

Evaluación de Modelos de Series de Tiempo para el Pronóstico del Turismo: El Caso de la Ocupación Hotelera en Colombia

Gisell Natalia Peña Agudelo

Especialización en Estadística Aplicada

Fundación Universitaria Los Libertadores

Aceptado: julio 03, 2025

Resumen

El turismo en Colombia constituye un sector estratégico para el desarrollo económico, con una notable capacidad de recuperación tras el impacto de la pandemia. Este estudio analiza su comportamiento a través de la tasa de ocupación hotelera, utilizada como variable representativa del dinamismo turístico. El objetivo principal es estimar y comparar modelos econométricos multivariados para pronosticar dicha tasa, empleando series temporales del PIB, IPC, visitantes no residentes y pasajeros nacionales entre los años 2012 y 2024. Se implementaron tres enfoques: el modelo VAR (Vector Autorregresivo), el ECM (Modelo de Corrección de Errores multivariado) y el VEC (Vector de Corrección de Errores multivariado). Los resultados indican que el modelo VEC presentó el mejor desempeño predictivo en términos de error cuadrático medio (RMSE), lo que resalta su capacidad para capturar relaciones de largo plazo entre variables cointegradas. No obstante, este modelo no superó completamente las pruebas de diagnóstico, especialmente en lo referente a normalidad y autocorrelación. Estos hallazgos reafirman la utilidad de los modelos VEC en contextos donde existen relaciones estructurales entre variables económicas, y permiten contrastar su desempeño frente a otros enfoques que, si bien superaron más pruebas estadísticas, no lograron una mejor capacidad de pronóstico. El estudio aporta así una herramienta valiosa para la toma de decisiones en el ámbito turístico, basada en análisis econométrico riguroso.

Palabras clave— turismo, tasa de ocupación, pronóstico, series de tiempo

Abstract

Tourism in Colombia is a strategic sector for economic development, with a remarkable capacity for recovery following the impact of the pandemic. This study analyzes its behavior through the hotel occupancy rate, used as a representative variable of tourism dynamics. The main objective is to estimate and compare multivariate econometric models to forecast this rate, using time series data on GDP, CPI, non-resident visitors, and domestic passengers from 2012 to 2024. Three approaches were implemented: the VAR (Vector Autoregressive) model, the ECM (Error Correction Model), and the VEC (Vector Error Correction) model. The results indicate that the VEC model showed the best predictive performance in terms of Root Mean Squared Error (RMSE), highlighting its ability to capture long-term relationships among cointegrated variables. However, this model did not fully pass all diagnostic tests, particularly those related to residual normality and autocorrelation. These findings reaffirm the usefulness of VEC models in contexts where structural relationships among economic variables exist, and allow for a comparison with other approaches which, although they passed more statistical tests, did not achieve better forecasting accuracy. Thus, the study provides a valuable tool for decision-making in the tourism sector, grounded in rigorous econometric analysis.

Keywords— Tourism, occupancy rate, forecast, time series

1 Introducción

El turismo constituye uno de los sectores estratégicos para el desarrollo económico y territorial en Colombia. (Brida et al., 2020). Su dinamismo, capacidad de generación de empleo y encadenamientos productivos lo posicionan como una actividad prioritaria dentro de las políticas de crecimiento sostenible. (Brida et al., 2008) Tras el impacto generado por la pandemia de COVID-19, el país ha evidenciado una recuperación progresiva del sector, reflejada en indicadores como el arribo de visitantes no residentes, los ingresos por alojamiento y, particularmente, la tasa de ocu-

pación hotelera, que sirve como variable representativa de la demanda turística formal. Según la definición de la Organización Mundial de Naciones Unidas (ONU), el turismo es un fenómeno social, cultural y económico que supone el desplazamiento de personas a países o lugares fuera de su entorno habitual por motivos personales, profesionales o de negocios. Esas personas se denominan viajeros (que pueden ser o bien turistas o excursionistas; residentes o no residentes) y el turismo abarca sus actividades, algunas de las cuales suponen un gasto turístico. Esta conceptualización resalta el carácter multidimensional del turismo y

su vínculo directo con variables económicas agregadas.

