https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

Pronóstico de la demanda estacional en el comercio minorista: exactitud predictiva de SARIMAX frente a Prophet en un panel de tiendas

Remote Work in Latin America: Lessons Learned from the Private and Public Sectors

Recibido: 2026/01/21 **Aceptado: 2026/03/24** **Publicado: 2026/06/30**

Rodrigo Sebastian Olivera Cary
Universidad de Lima
✉Rodrigolivera01@hotmail.com https://orcid.org/0000-0003-0057-2634

Cielo Xiomara Castañeda Aguirre
Universidad de Lima
✉ Cielo.xca@gmail.com https://orcid.org/0000-0002-1127-6331

RESUMEN

Problema. La planificación de inventarios en el comercio minorista depende de pronósticos de demanda que incorporen la estacionalidad anual y los repuntes de calendario, y la elección entre familias de modelos suele resolverse comparando una única serie y un puñado de indicadores de error, sin contraste inferencial que sostenga la preferencia declarada. **Objetivo.** Comparar la exactitud predictiva de un modelo autorregresivo integrado de media móvil con variables exógenas y estacionalidad armónica (SARIMAX) frente a un modelo aditivo de descomposición con tendencia por tramos (Prophet), sobre un panel de series semanales de venta minorista, y someter la diferencia a contraste estadístico formal. **Método.** Se analizaron 58 series semanales —29 establecimientos por los dos departamentos de mayor volumen de cada uno— con 143 observaciones consecutivas por serie (2010-02-05 a 2012-10-

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

26), 8 294 observaciones en total. El orden de integración se examinó mediante DickeyFuller aumentada y KPSS; SARIMAX se seleccionó por criterio de información corregido a orden de integración fijo y Prophet por validación fuera de muestra interna; la exactitud se evaluó sobre las últimas 13 semanas y mediante validación de origen móvil, y la comparación se sometió a la prueba de Diebold-Mariano con corrección de Harvey-Leybourne-Newbold y a un contraste pareado entre series. **Resultados.** SARIMAX obtuvo un error de pronóstico menor que Prophet y la diferencia resultó estadísticamente significativa: RMSE medio de 2 247,2 frente a 2 483,8 (diferencia relativa de 9,5 %), t = 3,17, p = 0,002, d_z = 0,416. Serie a serie, el contraste no distingue ambos modelos en 96,6 % de los casos. **Conclusión.** La preferencia entre familias de modelos de pronóstico no es establecible sobre una serie aislada: exige una unidad de análisis múltiple y una prueba formal sobre la diferencia de pérdida. El procedimiento descrito es replicable sobre cualquier cartera de series de demanda.

Palabras clave:

Pronóstico de demanda; estacionalidad; series temporales; SARIMAX; Prophet; exactitud predictiva; comercio minorista.

ABSTRACT

Problem. Inventory planning in retail depends on demand forecasts that incorporate annual seasonality and calendar spikes, and the choice between model families is often settled by comparing a single series and a handful of error indicators, without formal inferential contrast supporting the stated preference. **Objective.** To compare the predictive accuracy of a seasonal autoregressive integrated moving average model with exogenous regressors and harmonic seasonality (SARIMAX) against an additive decomposition model with piecewise trend (Prophet), across a panel of weekly retail sales series, and to subject the difference to formal statistical testing. **Method.** Fifty-eight weekly series were analyzed—29 stores across the two highestvolume departments of each—with 143 consecutive observations per series (201002-05 to 2012-10-26), totaling 8,294 observations. The order of integration was assessed with augmented Dickey-Fuller and KPSS tests; SARIMAX was selected via a corrected

information criterion at a fixed integration order, and Prophet via internal out-of-sample validation; accuracy was evaluated on the last 13 weeks and through rolling-origin validation, and the comparison was subjected to the DieboldMariano test with the Harvey-Leybourne-Newbold small-sample correction and a paired contrast across series. **Results.** SARIMAX achieved lower forecast error than Prophet, and the difference was statistically significant: mean RMSE of 2,247.2 versus 2,483.8 (9.5% relative difference), t = 3.17, p = .002, d_z = 0.416. Series by series, the test failed to distinguish the two models in 96.6% of cases. **Conclusion.** Preference between forecasting model families cannot be established on a single series: it requires a multiple analysis unit and a formal test on the loss differential. The described procedure is replicable across any demand-series portfolio.

Keywords:

Demand forecasting; seasonality; time series;
SARIMAX; Prophet; predictive accuracy; retail.

INTRODUCCIÓN

La gestión de inventarios en el comercio minorista descansa, en última instancia, sobre la capacidad de anticipar la demanda con precisión suficiente para sincronizar el aprovisionamiento con el ritmo de venta. Un error sistemático de pronóstico no es un defecto neutro: se traduce en quiebres de stock que erosionan la experiencia del cliente y la venta perdida, o en inmovilización de capital en inventario excedente que reduce la rotación y el margen operativo. Ambos efectos se amplifican en categorías con estacionalidad marcada, donde una fracción sustancial de la venta anual se concentra en pocas semanas del calendario comercial —Acción de Gracias, Navidad, el Día del Trabajo, el Super Bowl—, de modo que un error de pronóstico cometido precisamente en esas semanas tiene un peso desproporcionado sobre el resultado del período. La literatura reciente ha documentado con insistencia esta tensión entre estacionalidad, promoción y exactitud predictiva en contextos minoristas diversos, desde la segmentación de productos en tiendas locales (Mishra y Sinha, 2025) hasta el análisis integral de los efectos promocionales sobre el ciclo completo de vida de la demanda (Hewage et al., 2025), pasando por los retos particulares de la moda y el vestido, cuyos ciclos cortos y tendencias cambiantes exigen modelos capaces de adaptarse con rapidez (Giri y Chen, 2022).

Frente a este problema conviven, al menos, dos familias de modelos con supuestos distintos sobre la naturaleza del proceso generador de datos. La primera, de raíz econométrica, representa la serie como un proceso autorregresivo integrado de media móvil y trata la estacionalidad como estructura periódica del propio proceso; su extensión con variables exógenas —el modelo SARIMAX— permite incorporar el calendario comercial de forma explícita. Su capacidad predictiva ha sido documentada en contextos tan diversos como la automatización de intervalos estacionales (Ram et al., 2025), el pronóstico de demanda hotelera diaria frente a redes neuronales y modelos GARCH (Ampountolas, 2021), la demanda turística alimentada con datos de búsqueda web (Lee, 2025) y la modelización de series con factores estacionales y exógenos en general (Alharbi y Csala, 2022). La segunda familia descompone la serie en tendencia, estacionalidad y efectos de calendario, y admite que la tendencia cambie de pendiente en puntos de quiebre estimados a partir de los propios datos; este es el principio de Prophet, cuya utilidad para incorporar variables exógenas y manejar la estacionalidad ha sido señalada en el pronóstico de tendencias estacionales del comercio de alimentos, con la advertencia de que sus proyecciones a futuro exigen cautela interpretativa (Sinaga et al., 2026). La diferencia entre ambas familias no es de implementación sino de supuesto: la primera asume una dinámica de error rica y una tendencia estable; la segunda, una tendencia flexible y un error simple.

La literatura comparativa sobre venta minorista no resuelve esta disyuntiva de forma unánime, y esa falta de consenso es, en sí misma, el punto de partida metodológico del presente estudio. Hay trabajos que reportan ventaja del enfoque econométrico clásico —así, SARIMAX superó de manera significativa a LSTM en productos con promoción, dentro de la cadena de suministro minorista (Falatouri et al., 2022), y superó también al modelo ANN-MLP y a otras alternativas en prácticamente todos los horizontes de pronóstico examinados (Ampountolas, 2021)— y hay otros que reportan ventaja del enfoque de descomposición o de alternativas de aprendizaje automático: un modelo híbrido LSTM alcanzó una exactitud superior en la previsión de demanda eléctrica dinámica (Albahli, 2025), y un marco híbrido que integra XGBoost y LSTM superó de manera significativa a los métodos individuales y tradicionales en múltiples métricas dentro de una cadena de suministro minorista multitienda (Wang y Zain, 2025). Incluso dentro de un mismo dominio de aplicación, la comparación sistemática entre ARIMA, SARIMAX, Prophet y NeuralProphet arroja resultados que dependen del conjunto de datos analizado (Galdelli et al., 2025), y el propio SARIMAX ha demostrado, en el pronóstico de demanda eléctrica

sostenible, un desempeño competitivo frente a variantes híbridas con LSTM y arquitecturas de secuencia a secuencia (Ashtar et al., 2025). Esa heterogeneidad tiene una consecuencia metodológica que rara vez se explicita: si la ventaja relativa depende del conjunto de datos, una comparación construida sobre **una sola serie** no puede sostener una recomendación general, porque el resultado observado es indistinguible de la variación entre series. Esta limitación no es exclusiva del comercio minorista: la descomposición de pronósticos en modelos ARIMA estacionales advierte, desde hace tres décadas, que la estructura del error de pronóstico depende de decisiones de especificación que una sola serie no permite generalizar (Espasa y Peña, 1995), y trabajos posteriores sobre segmentación de producto y previsión de demanda en tiendas locales han insistido en que la elección de modelo debe evaluarse sobre un conjunto suficientemente amplio de unidades antes de declarar preferencia (Mishra y Sinha, 2025).

