

Análisis dinámico de series temporales multivariadas

Dynamic analysis of multivariate time series

^aHenry de Jesús Gallardo-Pérez, ^bMawency Vergel-Ortega, ^cJhan Piero Rojas-Suárez

 ^a Doctor en Educación, henrygallardo@ufps.edu.co, Universidad Francisco de Paula Santander, Cúcuta, Colombia

 ^b Doctora en Educación, mawencyvergel@ufps.edu.co, Universidad Francisco de Paula Santander, Cúcuta, Colombia

 ^c Magister en Ingeniería Civil, jhanpiero Rojas@ufps.edu.co, Universidad Francisco de Paula Santander, Cúcuta, Colombia

Recibido: Febrero 22 de 2020 **Aceptado:** Mayo 20 de 2020

Forma de citar: H.J. Gallardo-Pérez, M. Vergel-Ortega, J.P. Rojas-Suárez, “Análisis dinámico de series temporales multivariadas”, *Mundo Fesc*, vol. 10, no. 20, pp. 41-49, 2020

Resumen

El estudio de los sistemas dinámicos es un tema de gran interés en las ciencias básicas, económicas, empresariales, entre otras, permite realizar inferencia directa del comportamiento de los diferentes sistemas. El objetivo del trabajo consiste en la aplicación de modelos estadísticos de series temporales a la estimación empírica de ecuaciones que conforman sistemas dinámicos con el propósito de describir las relaciones entre factores capaces de producir alteraciones en un sistema económico, físico, social o ambiental y plantear ecuaciones de evolución para ese sistema. La metodología se enmarca en el paradigma cuantitativo y utiliza un razonamiento deductivo que, a partir del análisis de datos univariados y multivariados, recolectados a intervalos regulares de tiempo, permite establecer un derrotero para ajustar modelos dinámicos confiables que expliquen el comportamiento de variables aleatorias a lo largo del tiempo. Las aplicaciones revisten importancia para las diferentes ciencias puesto que permiten identificar y estimar modelos que describan sistemas dinámicos en variables estocásticas relacionadas.

Palabras clave: Serie temporal, modelo estadístico, sistema dinámico.

Abstract

The study of dynamic systems is a topic of great interest in basic sciences, economics, business, among others, allows direct inference of the behavior of different systems. The objective of this work is the application of statistical models of time series to the empirical estimation of equations that conform dynamic systems with the purpose of describing the relations between factors capable of producing alterations in an economic, physical, social or environmental system and to propose equations of evolution for that system. The methodology is framed in the quantitative paradigm and uses a deductive reasoning that, from the analysis of univariate and multivariate data, collected at regular intervals of time, allows establishing a path to adjust reliable dynamic models that explain the behavior of random variables over time. The applications are important for the different sciences since they allow identifying and estimating models that describe dynamic systems in stochastic related variables.

Keywords: Time series, statistical model, dynamic system.

Autor para correspondencia:

*Correo electrónico: henrygallardo@ufps.edu.co



Introducción

Un sistema está conformado por un conjunto de elementos en interacción; los sistemas dinámicos referidos en este trabajo son aquellos cuyos elementos están cuantificados por variables estocásticas que se observan a intervalos regulares de tiempo. El objetivo central del trabajo se enfoca en la aplicación de modelos estadísticos de series temporales a la estimación empírica de ecuaciones que describen sistemas dinámicos. Los fundamentos se basan en los modelos dinámicos y los modelos de series temporales, puesto que estos se utilizan en la construcción de modelos que describen los sistemas dinámicos.

En este trabajo se aplica el método de Engle-Graner [1] a la identificación de la cointegración de series temporales no estacionarias y se realizan procedimientos de estimación de modelos que relacionan las variables estocásticas que conforman esos sistemas. Se presentan dos aplicaciones, ambas con series no estacionarias, pero en una de ellas además se presenta efecto estacional. Se realiza estimación de los modelos univariados para cada caso, evaluación de la presencia de cointegración y estimación del modelo bivariado.

