ISSN-e 2683-2690

The Anáhuac Journal

Volume 24 | Number 1 | First Semester 2024

Facultad de Economía y Negocios
Universidad Anáhuac México

Edited by Jaime Humberto Beltrán Godoy

The Anáhuac
Journal

UNIVERSIDAD ANÁHUAC • VINCE IN BONO MALUM • MR
MÉXICO

60 ANIVERSARIO

The Anáhuac Journal es una revista de carácter científico de la Facultad de Economía y Negocios de la Universidad Anáhuac México, que se publica en los meses de enero y julio. Incluye artículos producto de investigación y estudios con resultados originales que tratan problemas de economía y negocios, así como temas vinculados a ellos. El Comité Editorial, con el apoyo de una amplia cartera de árbitros nacionales e internacionales, especializados en los temas que se publican, dictamina anónimamente los trabajos recibidos para evaluar su publicación y el resultado es inapelable. El contenido de los artículos que aparecen en cada número es responsabilidad de los autores y no compromete la opinión de los editores. Los trabajos que se presenten para su publicación deberán ser de carácter científico y ajustarse a los lineamientos que se incluyen al final de la revista.

The Anáhuac Journal

Volumen 24, Número 1, Primer Semestre (enero-junio de 2024), es una publicación semestral editada por la Universidad Anáhuac del Sur, S.C. (conocida como Universidad Anáhuac México) a través de la Facultad de Economía y Negocios. Av. de las Torres núm. 131, Col. Olivar de los Padres, C.P. 01780, Ciudad de México. Tel. +52 55 5628 8800
www.anahuac.mx/mexico

Editor responsable: Dr. Jaime Humberto Beltrán Godoy. Reserva de Derechos al Uso Exclusivo No. 04-2019-011513125600-203, ISSN-e: 2683-2690, ambos otorgados por el Instituto Nacional del Derecho de Autor.

Contacto
Universidad Anáhuac México
Av. Universidad Anáhuac # 46,
Col. Lomas Anáhuac
Huixquilucan, Estado de México
CP 52786 Tel. (55) 5627 0210

Contacto de asistencia
Rebeca del Carmen Soto Manrique
Coordinación Administrativa DGEPD,
Facultad de Economía y Negocios
Tel. (55) 5627 0210 x 7145
rsoto@anahuac.mx

Equipo editorial
Contacto principal
Editor
Dr. Jaime Humberto Beltrán Godoy
Tel. (55) 5627-0210 ext. 7536
jaime.beltrang@anahuac.mx

Soporte técnico
Isidro Agustín Moya López
Dirección de Sistemas y Tecnología,
Facultad de Economía y Negocios
isidro.moya@anahuac.mx

The Anáhuac Journal está indexado por Directorio Latindex, Cabell's, MIAR, Catálogo Latindex v1.0, CLASE, Sherpa Romeo, SciELO México, EBSCO Host, EconLit, RePEc y Google Scholar.

The Anáhuac
Journal

Volume 24 | Number 1 | First Semester 2024

Facultad de Economía y Negocios | Universidad Anáhuac México

Edited by Jaime Humberto Beltrán Godoy

CONTENTS

CONTENIDO

Letter from the Editor

Dear community,

The ideal of Universidad Anáhuac México is to form a humanistic and scientific community dedicated to the pursuit and sharing of truth and the advancement of knowledge through teaching, study, research, dialogue, and innovation. On our university's 60th anniversary, *The Anáhuac Journal* aims to contribute to this noble task by publishing cutting-edge research results that represent the ultimate knowledge aligned with the rigor and highest national and international academic standards.

This issue addresses the interaction between finance and technology, essential tools for building a prosperous and free society, as well as professional fields that should promote genuine human development through work. Artificial intelligence techniques, machine learning, and advanced algorithms that identify sentiments in financial markets enable the discovery of complex patterns and make predictions for a deeper analysis of the complex non-linear interrelationships between variables. In doing so, they help to mitigate risks and maximize the ever-present opportunities in financial markets.

The Anáhuac Journal presents its first issue entirely in English, in this new era, thus increasing its visibility to advance in the internationalization metrics, consistent with the requirements of prestigious indexing services.

We are grateful to Dr. José Antonio Núñez Mora, Guest Editor of this issue. His extensive experience in financial research and his dedication to training several generations of researchers in Mexico contributed to the consolidation of this edition.

In this mid-year volume, we are pleased to present ten interesting articles that contribute to local and global scientific community discussion and reaffirm our commitment to the advancement of the universal body of scientific knowledge.

Dr. Jaime Humberto Beltrán Godoy
Editor
The Anáhuac Journal

UNIVERSIDAD ANÁHUAC MR
VINCE IN BONO MALUM
MÉXICO
60 ANIVERSARIO

ARTICLES

Market Efficiency and Calendar Anomalies Post-COVID: Insights from Bitcoin and Ethereum

Eficiencia del mercado y anomalías de calendario pos-COVID: perspectivas de bitcoin y ethereum

Sonal Sahu

Tecnológico de Monterrey, Campus Guadalajara, Mexico

Received: March 3, 2024.
Approved: May 22, 2024.

Abstract

This study investigates day-of-the-week effects in the digital market, with a focus on Bitcoin and Ethereum, spanning from July 1st, 2020, to December 31st, 2023, in the post-COVID-19 period. Employing parametric and non-parametric tests alongside the GARCH (1,1) model, market dynamics was analized. The findings indicate the presence of a day-of-the-week effect in Ethereum, characterized by notable return variations across different days, while Bitcoin exhibits no discernible calendar anomalies, suggesting enhanced market efficiency. Ethereum's susceptibility to these effects underscores ongoing market complexities. Disparities in calendar anomalies stem from evolving market dynamics, methodological differences, and the speculative nature of cryptocurrency trading. Furthermore, the decentralized and global market complicates the accurate identification of market-wide effects. This study provides experimental findings on day-of-the-week effects in the digital market, facilitating investors in refining trading strategies and risk management. Further research is warranted to explore underlying mechanisms and monitor regulatory and technological developments for investor insights.

Keywords: *Cryptocurrencies, calendar anomalies, GARCH model, trading strategy, ANOVA.*
JEL Classification: *G14, G10, G41.*

Resumen

Este estudio investiga los efectos del día de la semana en el mercado digital, con un enfoque en bitcoin y ethereum, abarcando desde el 1º de julio de 2020 hasta el 31 de diciembre de 2023, en el período posterior al COVID-19. Empleando pruebas paramétricas y no paramétricas junto con el modelo GARCH (1,1), se analizó la dinámica del mercado. Los hallazgos indican un efecto significativo del día de la semana en ethereum, caracterizado por notables variaciones de rendimiento entre diferentes días, mientras que bitcoin no muestra anomalías de calendario discernibles, lo que sugiere una mayor eficiencia del mercado. La susceptibilidad de ethereum a estos efectos subraya las complejidades actuales del mercado. Las disparidades en las anomalías del calendario surgen de la evolución de la dinámica del mercado, las diferencias metodológicas y la naturaleza especulativa del comercio de criptomonedas. Además, el mercado descentralizado y global complica la identificación precisa de los efectos en todo el mercado. Este estudio proporciona evidencia empírica sobre los efectos del día de la semana en el mercado de criptomonedas, lo que facilita a los inversionistas refinar las estrategias comerciales y la gestión de riesgos. Se justifica realizar más investigaciones para explorar los mecanismos subyacentes y monitorear los desarrollos regulatorios y tecnológicos para obtener información de los inversionistas.

Palabras clave: *criptomonedas, anomalías de calendario, modelo GARCH, estrategia de trading, ANOVA.*
Clasificación JEL: *G14, G10, G41.*

1. Introduction

The cryptocurrency market has witnessed remarkable growth, establishing itself as a significant player within the financial landscape. This growth is evident through the soaring market capitalization of cryptocurrencies such as Bitcoin, Ethereum, and Dogecoin, which have surged to unprecedented levels (Stavrova, 2021). Key factors contributing to this surge include the decentralized nature of cryptocurrencies and the innovative blockchain technology underpinning them (Chen et al., 2019). Moreover, the integration of cryptocurrencies with traditional finance has sparked increased interest among investors (Volosovych et al., 2023).

The distinct decentralized structure of the cryptocurrency market, facilitated by blockchain technology, sets it apart from traditional financial markets. This structure enables peer-to-peer transactions without the need for intermediaries like banks (Andolfatto & Martin, 2022). Additionally, rapid technological innovation within the cryptocurrency sphere attracts diverse participants, consequently expanding the market infrastructure (Volosovych et al., 2023).

The regulatory landscape surrounding cryptocurrencies continues to evolve, adding layers of complexity to the market. Governments and regulatory bodies worldwide are increasingly focused on regulating cryptocurrencies to safeguard investor interests and ensure financial stability (Singh, 2021). As Pantielieieva et al. (2021) argue, regulatory scrutiny actively shapes the future adoption of virtual currencies.

Furthermore, various factors influence price movements, volatility, and investor sentiment within the cryptocurrency market. Heightened investor interest has led to increased market activity and trading volumes, with studies emphasizing the significance of comprehending price deviations and capital controls for exploiting arbitrage opportunities (Makarov & Schoar, 2020).

Volatility remains as a defining characteristic of the cryptocurrency market, with studies scrutinizing volatility co-movements among major cryptocurrencies such as Bitcoin and Ether (Katsiampa, 2019). External factors like the COVID-19 pandemic contribute to fluctuations in prices and market sentiment (Washington et al., 2023). Additionally, researchers have explored the influence of news media on virtual currency prices, analyzing the impact of news discourses on market dynamics (Coulter, 2022).

Given the inherent volatility in the cryptocurrency market, understanding and managing associated risks are imperative for investors. While offering the potential for substantial gains, the market also poses risks of significant losses (Zhao & Zhang, 2021). Challenges in forecasting cryptocurrency volatility persist due to market uniqueness and external factors such as the COVID-19 pandemic (Ftiti et al., 2021). Therefore, understanding and modelling cryptocurrency volatility are crucial for informed decision-making, with advanced techniques such as machine learning and GARCH models aiding in forecasting (Joshi & Sharma, 2022).

In traditional financial markets, investors note the day-of-the-week impact, which refers to discernible patterns in stock returns corresponding to specific days of the week, influencing their trading strategies and risk management (Tran, 2023). These patterns, influenced by psychological factors, underscore the intricacies of financial markets, necessitating investors to consider both fundamental and technical analysis (Țilică, 2021).

Investigating the day-of-the-week effect in the digital currency market holds significance amidst increasing investor interest. Recognizing these patterns can empower investors to tailor trading strategies and develop advanced algorithms and risk management strategies (Caporale & Plastun, 2019).

The present study aims to explore the implications of identified day-of-the-week effects for cryptocurrency investors. By understanding how returns and volatility vary across different days, investors can potentially capitalize on favorable market conditions and mitigate risks. Additionally, the study seeks to provide empirical evidence of the day-of-the-week pattern in the cryptocurrency market post-COVID-19, shedding light on evolving market dynamics. Focusing on Bitcoin and Ethereum from July 2020 to December 2023, this paper aims to investigate the day-of-the-week effect in these prominent cryptocurrencies, considering their significance in the market and the period post-COVID-19.

The subsequent sections of this paper are structured as follows: Section 2 reviews the theoretical framework; Section 3 presents the data and methodology; Section 4 analyzes empirical data and discusses findings; and Section 5 provides conclusions.

2. Theoretical Framework

Olivares-Sánchez et al. (2022) assert that market efficiency, a fundamental concept in finance, determines the extent to which asset prices reflect available information. The Efficient Market Hypothesis (EMH) states that asset prices fully integrate available information, rendering consistent outperformance of the market by investors impossible (Harabida et al., 2023). This theory describes three forms of market efficiency: weak efficiency, semi-strong efficiency, and strong efficiency, each defining the extent to which information is incorporated into asset prices (Souza & De França Carvalho, 2023).

Weak efficiency implies that all historical trading information is already incorporated into current equity prices, making achieving excess returns through historical data analysis challenging (Rossi & Gunardi 2018). In semi-strong efficiency, this idea extends to cover all information accessible to the public, suggesting that neither fundamental nor technical analysis can reliably produce outperformance (Liu et al., 2022). In the most stringent form, strong efficiency indicates that all information, regardless of its public or private nature, already factors into asset prices, making it impossible to gain an advantage even with insider information (Apergis, 2022).

Various empirical studies have evaluated the efficiency of traditional financial markets. However, the debate on market efficiency in cryptocurrency markets remains ongoing. Some studies support the weak-form efficiency of cryptocurrency markets, while others emphasize the impact of external factors, such as the pandemic COVID-19, on cryptocurrency market efficiency (Scherf et al., 2022). This ongoing discussion reflects the dynamic nature of cryptocurrency markets, with studies exploring factors like market liquidity, volatility, and the impact of geopolitical events on market efficiency (Fama, 1997).

To address these complexities, the adaptive market hypothesis (AMH) was proposed, which extends beyond the EMH by recognizing the limitations of the assumption of market efficiency and incorporating the role of behavioral biases and bounded rationality in market participants (Rehan & Gül, 2023). The AMH acknowledges that markets can be inefficient at times due to factors like investor sentiment, herding behavior, and information cascades (Okorie & Lin, 2021). By integrating insights from behavioral finance and evolutionary biology, the AMH provides a more nuanced understanding of market dynamics, highlighting the importance of adaptation, learning, and the interplay between rational and irrational behavior in shaping financial markets (Shahid, 2022).

Lo (2004) proposed the Adaptive Markets Hypothesis (AMH), which offers a valuable framework for understanding the dynamics of cryptocurrency markets. In the context of cryptocurrency trading, the presence of adaptive market participants is particularly pronounced. Cryptocurrency markets are characterized by high volatility and rapid price fluctuations, leading to a dynamic environment where market participants continuously adapt their strategies based on changing market conditions. The decentralized nature of cryptocurrencies and the absence of a central authority contribute to the adaptive behavior of market participants, who respond to news, regulatory developments, and technological advancements in real-time (Khuntia & Pattanayak, 2021).

Technological advancements play a significant role in shaping market behavior in cryptocurrency trading. The use of blockchain technology, algorithmic trading, and artificial intelligence has revolutionized the way transactions are conducted and information is processed in cryptocurrency markets (Davidson et al. 2018). These technological innovations have enabled faster execution of trades, increased market transparency, and facilitated the development of sophisticated trading strategies that respond to market signals and trends (Mikhaylov, 2020). However, they have also introduced new challenges related to market manipulation and cybersecurity (Ogunyolu & Adebayo, 2022).

The day-of-the-week effect is observed in capital markets where certain days exhibit distinct patterns in terms of volatility and returns (Luxianto et al., 2020). Researchers and investors have been interested in this effect as it can offer insights into market dynamics and potentially impact trading strategies (Zilca, 2017). Studies have shown that specific days of the week may experience higher or lower levels of market activity and price movements, indicating the day-of-the-week effect in both volatility and return equations (Chaouachi & Dhaou, 2020; Paital & Panda, 2018).

The day-of-the-week effect in the cryptocurrency market has garnered significant attention from researchers exploring anomalies within the realm of digital assets. Studies have demonstrated that specific days of the week may witness fluctuations in market activity and price movements, influencing both volatility and return equations. Caporale and Plastun (2019) conducted a thorough investigation into the day-of-the-week effect in the cryptocurrency market, shedding light on potential patterns and trends in price movements across different trading days. Theiri et al.