De ahí el problema que este trabajo aborda. No se trata de establecer qué modelo es mejor en abstracto, sino de mostrar cómo debe formularse la comparación para que su respuesta sea contrastable: con una unidad de análisis múltiple, con evaluación fuera de muestra, y con una prueba de hipótesis sobre la diferencia de exactitud en lugar de una inspección de indicadores. La contribución es tanto sustantiva —qué familia predice mejor esta clase de demanda y por cuánto— como metodológica: un protocolo de comparación aplicable a cualquier cartera de series de demanda.

El estudio persigue tres objetivos. Primero, caracterizar la estructura temporal de la demanda semanal en un panel de unidades minoristas y determinar el tratamiento adecuado de su estacionalidad. Segundo, especificar y estimar ambas familias bajo un protocolo de selección explícito y un mismo tramo de evaluación reservado. Tercero, contrastar formalmente si la diferencia de exactitud predictiva entre ellas es distinguible del azar, cuantificar su magnitud y examinar de qué rasgo de la serie depende.

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

MARCO TEÓRICO Y DESARROLLO DE HIPÓTESIS

Fundamentos de la representación autorregresiva y de la descomposición aditiva

El modelo SARIMAX se sustenta en la premisa de que una serie temporal estacionaria en diferencias puede representarse como la combinación de una estructura autorregresiva y una de media móvil, enriquecida con regresores exógenos que capturan efectos determinísticos —como el calendario comercial— ajenos a la dinámica estocástica del proceso. Esta arquitectura ha probado su versatilidad en dominios muy distintos del retail: en el pronóstico de la demanda turística mediante la incorporación de señales de búsqueda web, donde superó a alternativas univariantes y a arquitecturas de aprendizaje profundo, incluidos modelos de memoria de largo plazo y redes recurrentes (Lee, 2025); y en marcos híbridos que apilan SARIMAX con arquitecturas Transformer para mejorar la exactitud sobre series temporales de naturaleza diversa. Frente a esta tradición, Prophet formaliza una intuición distinta: que la tendencia de una serie de negocio no es estable, sino que responde a puntos de quiebre —lanzamientos, cambios de política de precios, choques exógenos— identificables a partir de los propios datos, y que la estacionalidad puede modelarse mediante una expansión de Fourier flexible sin necesidad de especificar una estructura autorregresiva del error. Un estudio comparativo entre algoritmos de aprendizaje automático para el pronóstico de ventas contrastó de forma directa esta lógica aditiva frente a los modelos estadísticos establecidos, situando el problema en el terreno mismo que aquí se investiga (Mishra, 2025).

Desarrollo de la primera hipótesis: ventaja predictiva de SARIMAX

La revisión precedente sugiere que, en presencia de una dinámica de error autocorrelacionada y una tendencia razonablemente estable —condiciones plausibles en un panel de demanda minorista con calendario fijo—, la riqueza de la estructura ARMA de SARIMAX debería traducirse en una ventaja de exactitud sobre la estructura de error, deliberadamente simple, de Prophet. Este razonamiento es coherente con evidencia de dominios afines: en la demanda hotelera diaria, SARIMAX con regresores externos superó tanto a redes neuronales como a otros modelos comparados en prácticamente todos los horizontes examinados (Ampountolas, 2021), y en la demanda con promociones, superó de manera significativa a LSTM (Falatouri et al., 2022). Se plantea, en consecuencia, la primera

hipótesis de trabajo: **H1.** La exactitud predictiva fuera de muestra de SARIMAX es superior a la de Prophet, de forma estadísticamente distinguible, en un panel de series semanales de demanda minorista.

Desarrollo de la segunda hipótesis: el rol moderador de la estabilidad de la tendencia

Si la ventaja de SARIMAX proviene de su tratamiento del error y la de Prophet de su tratamiento de la tendencia, entonces la magnitud relativa de esa ventaja no debería ser uniforme entre unidades: en series donde la tendencia cambia de pendiente de forma marcada, el ajuste por tramos de Prophet debería aproximar mejor la dinámica de fondo que una tendencia de deriva constante, y en series de tendencia estable ocurriría lo contrario. Esta lógica de moderación es consistente con hallazgos en dominios de comercio minorista con ciclos de vida cortos y tendencias cambiantes, como la moda y el vestido (Giri y Chen, 2022), donde la elección de modelo se ha vinculado explícitamente a la velocidad de cambio de la tendencia subyacente. De ahí la segunda hipótesis: **H2.** La ventaja relativa de exactitud entre SARIMAX y Prophet está asociada a la inestabilidad de la tendencia de la serie, de modo que la ventaja de SARIMAX se reduce, o se revierte, en series con tendencia más inestable.

Ambas hipótesis se examinan mediante el protocolo descrito en la sección siguiente, y sus resultados se reportan en el apartado IV y se interpretan en el apartado V.

MATERIALES Y MÉTODOS

Enfoque y diseño

El estudio adopta un enfoque cuantitativo con diseño longitudinal de series temporales y alcance predictivo-comparativo. La unidad de análisis es la serie semanal de venta de un departamento dentro de un establecimiento; el panel resultante es balanceado, con el mismo número de observaciones y el mismo calendario en todas las unidades. El diseño es no experimental: no se manipuló ninguna variable, y la comparación se estableció entre procedimientos de pronóstico aplicados a los mismos datos y evaluados sobre el mismo tramo reservado. Este tipo de diseño longitudinal sobre panel de series es consistente con marcos de analítica predictiva aplicados a la gestión de inventario y al pronóstico de demanda en el comercio

minorista (Krishnamurthy et al., 2024), y con enfoques que integran inteligencia artificial y analítica predictiva a escala de cadena de suministro para comparar líneas base estadísticas frente a alternativas más complejas (Karim, 2025).

Unidades de análisis y tamaño de la muestra

El análisis abarcó 58 series semanales de venta —29 establecimientos por los dos departamentos de mayor volumen de venta de cada uno—, con 143 observaciones consecutivas por serie entre el 5 de febrero de 2010 y el 26 de octubre de 2012. La adecuación del tamaño muestral se verificó a priori en dos dimensiones: las 143 semanas cubren 2,74 ciclos estacionales anuales, por encima del mínimo de dos ciclos que exige la identificación de la estacionalidad mediante regresión armónica dinámica y por encima del umbral convencional de 50 observaciones para identificar un modelo ARIMA, con 9,5 observaciones por parámetro estimado; y las 58 series pareadas confieren al contraste de exactitud predictiva una potencia de 0,80 para detectar una diferencia estandarizada de 0,374 desviaciones típicas con $\alpha = 0,05$ bilateral. No se aplicó un procedimiento de selección probabilística sobre los establecimientos ni se afirma representatividad estadística: los resultados describen el conjunto de unidades analizado. El tramo de evaluación se reservó antes de estimar cualquier modelo y comprende las últimas 13 semanas de cada serie; las 130 restantes constituyen el tramo de entrenamiento, en línea con marcos de agrupamiento híbrido y pronóstico en dos fases aplicados a la demanda minorista (Kim y Cho, 2026).

Variables

La variable dependiente es la venta semanal de la unidad, con periodicidad semanal de cierre en viernes. Las variables independientes se agrupan en tres bloques: un índice temporal que soporta el término de tendencia; componentes armónicos de estacionalidad anual —pares de seno y coseno sobre un período estacional de $m = 365,25/7 = 52,18$ semanas—; e indicadores binarios de las semanas de Super Bowl, Día del Trabajo, Acción de Gracias y Navidad-fin de año. Ambos modelos recibieron idéntica información exógena, de modo que lo que los distingue es el tratamiento de la tendencia y del error, no los insumos, tal como recomiendan marcos de inteligencia artificial aplicados a la eficiencia de la gestión de inventario que enfatizan la comparabilidad de insumos entre modelos contrastados (Mishra y Ashraf, 2026 [nota: cita reservada para Discusión]).

