



Pronóstico Probabilístico Basado en Predicción Conformal para Series Temporales: Comparación con Bootstrapping y DeepAR

Pedro José Leal Mesa

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá D.C., Colombia
2026

Pronóstico Probabilístico Basado en Predicción Conformal para Series Temporales: Comparación con Bootstrapping y DeepAR

Pedro José Leal Mesa

Tesis presentada como requisito parcial para optar al título de:

Magíster en Ciencias - Estadística

Director:
Ph.D. Mario E. Arrieta-Prieto

Línea de Investigación:
Análisis de Series de Tiempo y Predicción Conformal

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá D.C., Colombia
2026

Dedicatoria

A mi papá Antonio Leal, a mis compañeros y a mi director por su guía y apoyo constante en este proceso de formación.

“Essentially, all models are wrong, but some are useful.”

— George E. P. Box

Agradecimientos

Al Ph.D. Mario E. Arrieta-Prieto, director de esta tesis, por su invaluable orientación y paciencia. Quiero agradecerle especialmente por haberme incentivado a salir de mi zona de confort y presentar los resultados preliminares de este trabajo en el escenario internacional.

Un agradecimiento profundo y especial por la oportunidad de participar en el **45th International Symposium on Forecasting** en China. Esta experiencia marcó un hito en mi carrera profesional y fue posible gracias a un esfuerzo colectivo que jamás olvidaré: a la Universidad Nacional de Colombia por brindarme los fondos institucionales, y muy especialmente a mis compañeros, amigos, conocidos y familiares, quienes con un cariño inmenso organizaron y participaron en una rifa para completar el dinero necesario para este viaje. Su solidaridad fue el motor que me llevó al otro lado del mundo.

A mis compañeros de la Maestría en Estadística, con quienes compartí este camino académico, por las discusiones técnicas y el apoyo moral en los momentos de mayor reto.

A mi padre, Antonio Leal, por ser mi apoyo incondicional y por creer en este proyecto desde el primer día. Todo este esfuerzo es también suyo.

Con gratitud,
Pedro José Leal Mesa

Contents

Agradecimientos	vii
1 Introducción y Objetivos	1
1.1 Introducción	1
1.2 Planteamiento del problema	2
1.3 Justificación	3
1.4 Objetivos	4
1.4.1 Objetivo General	4
1.4.2 Objetivos Específicos	4
1.5 Estructura de la tesis	4

List of Figures

List of Tables

1 Introducción y Objetivos

En este primer capítulo se describe el contexto general de la investigación, enfocándose en la importancia de la cuantificación de la incertidumbre en el análisis de series temporales. Se presenta el planteamiento del problema, los retos actuales de la predicción probabilística y la justificación académica de este trabajo. Finalmente, se definen los objetivos general y específicos que guiarán el desarrollo de la tesis, se esbozan las hipótesis centrales y se presenta la estructura general del documento.

1.1 Introducción

La predicción es una tarea fundamental en diversas disciplinas, siendo de particular interés en el análisis de series de tiempo. Tradicionalmente, la literatura y la práctica se han centrado en estimaciones puntuales, como la media o el valor esperado de una variable futura. Sin embargo, este enfoque omite información crucial sobre la incertidumbre asociada a la predicción. Problemas clásicos, como la optimización de inventarios o la gestión de riesgos financieros, ilustran la necesidad de ir más allá de la media y considerar la distribución completa de los posibles resultados futuros para una toma de decisiones óptima.

La predicción probabilística aborda esta necesidad proporcionando no solo un valor central, sino una distribución de probabilidad o un conjunto de cuantiles sobre los valores futuros. En este contexto, surge la teoría de la *Predicción Conformal* (Conformal Prediction - CP) como un marco robusto para la construcción de intervalos y regiones de predicción con garantías de cobertura válidas bajo condiciones mínimas. A diferencia de otros métodos, la CP posee validez exacta y es aplicable a diversos modelos predictivos, lo que la convierte en una alternativa metodológica prometedora para el pronóstico probabilístico.

La relevancia comparativa de los métodos que se analizarán en este trabajo radica en sus características distintivas: mientras que la CP ofrece garantías teóricas de cobertura

bajo supuestos mínimos, los métodos de remuestreo como el *Bootstrapping* proporcionan flexibilidad no paramétrica para capturar la estructura de dependencia temporal, y los modelos de aprendizaje profundo como *DeepAR* (**Salinas2020**) destacan por su capacidad de modelar dependencias temporales complejas en múltiples series simultáneamente. Esta diversidad metodológica permite una evaluación integral de las fortalezas y limitaciones de cada enfoque.

No obstante, la aplicación de CP a series de tiempo presenta desafíos significativos debido a que los supuestos de intercambiabilidad de los datos suelen verse vulnerados por la dependencia temporal y la presencia de derivas distribucionales. Este trabajo busca explorar cómo adaptar estos métodos mediante mecanismos de ponderación temporal y contrastarlos frente a técnicas establecidas, contribuyendo así a cerrar una brecha importante en la literatura: la escasa aplicación y evaluación sistemática de CP con adaptaciones para no intercambiabilidad en escenarios de pronóstico de series temporales reales.