El turismo no solo en Colombia sino a nivel internacional, no ha sido estudiado de manera amplia (Voronkova, 2019). Pese a la creciente disponibilidad de información estadística, los estudios académicos sobre turismo suelen concentrarse en análisis descriptivos o de tipo exploratorio, siendo escasa la aplicación de modelos econométricos multivariados orientados al pronóstico. Esta limitación reduce la capacidad analítica para anticipar cambios en la actividad turística y limita la toma de decisiones basada en evidencia cuantitativa. El objetivo general de este trabajo es estimar y comparar modelos econométricos multivariados para pronosticar la tasa de ocupación hotelera en Colombia, considerada como un indicador representativo de la actividad turística nacional. Para alcanzar este propósito, se plantean los siguientes objetivos específicos: 1. Analizar la dinámica temporal de la tasa de ocupación hotelera y sus principales determinantes entre 2012 y 2024. 2. Estimar y validar modelos VAR, MEC y VEC a partir de series temporales del periodo 2012–2024. 3. Comparar el desempeño predictivo de los modelos y proyectar la tasa de ocupación hotelera para el primer trimestre de 2025.

2 Metodología

Este estudio adopta un enfoque cuantitativo y explicativo, orientado a modelar y pronosticar el comportamiento de la tasa de ocupación hotelera en Colombia mediante técnicas econométricas multivariadas. La investigación se basa en el análisis de series temporales mensuales y emplea modelos que permiten capturar tanto relaciones de corto como de largo plazo entre variables económicas y turísticas.

Se utilizaron tres enfoques metodológicos principales. En primer lugar, el modelo de Vectores Autorregresivos (VAR), propuesto por Sims (1980), permite analizar la dinámica conjunta de variables interrelacionadas y estacionarias sin imponer una estructura causal a priori. En segundo lugar, se aplicó un Modelo de Corrección de Errores (MEC) (Engle & Granger, 1987), que integra relaciones de equilibrio de largo plazo con ajustes dinámicos de corto plazo, a partir de una única relación de cointegración. Finalmente, se utilizó el modelo Vectorial de Corrección de Errores (VEC), desarrollado formalmente por (Johansen, 1988), que extiende el MEC para múltiples relaciones de cointegración entre variables no estacionarias.

El análisis comenzó con la identificación del orden de integración de las series a través de la prueba de Dickey-Fuller Aumentada (ADF), una herramienta estándar en análisis de series de tiempo (Dickey & Fuller, 1979). Para la detección de relaciones de cointegración, se aplicaron las pruebas de Phillips-Ouliaris y Johansen, ambas ampliamente aceptadas en la literatura econométrica (Phillips & Ouliaris, 1990); (Johansen, 1988). La selección del número óptimo de rezagos para los modelos VAR y VEC se basó en criterios de información como el Akaike Information Criterion (AIC), Hannan–Quinn Criterion (HQ), Schwarz Criterion o BIC (SC) y Final Prediction Error (FPE).

Una vez estimados los modelos, se aplicaron pruebas de diagnóstico correspondiente, para verificar la validez de

los supuestos estadísticos, algunas de estas: autocorrelación (Breusch-Godfrey), heterocedasticidad (ARCH test), normalidad (Jarque-Bera) y estabilidad estructural (CUSUM).

Cada modelo fue estimado sobre un periodo histórico común y utilizado para realizar pronósticos trimestrales. El desempeño predictivo se evaluó a través del error cuadrático medio (RMSE), lo que permitió comparar la precisión de los modelos y determinar cuál resulta más útil para la planificación del sector turístico colombiano.