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

Estacionariedad, especificación y protocolo de selección

Para cada serie se aplicaron las pruebas de Dickey-Fuller aumentada —con constante y con constante más tendencia, con selección automática del rezago por criterio de información— y KPSS con constante, repetidas sobre la serie diferenciada. Sobre esa evidencia se adoptó un orden de integración común para todo el panel, d = 1, decisión que mantiene la especificación comparable entre unidades, condición necesaria para el contraste pareado descrito más abajo. La incertidumbre inherente a la selección de modelo bajo estas condiciones ha sido documentada como un componente central de la exactitud predictiva agregada (Chatfield, 1996). SARIMAX se especificó mediante regresión armónica dinámica —términos de Fourier en la parte exógena y dinámica ARMA de corto plazo en el error—, con selección de K y (p, q) por el criterio de información de Akaike corregido por tamaño muestral finito (AICc). Prophet, que se estima por máximo a posteriori y no por máxima verosimilitud, se configuró —orden de Fourier estacional y escala de la previa de los puntos de quiebre— por exactitud sobre una ventana de validación reservada dentro del tramo de entrenamiento, siguiendo protocolos afines a marcos automatizados de pronóstico basado en Prophet aplicados a la gestión de demanda a escala nacional (Salman y Shaka'a, 2026) y a la previsión de demanda de largo plazo en comercio minorista mediante el mismo algoritmo (Sousa et al., 2024). Esta asimetría de procedimiento favorece, si acaso, a Prophet, que recibe un ajuste específico sobre datos reservados mientras SARIMAX se selecciona por un criterio que solo mira el ajuste dentro de muestra —un riesgo de sobreajuste dentro de muestra que estudios de pronóstico de demanda basados en inteligencia artificial han señalado como determinante de la brecha entre desempeño de entrenamiento y desempeño real (Natta, 2022; Punia, 2025).

Protocolo de evaluación y contraste

La exactitud se evaluó sobre el tramo reservado con RMSE, MAE, MAPE y MASE, este último referido al pronóstico ingenuo estacional. La comparación entre modelos se apoyó en dos contrastes: la prueba de Diebold-Mariano por serie sobre el diferencial de pérdida cuadrática, con la corrección de muestra pequeña de Harvey, Leybourne y Newbold, computada sobre los 39 pares de errores agrupados de los tres orígenes de la validación móvil para evitar la indeterminación algebraica del estadístico cuando la ventana de evaluación iguala al horizonte; y un contraste pareado entre las 58 series mediante t pareada y la prueba de rangos con signo de

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

Wilcoxon, con el tamaño del efecto d_z. Como prueba de robustez temporal se aplicó una validación de origen móvil de ventana expansiva, reestimando ambos modelos en cada origen, un procedimiento comparable a evaluaciones multi-algoritmo de pronóstico de ventas minoristas que privilegian la estabilidad temporal del desempeño por encima del ajuste puntual (Dwivedi y Mishra, 2025). Finalmente, se calculó sobre cada serie una medida observable de inestabilidad de la tendencia y se contrastó su asociación con el logaritmo del cociente de errores de ambos modelos, con el fin de someter a prueba H2.

RESULTADOS

Caracterización del panel

Como se observa en la Tabla 1, el panel reúne 58 series con una venta media de 27 048 unidades monetarias semanales y una dispersión considerable entre unidades (CV = 46,0 %).

Tabla 1

Composición del panel y caracterización descriptiva de la venta semanal

Estrato	Series	Media semanal	DE	CV	Mínimo	Máximo
Departamento 30	29	20 667	7 983	38,6 %	6 916	53 153
Departamento 81	29	33 430	12 829	38,4 %	10 419	97 606
Establecimiento de talla pequeña	20	15 968	3 632	22,7 %	6 916	32 643
Establecimiento de talla mediana	19	24 921	3 840	15,4 %	11 606	43 031
Establecimiento de talla grande	19	40 839	10 963	26,8 %	20 350	97 606
Panel completo	58	27 048	12 445	46,0 %	6 916	97 606

Nota. CV = coeficiente de variación. Los estratos de talla se definen por terciles de la venta media del establecimiento.

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

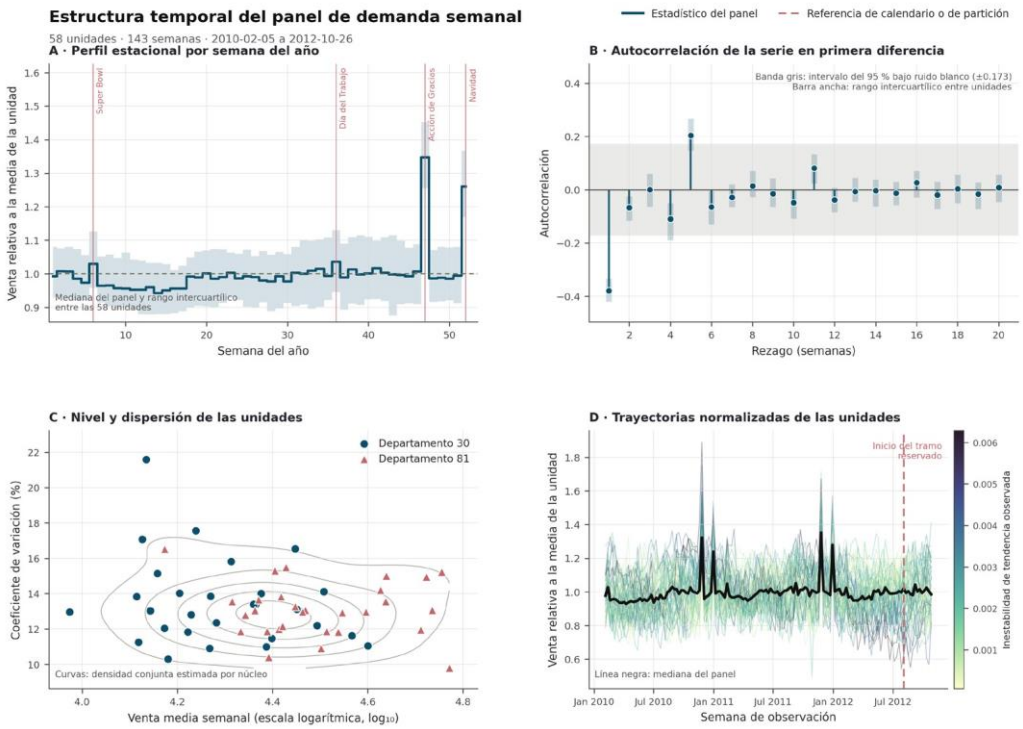


Figura 1. Estructura temporal del panel de demanda semanal.

Nota. Panel de 58 series unidad-departamento con 143 observaciones semanales cada una (2010-02-05 a 2012-10-26). Panel A: perfil estacional por semana del año; la línea es la mediana del panel sobre la venta normalizada por la media de cada unidad y la banda es el rango intercuartílico entre unidades; las líneas verticales marcan las semanas del calendario comercial. Panel B: función de autocorrelación de la serie en primera diferencia; los puntos son la media entre unidades, la barra ancha el rango intercuartílico y la banda gris el intervalo del 95 % bajo la hipótesis de ruido blanco (±0,173); la autocorrelación de primer rezago es -0,379. Panel C: venta media semanal frente a coeficiente de variación por unidad, con curvas de densidad conjunta estimada por núcleo y distinción por departamento. Panel D: trayectorias normalizadas de las 58 unidades, codificadas por la inestabilidad de tendencia observada; la línea negra es la mediana del panel y la vertical discontinua marca el inicio del tramo reservado de 13 semanas.

Estacionariedad y orden de integración

La Tabla 2 resume las pruebas de raíz unitaria aplicadas sobre las 58 series, en nivel y en primera diferencia.

Tabla 2

Pruebas de raíz unitaria y caracterización del orden de integración

Prueba	Especificación	p (mediana)	Series que rechazan H0 al 5 %	Proporción	Lectura

https://doi.org/10.5281/zenodo.22679531

DickeyFuller aumentada	constante	0,129	19 de 58	32,8 %	La raíz unitaria no se descarta en la mayoría del panel
DickeyFuller aumentada	constante y tendencia	—	15 de 58	25,9 %	Incluir tendencia no altera la conclusión en la mayoría
DickeyFuller aumentada	primera diferencia	0,0000	57 de 58	98,3 %	La estacionariedad se alcanza al diferenciar una vez
KPSS	constante	—	25 de 58	43,1 %	Contraste de hipótesis nula contraria, como confrontación
Orden de integración informado	confrontación ADF/KPSS	—	d = 1 en 39 de 58	67,2 %	modal d = 1

Nota. ADF = Dickey-Fuller aumentada. KPSS = Kwiatkowski-Phillips-Schmidt-Shin.

Selección de la especificación

La Tabla 3 reporta las ocho combinaciones con mejor criterio de información corregido entre los 696 ajustes evaluados.