Modelos dinámicos

Un modelo constituye una representación abstracta de un aspecto de la realidad. En su estructura intervienen los elementos que caracterizan la realidad y las relaciones existentes entre ellos. En particular, un modelo matemático, estático o dinámico, determinista o estocástico, está basado en la lógica matemática, sus elementos son variables y funciones y las relaciones matemáticas entre ellos están dadas por ecuaciones, inecuaciones y operadores lógicos [2].

En un modelo estático, la variable tiempo no desempeña un papel relevante. Por el contrario, en un modelo dinámico, los elementos que intervienen en la modelización no permanecen invariables, sino que se consideran como funciones del tiempo, describiendo trayectorias de variables estocásticas en el tiempo. Así, el análisis de un modelo dinámico tiene por objeto el estudio de la trayectoria temporal específica de sus elementos dando lugar a la teoría de los sistemas dinámicos, que involucra, entre otros, el análisis de series temporales utilizados frecuentemente en la explicación de fenómenos físicos, económicos y sociales [3].

El estudio de la dinámica está centrado en describir las causas que produce los movimientos y de la evolución del movimiento de un sistema en el tiempo; por tanto, puede entenderse como el estudio de las fuerzas que actúan sobre un cuerpo para producir movimiento y también el estudio de las interacciones en el tiempo de las variables estocásticas asociadas [4]; los sistemas económicos son dinámicos en el tiempo y sus variables presentan un comportamiento estocástico, sin embargo, estas pueden ser endógenas o exógenas según la fuente de la variación con respecto al modelo.

Series temporales

El estudio de series temporales en la explicación de fenómenos dinámicos, bien sean económicos, ambientales, empresariales, permite entre otros, comprender la variabilidad de la serie temporal, identificar los oscilaciones regulares y no regulares, describir las características de esas oscilaciones y comprender los procesos que dan origen a esas oscilaciones.

Una serie temporal se compone de un conjunto de realizaciones de variables aleatorias Z , cada una de ellas observadas en un periodo

de tiempo t , esto es una realización de un proceso estocástico en tiempo discreto: $\{Z_t\}_T = \{Z_1, Z_2, Z_3, \dots, Z_t, \dots\}$. Al valor observado del vector de variables aleatorias en el periodo t se le nota Z_t ; en cada periodo de tiempo se observa una sola realización de las variables aleatorias. Se asume que hay equiespaciamento entre las observaciones y que éstas corresponden a puntos discretos en el tiempo, así que los datos recolectados corresponden a sucesiones finitas de realizaciones de variables estocásticas [5-11]. Luego, una serie temporal consiste en una secuencia de observaciones ordenadas y equidistantes cronológicamente sobre una característica (serie univariante) o sobre varias características (serie multivariante) de una unidad observable en diferentes momentos de tiempo [12].

Existen diferentes métodos para el análisis de series temporales [13], sin embargo, en esta investigación se trabajará con la metodología de Box y Jenkins [14] que propone el modelo de Promedio Móvil Autoregresivo Integrado, ARIMA, explica el valor de la serie en función de la combinación de dos polinomios: el polinomio autoregresivo, AR, y el promedio móvil, MA (Ecuación 1); el ajuste del modelo ARIMA se realiza con base en la exploración de las funciones de autocorrelación y de autocorrelación parcial que se obtienen una vez ajustadas a una forma estacionaria a través de diferenciaciones simples o estacionales para estabilizar la media y transformaciones para estabilizar la varianza [12].

$$Z_t = c\mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

El propósito del análisis de series temporales consiste en elaborar un modelo estadístico que describa adecuadamente la serie de forma tal que las implicaciones teóricas del modelo sean compatibles con las pautas muestrales observadas en la serie temporal [14]. Para la estimación del modelo se asume

que la serie temporal es una realización particular finita de un proceso estocástico. Una vez se han estimado los parámetros y validado el modelo, este se utiliza para describir el comportamiento de la serie en el tiempo, pronosticar su evolución futura y contrastar teorías sobre las características o variables que componen la serie mediante evidencia empírica [13].