2.1 Datos

Este estudio se basa en una serie de tiempo mensual comprendida entre enero de 2012 y diciembre de 2024, construida a partir de fuentes oficiales como el Departamento Administrativo Nacional de Estadística (DANE) y Ministerio de Comercio, Industrio y Turismo (MINCIT). La variable dependiente seleccionada es la tasa de ocupación en establecimientos de alojamiento y hospedaje, utilizada como proxy del comportamiento turístico en Colombia. Esta variable refleja la proporción de plazas ocupadas frente al total disponible y es reconocida como un indicador clave para evaluar la dinámica del sector.

Las variables explicativas incluyen indicadores económicos y de movilidad asociados al turismo. El Índice de Seguimiento a la Economía (ISE) como la variable que recoge información del Producto Interno Bruto (PIB) de manera mensual, se incorpora como aproximación al nivel de actividad económica. Para efectos del presente análisis es tomada en cuenta como PIB, con información del sector Comercio al por mayor y al por menor; Reparación de vehículos automotores y motocicletas; Transporte y almacenamiento; Alojamiento y servicios de comida. El índice de precios al consumidor (IPC) se emplea como indicador de inflación y condiciones del entorno de consumo. Adicionalmente, se incluyen los Visitantes no Residentes que ingresan al país y el Número de Pasajeros nacionales por vía aérea, como aproximaciones a la demanda turística.

Inicialmente se realiza la eliminación de 3 filas de datos debido a que la variable Pasajeros Aéreos Nacionales presentaban valores en NA, por motivo de la pandemia. Acto seguido, las variables fueron transformadas a logaritmos naturales con el propósito de homogeneizar las unidades de medida, facilitar la interpretación y reducir posibles efectos de heterocedasticidad.

Tabla 1: Resumen de variables utilizadas

Variable	Descripción
Logocu_ts	Tasa mensual de ocupación hotelera
Logvis_ts	Número de Visitantes no Residentes
Logpib_ts	Índice de Seguimiento a la Economía
Logipc_ts	Índice de Precios al Consumidor
Logpas_ts	Pasajeros Aéreos Nacionales

2.2 Exploración de datos

Antes de realizar la modelación econométrica, se llevó a cabo un análisis exploratorio de las series temporales con

el objetivo de identificar patrones, tendencias y posibles rupturas estructurales.

Para ello, se graficaron las cinco variables transformadas en logaritmos:

- *Logocu_ts*: Tasa de ocupación hotelera.
- *Logvis_ts*: Número de visitantes no residentes.
- *Logpib_ts*: Índice de Seguimiento a la Economía.
- *Logipc_ts*: Índice de Precios al Consumidor.
- *Logpas_ts*: Pasajeros nacionales transportados por vía aérea.

En términos generales, todas las series presentan una evolución creciente a lo largo del tiempo, con caídas abruptas durante el año 2020 asociadas al impacto de la pandemia por COVID-19. Posteriormente, se observa una recuperación sostenida, especialmente en la tasa de ocupación y en la llegada de visitantes no residentes.

En la Figura 1 se compara la evolución relativa de las cinco variables desde 2012 hasta 2024. Todas han sido normalizadas a una base 100, es decir, parten del mismo valor inicial para facilitar la comparación de sus trayectorias.

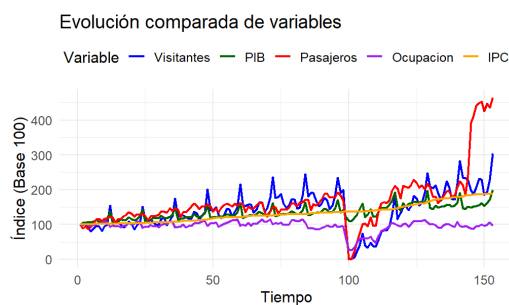


Figura 1: Evolución comparada de las variables utilizadas para el análisis del turismo en el periodo de 2012 a 2024 .

Se analizaron las características estadísticas básicas de cada serie, incluyendo la media, desviación estándar y cuantiles. También se aplicaron pruebas de raíz unitaria Augmented Dickey-Fuller (ADF), encontrando que las variables no son estacionarias, punto importante para iniciar el proceso de modelación.