(2022) contributed to this area of research by focusing on cryptocurrency liquidity during the Russia-Ukraine war, underscoring the crucial role of market liquidity in comprehending the day-of-the-week effect.

Tosunoğlu et al. (2023) advanced the literature by employing artificial neural networks to analyze the day-of-the-week anomaly in cryptocurrencies, offering insights into the predictability of various currencies. Furthermore, Bae and Kim (2022) explored robust anomaly scores in cryptocurrencies, highlighting the impact of network factors on cryptocurrency returns. Grobys and Junttila (2020) delved into speculation and lottery-like demand in cryptocurrency markets, shedding light on the short-term reversal effects observed in the cross-section of cryptocurrencies. These studies collectively contribute to our comprehension of the day-of-the-week effect and its implications for cryptocurrency markets.

The implications of the day-of-the-week effect for investors and trading strategies in cryptocurrency markets are significant. Understanding how specific days of the week influence market volatility and returns can help investors optimize their trading decisions and risk management strategies (Dangi, 2020). By leveraging insights from the day-of-the-week effect, investors may be able to identify potential opportunities for profit and adjust their trading activities accordingly. Furthermore, the day-of-the-week effect can inform the development of trading algorithms and strategies that incorporate the cyclicality and patterns observed in cryptocurrency market behavior (Miralles-Quirós & Miralles-Quirós, 2022).

In this paper, we also conducted both Parametric, Nonparametric, and OLS Regression models to find the effect of the day of the week on cryptocurrency market. This paper adds to the current literature by applying non-parametric tests alongside parametric tests, making it unique. By addressing the behavioral aspects driving the day-of-the-week effect in virtual currency markets, this paper provides deeper insights into investor sentiment and market dynamics, filling a gap in the existing literature. Additionally, the GARCH (1,1) model is commonly used for studying the day-of-the-week effect in cryptocurrencies. This model has been applied in various financial markets, including cryptocurrencies, to analyze volatility and the impact of specific days of the week on asset returns and market dynamics. Studies have shown that GARCH (1,1) models effectively capture time-varying volatility and examine the day-of-the-week effect in various markets (Katsiampa, 2017; Chu et al., 2017; Aggarwal & Jha, 2023; Ampountolas, 2022; Naimy et al., 2021).

3. Data and Methodology

Utilizing the daily closing prices of Bitcoin and Ethereum sourced from CoinMarketCap (https://coinmarketcap.com/coins/), this study covers the period from July 1st, 2020, to December 31st, 2023, enabling an examination of the post-COVID-19 period's impact.

Different quantitative methods, including both parametric and non-parametric tests, were applied to analyze the data. We used parametric tests such as the conventional regression model with dummy variables and ANOVA. Non-parametric tests like the Mood median test were also employed to address potential biases. Additionally, the Ordinary Least Squares (OLS) regression model with dummy variables and the GARCH (1,1) model were utilized.

The study commenced by applying descriptive statistics to characterize the returns distribution of the various days of the week for Bitcoin and Ethereum. We then used the Jacque-Bera (JB) test statistics and the Anderson-Darling (AD) test statistics to check for normality. Once the normality was conducted, we calculated returns by taking the log difference of consecutive daily closing prices of the cryptocurrencies, as described by Akyildirim et al. (2021). This process is expressed by the following equation:

$$R_n = (In\ CP_n - In\ CP_{(n-1)}) \times 100 \quad (1)$$

where R_n denotes returns on an nth day in percentage; CP_n denotes closing price on an nth day; $CP_{(n-1)}$ denotes closing price on the previous trading day; and *In* is a natural log.

Log returns for Bitcoin and Ethereum were then assessed using the Augmented Dickey-Fuller (ADF), Philips-Perron test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS), to confirm the stationarity of the series. These unit root tests have been utilized in various studies to analyze the stationarity of economic variables, environmental factors, and market indicators. The application of these tests provides insights into the behavior of time series data and aids in identifying trends, patterns, and potential relationships within the data (Ali et al., 2019; Haruna et al., 2022; Dao & Staszewski, 2021).

Following the assessment of stationarity, the study employed a dummy regression model that assumed constant return variance for cryptocurrencies. The equation for the Ordinary Least Squares (OLS) regression model is as follows:

$Return_t = \beta_1 MONDAY_t + \beta_2 TUESDAY_t + \beta_3 WEDNESDAY_t + \beta_4 THURSDAY_t + \beta_5 FRIDAY_t + \beta_6 SATURDAY_t + \beta_7 SUNDAY_t + \varepsilon_t$ (2)

where MONDAY, TUESDAY, WEDNESDAY, THURSDAY, FRIDAY, SATURDAY, and SUNDAY are dummy variables for each day of the week returns (e.g., if the day is Monday, then the dummy variable MONDAY will be 1 and 0 otherwise); β_1, β_2, β_3, β_4, β_5, β_6, and β_7 are coefficients; and ε_t is error term.

To prevent perfect multicollinearity, the intercept term was excluded, and dummy variables for all seven days of the week were included. The coefficients of these seven dummy variables represent the returns for each day of the week.

Following the least square regression analysis, the residuals were examined for autoregressive conditional heteroscedasticity using the ARCH test. If the residuals demonstrated an ARCH effect, indicating volatility clustering, the GARCH (1,1) model was employed. GARCH (1,1) serves as a mathematical framework utilized for both modelling and forecasting volatility in time series data, notably in cryptocurrencies (Kyriazis, 2019). This model is adept at capturing the inherent volatility clustering often observed in financial data, as it enables the modelling of both the mean and the variance of a time series (Kargar, 2021).

Research conducted by Micu and Dumitrescu (2022) further supports the effectiveness of the GARCH (1,1) model, highlighting its superior fit in modelling volatility across major cryptocurrencies. In the GARCH (1,1) model, the variance equation is given by:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (3)$$

Where

α is the coefficient of the lagged squared error term, representing the impact of past volatility shocks on current volatility.

β is the coefficient of the lagged conditional variance term, representing the persistence of volatility.

ω is the constant term representing the long-term average variance.

ε_{t-1}^2 = is the previous period ARCH term

σ_{t-1}^2 = is the previous period GARCH term

In the study, seven dummy variables were incorporated into the GARCH (1,1) model to investigate the day-of-week effect on cryptocurrency market volatility. These dummy variables allowed for assessing how the variance of asset returns changes across different days of the week. By including these dummy variables, any structural changes or anomalies in volatility associated with specific days were captured, providing deeper insights into market dynamics and investor behavior.

4. Analysis and discussion

Analyzing the daily prices of Bitcoin and Ethereum revealed insights into the day-of-the-week effect. Converting the prices of Bitcoin and Ethereum into return series provided data for further examination. Table 1 displays the basic statistics derived from these return series (see Table 1). Ethereum, in particular, stands out with its highest average returns, suggesting greater potential for profitability. The negative skewness observed in both cryptocurrencies indicates left-skewed distributions, implying a likelihood of small profits and minimal potential for significant losses. Ethereum's lower variability in returns compared to Bitcoin is evident from its low coefficient of variation (C.V). Additionally, the Jarque-Bera normality test, consistent with previous research, rejects the null hypothesis of normality for both cryptocurrencies. Interestingly, maximum returns for Bitcoin and Ethereum occur on Tuesdays.

Investigating variations across days of the week, one-way ANOVA and Mood's median tests were conducted. Additionally, the Anderson-Darling test for normality was performed. The p-values for both coins were less than 0.05, indicating rejection of the null hypothesis and non-normality of the data (see Table 2). Scrutinizing the one-way ANOVA results at a 95% confidence level revealed no significant differences in mean returns among days of the week. The Mood's-median test, a robust nonparametric test, was employed to examine median equality for log returns across seven days, as shown in Table 2. No coins yielded significant p-values, indicating no observed day-of-week effects, consistent with Kaiser's (2019) findings.

Table 1. Daily descriptive statistics for Bitcoin and Ethereum Post-COVID period

Bitcoin Returns								
Descriptive statistics	**Monday Returns**	**Tuesday Returns**	**Wednesday Returns**	**Thursday Returns**	**Friday Returns**	**Saturday Returns**	**Sunday Returns**	**Overall Returns**
Mean	-0.098	0.368	0.102	0.539	-0.244	-0.071	0.103	0.100
Maximum	9.314	17.603	8.179	11.966	13.774	11.622	9.148	17.603
Minimum	-10.170	-17.252	-11.533	-14.466	-43.371	-10.886	-8.989	-43.371
Standard Deviation	2.854	4.673	3.383	3.962	5.323	3.751	2.384	3.874
Coefficient of Variation	-29.056	12.682	33.152	7.355	-21.811	-52.933	23.224	38.813
Skewness	-0.311	0.023	-0.550	-0.159	-3.615	0.008	-0.200	-1.429
Kurtosis	4.856	5.242	4.166	4.868	31.138	4.217	6.076	19.683
Jarque-Bera	23.467	30.798	15.748	21.998	5169.707	9.066	58.927	12282.830

Ethereum Returns								
Descriptive statistics	**Monday Returns**	**Tuesday Returns**	**Wednesday Returns**	**Thursday Returns**	**Friday Returns**	**Saturday Returns**	**Sunday Returns**	**Overall Returns**
Mean	0.196	0.511	0.108	0.707	-0.345	0.003	0.471	0.236
Median	0.338	0.558	0.144	0.996	0.071	-0.176	0.647	0.348
Maximum	21.786	21.941	14.499	12.889	15.046	18.123	11.441	21.941
Minimum	-17.727	-18.782	-13.644	-30.520	-56.308	-16.209	-14.822	-56.308
Standard Deviation	4.331	5.943	4.235	5.428	6.827	5.181	3.919	5.209
Coefficient of Variation	22.100	11.626	39.296	7.682	-19.801	1775.665	8.323	22.092
Skewness	0.429	0.075	-0.024	-1.286	-3.911	-0.156	-0.172	-1.464
Kurtosis	8.430	4.886	4.196	9.280	32.559	4.079	5.171	18.420
Jarque-Bera	185.125	21.928	8.771	282.078	5726.383	7.726	29.599	10561.600

Source: Data elaborated by the author based on information gathered from coinmarketcap.com

Table 2. Results of the Parametric and Nonparametric Tests on Bitcoin and Ethereum

The table summarizes results from tests for Normality (Anderson-Darling — a parametric test), Central tendency (Mood's median test — a non-parametric test, One-way ANOVA — a parametric test), and Variance (Levene's and Bartlett's tests).

	Normality Test	**Central tendency Test**		**Variance test**	
	Anderson Darling test P-values	**Mood's median test P-values**	**One Way Anova P-values**	**Bartlett's test P-values**	**Bartlett's test P-values**
Bitcoin	<0.050	0.733	0.676	***0.000***	***0.000***
Ethereum	<0.050	0.674	0.675	***0.000***	***0.001***

Source: Elaborated by the author

Equal variances between days of the week were tested to assess variability and potential day-of-week effects. Bitcoin and Ethereum reject the null hypothesis at 95% confidence, indicating significant differences in variances among days. Both coins, with p-values below 0.05, are further analyzed to explore variance distribution. Table 3 reveals that the maximum variation for Bitcoin and Ethereum occurs on Tuesdays (see Table 3). It is noteworthy that the minimum variation is observed on Sundays. This observation aligns with the findings of Balcilar et al. (2017) and Dorfleitner and Lung (2018), suggesting that many traders abstain from weekend trading, possibly due to leisure activities or other commitments.

Utilizing both parametric and non-parametric tests can detect day-of-the-week effects, but integrating dummy variables into GARCH models presents a more refined approach. This method enables modelling of time-varying volatility patterns, resulting in improved forecasts and deeper insights into the influence of particular days on financial returns and volatility.

We checked the stationarity of the time series data by conducting unit root tests, utilizing the Augmented Dickey-Fuller, and Phillips-Perron tests, which are standard tools in time series analysis (Liao et al., 2021). The results, presented in Table 4 for the ADF test and PP test, consistently showed p-values below 0.05 (see Table 4). The time series data's stationarity was confirmed, and the null hypothesis was rejected at a 95% confidence level due to strong evidence.

Table 3. Results of Levene's test for Variance

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Bitcoin							
Lower Limit	0.747	1.022	0.865	0.934	0.970	0.895	0.644
Upper Limit	1.026	1.403	1.187	1.281	1.330	1.228	0.883
Standard Deviation	0.867	***1.185***	1.003	1.082	1.124	1.038	***0.746***
Ethereum							
Lower Limit	0.775	1.000	0.806	0.937	0.945	0.931	0.735
Upper Limit	1.063	1.372	1.106	1.285	1.296	1.277	1.008
Standard Deviation	0.898	***1.159***	0.934	1.086	1.095	1.079	***0.852***

Source: Derived and expanded upon by the author.

Table 4. Augmented Dickey-Fuller Test and Phillips-Perron Test results

	Augmented Dickey-Fuller Test Statistics		**Phillips-Perron Test statistic**	
	Bitcoin	**Ethereum**	**Bitcoin**	**Ethereum**
t-Statistic	-33.829	-34.677	-33.784	-29.931
P-value	0.000	0.000	0.000	0.000

Source: Derived and expanded upon by the author.

After conducting the ADF and PP tests to assess the stationarity of the time series data for both Bitcoin and Ethereum, the KPSS test was also conducted. The KPSS test serves as a complementary tool to the ADF and PP tests, offering additional insights into the stationarity properties of the data.

The KPSS test is particularly useful because it complements the ADF and PP tests by focusing on different aspects of stationarity. While the ADF and PP tests primarily detect trends in the data, the KPSS test is sensitive to detecting other forms of non-stationarity, such as level shifts, changes in variance, or sudden shocks. By running the KPSS test alongside the ADF and PP tests, a more comprehensive assessment of the stationarity of the time series data is ensured.

The results of the KPSS test, as shown in Table 5, indicate that the test LM statistics are less than the critical values at 99%, 95%, and 90% significance levels for both Bitcoin and Ethereum (see Table 5). This suggests that the null hypothesis of stationarity cannot be rejected, providing evidence that the time series data for both cryptocurrencies is stationary. Therefore, it can be concluded that the data does not exhibit significant non-stationarity, further validating the analysis and conclusions.

Table 5. Kwiatkowski-Phillips-Schmidt-Shin Test results

Bitcoin		**Ethereum**	
KPSS LM-Statistics	0.090	KPSS LM-Statistics	0.070
Critical value at 1%	0.216	Critical value at 1%	0.439
Critical value at 5%	0.146	Critical value at 5%	0.463
Critical value at 10%	0.119	Critical value at 10%	0.347

Source: Derived and expanded upon by the author.

Following the unit root tests, proceeded with Ordinary Least Squares (OLS) regression, incorporating dummy variables into the analysis. Subsequently, we scrutinized the OLS residuals for evidence of volatility clustering, employing Engle's ARCH test. Table 6 shows the results consistently yielded p-values below 0.05, compellingly rejecting the null hypothesis and accepting the existence of ARCH effects (see Table 6). After this GARCH(1,1) was applied and checked for robustness to predict the day-of-week-effect and volatility.

Table 6. Test results for Engle's Arch test

Bitcoin	F-statistics	0.142	F Probability	0.047
	Obs*R-squared	0.284	Probability Chi-Square	0.047
Ethereum	F-statistics	3.724	F Probability	0.025
	Obs*R-squared	7.416	Probability Chi-Square	0.025

Source: Derived and expanded upon by the author.