Un proceso estocástico es una secuencia de variables aleatorias, ordenadas, equidistantes cronológicamente de una unidad observable en diferentes momentos del tiempo. El proceso estocástico $\{Z_t\}_T$ es estacionario cuando las propiedades estadísticas de cualquier secuencia finita de componentes, $\{Z_{t_1}, Z_{t_2}, Z_{t_3}, \dots, Z_{t_n}\}, n \geq 1$, son semejantes a las de la secuencia en otro periodo desfasado en el tiempo $\{Z_{t_1+h}, Z_{t_2+h}, Z_{t_3+h}, \dots, Z_{t_n+h}\}, h = \pm 1, \pm 2, \dots$. El proceso no es estacionario cuando las propiedades estadísticas de al menos una secuencia finita de componentes son diferentes de las de la secuencia en al menos un periodo diferente en el tiempo. En un proceso estacionario se cumple que tanto la media como la varianza y la covarianza de la serie no están afectadas por cambios a lo largo del tiempo, esto es, no son funciones del tiempo. La estacionariedad de un proceso se determina por el test de raíz unitaria, uno de los más utilizados es el test de Dickey Fuller. La serie $\{Z_t\}_T$ es integrada de orden d , $Z_t \sim I(d)$, si $\Delta^d Z_t$ es estacionaria, es decir, la d -ésima diferencia del proceso es estacionaria, $I(0)$. La modelación de procesos estocásticos está dada por un conjunto de hipótesis bien definidas sobre las propiedades estadísticas del proceso [15].

El análisis de series temporales multivariadas implica establecer la relación funcional entre ellas para luego proceder a estimar una en función de otras explicativas, lo cual requiere verificación de existencia de cointegración entre ellas. Una vez cumplidas estas fases, se procede a estimar

los parámetros de modelos matemáticos que permitan realizar pronósticos confiables y con validez intertemporal de los valores de la serie explicada.

La noción de cointegración implica que dos o más series están cointegradas si se mueven conjuntamente a lo largo del tiempo y las diferencias entre ellas son estables, es decir, estacionarias, aun cuando cada serie en particular contenga una tendencia estocástica y sea por lo tanto no estacionaria. Esto implica que dos o más series temporales que son no estacionarias de orden I (1) están cointegradas si existe una combinación lineal de esas series que sea estacionaria o de orden I (0). El vector de coeficientes que se deriva de esta serie estacionaria es el vector cointegrante.

El procedimiento de cointegración de Engel-Granger [1] es un método en dos etapas basado en los residuos estimados, asume a priori que existe un solo vector de cointegración en el modelo y el resultado puede variar dependiendo de la variable que se seleccione como dependiente. El procedimiento parte de determinar el orden de integración de cada una de las variables a ser incluidas en el modelo, a continuación, se especifica la relación funcional a largo plazo y se contrasta si los residuos tienen raíz unitaria o no y se realiza prueba de cointegración a los residuos estimados y se estima el modelo de corrección de errores si las variables están cointegradas [16].

Materiales y métodos

La investigación se enmarca en el paradigma cuantitativo y utiliza el razonamiento deductivo que permite la aplicación del método de Box y Jenkins [13] al análisis de series temporales y el método de Engle-Granger [1] a la identificación de la cointegración de series temporales no estacionarias.

Los valores de las realizaciones de las variables aleatorias en cada instante de tiempo son exógenos al trabajo realizado y provienen de fuentes secundarias, ya que el investigador no tiene posibilidad de modificarlos, solamente de observarlos o de registrarlos. En este trabajo se realizan dos aplicaciones, una afín con el valor del producto interno bruto del departamento Norte de Santander y su relación con el consumo de energía eléctrica no residencial durante los años 2005 a 2018 y la otra con la precipitación pluvial y temperatura promedio mensual en la ciudad de Cúcuta durante los años 2010 a 2019.

El procedimiento se realiza en dos fases. La primera implica la identificación del modelo de series temporales que subyace en los datos recopilados, para ello se utiliza la metodología de Box y Jenkins [13] mediante el análisis de las funciones de autocorrelación y autocorrelación parcial de la serie una vez establecida su estacionariedad [17]. En segundo lugar, se identifica la presencia de autocorrelación entre las dos series analizadas mediante el procedimiento de Engle-Granger [1], el cual implica diferenciar las series hasta lograr estacionariedad, verificar que las dos tengan el mismo nivel de integración. Se establece entonces el modelo funcional lineal que establece la relación funcional entre las dos variables analizadas y finalmente se realiza prueba de raíz unitaria para establecer que la serie de residuales del modelo estimado es también no estacionaria. Cumplidos estos pasos, se procede a identificar la cointegración entre las series y evaluar la funcionalidad del modelo.

