




CC BY-NC 4.0

6

Análisis de anomalías de mercado usando Machine Learning: un caso de la Bolsa Mexicana de Valores, BMV, 2000-2020

Market anomalies analysis using Machine Learning: a case of Mexican Stock Market, MSM, 2000-2020

<http://doi.org/10.5281/zenodo.7415936> 

Recibido

2022/09/25

Aceptado

2022/10/15

Publicado

2022/12/05

ISSN electrónico

2631-2689

ISSN en línea

2953-6529

Carlos Omar Chávez Sánchez

Estudiante de doctorado en Ciencias de Administración FCA, UNAM. carcha@fca.unam.mx. <https://orcid.org/0000-0001-9057-868X>

Arturo Morales Castro

Profesor de tiempo completo en Posgrado de Facultad de Contaduría y Administración, UNAM, amorales@fca.unam.mx. <https://orcid.org/0000-0002-3159-5057>

Oswaldo García Salgado

Profesor de tiempo completo en la Universidad Autónoma del Estado de México UAE Mex, ogarcia@uaemex.mx. <https://orcid.org/0000-0002-8584-1006>

72

Línea de investigación

Ciencias económicas y empresariales.

Referencia

Chávez, C., Morales, A., & García, O. (2022, diciembre). Análisis de anomalías de mercado usando Machine Learning: un caso de la Bolsa Mexicana de Valores, BMV, 2000-2020. *Un Espacio Para la Ciencia*, 5(1), 72-82.

Reference

Chávez, C., Morales, A., & García, O. (2022, December). Market anomalies analysis using Machine Learning: a case of Mexican Stock Market, MSM, 2000-2020. *Un Espacio Para la Ciencia*, 5(1), 72-82.

Citación en el texto / In-Text Citation

Chávez et al. (2022)

(Chávez et al., 2022)

Código JEL:

C14, C45, C51, E37

Resumen

El presente trabajo tiene la finalidad de analizar la presencia de anomalías en el mercado de valores mexicano; particularmente, se concentra en determinar empíricamente el desempeño de un conjunto reducido: *momentum*, volatilidad, reversión a la media y efecto enero, fin de mes y fin de semana. Se proponen nuevos criterios de cálculo de las anomalías a partir de los movimientos de los precios con el objetivo de revisar sus efectos en los rendimientos, en contraste con el mercado para el periodo de 2000 a 2020. El estudio muestra los resultados en términos de rendimientos a partir de la identificación de una anomalía comparados con los promedios del principal índice del país: el S&P IPC (Índice de Precios y Cotizaciones), y es la serie de precios a analizar. Se utilizan metodologías de machine learning (Regresión logística, Perceptrón multicapa [MLP], Máquinas de soporte vectorial optimizada [SMO] y Método Logit) para analizar los modelos y los resultados se evalúan con datos dentro y fuera de la muestra (validación cruzada y partición de dos tercios para entrenar y restante para probar). Los resultados muestran que el *momentum* tiene mayor presencia, dado el comportamiento del índice, y reversión a la media y volatilidad son menos ocurrentes; el efecto enero presenta porcentajes ligeramente inferiores a los del *momentum*.

Palabras clave: Anomalías de mercado, Bolsa Mexicana de Valores, Reversión a la media en precios, Reversión a la media en volatilidad de precios, aprendizaje automático, predicción de rendimientos.

Abstract

The purpose of this paper is to analyze the presence of anomalies in the Mexican stock market; in particular, it focuses on empirically determining the performance of a reduced set: *momentum*, volatility, mean reversion, and January, month-end, and weekend effects. New criteria for calculating anomalies based on price movements are proposed with the aim of reviewing their effects on returns, in contrast to the market for the period from 2000 to 2020. The study shows the results in terms of returns from the identification of an anomaly in comparison to the averages of the country's main index: the S&P IPC (Index of Prices and Quotations), and it is the series of prices to be analyzed. Machine learning methodologies (Logistic Regression, Multilayer Perceptron (MLP), Support Vector Optimized Machines (SMO) and Logit Method) are used to analyze the models and the results are evaluated with data inside and outside the sample (cross-validation and partition). two-thirds to train and remaining to test). The results show that *momentum* has a greater presence, given the behavior of the index, and mean reversion and volatility are less frequent; the January effect presents percentages slightly lower than those of *momentum*.