The significant p-values of both the ARCH and GARCH terms, as shown in Table 7, indicate their importance in both Bitcoin and Ethereum. This significance implies that the returns on these cryptocurrencies exhibit continuous and time-varying volatility (see Table 7). Moreover, it suggests that the volatility of cryptocurrencies is heavily influenced by both recent historical data and projected future values.

For Bitcoin, the ARCH + GARCH terms being less than 1 indicate decaying volatility, suggesting a persistence of volatility over time. The daily returns show negativity for Friday and Saturday and positivity for other days, aligning with findings of previous studies (Lopez-Martin, 2022; Naz et al., 2023). Additionally, there are no significant p-values for any day-of-week effect. Prior to the COVID period, Bitcoin did not show a day-of-the-week effect, and it has grown increasingly effective with time. These findings of Bitcoin are consistent with the research of various authors (Tiwari et al., 2019; Aggarwal, 2019; Lade & Yi, 2020; Baur et al., 2019; Kinateder & Papavassiliou, 2021; Dumrongwong, 2021) but do not support the findings of others (Aharon & Qadan, 2019; Lopez-Martin, 2022; Naz et al., 2023).

Similarly, for Ethereum, the ARCH + GARCH terms being less than 1 also signify decaying volatility, indicating a persistence of volatility. The daily returns for all days are positive. Additionally, the significant p-value for Thursday's daily returns is noteworthy, which is in line with previous research (Lopez-Martin, 2022; Karaömer & Kakilli, 2023).

Table 7. GARCH (1,1) model estimation for return of Bitcoin and Ethereum

Dependent Variable: Bitcoin returns				
GARCH = C(8) + C(9)*RESID(-1)^2 + C(10)*GARCH(-1)				
Variable	**Coefficient**	**Std. Errors**	**z-Statistics**	**Probability**
MONDAY	0.108	0.243	0.444	0.657
TUESDAY	0.405	0.211	1.915	0.546
WEDNESDAY	0.132	0.213	0.620	0.535
THURSDAY	0.353	0.222	1.594	0.111
FRIDAY	-0.022	0.223	-0.097	0.923
SATURDAY	-0.072	0.237	-0.303	0.762
SUNDAY	0.191	0.272	0.703	0.482
Variance Equation				
Constant	0.312	0.170	1.833	0.067
ARCH Term	0.068	0.022	3.078	0.002
GARCH Term	0.924	0.020	45.430	0.000
Dependent Variable: Ethereum returns				
GARCH = C(8) + C(9)*RESID(-1)^2 + C(10)*GARCH(-1)				
Variable	**Coefficient**	**Std. Errors**	**z-Statistics**	**Probability**
MONDAY	0.173	0.365	0.475	0.635
TUESDAY	0.580	0.314	1.843	0.065

Dependent Variable: Ethereum returns				
GARCH = C(8) + C(9)*RESID(-1)^2 + C(10)*GARCH(-1)				
Variable	**Coefficient**	**Std. Errors**	**z-Statistics**	**Probability**
WEDNESDAY	0.016	0.333	0.047	0.963
THURSDAY	0.887	0.314	2.827	0.005
FRIDAY	0.089	0.321	0.279	0.780
SATURDAY	0.196	0.322	0.610	0.542
SUNDAY	0.526	0.399	1.320	0.187
Variance Equation				
Constant	1.943	0.766	2.537	0.011
ARCH Term	0.086	0.028	3.095	0.002
GARCH Term	0.842	0.044	18.990	0.000

Source: Derived and expanded upon by the author.

To assess the robustness of the GARCH (1,1) model for the study's time series, two diagnostic tests were applied. Firstly, the Nyblom stability test examined structural changes within the time series by testing whether the higher-order autocorrelations of the squared residuals are zero. This test, robust to heavy-tailed distributions and outliers, accepted the null hypothesis at a 95% confidence level, indicating stable behavior of the variables in the GARCH (1,1) model. Secondly, the Engle & Ng sign bias test detected misspecifications in conditional volatility models, such as nonlinearity or asymmetry in the conditional variance. Robust to heavy-tailed distributions and outliers, this test ensures the dependability of the GARCH model's results for forecasting and risk management purposes.

The study period being post-COVID reveals a day-of-week effect in Ethereum, the high-return cryptocurrency, while Bitcoin shows no calendar anomalies. This suggests that the most traded cryptocurrency, Bitcoin, is becoming efficient over time. Inconsistencies in cryptocurrency calendar anomalies stem from various factors, including the relatively new and less mature nature of the cryptocurrency market, methodological disparities among scholars, the speculative environment of the market, and its susceptibility to external factors such as news, rumors, socioeconomic trends, and political movements.

Furthermore, the decentralized and global nature of the cryptocurrency market presents challenges in identifying and quantifying market-wide effects. The interplay between market sentiment and the adoption of cryptocurrencies for commercial activities, coupled with shifts in government policies and regulations, further underscores the adaptive market hypothesis.

5. Conclusion

The study aimed to explore the presence of day-of-the-week effects in the virtual currency market post-COVID-19, focusing specifically on Bitcoin and Ethereum. Through a comprehensive analysis employing both parametric and non-parametric tests, alongside sophisticated econometric models like the GARCH (1,1) model, we uncovered valuable insights into the dynamics of these cryptocurrencies.

Our findings reveal that Bitcoin shows no evidence of calendar anomalies, while Ethereum exhibits a notable day-of-the-week effect, characterized by fluctuations in returns across different days. This suggests a trend towards efficiency in Bitcoin, the most traded cryptocurrency, over time. However, the susceptibility of Ethereum to day-of-the-week effects underscores the ongoing challenges and complexities within the cryptocurrency market.

The disparities in calendar anomalies across cryptocurrencies can be attributed to various factors, including the nascent and evolving nature of the market, methodological disparities among researchers, and the speculative environment intrinsic to cryptocurrency trading. Furthermore, the decentralized and global nature of the cryptocurrency market poses challenges in accurately identifying and quantifying market-wide effects.

By providing empirical evidence of day-of-the-week effects in the cryptocurrency market and shedding light on changing market dynamics, our work contributes significantly to the existing literature. This identification of day-of-the-week effects holds significant implications for investors' risk management strategies. By understanding and leveraging these effects, investors can enhance their risk management approaches, particularly in timing their trades and allocating resources more effectively. Incorporating day-of-the-week effects into risk management frameworks can aid in optimizing portfolio diversification strategies, ultimately assisting investors in achieving a more balanced risk-return profile. Hence, our study underscores the practical utility of considering day-of-the-week effects in cryptocurrency investment decision-making, providing investors with valuable tools for navigating the complexities of the market.

Moving forward, additional research is essential to explore other elements that may influence market dynamics and delve deeper into the fundamental mechanisms driving day-to-day effects in cryptocurrencies. Additionally, continuous monitoring of regulatory developments and technological advancements will be pivotal in understanding the evolving landscape of the cryptocurrency market and its implications for investors.

References

Aggarwal, D. (2019). Do bitcoins follow a random walk model? *Research in Economics*, *73*(1), 15-22. https://doi.org/10.1016/j.rie.2019.01.002

Aggarwal, K., & Jha, M. K. (2023). Day-of-the-week effect and volatility in stock returns: evidence from the Indian stock market. *Managerial Finance*, *49*(9), 1438-1452. https://doi.org/10.1108/mf-01-2023-0010

Aharon, David Yechiam, & Qadan, M. (2019). Bitcoin and the Day-of-The-Week Effect. *Finance Research Letters*, *31*. https://doi.org/10.1016/j.frl.2018.12.004

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, *297*(1-2), 3-36. https://doi.org/10.1007/s10479-020-03575-y

Ali, S., Li, G., Liu, Y., Ishaq, M., & Shah, T. (2019). The Relationship between Carbon Dioxide Emissions, Economic Growth and Agricultural Production in Pakistan: An Autoregressive Distributed Lag Analysis. *Energies*, *12*(24), 4644. https://doi.org/10.3390/en12244644

Ampountolas, Apostolos (2022). Cryptocurrencies Intraday High-Frequency Volatility Spillover Effects Using Univariate and Multivariate GARCH Models. *International Journal of Financial Studies*, *10*(3), 51. https://doi.org/10.3390/ijfs10030051

Andolfatto, D., & Martin, F. M. (2022). The Blockchain Revolution: Decoding Digital Currencies. *Review*, *104*(3). https://doi.org/10.20955/r.104.149-65

Apergis, N. (2022). COVID-19 and cryptocurrency volatility: Evidence from asymmetric modelling. *Finance Research Letters*, *47*, 102659. https://doi.org/10.1016/j.frl.2021.102659

Bae, G., & Kim, J. (2022). Observing cryptocurrencies through robust anomaly scores. *Entropy*, *24*(11), 1643. https://doi.org/10.3390/e24111643

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, *64*, 74-81. https://doi.org/10.1016/j.econmod.2017.03.019

Baur, D. G., Cahill, D., Godfrey, K., & Frank, Liu, Z. (2019). Bitcoin time-of-day, day-of-week and month-of-year effects in returns and trading volume. *Finance Research Letters*, *31*, 78-92. https://doi.org/10.1016/j.frl.2019.04.023

Caporale, G. M., & Plastun, A. (2019). The day of the week effect in the cryptocurrency market. *Finance Research Letters*, *31*. https://doi.org/10.1016/j.frl.2018.11.012

Chaouachi, O., & Dhaou, I. (2020). The Day of the Week Effect: Unconditional and Conditional Market Risk Analysis. *International Journal of Economics and Financial Issues*, *10*(6), 94-98. https://doi.org/10.32479/ijefi.10610

Chen, C., Després, R., Guo, L., & Renault, T. (2019). What Makes Cryptocurrencies Special? Investor Sentiment and Return Predictability During the Bubble. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3398423

Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). GARCH Modelling of Cryptocurrencies. *Journal of Risk and Financial Management*, *10*(4), 17. https://doi.org/10.3390/jrfm10040017

Coulter, K. A. (2022). The impact of news media on Bitcoin prices: modelling data driven discourses in the crypto-economy with natural language processing. *Royal Society Open Science*, *9*(4). https://doi.org/10.1098/rsos.220276

Dangi, V. (2020). Day of the Week Effect in Cryptocurrencies' Returns and Volatility. *Ramanujan International Journal of Business and Research*, *5*(1), 139-167. https://doi.org/10.51245/rijbr.v5i1.2020.221

Dao, P., & Staszewski, W. (2021). Lamb Wave Based Structural Damage Detection Using Stationarity Tests. *Materials*, *14*(22), 6823. https://doi.org/10.3390/ma14226823

Davidson, S., De Filippi, P. D., & Potts, J. (2018). Blockchains and the Economic Institutions of Capitalism. *Journal of Institutional Economics*, *14*(4), 639-658. https://doi.org/10.1017/s1744137417000200

Dorfleitner, G., & Lung, C. (2018). Cryptocurrencies from the perspective of euro investors: a re-examination of diversification benefits and a new day-of-the-week effect. *Journal of Asset Management*, *19*(7), 472-494. https://doi.org/10.1057/s41260-018-0093-8

Dumrongwong, K. (2021). Calendar Effects on Cryptocurrencies: Not so Straightforward. *Southeast Asian Journal of Economics*, 9(1), 1-26.

Fama, E. F. (1997). Market efficiency, long-term returns, and behavioral finance. *SSRN*, *49*(3), 283-306. https://doi.org/10.2139/ssrn.15108

Ftiti, Z., Louhichi, W., & Ben Ameur, H. (2021). Cryptocurrency volatility forecasting: What can we learn from the first wave of the COVID-19 outbreak? *Annals of Operations Research*. *330*(1-2), 665-690. https://doi.org/10.1007/s10479-021-04116-x

Grobys, K., & Junttila, J. (2020). Speculation and Lottery-Like Demand in Cryptocurrency Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3551948

Harabida, M., Radi, B., & Gueyie, J. (2023). ESG Indices Efficiency in Five MENA Countries: Application of the Hurst Exponent. *Theoretical Economics Letters*, *13*(02), 183-201. https://doi.org/10.4236/tel.2023.132011

Haruna, M., Hassan, S., & Ahmad, H. (2022). How responsive is the poverty to the foreign direct investment inflows in Nigeria? Evidence from linear and non-linear ARDL. *International Journal of Social Economics*, *50*(1), 73-96. https://doi.org/10.1108/ijse-08-2020-0530

Joshi, J., & Sharma, R. P. (2022). Network-centric Empirical Analysis of Bitcoins cryptocurrency organization. *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*. https://doi.org/10.1109/icicict54557.2022.9917885

Kaiser, L. (2019). Seasonality in cryptocurrencies. *Finance Research Letters*, *31*. https://doi.org/10.1016/j.frl.2018.11.007

Karaömer, Y., & Kakilli Acaravci, S. (2023). Adaptive Market Hypothesis: Evidence from the Cryptocurrency Market. *Iranian Journal of Management Studies*, *16*(1), 125-138. https:10.22059/IJMS.2022.336833.674889

Kargar, N. (2021). Generalized autoregressive conditional heteroscedasticity (GARCH) for predicting volatility in Stock Market. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(3), 73-75. https://doi.org/10.54660/.IJMRGE.2021.2.3.73-75

Katsiampa, P. (2017). Volatility estimation for bitcoin: a comparison of GARCH models. *Economics Letters*, *158*, 3-6. https://doi.org/10.1016/j.econlet.2017.06.023

Katsiampa, P. (2019). An empirical investigation of volatility dynamics in the cryptocurrency market. *Research in International Business and Finance*, *50*, 322-335. https://doi.org/10.1016/j.ribaf.2019.06.004

Khuntia, S., & Pattanayak, J. (2021). Adaptive calendar effects and volume of extra returns in the cryptocurrency market. *International Journal of Emerging Markets*, *17*(9), 2137-2165. https://doi.org/10.1108/ijoem-06-2020-0682

Kinateder, H., & Papavassiliou, V. G. (2021). Calendar effects in Bitcoin returns and volatility. *Finance Research Letters*, *38*, 101420. https://doi.org/10.1016/j.frl.2019.101420

Kyriazis, N. (2019). A Survey on Efficiency and Profitable Trading Opportunities in Cryptocurrency Markets. *Journal of Risk and Financial Management*, *12*(2), 67. https://doi.org/10.3390/jrfm12020067

Lade, S., & Yi, J. (2020). Does the South Korea Bitcoin Market Is Efficient? *International Journal of Management*, *11*(9), 1592-1597. https://ssrn.com/abstract=3713122

Liao, G., Liu, Q., Zhang, R., & Zhang, S. (2021). Rank test of unit-root hypothesis with AR-GARCH errors. *Journal of Time Series Analysis*, *43*(5), 695-719. https://doi.org/10.1111/jtsa.12635

Liu, G., Yu, C., Shiu, S., & Shih, I. (2022). The Efficient Market Hypothesis and the Fractal Market Hypothesis: Interfluves, Fusions, and Evolutions. *SAGE Open*, *12*(1), 215824402210821. https://doi.org/10.1177/21582440221082137

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, *30*(5), 15-29. https://doi.org/10.3905/jpm.2004.442611

Lopez-Martin, C. (2022). Ramadan effect in the cryptocurrency markets. *Review of Behavioral Finance*, *14*(4), 508-532. https://doi.org/10.1108/rbf-09-2021-0173