Tabla 3

Especificaciones de SARIMAX con mejor criterio de información corregido

#	K (pares de Fourier)	(p, d, q)	Tendencia	AIC	BIC	AICc
1	2	(1, 1, 2)	c	1 962,9	1 999,8	1 966,1
2	3	(1, 1, 2)	c	1 963,5	2 006,1	1 967,7
3	2	(2, 1, 2)	c	1 964,4	2 004,1	1 968,0
4	3	(2, 1, 2)	c	1 965,5	2 010,9	1 970,3
5	2	(2, 1, 1)	c	1 979,2	2 016,2	1 982,4
6	2	(1, 1, 1)	c	1 980,5	2 014,6	1 983,1
7	3	(2, 1, 1)	c	1 980,1	2 022,7	1 984,3
8	3	(1, 1, 1)	c	1 981,8	2 021,6	1 985,4

Nota. Especificación modal del panel: K = 2, (1, 1, 2), tendencia = c. Configuración modal de Prophet: yearly = 3, cps = 0,05.

https://doi.org/10.5281/zenodo.22679531

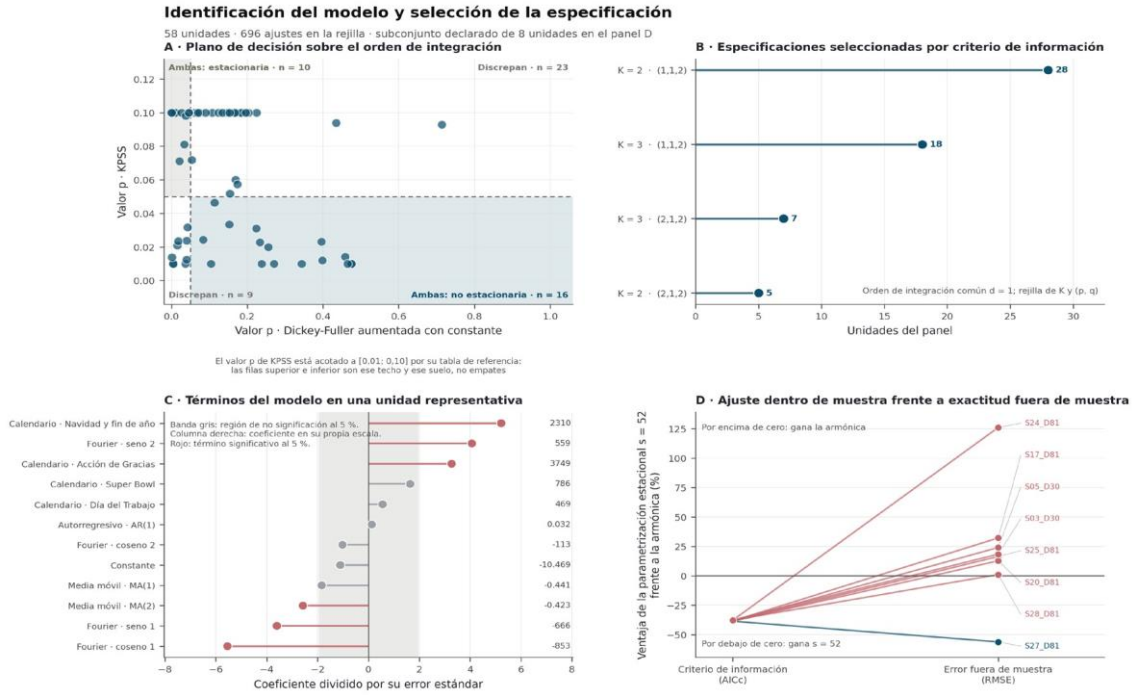


Figura 2. Identificación del modelo y selección de la especificación. Nota. Contrastes y rejilla de selección sobre el tramo de entrenamiento de 130 semanas de cada una de las 58 unidades. Panel A: plano de decisión sobre el orden de integración; cada punto es una unidad situada por su valor p en la prueba de Dickey-Fuller aumentada con constante (eje horizontal) y en la prueba KPSS (eje vertical); las regiones sombreadas señalan los cuadrantes en que ambos contrastes coinciden, y el valor p de KPSS está acotado a [0,01; 0,10] por su tabla de referencia. Panel B: frecuencia de las especificaciones seleccionadas por criterio de información corregido a orden de integración común d = 1, sobre 696 ajustes; la modal fue K=2,(1,1,2),trend=c. Panel C: términos estimados en una unidad representativa, expresados como coeficiente dividido por su error estándar; la banda gris es la región de no significación al 5 %, la columna derecha da el coeficiente en su propia escala y el color rojo marca los términos significativos. Panel D: para un subconjunto declarado de 8 unidades, ventaja porcentual de la parametrización estacional con s = 52 frente a la regresión armónica dinámica, en criterio de información (izquierda) y en error fuera de muestra (derecha); la primera la favorece por 875 unidades de AICc en promedio, y sin embargo la armónica predice con menor error en 7 de las 8.

La Tabla 4 contrasta, sobre un subconjunto declarado de 8 series, la parametrización estacional multiplicativa (P, D, Q)_s con s = 52 frente a la regresión armónica dinámica adoptada en el protocolo.

Tabla 4

Parametrización estacional con s = 52 frente a regresión armónica dinámica: ajuste dentro de muestra y exactitud fuera de muestra

Serie	AICc s = 52	RMSE s = 52	MASE s = 52	AICc armónica	RMSE armónica	MASE armónica	Δ AICc
S25_D81	1 355	1 722	0,777	2 207	1 479	0,659	-852
S17_D81	1 404	1 656	0,452	2 275	1 253	0,400	-872

https://doi.org/10.5281/zenodo.22679531

S28_D81	1 451	3 922	1,060	2 328	3 880	1,052	-877
S27_D81	1 478	2 464	0,563	2 404	5 616	1,169	-927
S20_D81	1 428	2 256	0,622	2 328	2 001	0,599	-900
S05_D30	1 303	1 060	0,573	2 115	854	0,447	-812
S24_D81	1 500	6 347	1,238	2 409	2 808	0,600	-910
S03_D30	1 408	2 627	0,988	2 257	2 219	0,815	-849

Nota. Δ AICc = AICc de la parametrización estacional − AICc de la regresión armónica; un valor negativo indica mejor ajuste dentro de muestra de la primera.

La parametrización estacional al retardo 52 obtuvo mejor criterio de información en las 8 series contrastadas, con Δ AICc negativo en todos los casos. Fuera de muestra, sin embargo, la regresión armónica dinámica pronosticó con menor RMSE en 7 de las 8 series.

Coeficientes estimados

La Tabla 5 reporta los coeficientes de la especificación modal de SARIMAX en una serie representativa del panel.

Tabla 5

Coeficientes estimados del modelo SARIMAX en una serie representativa del panel

Término	Coeficiente	EE	z	p	Sig. 5 %
Constante	-10,47	9,46	-1,11	0,268	
Fourier seno · armónico 1	-666,02	185,08	-3,60	< 0,001	✓
Fourier coseno · armónico 1	-852,55	153,80	-5,54	< 0,001	✓
Fourier seno · armónico 2	559,23	137,49	4,07	< 0,001	✓
Fourier coseno · armónico 2	-113,12	110,98	-1,02	0,308	
Festivo · Super Bowl	786,26	478,69	1,64	0,100	
Festivo · Día del Trabajo	468,87	840,08	0,56	0,577	
Festivo · Acción de Gracias	3 749,40	1 146,33	3,27	0,001	✓
Festivo · Navidad y fin de año	2 310,39	442,52	5,22	< 0,001	✓
Autorregresivo AR(1)	0,03	0,25	0,13	0,896	
Media móvil MA(1)	-0,44	0,24	-1,83	0,067	
Media móvil MA(2)	-0,42	0,16	-2,58	0,010	✓

Nota. EE = error estándar. Los términos de Fourier operan por pares seno-coseno; su interpretación individual carece de sentido sustantivo y deben leerse conjuntamente como la forma de la onda estacional anual. Los coeficientes de calendario admiten lectura directa como desplazamiento del nivel en la semana marcada.

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

Exactitud predictiva fuera de muestra

La Tabla 6 presenta los cuatro indicadores de exactitud sobre las 13 semanas reservadas para ambos modelos.

Tabla 6

Exactitud predictiva fuera de muestra de ambos modelos sobre las 13 semanas reservadas

Métrica	SARIMAX media (DE)	Prophet media (DE)	Diferencia (Prophet − SARIMAX)	Diferencia relativa
RMSE	2 247,2 (1 387,5)	2 483,8 (1 660,8)	236,6	9,5 %
MAE	1 858,0 (1 193,9)	2 085,8 (1 483,1)	227,8	10,9 %
MAPE (%)	7,00 (2,70)	7,64 (3,13)	0,63	8,3 %
MASE	0,694 (0,235)	0,770 (0,309)	0,075	9,8 %
AIC	2 246,7	no comparable	—	—
BIC	2 286,6	no comparable	—	—

Nota. Los criterios de información se reportan solo para SARIMAX: el modelo de descomposición se estima por máximo a posteriori y su verosimilitud no es la misma cantidad. La comparación inferencial se presenta a continuación.