1.2 Planteamiento del problema

La predicción en series de tiempo enfrenta tres retos fundamentales que motivan esta investigación:

1. **Dependencia Temporal:** La naturaleza secuencial de los datos viola el supuesto de independencia fundamental en muchos métodos estadísticos estándar.
2. **No Estacionariedad:** Las características de los datos (tendencias, estacionalidad o cambios estructurales) cambian con el tiempo, lo que complica el modelado a largo plazo.
3. **Calibración de la Incertidumbre:** Generar pronósticos probabilísticos que sean simultáneamente precisos (calibrados) e informativos (eficientes) es una tarea compleja.

Los métodos actuales presentan limitaciones que este estudio busca contrastar. Por un lado, los modelos clásicos de tipo autorregresivo dependen fuertemente de supuestos sobre la estructura del proceso y la distribución del ruido (usualmente Gaussiano). Por otro lado, técnicas de remuestreo como el *Bootstrapping*, aunque flexibles, pueden tener dificultades cerca de límites de no estacionariedad. Finalmente, modelos modernos de aprendizaje profundo como *DeepAR* (**Salinas2020**) ofrecen un alto poder predictivo, pero a menudo

operan como “cajas negras” y carecen de las garantías teóricas formales de cobertura que ofrece el enfoque conformal.

El problema central de este trabajo se resume en la siguiente pregunta: **¿Cómo adaptar y evaluar un modelo de predicción probabilística basado en la teoría conformal para series de tiempo que considere la dependencia temporal, y cuál es su desempeño comparativo frente a métodos de referencia como Bootstrapping y DeepAR?**

1.3 Justificación

La cuantificación precisa de la incertidumbre es esencial para la toma de decisiones informada en áreas como la economía, meteorología y planificación de recursos. Adaptar la Predicción Conformal al dominio temporal representa un avance metodológico importante, ya que ofrece garantías de cobertura en muestra finita y una aplicabilidad general (libre de distribución).

A pesar de los desarrollos recientes en predicción conformal, existe un vacío notable en la literatura respecto a su aplicación sistemática en series temporales con características de no intercambiabilidad. La mayoría de los trabajos se han centrado en la construcción de intervalos de predicción, relegando el desarrollo de métodos para predicción distribucional completa. Además, la comparación rigurosa de CP adaptada mediante ponderación temporal frente a métodos establecidos (Bootstrapping) y estado del arte (DeepAR) utilizando métricas probabilísticas apropiadas como el *Continuous Ranked Probability Score* (CRPS) ha sido escasamente explorada.

Este trabajo contribuirá a cerrar estas brechas metodológicas al:

- Desarrollar una adaptación práctica de CP para series temporales que incorpore ponderación para manejar deriva distribucional.
- Proporcionar una evaluación comparativa sistemática en escenarios simulados y reales.
- Ofrecer evidencia empírica sobre las condiciones bajo las cuales CP puede ser competitiva o superior a métodos establecidos.

La comparación sistemática con métodos estándar (Bootstrapping) y estado del arte

(*DeepAR*) proporcionará una perspectiva clara sobre las fortalezas y debilidades del enfoque conformal en escenarios reales y simulados, facilitando la selección informada de métodos de pronóstico probabilístico en aplicaciones prácticas.

1.4 Objetivos

1.4.1 Objetivo General

Formular un modelo fundamentado en la teoría conformal para realizar predicciones probabilísticas en el contexto de series de tiempo, evaluando su desempeño mediante una comparación con modelos establecidos como *Bootstrapping* y *DeepAR*.

1.4.2 Objetivos Específicos

- Desarrollar un modelo de predicción probabilística que integre la teoría de predicción conformal con técnicas de modelado de series temporales, incorporando mecanismos de ponderación para abordar la no intercambiabilidad.
- Diseñar un entorno de simulación para evaluar el comportamiento del modelo propuesto bajo diversos escenarios de estructuras temporales y condiciones de ruido.
- Comparar el desempeño del modelo propuesto frente a modelos de referencia (*DeepAR* y *Bootstrapping*) utilizando métricas específicas de predicción probabilística como el *Continuous Ranked Probability Score* (CRPS) y análisis de calibración.
- Implementar la metodología en un caso de estudio con datos reales para cuantificar su relevancia práctica en aplicaciones de pronóstico.

1.5 Estructura de la tesis

El presente documento se organiza de la siguiente manera:

Capítulo 2: Marco Teórico. Se presenta una introducción al pronóstico probabilístico, se explican las principales métricas para evaluar el desempeño de los modelos (incluyendo

CRPS, calibración y nitidez), se desarrollan los fundamentos de los sistemas de predicción conformal, y se incluye una discusión sobre la extensión de la demostración de consistencia universal en escenarios estacionarios y ergódicos.

Capítulo 3: Metodología de Simulación. Se detalla el diseño del entorno de simulación, incluyendo la definición de los procesos generadores de datos, la formulación de los modelos predictivos (CP con ponderación, Bootstrapping y DeepAR), los cambios necesarios para su aplicación en este trabajo, y las extensiones de las simulaciones para diferentes escenarios.

Capítulo 4: Resultados de Simulación. Se presentan y analizan los resultados de las simulaciones diseñadas en el Capítulo 3, incluyendo comparaciones de ECRPS, análisis de cobertura, evaluación de calibración y pruebas de significancia estadística.

Capítulo 5: Aplicaciones a Datos Reales. Se implementan los métodos desarrollados en casos de estudio con datos reales, se presentan los resultados obtenidos y se discute su relevancia práctica.

Capítulo 6: Conclusiones. Se sintetizan los hallazgos principales del trabajo, se discuten las limitaciones del estudio, y se proponen direcciones para investigación futura.