Luxianto, R., Arief, U., & Prasetyo, M. (2020). Day-of-the-Week Effect and Investors' Psychological Mood Testing in a Highly Mispriced Capital Market. *Journal of Indonesian Economy and Business*, *35*(3), 257. https://doi.org/10.22146/jieb.54377

Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, *135*(2), 293-319. https://doi.org/10.1016/j.jfineco.2019.07.001

Micu, R., & Dumitrescu, D. (2022). Study regarding the volatility of main cryptocurrencies. *Proceedings of the International Conference on Business Excellence*, *16*(1), 179-187. https://doi.org/10.2478/picbe-2022-0018

Mikhaylov, A. (2020). Cryptocurrency Market Analysis from the Open Innovation Perspective. *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(4), 197. https://doi.org/10.3390/joitmc6040197

Miralles-Quirós, J., & Miralles-Quirós, M. (2022). A new perspective of the day-of-the-week effect on bitcoin returns: evidence from an event study hourly approach. *Oeconomia Copernicana*, *13*(3), 745-782. https://doi.org/10.24136/oc.2022.022

Naimy, V., Haddad, O., Fernández-Avilés, G., & El Khoury, R. (2021). The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies. PLOS ONE *16*(1): e0245904. https://doi.org/10.1371/journal.pone.0245904

Naz, F., Sayyed, M., Rehman, R.-U., Naseem, M. A., Abdullah, S. N., & Ahmad, M. I. (2023). Calendar anomalies and market volatility in selected cryptocurrencies. *Cogent Business & Management*, *10*(1). https://doi.org/10.1080/23311975.2023.2171992

Ogunyolu, O., & Adebayo, A. (2022). An appraisal of ethical issues and the effect of artificial intelligence on the cryptocurrency market. *Global Journal of Engineering and Technology Advances*, *11*(2), 063-070. https://doi.org/10.30574/gjeta.2022.11.2.0083

Okorie, D., & Lin, B. (2021). Adaptive market hypothesis: the story of the stock markets and COVID-19 pandemic. *The North American Journal of Economics and Finance*, *57*, 101397. https://doi.org/10.1016/j.najef.2021.101397

Olivares-Sánchez, H., Rodríguez-Martínez, C., Coronel-Brizio, H., Scalas, E., Seligman, T., & Hernández-Montoya, R. (2022). An empirical data analysis of “price runs” in daily financial indices: dynamically assessing market geometric distributional behavior. *PLOS ONE*, *17*(7), e0270492. https://doi.org/10.1371/journal.pone.0270492

Paital, R. R., & Panda, A. K. (2018). Day of the Week and Weekend Effects in the Indian Stock Market. *Theoretical Economics Letters*, *08*(11), 2559-2568. https://doi.org/10.4236/tel.2018.811164

Pantielieieva, N. M., Rogova, N. V., Braichenko, S. M., Dzholos, S. V. & Kolisnyk, A. S. (2021). Current Aspects of Transformation of Economic Relations: Cryptocurrencies and their Legal Regulation. *Financial and Credit Activity Problems of Theory and Practice*, *4*(31), 410-418. https://doi.org/10.18371/fcaptp.v4i31.190962

Rehan, M., & Gül, M. (2023). Examining the efficiency of stock markets using multifractal detrended fluctuation analysis. empirical evidence from OIC (Organization of Islamic Cooperation) countries during the GFC and COVID-19 pandemic. *The Journal of Risk Finance*, *24*(5), 657-683. https://doi.org/10.1108/jrf-04-2023-0108

Rossi, M., & Gunardi, A. (2018). Efficient Market Hypothesis and Stock Market Anomalies: Empirical Evidence in Four European Countries. *Journal of Applied Business Research*, *34*(1), 183-192. https://doi.org/10.19030/jabr.v34i1.10111

Scherf, M., Matschke, X., & Rieger, M. O. (2022). Stock market reactions to COVID-19 lockdown: A global analysis. *Finance Research Letters*, *45*, 102245. https://doi.org/10.1016/j.frl.2021.102245

Shahid, M. (2022). COVID-19 and adaptive behavior of returns: evidence from commodity markets. *Humanities and Social Sciences Communications*, *9*(1). https://doi.org/10.1057/s41599-022-01332-z

Singh, A. (2021). The Regulatory Regime for Cryptocurrency in the Present Global Order. *SSRN*. https://doi.org/10.2139/ssrn.3978476

Souza, O., & De França Carvalho, J. (2023). Market efficiency assessment for multiple exchanges of cryptocurrencies. *Revista De Gestão*. https://doi.org/10.1108/rege-05-2022-0070

Stavrova, E. (2021). Banks' Digital Challenges. *Business Ethics and Leadership*, *5*(3),87-96. https://doi.org/10.21272/bel.5(3).87-96.2021

Theiri, S., Nekhili, R., & Sultan, J. (2022). Cryptocurrency liquidity during the Russia-Ukraine war: the case of bitcoin and ethereum. *The Journal of Risk Finance*, *24*(1), 59-71. https://doi.org/10.1108/jrf-05-2022-0103

Țilică, E. V. (2021). Financial Contagion Patterns in Individual Economic Sectors. The Day-of-the-Week Effect from the Polish, Russian and Romanian Markets. *Journal of Risk and Financial Management*, *14*(9), 442. https://doi.org/10.3390/jrfm14090442

Tiwari, A. K., Kumar, S., & Pathak, R. (2019). Modelling the dynamics of Bitcoin and Litecoin: GARCH versus stochastic volatility models. *Applied Economics*, *51*(37), 4073-4082. https://doi.org/10.1080/00036846.2019.1588951

Tosunoğlu, N., Abacı, H., Ateş, G., & Akkaya, N. (2023). Artificial neural network analysis of the day of the week anomaly in cryptocurrencies. *Financial Innovation*, *9*(1). https://doi.org/10.1186/s40854-023-00499-x

Tran, T. N. (2023). Day of Week Effect on Financial Market: Evidence in Vietnam during Normal Period and COVID-19 Pandemic. *KINERJA*, *27*(1), 29-45. https://doi.org/10.24002/kinerja.v27i1.6377

Volosovych, S., Sholoiko, A., & Shevchenko, L. (2023). Cryptocurrency Market Transformation During Pandemic COVID-19. *Financial and Credit Activity Problems of Theory and Practice*, *1*(48), 114-126. https://doi.org/10.55643/fcaptp.1.48.2023.3949

Washington, P. B., Gali, P., Rustam, F., & Ashraf, I. (2023). Analyzing influence of COVID-19 on crypto & financial markets and sentiment analysis using deep ensemble model. *PLOS ONE*, *18*(9), e0286541. https://doi.org/10.1371/journal.pone.0286541

Zhao, H., & Zhang, L. (2021). Financial literacy or investment experience: which is more influential in cryptocurrency investment? *International Journal of Bank Marketing*. https://doi.org/10.1108/ijbm-11-2020-0552

Zilca, S. (2017). The evolution and cross-section of the day-of-the-week effect. *Financial Innovation*, *3*(1). https://doi.org/10.1186/s40854-017-0077-6

About the author

Sonal Sahu is a seasoned professor at Tecnológico de Monterrey, Guadalajara, Mexico, with a decade-long tenure in the Department of Finance and Accounting. With over 11 years of work experience at Tecnológico de Monterrey, Sonal has demonstrated her expertise in finance and accounting education. Prior to her tenure at Tecnológico de Monterrey, Sonal held significant roles in the financial sector, accumulating over 10 years of experience working with prestigious institutions such as JP Morgan Chase, Deutsche Bank, ICICI Bank, and the Allianz Group. Currently pursuing a Ph.D. in Finance at EGADE Business School, her research focuses on cryptocurrencies and international investments, reflecting her dedication to understanding evolving financial landscapes. Sonal has showcased her scholarly prowess through publications in esteemed journals like the Risks Journal and presentations at conferences such as the European Conference on Games-Based Learning. Her contributions in academia and research underscore her commitment to advancing knowledge in finance and shaping future financial practices.

sonal.sahu@tec.mx
https://orcid.org/0000-0002-2755-0980

Analysis of the Impact of ESG Controversies on the Valuation of Shares in the Mexican Market

Análisis del impacto de las controversias por factores ASG en la valuación de acciones en el mercado mexicano

Martha Angélica León Alvarado

EGADE Business School, Tecnológico de Monterrey, Mexico

Eric Osio Cerón

Grenoble École de Management, Francia

Anahi Montserrat Revilla Antonio

Universidad Nacional Autónoma de México (UNAM), Mexico

Received: March 20, 2024.
Approved: May 27, 2024.

Abstract

There is extensive literature on the motivations of various investors to integrate environmental, social, and corporate governance (ESG) factors into their strategies. However, studies on this subject focus on developed markets. They do not explain with evidence whether these factors, if absent, significantly affect the valuation of financial instruments, such as stocks or bonds issued by companies. Thus, the study analyzes and demonstrates with statistical evidence that the controversies associated with ESG factors and related to shares in the Mexican market are yet to be integrated into the corresponding valuations.

Keywords: *Investments, ESG, controversies, environmental, social, corporate governance.*
JEL Classification: *C16, G10, Q59.*

Resumen

Existe una amplia literatura sobre las motivaciones que llevan a diversos inversionistas a integrar factores ambientales, sociales y de gobierno corporativo (ASG) en sus estrategias; sin embargo, los estudios sobre esta temática se centran en los mercados desarrollados y no detallan con evidencia si estos factores, a través de las potenciales controversias que generan en su ausencia, tienen efectos significativos en la valoración de instrumentos financieros, por ej., acciones o bonos emitidos por empresas. Por esta razón, el presente estudio analiza y demuestra con evidencia estadística que las controversias asociadas a factores ASG y relacionadas con las acciones en el mercado mexicano aún no se encuentran integradas en las valuaciones correspondientes.

Palabras clave: *inversiones, ASG, controversias, ambiental, social, gobierno corporativo.*
Clasificación JEL: *C16, G10, Q59.*

1. Introduction

The inclusion of environmental, social, and governance factors—ESG factors—has become increasingly relevant to companies, organizations, societies, and investors. Particularly for the latter, since decision-making becomes more relevant through their investments and risk mitigation processes.

The entire global community has also recognized the "new" mindset. For example, the World Economic Forum (WEF) Global Risks Report 2024 states that seven out of ten of the global most severe risks over the long term belong to environmental or societal related risks—five environmental and two societal (WEF, 2024, p. 8).

Another example is Krueger et al. (2020), who carried out a survey focused on identifying the relevance of climate factors for institutional investors in five key areas:

1) The role of climate risks in investment decisions.

2) Climate risk management.

3) Shareholder engagement related to climate risks.

4) The implications of climate risks for asset pricing, and

5) Investors' opinions on companies' disclosure of climate risks.

They concluded that for investors, reputation, moral or ethical considerations, legal or fiduciary duties, as well as financial motives, are the motivations that justify the incorporation of these factors.

Likewise, one of the revelations in that study—which will support part of the results of our analysis—is that respondents believe that stock valuations do not fully reflect the risks of climate change. Failing to include ESG factors like sustainability and corporate responsibility properly can affect financial performance in such a way that this could ultimately damage the value of companies and the financial instruments associated with them, i.e., bonds or stocks.

The literature refers potential benefits associated with investing in sustainable instruments or including practices of this nature. To cite an example: more than 20 years ago, market practice (mainly in Europe) regarding sustainability was limited to generating exclusion lists. This practice, which some asset managers maintain, excludes companies, sectors, or countries from the investment universe if they involve certain activities with defined criteria.

Among the most common exclusion criteria are activities related to weapons, pornography, and tobacco. In this context, the first studies on sustainable investment attributed higher capital costs associated with companies without "green" practices than those that did (Heinkel et al., 2001).

Another factor considered in the literature that could affect returns and, therefore, company valuations is the pricing power of some companies for being socially responsible. According to studies, this is due to the feeling of loyalty generated among customers (Albuquerque et al., 2020). Alternatively, some studies have determined that economic agents can receive different non-monetary benefits from simply holding or including assets with green characteristics in their investment strategies (e.g., Baker et al., 2022; Fama & French, 2007).

Some other studies, such as Cornell and Damodaran (2020), have tried to generate frameworks to determine the correct way to integrate ESG factors within the valuation of financial instruments. The above considering the growing pressure from politicians, regulators, interest groups, and even investors who commonly sugarcoat their speeches of inclusion of these factors with the promise that it will be good for their results and their shareholders.

In that same sense, we find a broad consensus in the literature on investors' sustainability preferences (Bauer et al., 2021; Ceccarelli et al., 2023; Heeb et al., 2023, among others). Empirical studies indicate that economic agents support sustainable investments, even if this implies sacrificing returns. Barber et al. (2021), Hartzmark and Sussman (2019), and Ceccarelli et al. (2023) document that investors find value in strategies associated with sustainability in different ways. The studies highlight the importance of interest in pension funds for sustainability issues, such as the Sustainable Development Goals (SDGs). Significantly, Bauer et al. (2021) provide evidence that support for sustainable investments is driven by strong social preferences of investors such as pension funds.

However, the studies cited, and others available in the current literature have focused their research on the economies of developed countries that, in addition to having much more solid economies, have more solid financial systems and more and better-quality information for all participants. All this allows the market to consider more variables and make more accurate decisions. Likewise, the evidence of preferences is still far from showing the impact of these decisions, so it is essential to analyze the effects they could have on measurable characteristics such as valuations.

Given the above, the objectives of this study are, on the one hand, to collect information on controversies associated with ESG factors in some Mexican companies that are listed on the stock exchanges in the country and, on the other hand, to identify whether said controversies have had significant effects in the share prices of these companies.

2. Description of the Methodology

A quantitative experimental research approach was adopted to conduct this study using historical data on Mexican stock prices and controversies related to ESG factors.

Based on these data, we used the Student's *t*-test with a 5% significance level to analyze the relationship between the returns associated with the shares before and after the ESG controversy and, in this way, verify if there were significant differences in the average of the returns obtained from the generation of the controversy. The methodology considers the following stages:

1) *Data Description and Collection.* We collected quantitative and qualitative data for this analysis. For quantitative data, we used historical stock price data for Mexican companies listed in Mexican stock exchanges (106 issuers with information from January 2013 to January 2024); historical index data for the "MEXBOL Index," which represents the all-cap index for Mexican capital markets, and "S&P/BMV Sovereign MBONOS 10-20 Year Bond Index," designed to measure the performance of the Mexican fixed income nominal-rate market—the constituents of the index must be Mexican government securities with maturities between 10 and 20 years and denominated in Mexican pesos).

   Regarding the qualitative data, controversies in ESG factors were collected, which include events such as strikes due to inadequate working conditions, corruption cases, environmental controversies (pollution, waste management, water stress, etcetera), criticism by communities, adverse data related to the corporate governance of companies, among others.

   It is essential to mention that platforms such as Bloomberg terminal and Morningstar Direct (through Sustainalytics) (Morningstar, n.d.) collect and disclose specific information on ESG factors. Considering that the databases, sources, and even the methodologies of the platforms could change, to obtain this information, the recommendation is to use the workspace with datapoints associated with ESG controversies within the Morningstar Direct terminal, as well as the "ESG controversies" section in the description of assets within the Bloomberg terminal. In any case, checking with specialists on each platform for specific queries will be necessary.