Resultados y discusión

El trabajo desarrollado se presenta en dos aplicaciones de los modelos de series temporales: en la primera, se estima un modelo para la serie multivariada compuesta

por el valor del producto interno bruto del departamento Norte de Santander y el consumo de energía eléctrica no residencial, para los años 2015 a 2018 y en la segunda, se muestra aplicación del análisis a la serie multivariada constituida por la precipitación pluvial mensual y la temperatura promedio mensual en la ciudad de Cúcuta durante los años 2010 a 2019.

Producto Interno Bruto y Consumo no residencial de Energía Eléctrica en Norte de Santander

En el departamento Norte de Santander se registra la producción anual mediante la variable económica definida por el valor del producto interno bruto anual, en billones de pesos, a precios constantes de 2015 [18]; la dinámica temporal de esta variable se pretende estimar a partir del conocimiento del valor de la demanda de energía eléctrica no residencial [19-20], esto es, la demanda de energía eléctrica para los sectores industrial, comercial y oficial en el mismo periodo de tiempo. En la Figura 1 se presenta la variación temporal de las dos series.

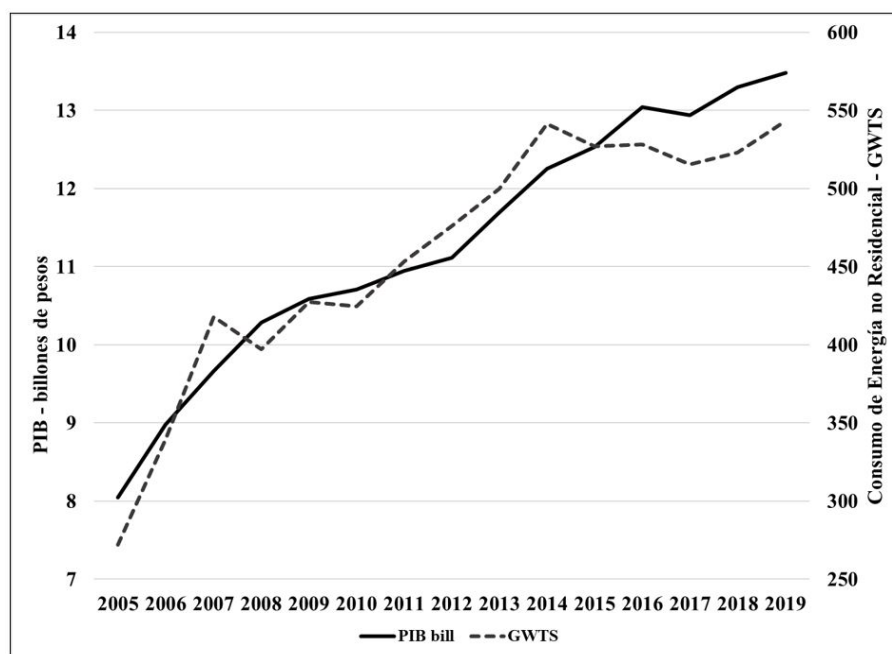


Figura 1. Producto Interno Bruto y Consumo de Energía no Residencial en Norte de Santander

Se aprecia en Figura 1 que las series son no estacionarias ya que presentan una tendencia creciente. Las dos series son tipo integrado de orden 1, $I(1)$. Se requiere entonces de una diferenciación para estabilizar nivel y establecer el grado de integración. La diferenciación de orden 1 estabiliza el nivel de la serie. La prueba de raíz unitaria permite concluir que las dos series diferenciadas son estacionarias.