Keywords: Anomalies, Mexican stock market, Mean reversion in prices, Mean reversion in prices volatility, machine learning, return prediction.

Introducción

Uno de los temas que sigue siendo de gran interés en el área financiera se relaciona con la conducta de los precios y sus rendimientos. Los enfoques de análisis para la determinación de los rendimientos se basan, comúnmente, en aquellos del área económico-financieras, como: las series de tiempo, donde se utilizan los históricos de precios de un único activo; análisis de datos tipo panel, para analizar el impacto de otras variables respecto a un activo; así como la estadística y la probabilidad. Los análisis son orientados a la estimación de precios o rendimientos o incluso para la categorización de estos. Por su lado, en el área de visión por computadora, de las ciencias de la computación, muchos de los problemas de segmentación y extracción de elementos de las escenas se resuelven programáticamente a partir de algoritmos creados de la observación del comportamiento de los datos, y es como se propone un cálculo para las anomalías de mercado. En este trabajo se proponen criterios con reglas creadas de la observación de los precios de las acciones para la identificación de anomalías de mercado: reversión a la media en precio y volatilidad, *momentum* y de calendario, analizando su presencia en la Bolsa Mexicana de Valores (BMV). Se emplean clasificadores de aprendizaje automático para evaluar el nivel de ajuste de los modelos creados con los rendimientos y las señales de anomalías propuestas. El criterio final para evaluación de las anomalías corresponde al máximo porcentaje de exactitud de la clasificación del modelo y el mayor rendimiento ajustado obtenido.

Objetivo

Mostrar el desempeño de las anomalías *momentum*, reversión a la media en precio y volatilidad y el efecto enero en el mercado mexicano usando clasificadores de aprendizaje automático.

Antecedentes

Actualmente existen en la literatura diversidad de paradigmas para el análisis de datos traídos de áreas de la matemática, física o incluso de las ciencias de la computación, como aprendizaje automático y la ya conocida inteligencia artificial (Hinojosa et al., 2021; Montalván et al., 2018; Rho et al., 2021), que han mostrado tener buen desempeño en tareas de predicción y clasificación (Damodaran, 2012; García et al., 2013; Semaan et al., 2014; Wong, 2020; Chopra & Sharma, 2021). Si bien, en el área financiera existen cuerpos teóricos fuertes cuyos principios indican la complejidad de la predicción de precios de los activos (Fama, 1970; Fama & French, 1992; Barberis & Thaler, 2022; Ansah et al., 2022), hoy en día se han sumado otros no menos relevantes que se orientan al análisis de eventos esporádicos o de baja ocurrencia, como la teoría de mercados ineficientes (Ashadi et al., 2012), y aquella que integra variables referentes a la conducta de los participantes de mercado, la inteligencia artificial y el

cómputo evolutivo intentando tener mejores aproximaciones (Ferreira & Gandomi, 2021; Weiwei, 2021; Ansah et al., 2022), más recientemente la teoría de mercados adaptativos (Lo, 2005; Khurssheed et al., 2020; Munir et al., 2022). Las anomalías de mercado son eventos no explicados por modelos como el CAPM (Jensen et al., 2006) o cuerpos teóricos, como la Hipótesis de mercados eficientes, HME (Tversky & Thaler, 1990; Green et al., 2013; Wong, 2020; Woo et al., 2020), han sido ampliamente estudiadas en la literatura con resultados divididos y han sido poco analizadas en el mercado mexicano (Cabello & Ortíz, 2003; Rodríguez & Morales, 2009; Duarte et al., 2013; Diaz-Ruiz et al., 2019).