2) *Selection of Companies and Controversies.* We selected Mexican companies[1] involved in controversies verified by an information platform specialized in ESG matters to guarantee the sample's representativeness. Also, the controversies were classified within the three pillars: environmental, social, or governance. In addition, the dates of the controversial ESG events were identified to determine the *a priori* and *a posteriori* differentiating effect.

3) *Statistical Analysis.* The Student's *t*-test was used to compare the means of the returns associated with stock prices before and after the controversial event. The tests were carried out at three time periods to avoid potential biases due to the temporal dilution of the events with daily observations[2]:

   a) 60 daily observations (approximately quarterly periods).

   b) 126 daily observations (approximately semiannual periods). And

   c) 255 daily observations (approximately annual periods).

   The construction of these statistical models to evaluate the impact of ESG controversies on stock returns allows us to isolate other possible determinants of the value of these stocks by using the controversial event as the epicenter and excluding it from the samples. Likewise, establishing these temporalities allows us to identify whether the impact of ESG factors can prevail in the short, medium, or long term.

   It is important to highlight that, to confirm the validity of the exercise, different checks were carried out, mainly to verify that systemic market movements did

[1] The names of the issuers are masked because the objective of the study is not to whistleblow on these issuers and their lack of consideration of ESG factors within their valuations. Also, the naming of the issuers could be interpreted as an assertion that the controversies should have affected the valuation and/or a recommendation for investment decisions, which is not the objective of this study.

[2] The shorter period is equivalent to 60 days to preserve the normality assumption necessary to perform the test. And although it could be assumed that the effect of the controversy would be diluted, it is argued that the effect on valuation should be structural and not only seen as short-term speculation.

not affect the exercise. The same statistical test was carried out, affecting the series of returns to generate some type of daily alpha. That is, a benchmark's performance discounted each stock's daily returns.

The two most relevant exercises correspond to the benchmarks: the MEXBOL Index and the S&P/BMV Sovereign MBONOS 10–20 Year Bond Index. Although we will focus on the baseline exercise in the study, the results of these additional tests will be available in Annex B of the study (see Annex B). Additionally, tests with a market beta—corresponding to stock market movements or the average risk-free rate—were discarded since they were considered constant. This would only shift the distribution of returns (down or up), reaching the same result as the statistical tests.

An alternative hypothesis test to the one proposed in this study could be to ensure that the returns a posteriori to the controversy are lower than the returns a priori to that event, i.e., a one-tailed Student's *t*-test. However, since it is impossible to determine what effect the controversies have on valuation, determining a downward trend in returns can be understood as a value judgment of the controversy itself and its effect. Nevertheless, Annex C of this document contains the test, and the results are consistent with the initially proposed test to strengthen the exercise (see Annex C).

4) *Control of Variables.* Stocks that:

a. Did not comply with at least one period of the required observations.

b. Did not have associated and verifiable controversies through the information platforms used. Or,

c. Did not have a public history of daily prices.

Likewise, in the case of associated disputes that lasted more than one day, the first day of the dispute was taken as the epicenter.

5) *Interpretation of Results.* The probabilities associated with the two-tailed Student's *t*-tests of comparison of means were analyzed, with a null hypothesis that aims to determine that the means of both samples, *a priori* and *a posteriori* of the controversy, are similar. This indicates that they do not contain a significant change. With this, we can identify if there is a significant relationship between controversies associated with ESG factors and stock valuations. Likewise, the strength and direction of this relationship were evaluated, revealing the implications for investors, companies, and the market.

3. Methodology Development

Initially, it was necessary to determine a pool of issuers that could be analyzed in terms of their daily valuation and potential controversies. Therefore, all capital issuers (shares) listed on the Mexican Stock Exchanges were selected as an initial set, resulting in 202 issuers. We stress the omission in our study of the capital of private companies due to the small amount of information available and the lack of homogeneity of said information. These requirements are considered essential to identify changes in market valuations.

The Bloomberg information platform provided the basis for the issuers' database. Subsequently, issuers with a broad historical range of publicly available prices were identified and selected for the following steps of the study. The selected history range ensures that comparability can be established between the prices of the instruments at different times, allowing for the construction of a consolidated information base of historical prices of Mexican capital issuers. It is important to mention that this database resulted in 106 issuers with information from January 2013 to January 2024.

Subsequently, using data from two information platforms—Bloomberg terminal and Morningstar Direct—ESG controversies were identified for Mexican capital issuers whose information was verifiable. We related this to the original base of issuers. For the relationship of ESG controversies with the different issuers, it was necessary to generate a dispute classification process where:

1) Those controversies or news that did not have clear and/or sufficient information or evidence were eliminated, and, on the contrary, those that could represent an impact finding would be given priority in the analysis, resulting in 10 issuers with 26 associated controversies[3].

2) The different controversies were classified according to each of the factors, which could be environmental, social, or corporate governance controversies, resulting in the following distribution: one environmental controversy, eight social controversies, and 17 corporate governance controversies. And

[3] Classified in 7 types of controversies related to the 3 ESG pillars: 1) labor accusations, 2) environmental impacts, 3) cyber-attacks and impacts on customer information, 4) monopolistic practices and competition, 5) strikes, 6) money laundering,7) bribery cases.

https://doi.org/10.36105/theanahuacjour.2024v24n1.02

3) The "epicenters" of each ESG controversy were identified within the historical database, with the epicenter being the initial date of creation or generation of the controversy or the first moment in which the investing public had access to the controversy information. After screening out the information, of the 26 epicenters, only 14 were considered for eight issuers with the distribution shown in Table 1 (see Table 1).

Table 1. Distribution of Epicenters per Issuer

ISSUER	Number of epicenters
ISSUER 1	2
ISSUER 2	3
ISSUER 3	2
ISSUER 4	1
ISSUER 5	1
ISSUER 6	1
ISSUER 7	3
ISSUER 8	1
Total	14

Source: Prepared by the authors.

For an initial analysis, localization of epicenters in the historical prices chart was done; nevertheless, no further information was obtained from this review. The resulting graphs will be available in Annex A of this study (see Annex A).

Regarding classifying controversies within each ESG pillar, we identified that, in the first analysis, the most significant number of controversies fell within the governance pillar.

Once the epicenters were identified, the following hypothesis test was proposed.

H_0: $\mu_i = \mu_j$

H_1: $\mu_i \neq \mu_j$

Such that:

μ_i : the average returns before the epicenter or ESG controversy

μ_j : the average returns after the epicenter or ESG controversy

The Student's *t*-test is appropriate for the exercise since the comparison of performance, in the case of our analysis, is a typical exercise for comparing statistical parameters of populations or samples (Department of Statistics and Data Science, n.d.). Likewise, to carry out this statistical procedure, the three necessary characteristics are determined as true (normality, independence, and heteroskedasticity), taking into account the following:

- *Normality*: The return associated with the price of each share in public markets is assumed to have a normal distribution.
- *Independence:* For each of the Student's t-tests, two returns samples are compared, and although they are from the same stock, they correspond to exclusive periods, such that there is no dependence between the samples, and they can be determined independently.
- Finally, in the case of *homoscedasticity:* When talking about the same action and the data population corresponds to the same series from which two samples are taken, it is assumed that the variance is constant among the observations.

Once the requirements were met, the statistical analysis was generated through the significance test, and it was determined that 95% would be the most appropriate confidence level to justify the analysis.

Student's *t*-tests were used to compare the means of the returns associated with the stock prices before and after the controversial event and, additionally, to avoid potential biases due to the temporality dilution of the events, the tests were carried out at three temporalities with daily observations: The periods are the following:

1. Quarterly periods (60 daily observations).

2. Semiannual periods (126 daily observations).

3. Annual periods (255 daily observations).

4. Analysis of the Results

From the proposed statistical model, information is derived at different levels, the most relevant for the study being the significance level with which the null hypothesis is not rejected in most cases (95%) as can be observed in Table 2. Taking this parameter into account, only for epicenter 1 of Issuer 5 it is possible to reject the null hypothesis, so we would not have enough information to say that the difference between the means of the *a priori* and *a posteriori samples* of that epicenter are significantly similar, meaning that in this case, the controversy affected in particular the average returns between both periods (see Table 2).

The event in which the null hypothesis was rejected corresponds to shareholder demands towards the company for an alleged bribery case, a controversy associated with the corporate governance classification. From the above, it is clear that the event, in addition to being controversial, had legal implications, and the materialization of the revaluation and its "permanence" to a medium temporality in the study is presumed to be due to this.

The second relevant result is associated with the different temporalities analyzed, so no trends are identified for the same event in different periods. This could be interpreted as the fact that the effects of controversies lack "permanence" within the behavior of asset valuations, in addition to not being significant. It should be noted that this result may be affected by "other" variables since other additional factors that could directly impact the valuations are not isolated. Nevertheless, even in the benchmarking-alpha analysis, there was no evidence of any trends or permanence among the different periods (for details, see Annex B).

Table 2. Probabilities of Non-Rejection of H_0 Associated with each Issuer for the Different Epicenters and Different Temporalities

Probability	EPICENTER 1		
	60	**126**	**255**
ISSUER 1	92.37 %	93.20 %	89.25 %
ISSUER 2	56.30 %	20.18 %	41.63 %
ISSUER 3	46.89 %	65.76 %	N/A
ISSUER 4	72.35 %	N/A	N/A
ISSUER 5	13.22 %	4.38 %	44.40 %
ISSUER 6	42.69 %	69.97 %	51.14 %

ISSUER 7	13.95 %	N/A	N/A
ISSUER 8	90.54 %	36.61 %	95.77 %

Probability	EPICENTER 2		
	60	126	255
ISSUER 1	52.30 %	N/A	N/A
ISSUER 2	79.23 %	62.93 %	75.32 %
ISSUER 3	31.67 %	9.38 %	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	69.90 %	47.80 %	58.63 %
ISSUER 8	N/A	N/A	N/A

Probability	EPICENTER 3		
	60	126	255
ISSUER 1	N/A	N/A	N/A
ISSUER 2	8.61 %	20.55 %	46.03 %
ISSUER 3	N/A	N/A	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	83.41 %	N/A	N/A
ISSUER 8	N/A	N/A	N/A

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

5. Conclusions

Based on the study results, we can infer that current controversies due to ESG factors have not been included substantially in the valuations of financial assets in public companies in the Mexican financial system. However, the growing global and local pressure from regulators, investors, and the market to include these factors in determining investment strategies should end up affecting the valuations of these assets, so it will be worth continuing to carry out this type of analysis, given that ESG factors have a significant presence within the Mexican financial system.

From the above, the following potential areas of opportunity are identified: the creation of approved databases on controversies and the development of clear and detailed methodologies to incorporate ESG factors into the valuation of different financial assets.

The latter is relevant because although this study focused on capital instruments, the steps to follow for other asset classes and types of instruments will need to be determined.

Finally, it is observed that for the local market, the issues related to the governance of companies are much more pressing, but without leaving aside social and certainly environmental issues. This phenomenon is attributed to the fact that Mexico is still a country classified as "emerging" by, for instance, the International Monetary Fund (IMF, 2024) and must adapt to the best international practices, so the adoption of ESG factors would allow a more orderly transition and, therefore, faster development into a developed economy.

References

Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. *Review of Corporate Finance Studies*, *9*(3), 593–621. https://doi.org/10.1093/rcfs/cfaa011

Baker, M., Bergstresser, D., Serafeim, G., & Wurgler, J. (2022). The pricing and ownership of US green bonds. *Annual Review of Finance Economics*, *14*, 415–437. https://doi.org/10.1146/annurev-financial-111620-014802

Barber, B., Morse, A., & Yasuda, A. (2021). Impact investing. *Journal of Finance Economics*, *139*(1), 162–185. https://doi.org/10.1016/j.jfineco.2020.07.008

Bauer, R., Ruof, T., & Smeets, P. (2021). Get real! Singles prefer more sustainable investments. *Review of Financial Studies*, *34*(8), 3976–4043. https://doi.org/10.1093/rfs/hhab037

Bloomberg L.P. (n.d.). [Mexican public equity daily prices, and ESG controversies 2013Q1–2024Q1 in MXN] [Data set]. Retrieved February 15, 2024, from SFU Bloomberg terminal.

Ceccarelli, M., Ramelli, S., & Wagner, A. (2023) [Forthcoming]. Low-carbon mutual funds. *Review of Finance*. [Finance Working Paper N° 659/2020]. https://dx.doi.org/10.2139/ssrn.3353239

Cornell, B., & Damodaran, A. (2020). Valuing ESG: Doing Good or Sounding Good? NY Stern School of Business. [SSRN Working Paper #3557432]. https://dx.doi.org/10.2139/ssrn.3557432

Department of Statistics and Data Science. (n.d.). *Comparison of two means*. Yale University. http://www.stat.yale.edu/Courses/1997-98/101/meancomp.htm

Fama, E., & French, K. (2007). Disagreement, tastes, and assets prices. *Journal of Finance Economics*, *83*(3), 667–689. https://doi.org/10.1016/j.jfineco.2006.01.003

Hartzmark, S., & Sussmann, A. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance*, *74*(6), 2789–2837. https://doi.org/10.1111/jofi.12841

Heeb, F., Kölbel, J., Paetzold, F., & Zeisberger, S. (2023). Do investors care about impact? *Review of Finance Studies*, *36*(5), 1737–1787. https://doi.org/10.1093/rfs/hhac066

Heinkel, R., Kraus, A., & Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis*, *36*(4), 431–449. https://doi.org/10.2307/2676219

International Monetary Fund (IMF). (2024). *World Economic Outlook Update, January 2024: Moderating inflation and steady growth open path to soft landing*. https://www.imf.org/en/Publications/WEO/Issues/2024/01/30/world-economic-outlook-update-january-2024

Krueger, P., Sautner, Z., & Starks, L. (2020). The importance of climate risks for institutional investors. *Review of Finance Studies*, *33*(3), 1067–1111. https://doi.org/10.1093/rfs/hhz137

Morningstar Inc. (n.d.). [Mexican public equity funds, ESG controversies through workspace 2013Q1–2024Q1 in MXN] [Data set]. Retrieved February 15, 2018, from Morningstar Direct.

World Economic Forum (WEF). (2024). *The Global Risks Report 2024. Insight report* (19th ed.). https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

Annex A

Graph A1. Daily Returns Index (issuer 1) Showing its ESG Epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Graph A2. Daily Returns Index (issuer 2) Showing its ESG Epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.02

Graph A3. Daily Returns Index (issuer 3) Showing its ESG Epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Graph A4. Daily Returns Index (issuer 4) Showing its ESG Epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Graph A5. Daily Returns Index (issuer 5) Showing its ESG Epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Graph A6. Daily Returns Index (issuer 6) Showing its ESG Epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Graph A7. Daily returns index (issuer 7) showing its ESG epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Graph A8. Daily returns index (issuer 8) showing its ESG epicenters

Source: Prepared by the authors with data from Bloomberg L.P. (n.d.).