Si se analiza únicamente la serie temporal correspondiente al producto interno bruto

para Norte de Santander se encuentra que, una vez diferenciada, los valores de la serie tienen muy baja autocorrelación, lo cual no permite el ajuste de un modelo ARIMA. La mejor aproximación es desde un modelo ARIMA autorregresivo de orden 1 sobre la base de u na diferencia; el modelo estimado sería del tipo: $Z_t = 0,4 + 1,3 Z_{t-1} + 0,3 Z_{t-2}$. Sin embargo, este modelo es poco confiable para uso en predicciones por cuanto el nivel de significación de los parámetros es bajo, se tiene un p-valor de 0,26. Por tal motivo,

se hace necesario recurrir a otra serie cuyos valores observados se obtengan con mayor prontitud y ajustar un modelo estadístico de regresión, previo establecimiento de una relación de cointegración entre las dos series.

Dado que las dos series presentadas en la Figura 1 están cointegradas, se procede entonces a estimar el modelo de regresión lineal entre las dos variables, se escoge como variable dependiente el producto interno bruto, el cual se pretende estimar desde el conocimiento de la demanda eléctrica no residencial.

El modelo ajustado es el siguiente: $PIB = 2,28 + 0,02 * DEnR$, con un nivel de determinación de 92%. Lo cual indica que variaciones en la demanda de energía no residencial explican el 92% de la variación en el producto interno bruto de la región.

El análisis de residuales del modelo indica que son estacionarios, esto es son $I(0)$, y además tienen varianza constante. La prueba de raíz unitaria permite reafirmar la estacionariedad de los residuales. En consecuencia, puede decirse que las dos series están cointegradas, por tanto, si se tiene el valor del consumo de energía no residencial, es posible estimar con alta confiabilidad el valor de la producción de ese mismo periodo.

Comportamiento térmico ambiental y precipitación pluvial en Cúcuta

Para la ciudad de Cúcuta, capital del departamento Norte de Santander, Colombia, se presenta en Figura 2 la precipitación mensual y la temperatura mensual promedio durante los años 2010 a 2019, registrados en la estación meteorológica del aeropuerto Camilo Daza de la ciudad.

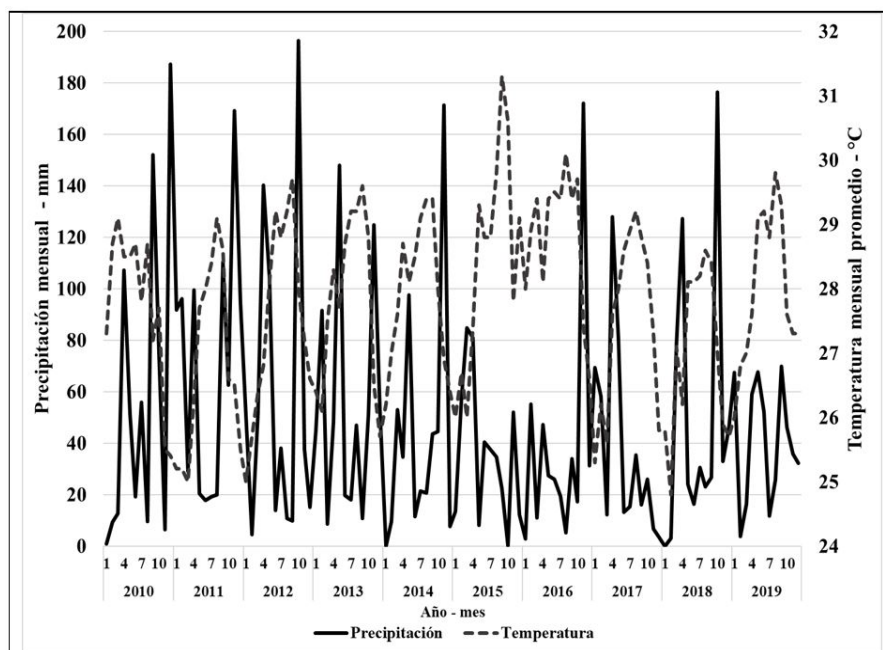


Figura 2. Precipitación mensual y temperatura mensual promedio en Cúcuta

Se observa el comportamiento de las dos series en la Figura 2, nótese que existe

correspondencia entre las dos series en el sentido de que por lo general en los meses con temperatura promedio alta, se encuentra baja precipitación y viceversa.