Metodología

Se analizan las series de precios del S&P IPC correspondientes al periodo comprendido del año 2000 al 2020 de cotizaciones diarias, semanales y mensuales (Yahoo! Finanzas, s.f.). Los rendimientos se calculan con los precios al cierre utilizando la Ecuación 1.

$$R_{log} = \ln \left(\frac{V_t}{V_{t-1}} \right) \quad (\text{Ecuación 1})$$

Donde

R_{log} : Rendimiento logarítmico o compuesto continuamente.

t : Periodo de tiempo.

V_t : Precio al cierre en tiempo t .

El estudio se reduce a un sistema de reconocimiento de patrones para clasificación usando aprendizaje supervisado. Las variables dependientes corresponden a ventanas de rendimientos más una variable booleana *dummy* que indica si la ventana presenta señal de la anomalía (Ecuación 2).

$$C_i = \delta_t + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_i; \quad (\text{Ecuación 2})$$

$$X_t = \frac{1}{N} \sum_{i=1}^N x_{t-i}$$

$$C_i = \{1 \text{ Si } \delta_t=1 \text{ y } Y_t - X_t > 00 \text{ } Y_t - X_t \leq 0, \text{ promedio}$$

Donde

p : Periodos de rendimiento.

i : Número de clases.

C_i : Clase.

α_i : Impacto del rendimiento.

Y_t : Serie de rendimientos.

δ_t : Señal.

n : Periodos del rendimiento objetivo.

t : Periodo, donde t_0 es el periodo actual o de estudio.

ϵ_t : Error.

N : Rendimientos del IPC cuyo promedio representa al mercado.

X_t : Rendimiento promedio del IPC.

x_t : Serie de rendimientos del IPC.

Las clases C_p representan los componentes cuyos rendimientos Y_t superan al mercado X_t . El análisis es realizado para el caso donde los precios están con tendencia o pendiente positiva y para el caso negativo. Es decir, los criterios anteriores se utilizan en caso opuesto para evaluar las anomalías cuando los precios van a la baja, o en términos financieros, son candidatos de una posición corta.

El procedimiento realizado es:

1. Calcular los rendimientos.
2. Obtener las anomalías.
3. Calcular los rendimientos entre el precio de cierre dada la señal de la anomalía al periodo n .
4. Etiquetar el periodo n con 1 si el rendimiento obtenido es mayor al del mercado, de lo contrario asigna 0.
5. Ordenar los conjuntos y evaluar con cada metodología y para cada método de evaluación de desempeño del modelo: validación cruzada y para partición del 66%.
6. Asentar el resultado de la evaluación en una matriz.
7. Repetir el procedimiento incrementando el periodo n para evaluar el número de periodos que consigue mayor rendimiento manteniendo la posición, para $n = 0, 1, 2, 3, \dots, 10$.

El procedimiento completo se realiza como se describe en el diagrama de flujo de la Figura 1.