Tabla 2: Estadísticas descriptivas (2012–2024)

Variable	Media	Mediana	Desv. Est.
PIB	108.1	104.1	16.1
IPC	101.8	99.2	20.0
Ocupación	51.1	52.4	8.4
Visitantes	294794	298009	113589
Pasajeros	2108878	1939622	1022229
Variable	Coef. V	Mínimo	Máximo
PIB	14.88	79.92	159.76
IPC	19.61	76750	144.8
Ocupación	16.41	13.57	61.71
Visitantes	38.53	2968	629448
Pasajeros	47.52	598	5930456

Adicionalmente, se realiza la aplicación del test de granger para determinar la forma cómo se causan las variables entre ellas y así poder complementar la identificación de la variable explicada.

Los resultados de la prueba de causalidad de Granger aplicada a las variables logarítmicas del proyecto muestran que la tasa de ocupación hotelera (*Logocu*) mantiene relaciones dinámicas significativas con las distintas variables analizadas, lo cual respalda su uso como variable dependiente en los modelos econométricos. En primer lugar, se evidencia una relación de causalidad bidireccional entre *Logocu* y los visitantes no residentes (*Logvis*), con valores $p < 0.01$ en ambos sentidos, lo que indica una fuerte interdependencia entre el flujo turístico internacional y la ocupación. De igual forma, el PIB del sector turístico (*Logpib*) presenta también causalidad en ambas direcciones con *Logocu*, lo que sugiere un vínculo estrecho entre la dinámica económica del sector y la actividad hotelera. En cuanto al IPC turístico (*Logipc*) y los pasajeros nacionales (*Logpas*), si bien no presentan evidencia estadística de causar a *Logocu*, sí resultan afectados por ella, lo que sugiere que su relación con la ocupación podría manifestarse más claramente en modelos donde estas variables sean tratadas como dependientes. En conjunto, los resultados respaldan empíricamente una estructura en la que *Logocu* actúa como variable endógena explicada por determinantes turísticos y económicos clave.

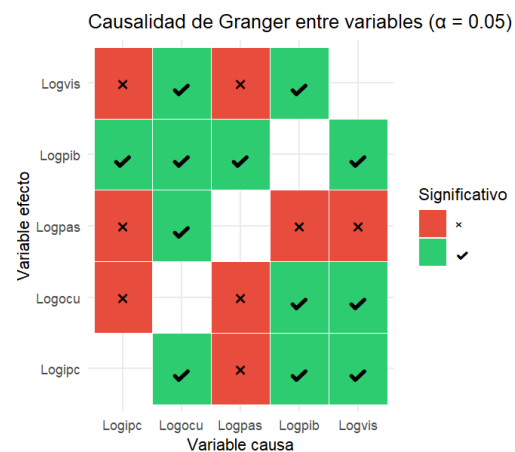


Figura 2: Mapa de significancia de la causalidad de Granger

Continuando, en la Figura 3 se presenta un diagrama de boxplot de valores atípicos comparado para las diferentes variables.

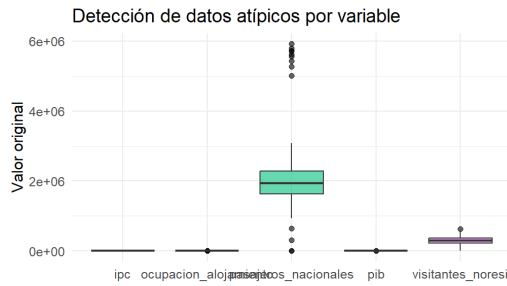


Figura 3: Diagrama exploratorio de datos atípicos de las cinco variables originales.

se observa la presencia de valores atípicos notables en las variables pasajeros nacionales y, en menor medida, en visitantes no residentes. Esto sugiere una alta dispersión en el número de pasajeros registrados, posiblemente asociada a eventos excepcionales como temporadas altas o choques externos (por ejemplo, la pandemia o aperturas turísticas). Las demás variables —IPC, PIB y Ocupación hotelera— presentan una distribución mucho más concentrada, sin evidencias significativas de valores extremos. Esta heterogeneidad en escalas y dispersión refuerza la decisión de transformar las variables a logaritmos para estabilizar la varianza y mitigar la influencia de valores extremos en los modelos.