Por otra parte, las pruebas estadísticas indican la presencia de no estacionariedad en las dos series. Sin embargo, la prueba de raíz unitaria indica que ambas series en su primera diferencia y primera diferencia estacional con retardo 12 son estacionarias y por tanto están cointegradas; más aún, existe autocorrelación entre sus valores, luego puede aplicarse el procedimiento de Box y Jenkins para estimar un modelo de serie temporal univariado.

El modelo ajustado para la serie de precipitación mensual es de tipo ARIMA(1,1,0)x(1,2,0)12 y está dado por la ecuación $\hat{W}_t = -0,511 \hat{W}_{t-1} - 0,484 \hat{W}_{t-12}$, en la cual, W_t está definido en función de Z con el operador de retardo B , así $\hat{W}_t = (1-B^k) Z_t = Z_t - Z_{t-k}$ [5]. Por ello el modelo ajustado en función de la variable original, precipitación, está dado por: $Z_t = 0,488 Z_{t-1} - 0,512 Z_{t-2} - 0,484 Z_{t-12} + 0,484 Z_{t-24}$

El modelo ajustado para la temperatura promedio es de tipo ARIMA(1,1,0)x(1,2,0)12 y está dado por la ecuación $\hat{W}_t = -0,286 \hat{W}_{t-1} - 0,403 \hat{W}_{t-12}$, siguiendo el mismo procedimiento se obtiene el modelo ajustado en función de la variable original, temperatura, está dado por: $Z_t = 0,714 Z_{t-1} - 0,286 Z_{t-2} - 0,403 Z_{t-12} + 0,403 Z_{t-24}$

Las dos series, precipitación mensual y temperatura promedio mensual, en la ciudad de Cúcuta requieren el mismo nivel de diferenciación para obtener las series estacionarias correspondientes, lo cual es un indicador de que las dos series estén cointegradas; para ello se construye un modelo de regresión lineal y se analizan los residuales del modelo. Es de esperarse que, en una misma región durante periodos

temporales similares, la temperatura esté en función de la precipitación, por tanto, se elige la temperatura como variable explicada o dependiente y la precipitación como variable explicativa o independiente.

La ecuación del modelo estimado es: $T_t = 28,2 - 0,01 P_t$, en la que T_t representa la temperatura en el periodo t y P_t la precipitación en ese periodo. La prueba de raíz unitaria indica que los residuales de este modelo son estacionarios. Por tanto, se puede concluir que las dos series están cointegradas. Sin embargo, el coeficiente de determinación del modelo es tan solo del 8%. Por tanto, aun cuando hay cointegración entre las dos series temporales, esto es, que tienen comportamiento similar en el tiempo, no es conveniente utilizar una para estimar el comportamiento de la otra pues el error de pronóstico sería bastante grande.

Una de las posibles razones se puede dar por la variabilidad de los datos o bien por la magnitud y velocidad de los vientos en la ciudad, aunado a su cambio, generalmente aleatorio y continuo, de dirección. La incorporación de esta variable será fundamental para la realización de una nueva fase de la investigación; sin embargo, se está a la espera de la obtención de la información pertinente para proceder al análisis de datos y estimación de nuevos modelos

Conclusión

Las variables estocásticas en el tiempo producen diferentes fenómenos que pueden ser estudiados desde la física o la estadística. Los valores observados en una serie de tiempo son irrepetibles y si no se toma el valor o se comete un error en su medición, no hay forma de volver atrás en el tiempo para repetir la observación. El método aquí ilustrado permite la realización de predicciones de la variable de interés con base en la observación

de otras variables. Se debe si, determinar que las series estén cointegradas, pues de lo contrario perdería validez la estimación. La presencia de cointegración supone que una parte de la componente de tendencia de las series es un componente común y existe una combinación lineal de las series que carece de esta característica común. Otra forma de entender la cointegración es que existe un equilibrio a largo plazo entre las series, de manera que las desviaciones de este equilibrio tienden a desaparecer a corto plazo.