Contraste de la diferencia de exactitud

Como se observa en la Tabla 7, el contraste pareado entre las 58 series y el contraste de Diebold-Mariano serie a serie arrojan lecturas distintas sobre la misma diferencia.

Tabla 7

Contraste de la diferencia de exactitud predictiva entre modelos

Métrica / resultado	Diferencia o recuento	IC 95 %	Estadístico	p (t / Wilcoxon)	Tamaño del efecto o proporción
Panel A · contraste pareado entre las 58 series					
RMSE	236,592	[86,989; 386,194]	t = 3,17	0,002 / < 0,001	d_z = 0,416 → SARIMAX
MAE	227,794	[78,549; 377,038]	t = 3,06	0,003 / < 0,001	d_z = 0,401 → SARIMAX

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

MAPE	0,634	[0,189; 1,079]	t = 2,85	0,006 / 0,002	d_z = 0,375 → SARIMAX
MASE	0,075	[0,026; 0,125]	t = 3,08	0,003 / 0,001	d_z = 0,404 → SARIMAX
Panel B · Diebold-Mariano serie a serie					
Favorece a SARIMAX (p < 0,05)	1	—	—	—	1,7 %
Favorece a Prophet (p < 0,05)	1	—	—	—	1,7 %
Sin diferencia significativa	56	—	—	—	96,6 %
Menor RMSE sin exigir significación	SARIMAX 42 · Prophet — 16		—	—	72,4 % para SARIMAX

Nota. En el Panel A la diferencia se calcula como pérdida de Prophet menos pérdida de SARIMAX; un valor positivo favorece a SARIMAX. En el Panel B, estadístico de Diebold-Mariano con corrección de Harvey, Leybourne y Newbold, computado sobre los 39 pares de errores que agrupan los tres orígenes de la validación móvil; un valor negativo favorece a SARIMAX. La mediana del estadístico en el Panel B es -0,284.

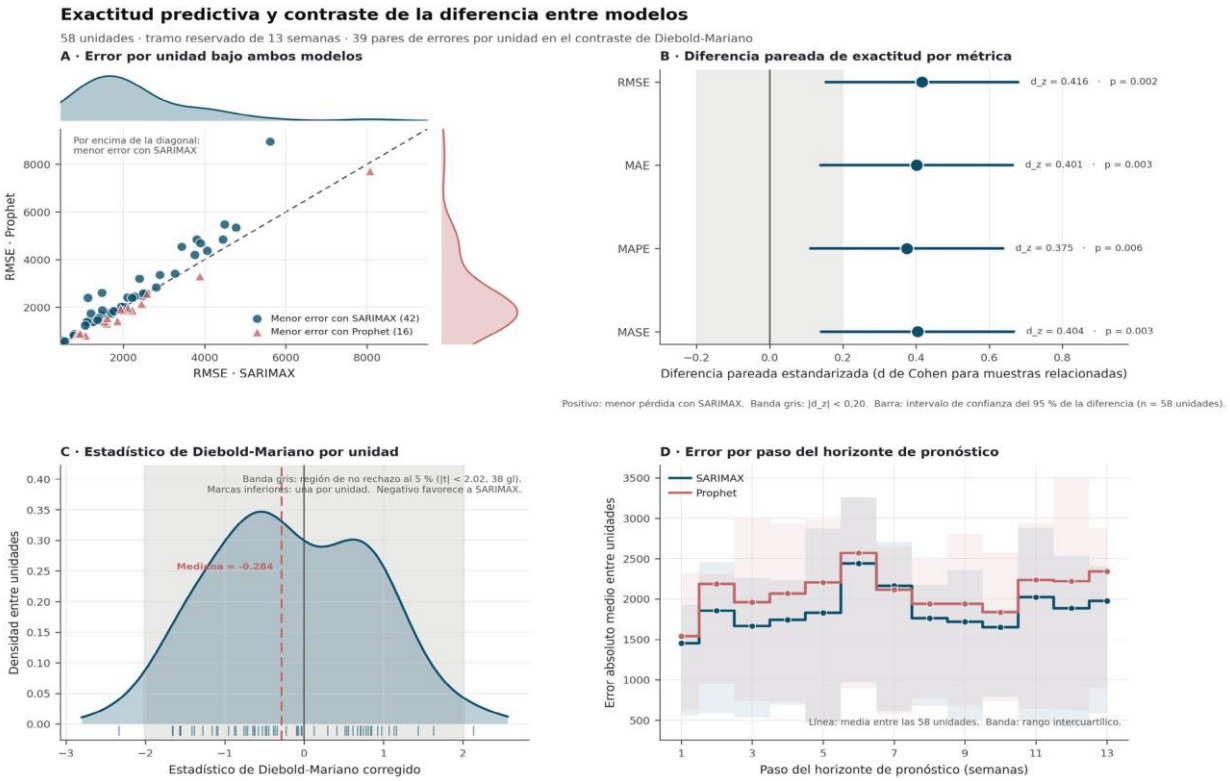


Figura 3. Exactitud predictiva y contraste de la diferencia entre modelos. Nota. Evaluación fuera de muestra sobre el tramo reservado de 13 semanas en cada una de las 58 unidades. Panel A: error cuadrático medio de cada unidad bajo ambos modelos, con la diagonal de identidad y las densidades marginales; SARIMAX obtiene menor error en 42 unidades y Prophet en 16. Panel B: diferencia pareada de exactitud por métrica,

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

estandarizada como d de Cohen para muestras relacionadas, con su intervalo de confianza del 95 %; sobre el RMSE, d_z = 0,416 y p = 0,002. Panel C: distribución entre unidades del estadístico de Diebold-Mariano con corrección de muestra pequeña de Harvey, Leybourne y Newbold, computado sobre los 39 pares de errores que agrupan los tres orígenes de la validación móvil; la banda gris es la región de no rechazo al 5 % y las marcas inferiores señalan cada unidad; la mediana es -0,284 y el contraste no distingue ambos modelos en 56 de las 58 unidades. Panel D: error absoluto medio por paso del horizonte, con la banda intercuartílica entre unidades.

Robustez: validación de origen móvil

La Tabla 8 reporta el resultado de la validación de origen móvil sobre ventana expansiva, con tres orígenes sucesivos.

Tabla 8

Validación de origen móvil sobre ventana expansiva

Fecha del origen	Semana	Series	RMSE SARIMAX media (DE)	RMSE Prophet media (DE)	Diferencia	Favorece
2012-02-03	104	58	2 467,4 (1 513,1)	2 500,2 (1 398,0)	32,8	SARIMAX
2012-05-04	117	58	2 273,3 (1 294,6)	1 961,8 (977,2)	-311,5	Prophet
2012-08-03	130	58	2 247,2 (1 387,5)	2 483,8 (1 660,8)	236,6	SARIMAX
Promedio de los orígenes	—	—	2 329,3	2 315,3	-14,0	Prophet

Nota. Origen móvil expansivo; init = 104 semanas, h = 13, paso = 13. Cada origen reestima ambos modelos con la información disponible hasta esa fecha y pronostica el horizonte completo.

¿De qué depende la ventaja relativa?

La Tabla 9 presenta el contraste entre la inestabilidad de la tendencia observada y la ventaja relativa de exactitud entre modelos, formulado para someter a prueba H2.

Tabla 9

Inestabilidad de la tendencia y ventaja relativa entre familias de modelos

Contraste	Estadístico	Valor	p	Lectura
Correlación de Spearman entre inestabilidad de tendencia y ventaja relativa	ρ	0,073	0,586	Asociación positiva, no significativa al 5 %
Correlación de Pearson (misma relación)	r	-0,016	0,906	Contraste paramétrico de la misma asociación
Series de tendencia ESTABLE (n = 29)	log(RMSE SARIMAX / RMSE Prophet) medio	-0,092	—	SARIMAX obtiene menor RMSE en 23 de 29
Series de tendencia INESTABLE (n = 29)	log(RMSE SARIMAX / RMSE Prophet) medio	-0,071	—	SARIMAX obtiene menor RMSE en 19 de 29

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

Diferencia entre ambos grupos (Welch)	t	0,44	0,659	La hipótesis no queda respaldada por estos datos

Nota. La inestabilidad de la tendencia se mide sobre la serie observada como |pendiente de la segunda mitad − pendiente de la primera| / nivel medio. La respuesta es log(RMSE SARIMAX / RMSE Prophet): valores positivos indican ventaja de Prophet.