2.3 Construcción de modelos

En esta sección se estiman y comparan tres modelos econométricos —VAR, MEC (Engle-Granger) y VEC (Johansen)— con el propósito de analizar y pronosticar la tasa de ocupación hotelera en Colombia, a partir de variables económicas y turísticas seleccionadas. Cada modelo es evaluado con pruebas estadísticas y criterios de desempeño predictivo.

1. Modelo VAR

Se estimó un modelo VAR con 12 rezagos sobre las variables diferenciadas, seleccionados con base en el criterio AIC, SC HQ, FPE. La ecuación estimada para la variable dependiente (Logocu) es la siguiente:

$$\begin{aligned} \log(\text{ocup}_t) = & \alpha + \sum_{i=1}^{12} \beta_{1i} \log(\text{ocup}_{t-i}) + \sum_{i=1}^{12} \beta_{2i} \log(\text{vis}_{t-i}) \\ & + \sum_{i=1}^{12} \beta_{3i} \log(\text{pib}_{t-i}) + \sum_{i=1}^{12} \beta_{4i} \log(\text{ipc}_{t-i}) \\ & + \sum_{i=1}^{12} \beta_{5i} \log(\text{pas}_{t-i}) + \varepsilon_t \end{aligned}$$

Para garantizar la estacionariedad de las series, se aplicaron transformaciones en logaritmos y dos diferenciaciones a cada variable. Sin embargo, con el fin de simplificar la notación, la ecuación del modelo VAR(12) se presenta en términos de logaritmos sin detallar explícitamente las diferencias aplicadas.

Se detectó autocorrelación en los residuos, lo que afecta la validez de las inferencias dinámicas. La prueba de

Jarque-Bera multivariada rechazó la hipótesis de normalidad ($p < 0.05$). No se detectó heteroscedasticidad significativa ($p = 0.05$), por lo cual se asume homocedasticidad marginal. Además, el modelo cumple con la condición de estabilidad, ya que todas las raíces del polinomio característico están dentro del círculo unitario.

2. Modelo de Corrección de Errores (MEC)

Con base en la metodología de Engle y Granger (1987), se estimó un modelo de corrección de errores (MEC) para capturar la dinámica de corto plazo entre la tasa de ocupación hotelera y sus principales determinantes: visitantes no residentes y producto interno bruto. Debido a que la variable IPC resultó ser integrada de orden dos, fue excluida del modelo MEC. Del mismo modo, la variable Pasajeros Aéreos Nacionales fue descartada por no ser estadísticamente significativa.

La evidencia de cointegración se estableció mediante la prueba ADF aplicada a los residuos del modelo en niveles, los cuales resultaron estacionarios en todas las especificaciones consideradas (sin constante, con deriva y con tendencia), confirmando la existencia de una relación de equilibrio de largo plazo.

Una vez surtido el ajuste del modelo, es estimado por medio de la siguiente ecuación:

$$\Delta \log(\text{ocup}_t) = \alpha + \beta_1 \Delta \log(\text{vis}_t) + \beta_2 \Delta \log(\text{pib}_t) + \gamma \cdot ECT_{t-1} + \varepsilon_t$$

El modelo MEC resulta adecuado para capturar tanto los efectos de corto plazo como los mecanismos de ajuste de largo plazo en la tasa de ocupación hotelera. Es el único modelo que supera todas las pruebas de diagnóstico esenciales —excepto la normalidad— y se ajusta a los supuestos requeridos por el método de Engle y Granger. Esto lo convierte en una alternativa robusta para análisis explicativo y para pronósticos de corto plazo.

3. Modelo VEC

Dado que las variables sobre Ocupación hotelera, Visitantes no Residentes, PIB y Pasajeros Aéreos Nacionales son integradas de orden uno, se evaluó la existencia de relaciones de cointegración mediante el enfoque de Johansen. La prueba $po.test$ confirmó cointegración ($p = 0.01$), y los resultados de la prueba de trazas indicaron la presencia de dos relaciones de cointegración cuando se utilizó la especificación con tendencia transitoria y 12 rezagos en niveles.