Annex B

Table B1. Statistical tests results for alpha analysis—Benchmark MEXBOL Index

Probability	EPICENTER 1		
	60	126	255
ISSUER 1	61.40%	49.27%	64.34%
ISSUER 2	44.86%	19.43%	22.29%
ISSUER 3	63.39%	80.94%	N/A
ISSUER 4	21.12%	N/A	N/A
ISSUER 5	7.49%	11.81%	38.94%
ISSUER 6	57.75%	51.04%	25.90%
ISSUER 7	69.27%	N/A	N/A
ISSUER 8	90.63%	42.80%	96.29%

Probability	EPICENTER 2		
	60	126	255
ISSUER 1	17.39%	N/A	N/A
ISSUER 2	70.54%	64.05%	60.27%
ISSUER 3	30.17%	73.21%	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	35.09%	43.30%	85.17%
ISSUER 8	N/A	N/A	N/A

Probability	EPICENTER 3		
	60	126	255
ISSUER 1	N/A	N/A	N/A
ISSUER 2	6.12%	2.98%	39.09%
ISSUER 3	N/A	N/A	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	72.87%	N/A	N/A
ISSUER 8	N/A	N/A	N/A

Source: Prepared by the authors.

Table B2. Statistical tests results for alpha analysis—Benchmark S&P/BMV Sovereign MBONOS 10-20 Year Bond Index

Probability	EPICENTER 1		
	60	126	255
ISSUER 1	79.65%	78.29%	63.04%
ISSUER 2	45.98%	21.90%	60.96%
ISSUER 3	86.10%	71.84%	N/A
ISSUER 4	37.31%	N/A	N/A
ISSUER 5	18.85%	8.37%	46.89%
ISSUER 6	35.71%	60.23%	40.60%
ISSUER 7	46.48%	N/A	N/A
ISSUER 8	77.02%	50.27%	77.52%

Probability	EPICENTER 2		
	60	126	255
ISSUER 1	26.52%	N/A	N/A
ISSUER 2	48.61%	41.59%	69.87%
ISSUER 3	31.91%	43.77%	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	54.27%	31.53%	28.55%
ISSUER 8	N/A	N/A	N/A

Probability	EPICENTER 3		
	60	126	255
ISSUER 1	N/A	N/A	N/A
ISSUER 2	5.14%	12.38%	46.33%
ISSUER 3	N/A	N/A	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	35.47%	N/A	N/A
ISSUER 8	N/A	N/A	N/A

Source: Prepared by the authors.

Annex C

Student's *t*-test for one tail

Student's *t*-test, assuming that the mean average returns *a posteriori* are less than the mean average returns *a priori*, with a significance level of 5%.

H_0: $\mu_i > \mu_j$

H_1: $\mu_i \leq \mu_j$

Such that:

μ_i : the average returns before the epicenter or ESG controversy

μ_j : the average returns after the epicenter or ESG controversy

Results:

The values for *t-statistic* at a 5% level of significance at the different degrees of freedom for the different periods are as follows:

Days	60	126	255
t-statistic (5%)	-1.67064886	-1.65703698	-1.65085109

Considering these *t*-statistics, we fail to reject H0 according to the following tables:

Hypothesis result	EPICENTER 1		
	60	126	255
ISSUER 1	H1	H1	H1
ISSUER 2	H1	H1	H1
ISSUER 3	H1	H1	N/A
ISSUER 4	H1	N/A	N/A
ISSUER 5	H1	H0	H1
ISSUER 6	H1	H1	H1
ISSUER 7	H1	N/A	N/A
ISSUER 8	H1	H1	H1

Hypothesis result	EPICENTER 2		
	60	126	255
ISSUER 1	H1	N/A	N/A
ISSUER 2	H1	H1	H1
ISSUER 3	H1	H1	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	H1	H1	H1
ISSUER 8	N/A	N/A	N/A

Hypothesis result	EPICENTER 3		
	60	126	255
ISSUER 1	N/A	N/A	N/A
ISSUER 2	H1	H1	H1
ISSUER 3	N/A	N/A	N/A
ISSUER 4	N/A	N/A	N/A
ISSUER 5	N/A	N/A	N/A
ISSUER 6	N/A	N/A	N/A
ISSUER 7	H1	N/A	N/A
ISSUER 8	N/A	N/A	N/A

About the authors

Martha León has a PhD in Financial Sciences from Tecnológico de Monterrey, a Fintech certification from the University of Chicago, as well as a specialty in applied statistics at the Instituto de Investigación en Matemáticas Aplicadas y en Sistemas (IIMAS) from the Universidad Nacional Autónoma de México (UNAM). She also has a background in actuarial science at the UNAM. León is the current Financial Vice-president at the National Commission of the Retirement Savings System (CONSAR). She has 18 years of professional experience, focusing on financial institutions in the public sector. In the private sector, she has experience in finance institutions, insurance, financial risk management, and business intelligence. She is a professor at the EGADE Business School of the Tecnológico de Monterrey.

martha_leon@tec.mx
https://orcid.org/0000-0001-5632-4512

Eric Osio serves as General Director of Risk Management at the National Commission of the Retirement Savings System (CONSAR), where he has actively participated, among many other projects and topics, in the migration of the Retirement Savings System to Target Date Funds to improve how pension resources are managed, as well as in the integration of ESG criteria within the regulations applicable to the investment and risk processes of the Retirement Fund Administrators (AFORES). Previously, Eric was part of the team responsible for providing attention to the investment and risk teams within the Mexican Association of Retirement Fund Administrators (AMAFORE), and before that, the Suspicious Activities Investigation team for the prevention of money laundering and financing of terrorism at HSBC. Eric is an Actuary from Anahuac University, has a master's degree in finance from Grenoble École de Management in France, and has the CESGA certification issued by the European Federation of Financial Analysts Societies (EFFAS).

eosio.c@hotmail.com
https://orcid.org/0009-0001-3218-0347

Anahi Revilla has a Bachelor's degree in Economics from the Universidad Nacional Autónoma de México (UNAM) and currently serves as supervisor of the Retirement Savings System (SAR) in the General Directorate of Risk Management at the National Commission of the Retirement Savings System (CONSAR), where she is involved in monitoring and performing analysis on risk management related to the investment regime of the AFORES as well as the development of regulation and integration of ESG criteria. Anahi is involved in various courses in different financial areas and has language skills in English, French, and basic Korean.

anahirean@gmail.com
https://orcid.org/0009-0007-1416-4718

Corporate Sustainability Reporting Compliance Factors: A Panel Data Study of Listed Companies in Peru

Factores de cumplimiento de reportes de sostenibilidad corporativa: un estudio de datos de panel de empresas cotizadas en Perú

Julio Hernández Pajares
Universidad de Piura-Campus Lima, Peru

Yulliana Llauce Ontaneda
Universidad de Piura-Campus Piura, Peru

Macarena Mansilla Mahmud
Universidad de Piura-Campus Lima, Peru

Received: April 8, 2024.
Approved: May 24, 2024.

Abstract

In 2016, the Peruvian Stock Market Superintendency (SMV, by its Spanish acronym) mandated listed companies to publish a Corporate Sustainability Report in the Peruvian stock market as part of corporate governance policies. This research aims to analyze the evolution of compliance levels with the requirements for the sustainability reports established by the SMV as well as the corporate determinants for listed companies in Peru. This quantitative and explanatory research was conducted using a panel data linear regression analysis with fixed effects for the period 2018–2022, to assess the determinant effect of the independent variables on levels of corporate sustainability reporting compliance. The results show an unsteady increase in the level of sustainability reporting compliance until 2022, mainly in social performance. For the regression model with fixed effects on the year variable, the following are determinants of the level of sustainability reporting compliance: company size, belonging to sectors with higher sustainability impact, and voluntary sustainability reporting.

Keywords: *determinants of sustainability reporting, corporate governance, listed companies, Peru, sustainability report.*
JEL Classification: *M14, M48, Q56.*

Resumen

Desde el año 2016, la Superintendencia del Mercado de Valores del Perú (SMV) ha establecido la obligatoriedad para las empresas cotizadas de publicar el reporte de sostenibilidad corporativa en el mercado de valores peruano como parte de las políticas de gobierno corporativo. Esta investigación tiene como objetivo analizar la evolución del nivel de cumplimiento de los requerimientos para la elaboración de los reportes de sostenibilidad exigidos por la SMV y analizar los determinantes corporativos para las empresas cotizadas en Perú. Este estudio, cuantitativo y explicativo, se realizó mediante un análisis de regresión lineal de datos de panel con efectos fijos para el periodo 2018-2022, con el fin de evaluar el efecto determinante de las variables independientes sobre el nivel de cumplimiento de los reportes de sostenibilidad corporativa. Los resultados muestran un aumento no sostenido en el nivel de cumplimiento de los informes de sostenibilidad hasta 2022, principalmente en el desempeño social. Para el modelo de regresión con efectos fijos sobre la variable año, los determinantes del nivel de cumplimiento de reporte de sostenibilidad son los siguientes: el tamaño de la empresa, la pertenencia a sectores con mayor impacto en la sostenibilidad y la elaboración voluntaria de reportes de sostenibilidad.

Palabras clave: *determinantes de reporte de sostenibilidad, gobierno corporativo, empresas cotizadas, Perú, reportes de sostenibilidad.*
Clasificación JEL: *M14, M48, Q56.*

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.03

1. Introduction

In 2016, the Peruvian Stock Market Superintendency (SMV, by its Spanish acronym), the regulatory body for companies listed on the Lima Stock Exchange (BVL, by its Spanish acronym), mandated the publication of a Corporate Sustainability Report as part of the non-financial information requirements for environmental and social performance and compliance with sustainability standards. Other countries also have such a reporting requirement set by regulatory bodies, given the importance of transparency for stakeholders. Furthermore, much academic research has been done on levels of transparency and the quality of sustainability disclosure in the stock markets (Christensen et al., 2021; Gerwing et al., 2022; Venturelli et al., 2017). Research on mandatory sustainability reporting indicates that its purpose is to provide more objective, comparable, and verifiable information, enabling stock market users to make more efficient decisions, as pointed out by studies in stock market-regulated countries (Gulenko, 2018; Matuszak & Rozanska, 2017; Roy et al., 2022).

The literature considers institutional theory as explaining the motivations for sustainability reporting influenced by public and private regulatory bodies, where, as a result of coercive and normative isomorphisms, companies seek to emulate other companies in terms of disclosure (Powell & DiMaggio, 1991; Scott, 1987; Zucker, 1987). Mandatory non-financial reporting has led to a rise in the publication of separate or integrated sustainability reports that, based on an institutional influence, follow regulatory compliance models, and seek to respond to the economic, social, or environmental interests of stakeholders, as shown in the studies of Agostini et al. (2022), Carungu et al. (2021), Posadas & Tarquinio (2021), and Venturelli et al. (2017).

However, other research, such as that conducted by Costa & Agostini (2016), Larrinaga et al. (2002), and Luque-Vílchez & Larrinaga (2016), argues that mandatory reporting has not improved the quality and relevance of sustainability information. Rather, it has resulted only in a regulatory and compliance reform, not in organizational and sustainability management change or in engagement with stakeholders. Furthermore, while sustainability reporting should consider both financial and sustainability materiality, there is a risk that the reporting is limited to investors, with less accountability to other stakeholders (Carrasco et al., 2022; Macias & Ficco, 2022).

In Latin America, the adoption of voluntary corporate sustainability reporting and practices has developed significantly in recent years. However, research findings show that reporting levels have not reached international compliance standards due to a lack of sustainability strategies in companies, an incipient institutional normative, as well as the coercive influence of stock market regulatory bodies that have not made adequate information available to assess sustainability risks and opportunities in stakeholder decisions (Aranguren & Maldonado, 2019; Hernández-Pajares, 2023; Moneva et al., 2019, Sepúlveda-Alzate et al., 2022).

In Peru, sustainability reporting studies are still in development, and have mainly focused on voluntary sustainability reporting in accordance with the Global Reporting Initiative (GRI) standards (Aubert & Venegas, 2022; Bunclark & Barcellos-Paula, 2021; Díaz-Becerra et al., 2021; Pocomucha & Venegas, 2021). Moreover, mandatory non-financial sustainability reporting studies have focused on analyzing reporting factors. It is thus considered important to continue this line of research regarding the evolution of sustainability transparency by listed and regulated companies in Peru (Hernández-Pajares & Pocomucha, 2021; Caballero et al., 2019).

The objective of this research is therefore to analyze the evolution of compliance with the requirements of the sustainability report established by the SMV and corporate determinants for the period 2018-2022 for listed companies in Peru.

2. Theoretical Background

2.1 Institutional Influence

Regulation of sustainability information in public markets by regulatory and governmental bodies, such as the European Union (EU), has had a significant institutional influence on companies' levels of transparency. However, although the amount of information provided has increased, it has not reached the levels of quality voluntary information established by, for example, the GRI standards (Agostini et al., 2022; Caputo et al., 2020; Carungu et al., 2021; Ottenstein et al., 2022).

The first type of institutional influence on sustainability reporting is normative. In keeping with this, companies improve their voluntary sustainability reporting based on normative compliance, similar to other companies, in accordance with international standards, and compete to achieve stakeholder legitimacy and influence (Carini et al., 2018; Christensen et al., 2021; Masoud & Vij, 2021). This has been impacted locally by professional associations and internationally, by, for example, the GRI standards (Fortanier et al., 2011; Neu et al., 1998).

Furthermore, mandatory sustainability reporting has had a coercive institutional influence, as companies comply with specific reporting requirements due to pressure from regulatory bodies to improve accountability, essentially in response to investment risk assessment decisions by stock market investors. Nevertheless, improvements in sustainability strategies and practices by companies for better transparency are still required (Carungu et al., 2021; De Villiers & Alexander, 2014; Gulenko, 2018; Jackson et al., 2019; Posadas & Tarquinio, 2021).

2.2 Reporting Factors

In addition to institutional influences on the mandatory sustainability reporting of regulated companies, studies have also addressed other corporate factors, such as levels of resource investment, performance, and relevance of the influence of previous disclosure experiences, as determinants of such reporting (Bergmann & Posch, 2018; Costa & Agostini, 2016). In this sense, studies, such as those conducted by Aragón-Correa et al. (2020) and Balluchi et al. (2021), have found that regulatory pressures, in conjunction with voluntary pressures, positively influence improvement and innovation in sustainability strategies and performance in environmental aspects rather than these merely being a response to prevent sanctions.

Besides corporate aspects of company size and profitability, other research has found additional influencing factors, such as stock index rating, belonging to business sectors with a higher sustainability impact, and voluntary reporting experience, as seen in the studies of Mion & Loza (2019); Gerwing et al. (2022); Radu et al. (2023), Venturelli et al. (2017) and Wang et al. (2018).

Regarding company size, the greater the resources and capacity available for sustainability management, the more likely companies will implement sustainability aspects and reporting to improve their legitimacy and reputation with stakeholders. Larger companies seek to enhance their global reputation and consider mandatory sustainability reporting as an opportunity to boost the confidence of their investors and reduce information asymmetry and agency costs, as well as to improve their reputation in the market (Bergmann & Posch, 2018; Gerwing et al., 2022; Mio et al., 2020; Mion & Loza, 2019; Roy et al., 2022).

Research on the relationship between economic performance and levels of sustainability disclosure within a context of mandatory regulation is not conclusive. In some cases, a positive and significant relationship between profitability and levels of mandatory sustainability reporting as a form of legitimacy for stakeholders have been found (Abdul Rahman & Alsayegh, 2021; Hernández-Pajares & Pocomucha, 2021). However, in other cases, a negative relationship is found, that is, transparency is not due to better performance but to the search for reputation (Mion & Loza, 2019; Masoud & Vij, 2021). Finally, no significant relationship has also been found between performance and reporting levels, that is, better-performing companies do not necessarily seek sustainability transparency to ensure financing through stock market investors (Balluchi et al., 2021; Gerwing et al., 2022; Mio et al., 2020).