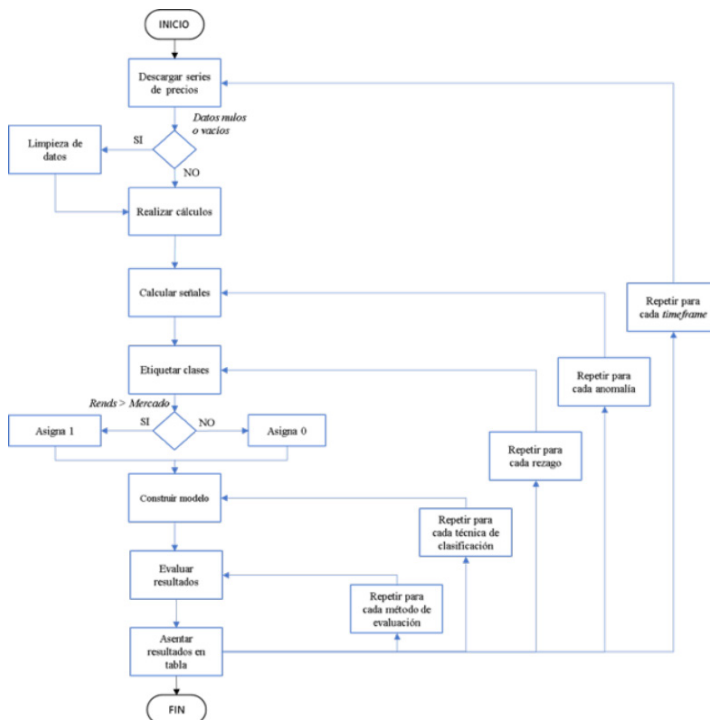
Resultados

En este trabajo las metodologías de aprendizaje automático (*machine learning* en inglés) son usadas para aprovechar las relaciones no lineales que pudieran existir entre los factores analizados y con ello obtener mayor convergencia y ajuste de los modelos. El porcentaje de exactitud se utiliza como el grado de presencia de las anomalías respecto a los rendimientos futuros. Es decir, entre más cercano esté el modelo al 100% significará que tiene mayor ajuste y, por lo tanto, la anomalía explica los rendimientos ajustados. Por el contrario, si presenta porcentajes bajos, se traduce en la degradación de confiabilidad, por lo menos en términos de obtención de rendimientos superiores a los del mercado.

Las gráficas de este apartado muestran y hacen referencia a los resultados obtenidos con el Perceptrón Multicapa, MLP, que fue el que presentó mejores porcentajes de recuperación con validación cruzada.

Figura 1

Diagrama de flujo del procedimiento de creación y ejecución de los modelos



El IPC es analizado para identificar la presencia de las anomalías sobre la misma serie, es decir, se utilizan sus rendimientos promedio en contraste con los rendimientos dados a partir de la señal y hasta diez periodos o rezagos siguientes.

En las Figuras 2, 3, 4 y 5 se observan las gráficas de la evaluación de las anomalías sobre las series de las distintas ventanas de tiempo: diaria (Dy del inglés *day*), semanal (Wk del inglés *week*) y mensual (Mon del inglés *month*).

En la Figura 6 se muestran los rendimientos ajustados promedio del análisis del momentum por tipo de serie (cotizaciones diarias, etiquetada con PROM_D, semanales con PROM_W, y mensuales con PROM_M).

En la serie de rendimientos mensuales en tendencia positiva se aprecia el periodo 9 como máximo; en la de rendimientos semanales y diarios, el máximo es el periodo 10.

En tendencia negativa el rendimiento máximo para cotizaciones mensuales se aprecia en el periodo 5, mientras que para semanales y diarios es el periodo 10, similar al de tendencia positiva.