Una vez estimado el modelo se evidenció que la variable Pasajeros Aéreos Nacionales no aportaba capacidad significativa explicativa al modelo, por tanto, fue retirada. Quedando la tasa de ocupación hotelera como variable dependiente del PIB y los Visitantes no Residentes. En esta medida se evidenció que el PIB es la variable que más ajusta los desequilibrios de largo plazo, mientras que la ocupación hotelera responde fuertemente al PIB en el corto plazo. Por lo demás, se puede considerar que la especificación es adecuada y permite capturar relaciones relevantes del turismo.

El modelo VEC fue estimado especificando dos vectores de cointegración ($r = 2$) y 12 rezagos en niveles, los cuales habían sido seleccionados previamente mediante criterios de información.

Este modelo es apropiado para capturar tanto la dinámica de corto plazo como las relaciones de equilibrio de largo plazo entre las variables analizadas. Si bien su desempeño predictivo es razonable y cumple con la condición de homoscedasticidad, es importante tener en cuenta que presenta autocorrelación y falta de normalidad en los residuos.

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \mu + \varepsilon_t$$

2.4 Evaluación

A continuación se presentan los resultados de los pronósticos para los 12 meses del año 2025 de los tres modelos especificados y la comparación con los datos reales de la tasa de ocupación hotelera, tomados de la Encuesta Mensual de Agencias (EMA) realizada por el DANE:

Mes	VAR	MEC	VEC	Real
Ene 2025	56.12	53.87	52.27	50.50
Feb 2025	56.26	56.53	52.31	50.80
Mar 2025	48.68	59.32	50.46	49.50
Abr 2025	45.51	62.25	44.18	46.80
May 2025	59.31	65.33	43.35	—
Jun 2025	64.41	68.56	44.63	—
Jul 2025	66.25	71.94	47.84	—
Ago 2025	56.14	75.49	49.00	—
Sep 2025	53.18	79.22	47.54	—
Oct 2025	56.45	83.14	49.34	—
Nov 2025	62.21	87.24	53.71	—
Dic 2025	50.68	91.55	48.29	—

Tabla 3: Comparación de pronósticos 2025 con valores reales observados (enero–abril)

Con el fin de identificar el mejor modelo desde la perspectiva de los pronósticos realizados, se realiza un análisis comparativo de los modelos VAR, MEC y VEC, empleando el Error Cuadrático Medio (RMSE) como criterio de evaluación de precisión predictiva. Este indicador, ampliamente utilizado, mide la magnitud promedio del error entre los valores pronosticados y los valores reales, penalizando más fuertemente los errores grandes. En este caso, se calculó el RMSE para los primeros cuatro meses de 2025, periodo para el cual se dispone de datos observados de la tasa de ocupación hotelera. Los resultados indican que el modelo VEC presentó el menor RMSE (1.82), lo que evidencia su alta capacidad para capturar la dinámica conjunta de largo y corto plazo entre las variables del sistema. El modelo VAR mostró un error moderado (3.99), con cierta sobreestimación en los meses iniciales, mientras que el modelo MEC presentó el mayor error (9.74), reflejando una menor precisión en sus pronósticos. Estos resultados respaldan técnicamente la superioridad del modelo VEC en términos de ajuste predictivo y robustez, lo cual es coherente con la presencia de cointegración detectada entre las variables y la especificación estructural del modelo.