Studies on companies that implement activities with a high environmental and social impact show that these companies report more sustainability information to legitimize their activities with stakeholders affected by activities such as manufacturing, power generation, mining, and other natural resource consumption activities. Such companies seek to legitimize their operations, resulting in this factor being more representative than institutional influence (Aranguren & Maldonado, 2019; Balluchi et al., 2021; Caputo et al., 2020; Mio et al., 2020).

Studies of mandatory regulation, such as those conducted by Balluchi et al. (2021), Doni et al. (2020), Matuszak & Rozanska (2017), and Schröder (2022), indicate that previous experience of sustainability management and voluntary sustainability reporting, in most cases under GRI standards, results in a better level of mandatory reporting compliance and an increased capacity to adapt and prepare quality mandatory sustainability information. In addition, membership of companies in certain stock indexes, such as the S&P and sustainable indexes, is a determinant of the level of mandatory sustainability reporting (Loza-Adaui, 2020; Ioannou & Serafeim, 2017).

Based on the previous theoretical background, we propose the following hypotheses:

H1. Company size is a positive and significant factor in levels of corporate sustainability reporting compliance.

H2. Company profitability is a positive and significant factor in levels of corporate sustainability reporting compliance.

H3. Belonging to sectors with a higher sustainability impact is a positive and significant factor in levels of corporate sustainability reporting compliance.

H4. Voluntary sustainability reporting is a positive and significant factor in levels of corporate sustainability reporting compliance.

H5. Membership in the S&P index in the BVL is a positive and significant factor in levels of corporate sustainability reporting compliance.

3. Methodology

Our research employed a quantitative approach of explanatory scope. The ordinary least squares (OLS) panel data regression estimation method was implemented, using a cluster at the company level for the period 2018-2022, to assess the determinant effect of the corporate variables of size, profitability, business activity with sustainability impact, voluntary sustainability reporting, and S&P membership on level of corporate sustainability reporting compliance. The year-fixed effect model regression analysis was also employed. The factor analysis in this study used STATA 17 software. A qualitative content analysis approach was also used on the compliance of disclosed aspects of sustainability performance by the companies regarding the new requirements for the report established in 2020.

Company size was considered the first independent variable for the explanatory study, measured by the natural logarithm of the value of assets for the periods prior to the reporting period (Braam et al., 2016; Gerwing et al., 2022). The variable of profitability was measured by the value of return on assets (ROA) for the 2017-2021 periods prior to the reporting period (Aboagye-Otchere et al., 2020; Braam et al., 2016; Mio et al., 2020). Type of activity reflected the BVL sectors: financial, industrial, mining, energy, services, and trade. These were classified into a categorical variable of two groups, one including the sectors with significant sustainability impact, such as industry, mining, and energy, with a value of 1, and a second that included sectors without significant sustainability impact, such as services, finance and trade, with a value of 0 (Caputo et al., 2020; Mio et al., 2020). The variable of voluntary reporting was calculated using a dichotomous variable with a value of 1 if voluntary reporting was done (GRI or non-GRI) for the study periods and 0 if voluntary reporting was not done (Balluchi et al, 2021; Schröder, 2022). Membership in the S&P in the BVL was valued using a dichotomous variable with a value of 1 if the company was listed in the S&P categories for the study periods and 0 if it was not listed (Loza-Adaui, 2020).

The dependent variable is the sustainability reporting compliance level for 2019-2021. First, a categorical variable was measured, with a value of 0 if the requirement of sustainability reporting indicators was not reported and 1 if it was reported. The dependent variable was determined for the regression analysis by the average influence rate of total required indicators for each company (Braam et al., 2016; Carini et al., 2018; Jackson et al., 2019).

To better understand the variable of reporting compliance level, it is necessary to explain the content of the SMV-issued regulations, whose requirements are classified into environmental performance, suppliers, other stakeholders, labor performance, human rights, and standards. The main difference with the new 2020 report is compliance with environmental performance, labor performance, human rights, and voluntary reporting. Table 1 shows the requirements for the reports issued in 2020 and in 2016 and their equivalencies for each category (see Table 1).

Table 1. Comparison of Requirements for 2020 and 2016 Reports

Performance	2020 Report	2016 Report
Environmental	**Environmental management system or environmental policy**	**Environmental impact-related corporate policy**
	Details of investigation, imposition of corrective measures, affecting environmental standards	
	• Greenhouse gas (GHG) emission measurement	
	• GHG emissions reduction targets	• Quantification of GHG emissions
	• Water consumption measurement	• Quantification and documentation of total water consumed
	• Water footprint measurement	
	• Water consumption reduction targets	
	• Effluent control mechanisms	
	• Energy consumption measurement	• Quantification and documentation of total energy use
	• Energy consumption reduction targets	
	• Solid waste measurement	• Quantification and documentation of waste generated
	• Waste management targets and goals	

Performance	2020 Report	2016 Report
Suppliers	• Inclusion of social, environmental, and corporate governance aspects in supplier selection criteria	• Updated supplier registry
		• Criteria for selecting suppliers and complying with labor legislation
		• Policy for selecting suppliers that comply with environmental standards • Policy for managing the relationship with suppliers
Other stakeholders	• Identification of risks and opportunities related to stakeholders • Details of any significant controversy with any stakeholders	• Work in collaboration with the community
		• Investment in social programs where the main activities are carried out
		• Community interaction policies • Specify the social conflicts in the community where it operates
		• Customer management policy
		• Customer complaint registry
		• Permanent public service channels or means
		• Recognition of quality in the service provided to customers

Performance	2020 Report	2016 Report
Labor	• Details of labor policy	• Labor policy and fundamental employee rights
	• Details of investigation, imposition of corrective measures or fines related to non-compliance with labor, health and safety, or child labor regulations	
	• Evaluation of compliance with occupational health and safety regulations	
	• Accidents at work registry	• Occupational accident registry
	• Measuring your work environment	• Work environment evaluation
	• Worker talent management policy	• Worker training plan
	• Procedures for detecting and sanctioning workplace hostility and sexual harassment	
Human rights	• Compliant or grievance handling policy or system to reduce the impact on human rights	
	• Training plan on human rights topics	
Standards and reporting	• International corporate sustainability-related certification	• Sustainability standards
	• Corporate Sustainability Report other than mandatory report	

Source: Prepared by the authors, based on SMV (2016) and SMV (2020).

Based on the above variables, a multiple linear panel data regression model was designed considering a fixed effect per year. For the companies, the correlation of observations for the five consecutive years was monitored using a cluster at the company level observed five times. Thus, the standard errors are robust and do not show significant changes. The regression model was as follows:

$$\begin{aligned} \textit{Reporting rate}_{i,t} \\ &= \beta_0 + \beta_1\,\textit{Size}_{i,\,t\text{-}1} + \beta_2 \textit{ROA}_{i,t\text{-}1} + \beta_3\,\textit{Type of industry}_{i,\,t} \\ &\quad + \beta_4\,\textit{Voluntary reporting}_{i,t} + \beta_5 \textit{S\&P}_{i,t} + \varepsilon_{i,t} \end{aligned}$$

Where: i = company, t = year (2022, 2021, 2020, 2019, 2018), t-1= year prior to t (2021, 2020, 2019, 2018).

The Hausman test was performed to determine whether the panel data regression model fits a year-fixed effect. The results showed that the coefficients of the independent variables remain constant for each individual company in the panel data and do not vary with time and the observed units. The individual heterogeneity of each case in the model was monitored by applying the fixed effects model in panel data regressions for the analysis of sustainability disclosure factors, such as in the study by Wang (2017).

The sample calculation was based on a total population of 260 companies listed on the BVL in 2022. This population was adjusted by excluding investment fund companies, securitization process management companies, and companies that did not publish sustainability reports, leaving a total of 220 companies. From the adjusted population, stratified random sampling by sector was carried out with a confidence level of 95% and a margin of error of 5%, resulting in a sample of 116 companies, as detailed in Table 2. It can be seen that the most representative companies belong to the financial, industrial, and energy sectors (see Table 2).

Table 2. Sample of companies by sector

Business sector	**Frequency**	**Percentage**
Pension management companies	3	2 %
Agro-industrial	9	8 %
Banks and finance companies	26	22 %
Trade	4	4 %
Energy and oil	13	12 %
Industrial	18	16 %
Real estate/Construction	9	7 %
Mining	11	10 %
Insurance	12	10 %
Services	11	9 %
Total	116	100 %

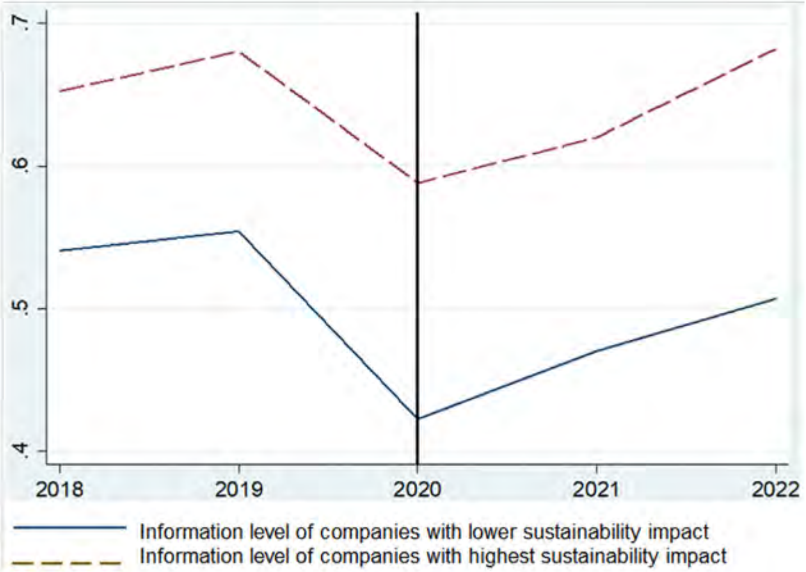
Source: Prepared by the authors.

4. Results

4.1 Descriptive Analysis

Figure 1 shows that the mean of compliance with sustainability reporting requirements does not show a steady increase (see Figure 1). The compliance level rose from 2018 to 2019, decreased for 2020, and increased again for 2021 and 2022. This represents an improvement in reporting levels, with the first report in 2019 and a second in 2022, with the new report comprising more reporting requirements for each environmental, labor, and human rights category. It is worth noting that reporting compliance was higher for companies with significant sustainability impact, such as mining, energy, and industrial companies, compared to trade, finance, and service companies, with a marked increase in 2022 compared to 2021.

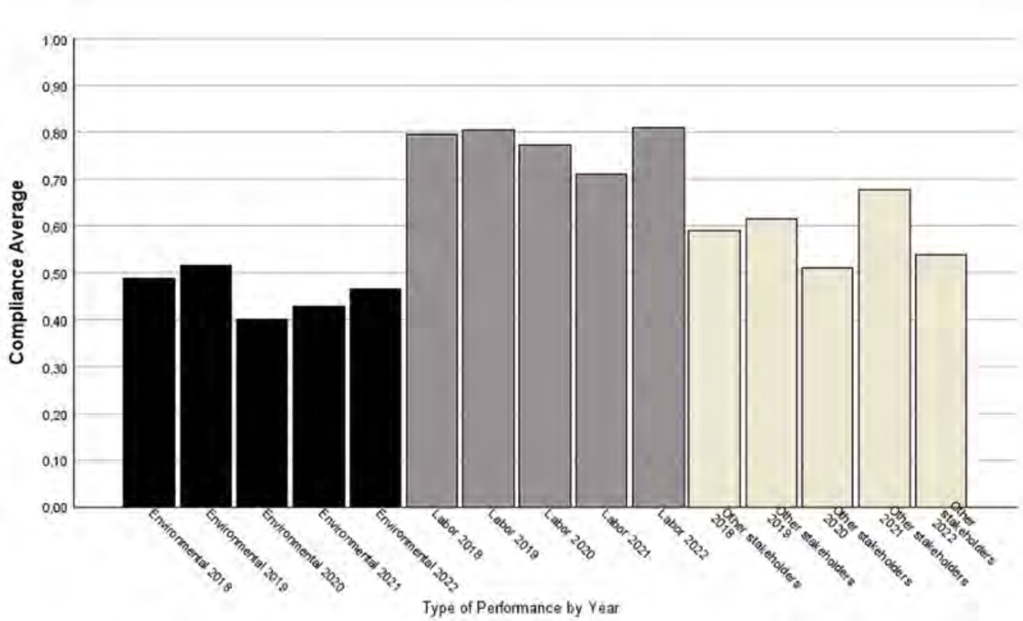
Figure 1. Annual Evolution of the Mean of Sustainability Reporting Compliance



Source: Prepared by the authors.

The analysis of the evolution of reporting compliance by environmental, labor, and other stakeholder categories, shown in Figure 2, did not find a sustained increase during the period under study (see Figure 2). The means of compliance with environmental practices were lower than for labor performance and other stakeholders, such as suppliers, customers, and the community. Although the environmental performance reporting requirements did not show the highest compliance, it is worth highlighting that energy, water, and waste savings management, as well as carbon or water footprint measurements were the least disclosed.

Figure 2. Means of Reporting Compliance by Sustainability Performance Categories

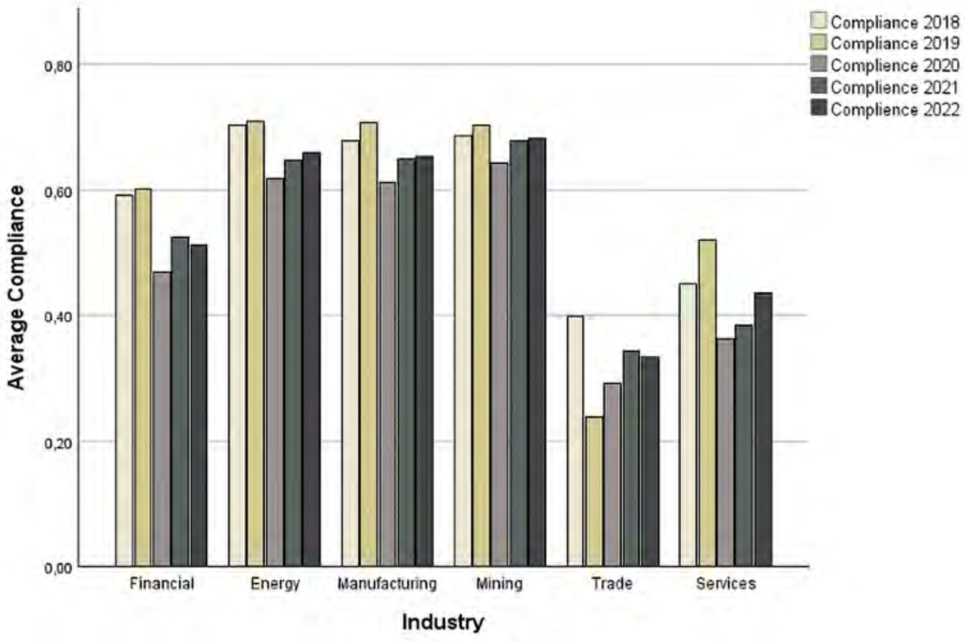


Source: Prepared by the authors.