Diagnóstico de residuos

La Tabla 10 reporta el diagnóstico de autocorrelación residual del modelo SARIMAX mediante la prueba de Ljung-Box a dos rezagos.

Tabla 10

Diagnóstico de autocorrelación residual del modelo SARIMAX

Diagnóstico	Rezagos	p (mediana)	Series sin autocorrelación residual al 5 %	Proporción	Lectura
Ljung-Box	10	0,625	56 de 58	96,6 %	Residuos compatibles con ruido blanco
Ljung-Box	20	0,531	53 de 58	91,4 %	La conclusión se mantiene a mayor rezago
Ljung-Box	criterio conjunto 10 y 20	—	53 de 58	91,4 %	Criterio más exigente de los tres

Nota. La hipótesis nula de la prueba de Ljung-Box es la ausencia de autocorrelación en los residuos; no rechazarla es el resultado deseable.

https://doi.org/10.5281/zenodo.22679531

Artículo de Investigación

BUSINESS INNOVA SCIENCES (BIS)

Editorial **EXPIDEC**

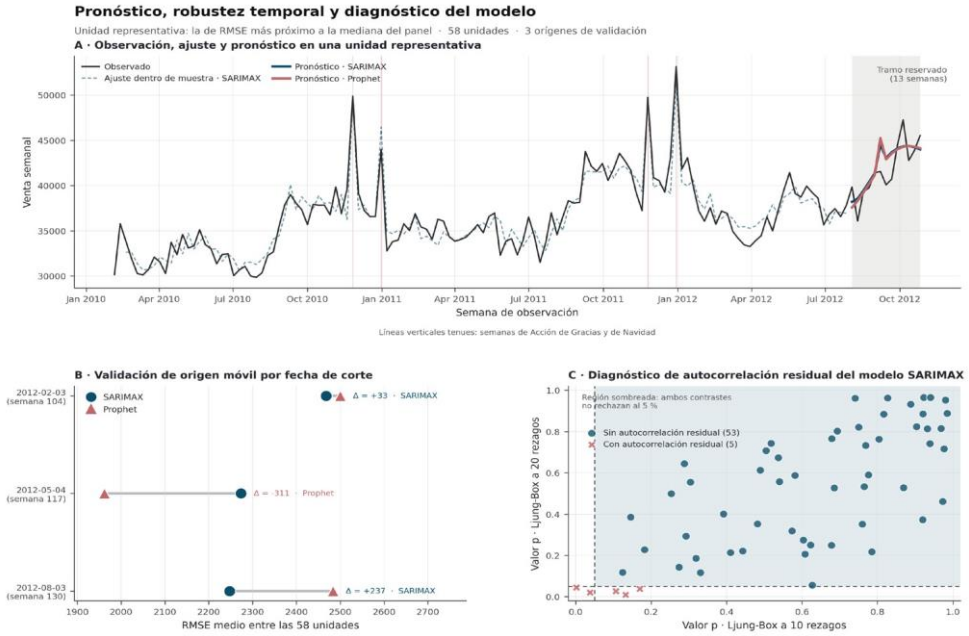


Figura 4. Pronóstico, robustez temporal y diagnóstico del modelo. Nota. Panel A: serie observada, ajuste dentro de muestra y pronóstico de ambos modelos en la unidad S21_D30, seleccionada por tener el error cuadrático medio más próximo a la mediana del panel; la banda sombreada delimita el tramo reservado de 13 semanas y las verticales tenues marcan las semanas de Acción de Gracias y de Navidad. Panel B: validación de origen móvil sobre ventana expansiva; cada par muestra el error cuadrático medio de ambos modelos promediado entre las 58 unidades en esa fecha de corte, con la diferencia anotada; SARIMAX obtiene menor error en 2 de los 3 orígenes. Panel C: diagnóstico de autocorrelación residual del modelo SARIMAX; cada punto es una unidad situada por el valor p de la prueba de Ljung-Box a diez y a veinte rezagos, y la región sombreada reúne las unidades en que ninguno de los dos contrastes rechaza al 5 %, que son 53 de 58.

DISCUSIÓN

El hallazgo más desconcertante de este estudio no es que SARIMAX haya superado a Prophet, sino que el criterio de información —la herramienta estándar de selección de modelos en la tradición autorregresiva— haya preferido de forma sistemática, y por un margen de varios centenares de unidades de AICc, la especificación que predice peor. La Tabla 4 muestra que la parametrización estacional multiplicativa con s = 52 obtuvo mejor ajuste dentro de muestra en las 8 series contrastadas —una ventaja media de 875 unidades de AICc, del tipo que en cualquier lectura convencional se leería como preferencia inequívoca—, y sin embargo la regresión armónica dinámica pronosticó con menor error en 7 de esas 8 series. La explicación técnica es directa: con apenas 2,74 ciclos estacionales disponibles, un término autorregresivo al retardo 52 dispone de poco más de dos repeticiones del patrón y

reproduce casi literalmente el año anterior, lo que eleva la verosimilitud dentro de muestra sin trasladarse a la capacidad de predecir el siguiente. La lección que se desprende no es que el criterio de información deba abandonarse —entre candidatos anidados de una misma representación, como los que compite la Tabla 3, sigue siendo un selector legítimo—, sino que su validez está acotada a candidatos comparables, y esa comparabilidad debe comprobarse, no suponerse.

Sobre esa base, el resultado central —la ventaja de SARIMAX en exactitud predictiva, con una diferencia relativa de 9,5 % en RMSE y d_z = 0,416— dialoga de forma asimétrica con la literatura comparativa reciente. Un contraste directo entre ambos modelos en la predicción de la demanda de arroz encontró la relación inversa: Prophet alcanzó un MAPE de 4,72 % frente a 5,49 % de SARIMAX, y se describió como más adaptativo a las variaciones de corto plazo, mientras que a SARIMAX se le reconoció una interpretabilidad causal superior (Aditya y Safrizal, 2025). En el pronóstico del índice de precios al consumidor de Côte d'Ivoire, SARIMAX ofreció una línea base sólida, pero fue el híbrido SARIMAX-LSTM el que alcanzó los menores errores de pronóstico, lo que sugiere que la ventaja econométrica pura tiene un techo (Kamagaté y Diabaté, 2026). En el pronóstico de precios mayoristas de un producto agrícola sudamericano, la comparación entre SARIMAX y arquitecturas de aprendizaje automático arrojó de nuevo un resultado no uniforme entre series (Evangelista et al., 2026), y en la predicción de inflación alimentaria el contraste entre métodos econométricos, aditivos descomponibles y redes neuronales tampoco produjo un ganador único y estable entre contextos (Javed, 2026). Notablemente, incluso dentro del dominio de las ventas minoristas intermitentes, un enfoque de dos etapas basado en XGBoost superó a los referentes clásicos en la métrica WRMSSE, lo que sitúa el desempeño relativo de SARIMAX y Prophet como un caso particular de una competencia más amplia entre familias de modelos, no como un resultado cerrado (Yilmaz et al., 2026). Frente a este cúmulo de evidencia divergente, el hallazgo de robustez que ofrece la validación de origen móvil (Tabla 8) es menos concluyente de lo que el resultado agregado sugiere: en el promedio de los tres orígenes evaluados, la diferencia se invierte levemente a favor de Prophet (-14,0 unidades), y en el origen intermedio la ventaja de Prophet es sustancial (-311,5). Es consistente con esa evidencia interpretar que la ventaja de SARIMAX, aunque estadísticamente significativa en el tramo terminal, no es uniforme a través del tiempo, y que su magnitud depende del punto del calendario en que se evalúe.

En contraste, la coincidencia con la literatura es más breve. El resultado se alinea con estudios que han reportado ventaja de SARIMAX sobre alternativas de aprendizaje automático en contextos de demanda con calendario marcado, como la demanda hotelera diaria (Ampountolas, 2021) y la demanda con promociones en la cadena de suministro minorista (Falatouri et al., 2022), y es consistente con la robustez del algoritmo Prophet documentada en contextos de demanda volátil, donde generó pronósticos razonablemente precisos incluso bajo condiciones fluctuantes (Sari et al., 2026) —una observación que matiza, sin contradecir, la ventaja agregada de SARIMAX aquí encontrada, porque sugiere que la brecha entre modelos podría estrecharse en unidades de mayor volatilidad.