La aplicación presentada en este artículo ilustra la posibilidad de encontrar relaciones entre variables aparentemente independientes en realidad. Sin embargo, esta posibilidad se presenta cuando las dos variables están autocorrelacionadas y son no estacionarias, lo cual refleja la presencia de la propiedad de cointegración entre variables estocásticas observadas en los mismos periodos de tiempo.

El propósito fundamental de este trabajo fue la aplicación de métodos estadísticos de series temporales bivariadas a la estimación de relaciones entre las variables con el fin de obtener estimaciones confiables de su comportamiento temporal. Existen otros métodos multivariantes, en los que la propuesta desarrollada aquí puede generalizarse después de cumpliendo con los supuestos de cointegración de las variables y la estimación de un modelo de series de tiempo multivariado, entre ellos se pueden citar la transformación de Fourier o el método de Johansen que es aplicable a sistemas de ecuaciones, a diferencia del aquí tratado cuya mayor aplicación está en modelos uniecuacionales.

La relevancia del trabajo está dada por el hecho de que la metodología desarrollada permite realizar estimaciones de modelos matemáticos para la explicación del

comportamiento de variables estocásticas que varían en el tiempo a partir de datos empíricos, lo cual resulta de gran utilidad tanto en pruebas controladas como en la modelación de sistemas económicos, ambientales, naturales y generales en los que se aplique.

Referencias

- [1] L. Maldonado. El modelamiento matemático en la formación del ingeniero, Bogotá: Ediciones Universidad Central, 2013
- [2] M. Ginovart. “¿Qué pueden ofrecer los modelos basados en agentes vivos en el contexto docente?” *Modelling in Science Education and Learnig*, vol. 8, no. 2, pp. 5-25, 2015
- [3] R. Hibbeler. Mecánica vectorial para ingenieros: dinámica. México: Pearson Educación, 2004
- [4] H. Gallardo, J. Rojas y O. Gallardo. Modelación de Series Temporales en el Sector Productivo del Norte de Santander. Bogotá: ECOE, 2019
- [5] H. Gallardo, M. Vergel y J. Rojas. “Dynamic and sequential update for time series forecasting” *Journal of Physics: Conference Series*, vol. 1587 art. 012016, 2020
- [6] J. Mauricio. Introducción al Análisis de Series Temporales. Madrid: Universidad Complutense de Madrid, 2007
- [7] D. Peña. Análisis de series temporales. Barcelona: Alianza Editorial, 2010
- [8] V. Guerrero. Análisis estadístico de series de tiempo económicas. México: Universidad Autónoma Metropolitana, 2003

- [9] A. Nova. Procesamiento de series de tiempo. México: Fondo de Cultura Económica, 2013
- [10] D. Peña. Estadística Modelos y Métodos. Barcelona: Alianza Editorial, 1990
- [11] J. Hamilton. Time series Analysis. New Jersey: Princeton University Press, 1994
- [12] J. Abril. “Análisis de la evolución de las técnicas de series tiempo. Un enfoque unificado” *Estadística*, vol. 63, no. 181, pp. 5-56, 2011
- [13] G. Box y G. Jenkins. Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day, 1969
- [14] H. Gallardo, O. Gallardo y J. Rojas. “Estimation of models and cycles in time series applying fractal geometry” *Journal of Physics: Conference Series*, vol. 1329, art. 012018, 2019
- [15] P. Brockwell y R. Davis. Introduction to Time Series and Forecasting. New York: Springer, 2002
- [16] R. Engle y W. Granger. “Cointegration and error correction representation, estimation and testing” *Econometrica*, vol. 55, pp. 251-276, 1987
- [17] J. Rosel, P. Jara y J. Oliver. “Cointegración en series temporales multivariadas” *Psicothema*, vol. 11, no. 2, pp. 409-419, 1999
- [18] M. Nores y M. Díaz, “Construcción de modelos GEE para variables con distribución simétrica” *Revista de la Sociedad Argentina de Estadística*, vol. 9, pp. 43-63, 2005
- [19] Departamento Administrativo Nacional de Estadística. Cuentas Nacionales. Bogotá: Departamento Administrativo Nacional de Estadística, 2020
- [20] Unidad de Planeación Minero-Energética. Sistema de Información Minero Energético Colombiano. Bogotá: Ministerio de Minas y Energía, 2019