Figura 2

Gráficas de ajuste del momentum

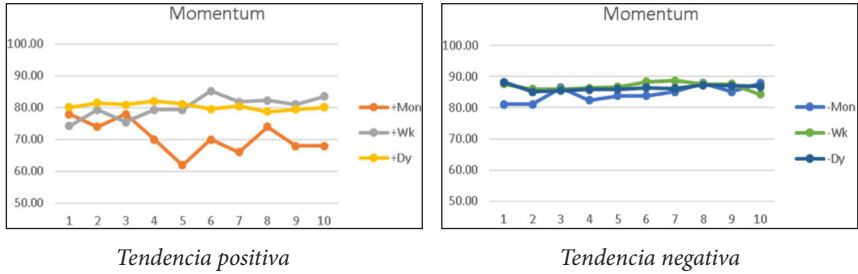


Figura 3

Gráficas de ajuste de la reversión a la media en precio

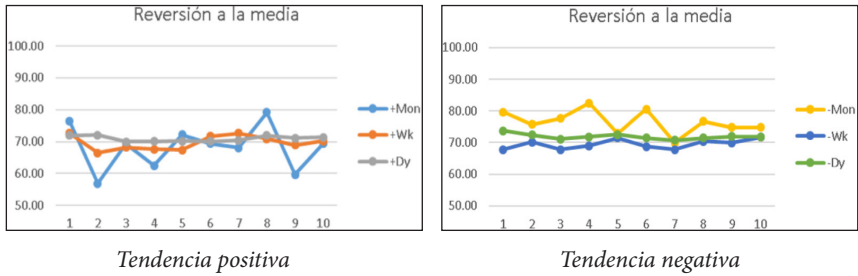


Figura 4

Gráficas de ajuste de la reversión a la media en volatilidad del precio

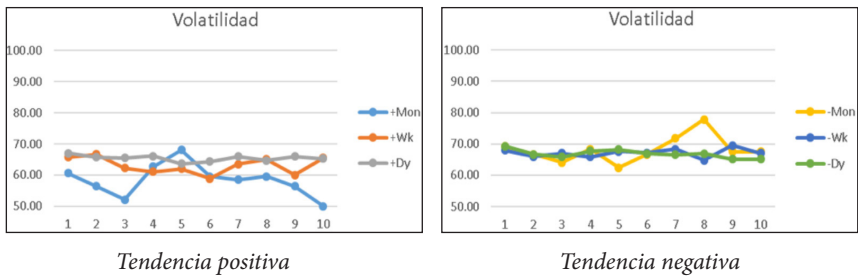


Figura 5

Gráficas de ajuste de las anomalías de calendario: efecto enero, de fin de mes y fin de semana

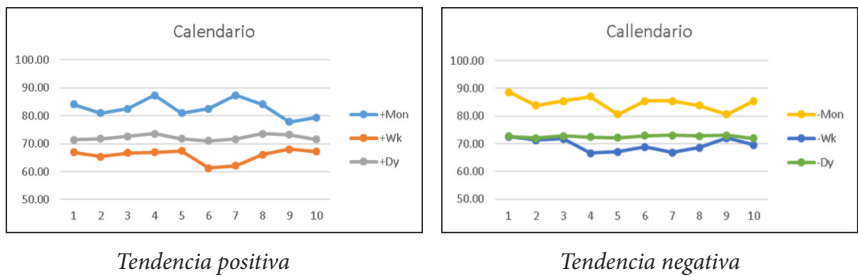
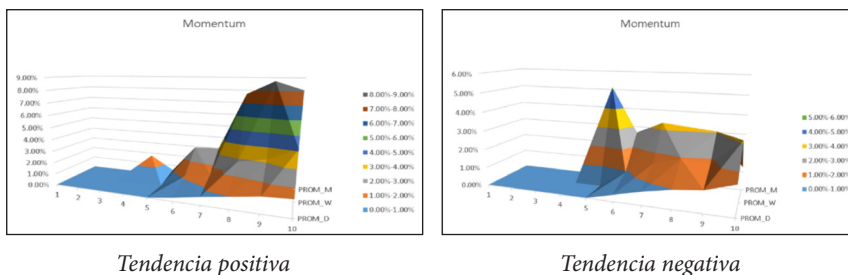


Figura 6

Profundidad de los rendimientos del momentum por tipo de serie



Discusión

De manera general, se puede observar que resaltan los resultados del *momentum* para ambos casos, para precios positivos y negativos, con porcentajes entre el 85% con superiores en los periodos del 4 al 8 (Figura 2). La reversión a la media y volatilidad presentan porcentajes entre 70% y 75%, con poco menos la serie mensual y con superiores entre 2 y 8 periodos (Figura 3 y 4). En las anomalías de calendario el efecto enero, serie de datos mensuales, se encuentran hasta el 90%, superando el 80% y 70% de efecto de fin de semana y fin de mes, serie diaria y serie semanal respectivamente (Figura 5). No se observan porcentajes sobresalientes como para asegurar la explicación de los rendimientos ajustados a un periodo en específico donde el modelo sea determinante.

Cabe mencionar que los resultados para el caso donde la tendencia es negativa, es decir, que los precios al cierre son cada vez menores, los resultados son más uniformes y con menos diferencias de picos entre cada periodo. Esto pudiera deberse a que estos casos son escasos en este activo y los que se presentaron tienen pendiente pronunciada y sin cambios radicales favoreciendo especialmente al *momentum*, que tuvo porcentajes mayores.