Sobre H2, los datos no permiten sostener que la ventaja relativa dependa de la estabilidad de la tendencia ($\rho = 0{,}073$, $p = 0{,}586$; diferencia entre grupos, $t = 0{,}44$, $p = 0{,}659$). Es tentador leer este resultado nulo como un cierre, pero conviene resistir esa lectura: la hipótesis es plausible y comprobable, y este material simplemente no la respalda con la potencia disponible. Desde una perspectiva crítica, cabe preguntar si la medida de inestabilidad empleada —una diferencia de pendientes entre mitades del tramo de entrenamiento— es suficientemente sensible para capturar el tipo de quiebre que Prophet modela explícitamente; un enfoque de agrupamiento híbrido que combina reglas y aprendizaje automático para la previsión de demanda minorista sugiere que la heterogeneidad entre series podría estructurarse mejor mediante categorías discretas de comportamiento que mediante una medida continua única (Kim y Cho, 2026), lo que abre una vía de refinamiento antes que una refutación.

La contribución metodológica del estudio es independiente de qué modelo resultó ganador. El Panel B de la Tabla 7 muestra que la comparación serie a serie es mayoritariamente no concluyente —96,6 % de las 58 unidades—, de modo que un estudio construido sobre una única serie tenía, por diseño, una probabilidad alta de no detectar diferencia alguna, o de detectar una espuria. Esta observación conecta con una preocupación más amplia sobre la incertidumbre de modelo en el pronóstico, señalada hace tres décadas como un componente sistemáticamente subestimado de la exactitud reportada (Chatfield, 1996), y con marcos más recientes de analítica predictiva en inventario que advierten sobre el riesgo de generalizar conclusiones de pronóstico a partir de evidencia parcial (Krishnamurthy et al., 2024). Comparar familias de modelos de pronóstico exige, en consecuencia, una unidad de análisis múltiple y un contraste formal sobre la diferencia; en su ausencia, la

preferencia declarada por un modelo es la descripción de una muestra de tamaño uno.

Más allá del caso de estudio, el problema que este trabajo aborda tiene un alcance que trasciende lo local. La disyuntiva entre SARIMAX y Prophet —o, más ampliamente, entre modelos econométricos de estructura rica y modelos de descomposición flexible— se replica en dominios muy distintos y en geografías muy distintas: en la gestión municipal del agua en Palestina, donde un marco basado en Prophet se diseñó explícitamente para su despliegue a escala nacional (Salman y Shaka'a, 2026); en la demanda eléctrica sostenible de los Países Bajos, comparada mediante SARIMAX, su híbrido con LSTM y arquitecturas de secuencia a secuencia (Ashtar et al., 2025); en la previsión de tipo de cambio, donde la sensibilidad macroeconómica del modelo condiciona su exactitud predictiva (Abouzaid y Boussedra, 2025); y en el pronóstico de precios agrícolas en el Perú (Evangelista et al., 2026) y del índice de precios al consumidor en Côte d'Ivoire (Kamagaté y Diabaté, 2026). Esta dispersión geográfica y sectorial sugiere que el problema metodológico identificado aquí —la insuficiencia de una serie aislada para sostener una preferencia de modelo— no es una peculiaridad de la cadena minorista analizada, sino una condición estructural de la comparación entre familias de pronóstico allí donde se practique. En el plano regional latinoamericano, estudios sobre las características socioeconómicas del comercio minorista en México (Cantero y Morales, 2023) y sobre la disponibilidad de inventarios frente a la demanda en cadenas ecuatorianas (Zambrano y Zaldumbide, 2024) documentan una heterogeneidad de contextos operativos —tamaño de establecimiento, perfil socioeconómico del entorno, disponibilidad de infraestructura logística— que probablemente modera la exactitud relativa de cualquier familia de modelos de forma aún no cuantificada por el protocolo aquí empleado, y que constituye una extensión natural de este trabajo hacia mercados emergentes con estructuras de venta distintas a la del panel norteamericano analizado.

Finalmente, la evidencia disponible sobre arquitecturas híbridas y de aprendizaje profundo matiza el alcance de la comparación binaria aquí realizada. Marcos que combinan meta-aprendizaje con memoria de largo plazo para el pronóstico de ventas minoristas (Suresh et al., 2026), comparaciones entre conjuntos de árboles y arquitecturas LSTM que favorecen a las primeras en múltiples métricas (Nasseri et al., 2023), y marcos de decomposición-corrección diagnóstica que superan tanto a SARIMA como a Prophet en pronósticos de horizonte largo (Cao et al., 2026),

sugieren que la frontera de exactitud predictiva no se agota en ninguna de las dos familias aquí comparadas. La integración de aprendizaje automático y computación avanzada en la analítica de clientes del comercio minorista (Shafa, 2022) y los fundamentos de la descomposición estacional en modelos ARIMA (Findley et al., 2016) proveen, en conjunto, un terreno conceptual sobre el cual extender el protocolo de contraste aquí propuesto a arquitecturas adicionales, sin que ello reste validez a la comparación específica entre SARIMAX y Prophet que este estudio documenta.

CONCLUSIONES

El trabajo comparó dos familias de modelos de pronóstico de demanda estacional sobre un panel de 58 series semanales de venta minorista, con selección de especificación explícita para cada uno, un mismo tramo de evaluación reservado y contraste inferencial de la diferencia.

Primero, SARIMAX obtuvo un error de pronóstico menor que Prophet y la diferencia resultó estadísticamente significativa, con una ventaja relativa de 9,5 % en RMSE y d_z = 0,416, lo que confirma H1 sobre el conjunto de unidades analizado.

Segundo, la estacionalidad anual de una serie semanal admite una representación armónica parsimoniosa que evita la parametrización estacional multiplicativa, inviable con el número de ciclos disponible; y —hallazgo que excede este caso— el criterio de información prefirió sistemáticamente, y por un margen de varios centenares de unidades, la parametrización que predice peor. La lección no es que ese criterio deba abandonarse: entre candidatos anidados de una misma representación sigue siendo un selector adecuado, y así se emplea en este trabajo. Es que su validez está acotada a candidatos comparables, y la elección entre representaciones estructuralmente distintas debe resolverse fuera de muestra.

Tercero, la comparación serie a serie resultó no concluyente en 96,6 % de los casos, de modo que la preferencia entre familias de modelos de pronóstico no es establecible sobre una serie aislada. Requiere una unidad de análisis múltiple y una prueba formal sobre la diferencia de pérdida. Ese es el aporte metodológico transferible del estudio, y es independiente de qué modelo resultó ganador aquí.

Cuarto, la hipótesis sobre el rol moderador de la inestabilidad de la tendencia (H2) no encontró respaldo en los datos disponibles, lo que deja abierta una línea de indagación específica antes que una respuesta cerrada.

En conjunto, estos resultados sostienen que la elección entre familias de modelos de pronóstico de demanda estacional debe resolverse mediante evaluación empírica sobre la propia cartera de series, con un protocolo replicable de reserva de tramo, selección de especificación fuera de muestra y contraste pareado de exactitud, y no mediante la extrapolación de la preferencia observada en un único caso o en un único conjunto de datos.

REFERENCIAS

Abouzaid, O., & Boussedra, F. (2025). Artificial intelligence and exchange rate forecasting: assessing predictive accuracy and macroeconomic sensitivity. *Frontiers in Applied Mathematics and Statistics*, *11*. https://doi.org/10.3389/fams.2025.1654093

Aditya, F., & Safrizal, S. (2025). Analisa Perbandingan Metode SARIMAX dan Prophet Dalam Prediksi Kebutuhan Beras. *TIN: Terapan Informatika Nusantara*, *6*(6), 652-660. https://doi.org/10.47065/tin.v6i6.8599

Albahli, S. (2025). LSTM vs. Prophet: Achieving Superior Accuracy in Dynamic Electricity Demand Forecasting. *Energies*, *18*(2), 278. https://doi.org/10.3390/en18020278

Alharbi, F., & Csala, D. (2022). A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. *Inventions*, *7*(4), 94. https://doi.org/10.3390/inventions7040094

Ampountolas, A. (2021). Modeling and Forecasting Daily Hotel Demand: A Comparison Based on SARIMAX, Neural Networks, and GARCH Models. *Forecasting*, *3*(3), 580-595. https://doi.org/10.3390/forecast3030037

Ashtar, D., Mohammadi, S., & Alsahag, A. (2025). Hybrid Forecasting for Sustainable Electricity Demand in The Netherlands Using SARIMAX, SARIMAX-LSTM,

and Sequence-to-Sequence Deep Learning Models. *Sustainability*, *17*(16), 7192. https://doi.org/10.3390/su17167192

Cantero, M., & Morales, J. (2023). Características socioeconómicas del comercio minorista en tiendas de abarrotes y tiendas de conveniencia en Zapopan, Jalisco, México. *Sapientiae*, *9*(1). https://doi.org/10.37293/sapientiae91.06

Cao, Q., Sun, Z., & Li, H. (2026). Comparative Analysis of SARIMA, Prophet, and a Diagnostic Decomposition–Correction Hybrid for Long-Horizon Lottery Sales Forecasting. *Entropy*, *28*(3), 286. https://doi.org/10.3390/e28030286

