer.

Wehrl, A. (1978). General properties of entropy. *Rev. Mod. Phys.*, 50:221–260.

Williams, J. (2014). *The Theory of Investment Value*. BN Publishing.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2nd edition.

Yu, J., Zhang, J., Shin, H. E., and Kong, J. (2019). Revisiting the economic crisis after a decade: Statistical and machine learning perspectives. *Annals of Dunarea de Jos University of Galati. Fascicle I. Economics and Applied Informatics*, 25(2):14–19.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.