Conclusiones

Dada la conducta del IPC resulta notorio que las anomalías relacionadas a la explotación de tendencia pueden ser más beneficiosas, como se ha mostrado con los porcentajes del *momentum*. Por el contrario, las anomalías con detección de cambios en tendencia o relacionadas a las oscilaciones suaves en los precios no son de gran utilidad como se mostró con la obtención de porcentajes de ajuste bajos. Las anomalías de periodos fijos o de estación como el efecto enero, aprovechan parte de la tendencia, similar al *momentum*.

Las anomalías estudiadas comprobaron rechazar la EMH al mostrar presencia en el mercado mexicano:

- momentum tiene presencia media-alta
- Reversión a la media en precio tiene baja presencia, lo mismo que la volatilidad.
- momentum arroja mejores porcentajes en tendencia positiva y media en negativa para las series mensual semanal y diaria, confirma a (Díaz-Ruiz et al., 2020).
- Los períodos identificados de 6 al 10 son los que ofrecen los mejores rendimientos promedio ajustados al riesgo. Utilizando una estrategia de comprar y mantener teniendo en cuenta los periodos resultaría rentable.
- El mercado mexicano cuenta con información histórica limitada, incluso de los proveedores privados. Es complicado probar más empresas en periodos largos.

Como se ha podido observar el mercado mexicano a pesar de ser un mercado nuevo y con registros con poca historia es posible obtener análisis de relevancia para el aprovechamiento de las anomalías. Es posible someterlo para probar otras asimetrías con el objeto de tener y formar mejores indicadores. Este trabajo confirma y extiende análisis anteriores respecto a las anomalías de tipo calendario mostrando su presencia no tan marcada. Sin bien los resultados muestran porcentajes aceptables de presencia estos no son contundentes en la explicación de los rendimientos ajustados futuros.

Referencias

- Ansah, K., Denwar, I. W., & Appati, J. K. (2022). Intelligent Models for Stock Price Prediction: A Comprehensive Review. *Journal of Information Technology Research (JITR)*, 15(1), 1-17. <https://doi.org/jgtj>
- Asadi, S., Hadavandi, E., Mehmanpazir, F., & Nakhostin, M. (2012). Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction. *Knowl. Based Syst.*, 35, 245-258. <https://bit.ly/3CAteXl>
- Barberis, N., & Thaler, R. (2022, September). A Survey of Behavioral Finance. In G.M. Constantinides, M. Harris & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (pp. 294-305). Elsevier. <https://bit.ly/3VmHFOu>
- Cabello, A., & Ortíz, E. (2003). Day of the week and month of the year anomalies in the Mexican stock market. *Revista Mexicana de Economía y Finanzas*, 2(3), 217-241. <https://bit.ly/3gbXuHB>
- Chopra, R., & Sharma, G. D. (2021, November 4). Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review and Research Agenda. *Journal of Risk and Financial Management*, 14(11), 1-34. <https://doi.org/jgk>