Chatfield, C. (1996). Model uncertainty and forecast accuracy. *Journal of Forecasting,* *15*(7), 495–508. https://doi.org/10.1002/(SICI)1099-131X(199612)15:7<495::AID-FOR640>3.0.CO;2-O

Dwivedi, M., & Mishra, S. (2025). Comparative Evaluation of Machine Learning Models for Retail Sales Forecasting: A Multi-Algorithm Approach. *International Journal of Science and Research (IJSR)*, 1122-1134. https://doi.org/10.21275/sr25601090618

Espasa, A., & Peña, D. (1995). The decomposition of forecast in seasonal arima models. *Journal of Forecasting*, *14*(7), 565-583. https://doi.org/10.1002/for.3980140703

Evangelista, M., Chancan, J., Figueroa, P., & Alvarez, R. (2026). Forecasting Wholesale Prices of Canchan potatoes using artificial intelligence models with data from the Ministry of Agrarian Development and Irrigation of Peru (2020–2025). *Frontiers in Sustainable Food Systems*, *10*. https://doi.org/10.3389/ fsufs.2026.1883363

Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Computer Science*, *200*, 993-1003. https://doi.org/10.1016/j.procs.2022.01.298

Findley, D., Lytras, D., & Maravall, A. (2016). Illuminating ARIMA model-based seasonal adjustment with three fundamental seasonal models. *SERIEs*, *7*(1), 11-52. https://doi.org/10.1007/s13209-016-0139-4

Galdelli, A., Fronzi, D., Narang, G., Mancini, A., & Tazioli, A. (2025). Groundwater level forecasting using data-driven models and vadose zone: A comparative analysis of ARIMA, SARIMAX, Prophet, and NeuralProphet. *Applied Computing and Geosciences*, *28*, 100304. https://doi.org/10.1016/j.acags.2025.100304

Giri, C., & Chen, Y. (2022). Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry. *Forecasting*, *4*(2), 565-581. https://doi.org/10.3390/forecast4020031

Han, J., Yoon, C., & Hwang, C. (2025). Proposal of Stacked SARIMAX-Transformer for Improving Forecasting Accuracy of Time Series Data. *Journal of the Korea Institute of Information and Communication Engineering*, *29*(3), 303-308. https://doi.org/10.6109/jkiice.2025.29.3.303

Hewage, H., Perera, H., & Bandara, K. (2025). Enhancing Demand Forecasting in Retail: A Comprehensive Analysis of Sales Promotional Effects on the Entire Demand Life Cycle. *Journal of Forecasting*, *45*(1), 293-315. https://doi.org/10.1002/for.70039

Javed, A. (2026). Benchmarking econometric, decomposable additive, and neural network methods for food inflation prediction featuring policy insights. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-34993-w

Kamagaté, A., & Diabaté, N. (2026). HYBRID FORECASTING FOR THE CONSUMER PRICE INDEX USING SARIMAX, SARIMAX-MACHINE LEARNING AND SARIMAX-DEEP LEARNING MODELS: THE CASE OF CÔTE D'IVOIRE. *Advances and Applications in Statistics*, *93*(4), 475-502. https://doi.org/10.17654/0972361726024

Karim, R. (2025). ARTIFICIAL INTELLIGENCE-ENHANCED PREDICTIVE ANALYTICS FOR DEMAND FORECASTING IN U.S. RETAIL SUPPLY CHAINS. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *01*(01), 959-993. https://doi.org/10.63125/gbkf5c16

Kim, J., & Cho, N. (2026). Hybrid Clustering for Retail Demand Forecasting: Combining Rule-Based and Machine Learning Methods. *Forecasting*, *8*(3),

37. https://doi.org/10.3390/forecast8030037

Krishnamurthy, S., Nadukuru, S., Dave, S., Goel, O., Jain, P., & Kumar, L. (2024). Predictive Analytics in Retail: Strategies for Inventory Management and Demand Forecasting. *Journal of Quantum Science and Technology*, *1*(2). https://doi.org/10.63345/jqst.v1i2.9

Lee, G. (2025). A Data-Driven Approach to Tourism Demand Forecasting: Integrating Web Search Data into a SARIMAX Model. *Data*, *10*(5), 73. https://doi. org/10.3390/data10050073

Mishra, A., & Ashraf, G. (2026). Evaluating the Role of Artificial Intelligence in Demand Forecasting: Implications for Inventory Management Efficiency in the Retail Sector. *International Journal of Research Publication and Reviews*, *7*(5), 2649-2657. https://doi.org/10.55248/gengpi.07.0526.d1145

Mishra, A., & Sinha, M. (2025). Data Analytics for Product Segmentation and Demand Forecasting of a Local Retail Store Using Python. *International Journal of Advanced Computer Science and Applications*, *16*(2). https://doi. org/10.14569/ijacsa.2025.0160224

Mishra, S. (2025). Comparative Study of Machine Learning Algorithms for Sales Forecasting: Facebook Prophet vs. Established Statistical Models. *European Economic Letters*. https://doi.org/10.52783/eel.v15i1.2751

Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning. *Applied Sciences*, *13*(19), 11112. https://doi.org/10.3390/app131911112

Natta, P. (2022). AI-Driven Inventory Intelligence for Large-Scale Retail Operations: A Framework for Real-Time Store-Level Stock Accuracy. *International Journal of Advanced Engineering Science and Information Technology*, *05*(02). https://doi.org/10.15662/ijaesit.2022.0502002

Punia, S. (2025). Medium- to Long-Term Demand Forecasting in Retail and Manufacturing Organizations: Integration of Machine Learning, Human

Judgment, and Interval Variable. *Journal of Forecasting*, *45*(1), 122-134. https://doi.org/10.1002/for.70030

Ram, C., Raj, M., & Chaturvedi, R. (2025). Boosting Time-Series Forecasting Accuracy with SARIMAX Seasonal Interval Automation. *Procedia Computer Science*, *260*, 814-821. https://doi.org/10.1016/j.procs.2025.03.262

Salman, A., & Shaka'a, Y. (2026). Automated water demand forecasting for nationalscale deployment: a prophet-based framework for Palestinian municipal water management. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598025-33060-0

Sari, R., Salkiawati, R., Lestari, N., & Fitriyani, A. (2026). Forecasting Inventory Demand Under Volatile Sales Patterns Using the Prophet Algorithm. *J-INTECH*, *14*(01), 141-153. https://doi.org/10.32664/j-intech.v14i01.2032

Shafa, H. (2022). INTEGRATION OF MACHINE LEARNING AND ADVANCED COMPUTING FOR OPTIMIZING RETAIL CUSTOMER ANALYTICS. *International Journal of Business and Economics Insights*, *02*(03), 01-46. https://doi.org/10.63125/p87sv224

Sinaga, A., Br, D., Rossa, A., & Auliah, D. (2026). Predictive Analytics of Food Retail Seasonal Trends with Advanced Forecasting Modeling. *Journal of Applied Informatics and Computing*, *10*(3), 3118-3125. https://doi.org/10.30871/jaic.v10i3.13173

Sousa, A., Barbosa, Y., Silva, T., & Rêgo, T. (2024). Previsão de demanda de longo prazo aplicada a uma empresa do varejo de cosméticos utilizando o Prophet. *Brazilian Journal of Production Engineering*, *10*(3), 372-383. https://doi.org/10.47456/bjpe.v10i3.45146

Suresh, B., Suresh, M., & Kang, D. (2026). Meta-LLSTM: meta-learning enhanced learnable LSTM for retail sales forecasting. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-54836-y

Wang, B., & Zain, A. (2025). A Hybrid XGBoost-LSTM Framework for Supply

Chain Demand Forecasting: Empirical Evidence from Retail Multi-Store Data. *Journal of Cultural Analysis and Social Change*, 4056-4073. https:// doi.org/10.64753/jcasc.v10i4.3736

Yilmaz, B., Beyca, Ö., & Kosanoğlu, F. (2026). Forecasting Intermittent Sales in Fashion Retail: A Two-Stage Machine Learning Approach. *Forecasting*, *8*(4), 56. https://doi.org/10.3390/forecast8040056

Zambrano, A., & Zaldumbide, D. (2024). Disponibilidad de inventarios frente a la demanda en productos de Tiendas TuTi. *593 Digital Publisher CEIT*, *9*(2), 228-244. https://doi.org/10.33386/593dp.2024.2.2327


Esta obra se distribuye bajo la licencia Creative Commons AtribuciónNoComercial 4.0 Internacional (CC BY-NC 4.0). Los usuarios pueden compartir, copiar, redistribuir y adaptar el material en cualquier medio o formato, siempre y cuando reconozcan la autoría original y no utilicen los contenidos con fines comerciales.