3 Resultados y discusión

Criterio	Modelo VAR	Modelo ECM	Modelo VEC
Variables	Logocu, Logvis, Logpib, Logpas, Logipc	Logocu, Logvis, Logpib	Logocu, Logvis, Logpib
Transformación	Dos diferencias	Primeras diferencias + ECT	Diferencias + ECT (r=2)
Rezagos	12	1	12 en niveles (11 en diferencias)
Variable dependiente	$\Delta^2 \text{Logocu}_{ts}$	$\Delta \text{Logocu}_{ts}$	$\Delta \text{Logocu}_{ts}$
ECT	No aplica	Negativo y significativo	ECT1 cercano a significativo (p = 0.0864)
Vbles significativas	Muchas sin significancia individual	ΔLogvis , ΔLogpib y ECT	$\Delta \text{Logocu}_{ts11}$, $\Delta \text{Logpib}_{ts11}$, $\Delta \text{Logocu}_{ts111}$
R-cuadrado ajustado	0.4735	0.6065	0.4611
Prueba de autocorrelación	No pasa (PT. p < 0.0001)	Pasa (BG, p = 0.8083)	No pasa (PT. p < 0.0001)
Heterocedasticidad (ARCH)	No pasa (p = 0.0006)	Pasa (p = 0.6151)	Pasa (p = 0.3838)
Normalidad	No pasa (P < 0.01)	No pasa (KS p = 0.0122)	No pasa (P < 0.01)
Consistencia teórica	No permite evaluar relaciones de largo plazo	Captura equilibrio de largo plazo	Captura relaciones de largo y corto plazo

Figura 4: Comparación de los tres modelos especificados

Aunque el modelo VEC mostró el mejor desempeño en términos de precisión predictiva, es preciso señalar que no cumplió satisfactoriamente con todos los supuestos de diagnóstico estadístico. En particular, el modelo VEC presentó evidencia de autocorrelación y falta de normalidad, lo que puede comprometer la validez de inferencias estadísticas, aunque no necesariamente su capacidad de predicción en el corto plazo. Por su parte, el modelo ECM, a pesar de mostrar el mayor RMSE y un desempeño débil en términos de ajuste, fue el único que cumplió con todos los supuestos de autocorrelación, homocedasticidad y estabilidad estructural, lo que lo hace más sólido desde un punto de vista econométrico, aunque menos competitivo en términos de pronóstico. El modelo VAR mostró un comportamiento intermedio: obtuvo un RMSE aceptable y pasó algunas pruebas, pero también evidenció problemas con las pruebas aplicadas.

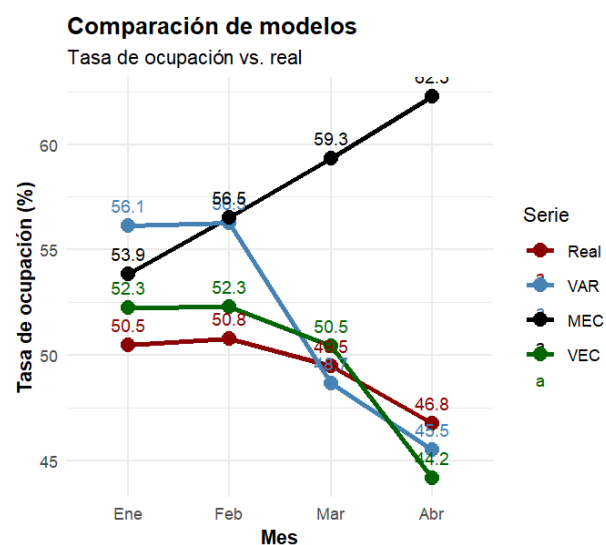


Figura 5: Comparación entre los modelos VAR, ECM y VEC frente a los datos reales de ocupación (enero-abril 2025).

Los resultados de la comparación desarrollada a lo largo del presente documento muestran que el modelo VEC