- Damodaran, A. (2012, April). *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset*. Wiley Finance.
- Diaz-Ruiz, P., Herrerías, R., & Vasquez, A. (2020, July). Anomalies in Emerging Markets: The Case of Mexico. *The North American Journal of Economics and Finance*, 53, 101188. <https://doi.org/gkx36k>
- Duarte, J., Sierra, K., & Garcés, L. (2013, 2-4 de octubre). *Anomalías de Mercado en la Bolsa de Valores de México* [Ponencia]. XVIII Congreso Internacional de Contaduría, Administración e Informática, México D. F., México. <https://bit.ly/3S1OzWI>
- Fama, E. (1970, May). Efficient Capital Markets: A Review of Theory and Empirical Work. *Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, December, 28-30*, 25(2), 383-417. <https://doi.org/b3kldr>
- Fama, E., & French, K. (1992, June). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), 427-465. <https://doi.org/bpg8>
- Ferreira, F., Gandomi, A., & Cardoso, R. (2021). Artificial Intelligence Applied to Stock Market Trading: A Review. *IEEE Access*, 9, 30898-30917. <https://bit.ly/3EFW6XZ>
- García, M., Jalal, A., Garzón, L., & López, J. (2013, julio-diciembre). Métodos para predecir índices bursátiles. *Ecos de Economía*, 17(37), 51-82. <https://bit.ly/3S6Zezc>
- Green, J., Hand, J. R., & Zhang, F. (2013, September). The supraview of return predictive signals. *Review of Accounting Studies*, 18(3), 692-730. <https://bit.ly/3evwNNU>
- Hinojosa, R., Trejo, L., Hervert-Escoba, L., Hernández-Gress, N., & González, E. (2021, October). *Mexican Stock Return Prediction with Differential Evolution for Hyperparameter Tuning*. Proceedings, Advances in Computational Intelligence, MICAI 2021, Mexico City, Mexico, October 25-30, 2021, pp. 355-368. <https://bit.ly/3PFrqtD>
- Jensen, M., Black, F., & Scholes, M. (2006, June 13). The Capital Asset Pricing Model: Some Empirical Tests. In M. Jensen, *Studies in the theory of capital markets*, (pp. 79-121). Praeger Publishers. <https://bit.ly/3S5IaK1>
- Khursheed, A., Naeem, M., Ahmed, A., & Mustafa, F. (2020, January 30). Adaptive market hypothesis: An empirical analysis of time-varying market efficiency of cryptocurrencies. *Cogent Economics & Finance*, 8(1). <https://doi.org/jgth>
- Lo, A. W. (2005, March 8). Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. *Journal of Investment Consulting*, 7(2), 21-44. <https://bit.ly/3CW44uW>

- Montalván, D., Barrón, R., & Godoy, S., (2018). Generating Trading Strategies in the Mexican Stock Market: A Pattern Recognition Approach. *Research In Computer Science*, 147(12), 107-114. <https://bit.ly/3FNgKo0>
- Munir, A. F., Sukor, M. E. & Shaharuddin, S. S., (2022, January-March). Adaptive Market Hypothesis and Time-varying Contrarian Effect: Evidence from Emerging Stock Markets of South Asia. *SAGE Open journals*, 12(1),1-16. <https://doi.org/jgtn>
- Rho, C., Fernández, R., & Palma, B. (2021, May). *A Sentiment-based Risk Indicator for the Mexican Financial Sector* [Documento de Investigación]. Banco de México. <https://bit.ly/3PMXkV8>
- Rodríguez, D., & Morales, A. (2009). *Los efectos “día de la semana” y “mes del año” para los rendimientos de una muestra de acciones que cotizan en la bolsa mexicana de valores*. XV Congreso Internacional de Contaduría, Administración e Informática. <https://bit.ly/3fYVMt6>
- Semaan, D., Harb, A., & Kassem, A. (2014). Forecasting exchange rates: Artificial neural networks vs regression. *Third International Conference on e-Technologies and Networks for Development (ICeND)*, 156-161. <https://doi.org/jgtp>
- Tversky, A., & Thaler, R. H. (1990, Spring). Anomalies: Preference Reversals. *Journal of Economics Perspectives*, 4(2), 201-211. <https://bit.ly/3MzCIDw>
- Weiwei, J. (2021, December 1). Applications of Deep Learning in Stock Market Prediction: Recent Progress. *Expert Systems with Applications*, 184, 115537. <https://doi.org/gn38jj>
- Wong, W.-K. (2020, February). Editorial Statement and Research Ideas for Efficiency and Anomalies in Stock Markets [Editorial]. *Economies, MDPI Journal*, 8(1). 1-4. <https://bit.ly/3rXCXt9>
- Woo, K.-Y., Mai, C., McAleer, M. & Wong, W.-K. (2020, March 12). Review on Efficiency and Anomalies in Stock Markets. *Economies, MDPI Journal*, 8(20), 1-51. <https://bit.ly/3VrMW7v>
- Yahoo! Finanzas. (s. f.). *Estadísticas financieras en línea*. Recuperadas el 10 de mayo de 2020, de <https://bit.ly/2J9KvTL>