(Vector de Corrección de Errores) presentó el mejor desempeño predictivo para estimar la tasa de ocupación hotelera en Colombia durante los primeros meses de 2025, tanto en términos gráficos como en el menor error cuadrático medio (RMSE). Este resultado refleja la capacidad del VEC para capturar relaciones de largo plazo entre las variables cointegradas. No obstante, el modelo no superó todas las pruebas de diagnóstico (autocorrelación y normalidad), lo que sugiere limitaciones en su especificación dinámica. Por su parte, el modelo VAR mostró un rendimiento aceptable en los primeros dos meses, pero perdió precisión a medida que avanzó el horizonte de pronóstico. En contraste, el modelo de corrección de errores (MEC), aunque metodológicamente sólido y estadísticamente más robusto en las pruebas de diagnóstico, sobreestimó sistemáticamente la ocupación, lo que lo ubicó como el menos preciso en términos predictivos. En conjunto, estos hallazgos sugieren que incorporar relaciones de cointegración mejora la capacidad predictiva de los modelos, pero también es fundamental considerar la validez de los supuestos estadísticos para asegurar su aplicabilidad práctica.

4 Conclusiones

El estudio permitió analizar la dinámica temporal de la tasa de ocupación hotelera en Colombia entre 2012 y 2024, observando comportamientos estacionales y efectos coyunturales asociados a variables como el PIB, los visitantes no residentes y los pasajeros nacionales. Las series mostraron tendencias compartidas y relaciones de largo plazo, lo cual justificó el uso de modelos con estructura de corrección de errores. A través de pruebas de raíz unitaria y cointegración, se identificaron las propiedades estocásticas necesarias para especificar y estimar correctamente modelos VAR, MEC y VEC. Además, se aplicaron criterios econométricos rigurosos para validar la robustez de cada modelo, incluyendo pruebas de autocorrelación, heterocedasticidad, normalidad y estabilidad.

Se estimaron los tres modelos (VAR, MEC y VEC) y se utilizaron para proyectar la tasa de ocupación hotelera durante el primer cuatrimestre de 2025. El modelo VEC logró el mejor desempeño en términos de precisión predictiva (menor RMSE), gracias a su capacidad de incorporar relaciones de cointegración entre las variables. Sin embargo, este modelo no cumplió con todos los supuestos estadísticos en las pruebas de diagnóstico.

Los resultados de esta investigación ofrecen un marco útil para la modelación de series turísticas, demostrando que la inclusión de relaciones de largo plazo mejora la calidad de los pronósticos. Sin embargo, también se evidencia que la validez de los modelos no puede medirse exclusivamente por su capacidad predictiva, sino también por el cumplimiento de los supuestos estadísticos que aseguran su confiabilidad. Entre las principales limitaciones se encuentra la necesidad de considerar otras variables potencialmente relevantes, así como, la incorporación de choques exógenos como eventos climáticos, reformas políticas o conflictos sociales, que pueden afectar el turismo. Para investigaciones futuras, se sugiere ampliar el horizonte temporal e incorporar variables estructurales adicionales. Finalmente, la generación de pronósticos más confiables puede contribuir a la planificación estratégica del sector turístico colombiano, en especial en el contexto de reactivación post-pandemia y cambios en la demanda global.

Conflicto de intereses

Se precisa que no existe conflicto de interés alguno en relación con los datos recopilados. Estos fueron obtenidos de fuentes como el Ministerio de Industria, Comercio Y Turismo y el Departamento Administrativo Nacional de Estadística. De igual forma, durante el uso de los datos se siguieron los protocolos correspondientes para garantizar la integridad de la información y del estudio realizado.

Referencias

- Brida, J. G., Pereyra, J. S., Devesa, M. J. S., & Aguirre, S. Z. (2008). La contribución del turismo al crecimiento económico. *Cuadernos de turismo*, (22), 35-46.
- Brida, J. G., Rodríguez Brindis, M. A., & Mejía-Alzate, M. L. (2020). La contribución del turismo al crecimiento económico de la ciudad de Medellín – Colombia [Recibido: 10 de abril de 2020 – Aceptado: 29 de agosto de 2020]. *Revista de Economía del Rosario*. <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.8926>
- Dickey, D. A., & 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.
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251-276.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231-254.
- Phillips, P. C. B., & Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58(1), 165-193.
- Voronkova, L. P. (2019). Tendencias del turismo en América Latina [Disponible en: <https://iberoamericajournal.ru/sites/default/files/2019/2/voronkova.pdf>]. *Iberoamérica*, (2).