Social reporting on performance and workers in occupational health and safety, and labor climate management aspects had the highest levels of compliance. Reporting on management and other stakeholders, such as suppliers, in environmental and labor compliance evaluations are particularly noted due to regulatory influence and because they are direct stakeholders in the operational management of companies (Carini et al., 2018; Gulenko, 2018). Labor performance is fundamental as a sustainable practice in almost all types of activities, as pointed out in the studies of Aranguren & Maldonado (2019), Christensen et al. (2021), and Korca et al. (2021).

Figure 3 shows levels of compliance with sustainability reporting requirements by business activity (see Figure 3). Regarding environmental performance, we note that the sectors with the highest level of compliance are energy, industry, and mining, corresponding to companies with a higher environmental and social impact that are also strongly oriented to sustainability performance and sustainability reporting on environmental practices by their productive activity, as pointed out in the studies of Balluchi et al. (2021), Carini et al. (2018), and Caputo et al. (2020).

Figure 3. Means of Sustainability Category Reporting Compliance by Business Activity



Source: Prepared by the authors.

The significant reporting compliance for companies with high sustainability impact is mainly due to the labor performance of their large number of workers. However, the financial sector also has a significant level of reporting on performance with workers (Korca et al., 2021; Schröder, 2022). For sustainability reporting of performance with other stakeholders, sustainable value chain management with suppliers in productive activities should be highlighted, especially management with communities for mining, industrial, and energy companies, for which the social impact and search for legitimacy is significant, as found in studies conducted by Ivic et al. (2021), Loza-Adaui (2020) and Pocomucha & Venegas (2021).

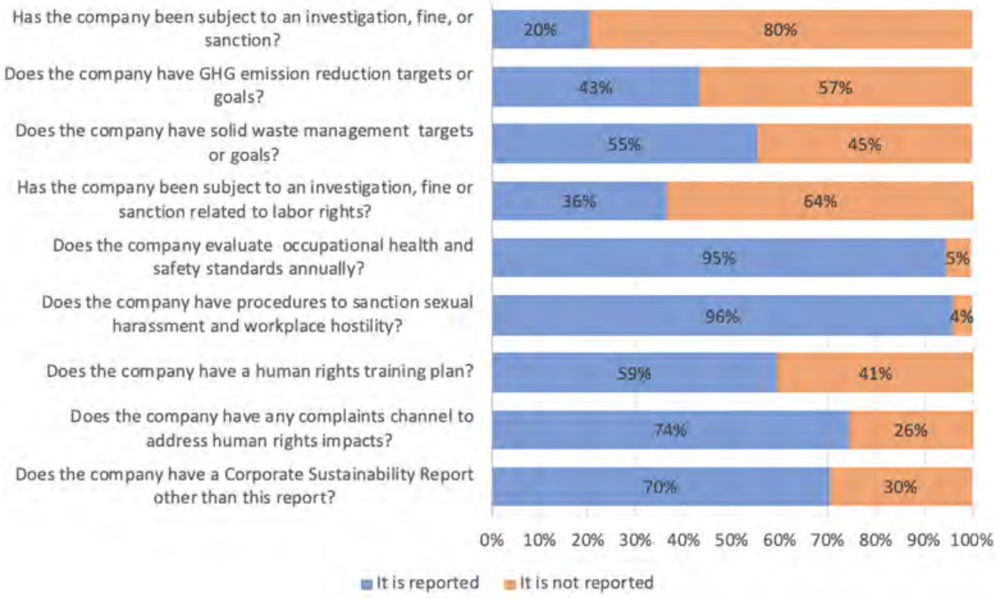
To better analyze the evolution of sustainability reporting during the period under study, we analyzed the information content that resulted from the changes required by the new 2020 report. Figure 4 shows the mean of compliance with the new reporting requirements for the period 2021, grouped by key axes of the new sustainability report (see Figure 4). The means for compliance with environmental performance reporting are the lowest. It was found that most companies have not defined greenhouse gas (GHG) reduction or solid waste management targets or goals.

There are also companies that only report a series of initiatives they have adopted to contribute to the care of the environment, such as energy savings, recycling of non-hazardous materials, efficient use of resources, and awareness of their employees on the responsible use of resources. It is important to mention that while some companies claim to have implemented management control systems, they do not report indicators related to environmental management, specifically GHG reduction and solid waste management. Commitments are reported, but there are no specific goals for or measurements of environmental matters. In addition, it is noted that companies report targets with little mention of measurable and relevant compliance practices that should be disclosed in more detail in the reports.

Regarding social and labor performance disclosure, almost all companies report performing annual evaluations of compliance with occupational health and safety policies, which can be evidenced by reported legal compliance. Additionally, companies report to have implemented procedures to identify and sanction sexual harassment and workplace hostility. This is also due to regulatory compliance rather than to objectives of labor management with employees.

As shown in Figure 4, human rights and labor rights performance are the most significant. Ninety-six percent of the companies report procedures to sanction harassment and hostility at work, 59% report that they have a human rights training plan, and 74% report that they have a channel for complaints regarding human rights violations. On the other hand, 95% of the companies disclose annual evaluations of safety and health occupational standards. Finally, 70 % of the companies report that they prepare voluntary sustainability reports, with some indicating that these are prepared in accordance with the GRI (see Figure 4).

Figure 4. Compliance with New Mandatory Sustainability Reporting Requirements for the Period 2021



Source: Prepared by the authors with data from the report content analysis.

4.2 Regression Analysis

Table 3 shows the descriptive statistics of the variables for the 116 companies studied and for each year of the regression model (see Table 3). The analysis of the variables reveals different patterns for the reporting level, ROA, and asset size. The degree of reporting compliance shows some variability over the years, with means that do not reflect full compliance. The reporting level has changed due to the need for companies to adapt to each report by learning and implementing sustainability and communication objectives. Furthermore, ROA shows a notable fluctuation over time. For example, 2020 has extremely negative values, possibly reflecting significant financial problems experienced by certain companies during the COVID-19 pandemic (Rosales et al., 2021). In contrast, the company size, as measured by the logarithm of the volume of assets, appears to remain relatively stable over the years, with consistent means and modest standard deviations, indicating stability in the size of company assets during the period considered.

Table 3. Regression Descriptive Statistics

Variable	N	Minimum	Maximum	Mean	Standard deviation
Reporting level 2018	116	0	1	0.5985502	0.2739751
Reporting level 2019	116	0	1	0.6195468	0.2652678
Reporting level 2020	116	0	1	0.5082615	0.2606557
Reporting level 2021	116	0	1	0.5477730	0.2727706
Reporting level 2022	108	0	1	0.5977832	0.2360786
ROA 2018	116	-0.268779	0.697357	0.0553307	0.1031582
ROA 2019	116	-0.418815	0.287353	0.0366210	0.0856114
ROA 2020	116	-50.30010	0.225695	-0.6375204	5.2045650
ROA 2021	116	-2.202250	0.331103	0.0188849	0.2576616
ROA 2022	116	-2.202250	0.331103	0.0188849	0.2576616
Asset size 2018	116	8.357024	18.75715	13.827200	1.7732600
Asset size 2019	116	8.567316	18.82615	13.864760	1.7699830
Asset size 2020	116	8.228711	19.11383	13.858440	1.9093500
Asset size 2021	116	8.254008	19.06253	13.916900	1.9409000
Asset size 2022	116	8.780941	19.00847	13.94586	1.9402960

Source: Prepared by the authors with data obtained from information in financial reports using STATA 17 software.

Table 4 shows the correlation analysis used to evaluate whether the independent variables of the model showed multicollinearity problems (see Table 4). The test results showed correlation coefficients that exceeded the 0.5 threshold, suggesting a low linear association between the variables.

Table 4. Correlation of Regression Variables

	ROA	Size	S&P	Sector type	Reporting level	Year
ROA	1.0000					
Size	0.1652	1.0000				
S&P	0.0239	0.2540	1.0000			
Sector type	-0.0084	-0.1144	0.2864	1.0000		
Reporting level	0.0607	0.4652	0.3432	0.1770	1.0000	
Year	-0.0055	0.0220	-0.0000	0.0000	0.0000	1.0000

Source: Prepared by the authors with data obtained from STATA 17 software.

Table 5 shows the regression results of the panel data regressions under two models (see Table 5). The first column presents the OLS model without fixed effects. It can be observed that the asset size has a significant and positive effect with a coefficient of 0.0243, meaning that larger companies tend to have a higher sustainability reporting rate. The first hypothesis is thus accepted with a significance level of less than 0.001.

However, ROA has a negative and significant coefficient of -0.0061, meaning that unprofitable companies have a higher reporting level than better-performing companies. The second hypothesis of a positive relationship between profitability and reporting rate was therefore rejected.

As observed in the descriptive analysis, the variable of the business sector with higher sustainability impact has a significant and positive influence with a coefficient of 0.0920 on the reporting level. The third hypothesis was thus accepted with a positive significance level of less than 0.001, meaning that companies with higher environmental and social impact have greater disclosure.

Table 5. Multiple Linear Regression Panel with Fixed Effects

	(1) Reporting rate (MCO)	**(2) Reporting rate (Fixed effects)**
Log (Asset size)	0.0243*** (4.36)	0.0129* (1.04)
ROA	-0.00608*** (-6.00)	-0.00107** (-0.49)
Sustainability impact sector	0.0920***	0.1490***
(1=belongs)	(4.93)	(4.95)
Voluntary reporting (1=Reports)	0.243*** (11.93)	0.0798* (11.36)
S&P(1=membership)	0.0998*** (4.79)	0.0244 (3.78)
Year	-0.0106 (-1.85)	-0.0161*** (-5.18)
Constant	21.51 (1.86)	32.87*** (5.26)
N	572	572
R^2	0.444	0.536

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Prepared by the authors with data from the content analysis of sustainability and financial reports using STATA 17 software.

Likewise, voluntary reporting by companies was a positive and significant factor of the sustainability reporting rate. The fourth hypothesis is accepted with a significance level of less than 0.001, meaning that the companies with voluntary sustainability reporting, such as the GRI, exhibited higher mandatory reporting compliance. Similarly, the variable of membership in the S&P index has a significant and positive impact. The fifth hypothesis is thus accepted with a significance level of less than 0.001, meaning that companies with membership in the S&P index have a higher level of sustainability reporting.

The second column presents the regression model with fixed effects. The results differ slightly but confirm those obtained in the OLS model. The fixed effect model was more representative as it considered the heterogeneity of the companies.

Company size continued to show a positive and significant effect, with a lower coefficient of 0.0129. As a result, acceptance of the first hypothesis, that larger companies had higher disclosure, is maintained, with a significance of less than 0.05. As the coefficient for ROA continues to be negative and less significant at -0.0011, the second hypothesis is still rejected.

The sustainability impact of the sector maintained a significant, positive coefficient greater than 0.1490, so the third hypothesis is accepted with a significance level of less than 0.001. In addition, voluntary reporting also maintained a positive and significant effect on the level of disclosure with a coefficient of 0.0798. The fourth hypothesis is thus still accepted with a significance level of 0.05. The variable of S&P is the only variable that was no longer statistically significant in the fixed effect model, inhibiting acceptance of the fifth hypothesis.

5. Discussion of Results

The results of this research found an unsteady increase in levels of sustainability reporting compliance for Peruvian listed companies as demonstrated in the descriptive statistics in Table 2. As shown in Figure 2, the greatest disclosure of information on sustainability corresponded to social performance, mainly related to benefits with workers, with an average compliance of 70% to 80%, and management with other stakeholders, with an average compliance of 50% to 70%, related to operating activities. However, information on environmental practices had lower levels of disclosure, with an average of only 40% to 50%, focused mainly on water and energy saving management and waste management. This indicates that greater resources and knowledge are required for the implementation of environmental strategies and goals for better environmental information levels that do not only correspond to compliance with reporting requirements (Balluchi et al., 2021; Caputo et al., 2020; Posadas & Tarquinio, 2021).

Thus, for the new 2020 sustainability reporting requirements, it was found that companies have better implemented and disclosed their policies and procedures related to labor and human rights (Carini et al., 2018; Gulenko, 2018). However, as shown in Figure 2, the lowest levels of compliance corresponded to environmental requirements for companies still in the learning phase in terms of environmental performance and information. Companies with greater environmental impact and information did not report environmental objectives or management in sufficient detail in the sustainability report (Aranguren & Maldonado, 2019; Sepúlveda-Alzate et al., 2022). A minimal approach to reporting compliance requirements was found, based on coercive isomorphism resulting from institutional influence on mandatory reporting. Greater proactivity in sustainability performance is required to ensure better transparency (Aragon-Correa et al., 2020; Bergman & Posch, 2018; Radu et al., 2023).

Considering the above, not only did institutional influence impact levels of disclosure, but type of industry was also a determinant of sustainability reporting compliance. Companies with high environmental and social impact activities, such as mining, energy, and manufacturing, justified a higher level of disclosure to legitimize their activities with the information required by mandatory reporting with stakeholders (Aranguren & Maldonado, 2019; Bergman & Posch, 2018; Fortanier et al., 2011; Mion & Loza, 2019; Sepúlveda-Alzate et al., 2022).

Company size was a determinant of sustainability reporting compliance for the companies studied. According to the agency theory, large companies, with better share prices and corporate governance policies, are expected to have better sustainability performance and information as a way of legitimizing the financial decisions of investors (Christensen et al., 2021; Gerwing et al., 2022; Mio et al., 2020; Roy et al., 2022). The profitability effect in this study was negative on the reporting level. This result can be explained by the fact that highly profitable companies consider mandatory sustainability reporting as a regulatory compliance due to coercive institutional influence rather than as a way to obtain financing in the stock market (Hernández-Pajares & Pocomucha, 2021; Loza-Adaui, 2020), or because less profitable companies seek legitimacy in the eyes of their stakeholders with more sustainability information (Masoud & Vij, 2021; Mio et al., 2020).

Finally, previous experience in voluntary sustainability management and sustainability reporting following the GRI standards or others as a determinant of the level of mandatory sustainability reporting was confirmed. The listed companies with more experience in environmental and social management and voluntary reporting have a better level and quality of mandatory sustainability reporting (Bergman & Posch, 2018; Doni et al., 2020; Mion & Loza, 2019; Schröder, 2022).

6. Conclusions

This study analyzed the evolution of compliance with mandatory sustainability reporting requirements and corporate determinants for listed companies in Peru.

The results did not show full compliance with sustainability reporting requirements during the period studied. This information impacts the declaration of environmental and social commitments. It shows low levels of disclosure of sustainability strategies and management due to a lack of experience or institutional influence. Compliance tends to be limited to minimum reporting requirements (Loza-Adaui, 2020; Mion & Loza, 2019; Venturelli et al., 2017).

Although improvements are evident with the new report in terms of labor and human rights information, sustainability reports present little detail on strategic sustainability management. This may be due to the lack of requirements for this reporting that include a compliance focus and description of practices. Sustainability reporting may not be a major driver of change at organizations in sustainability performance in the period studied. This confirms mainly an institutional influence based on coercive isomorphism for compliance with reporting requirements (Carini et al., 2018; Posadas & Tarquinio, 2021).

The low levels of sustainability information provided by the companies suggest that Peruvian companies may have less incentive to provide sustainability information for users who need to make financial investment decisions and evaluate risks, as pointed out in the studies conducted by Roy et al. (2022) and Wang & Li (2016). It is considered that there is a need for greater institutional influence of the regulatory bodies in the Peruvian stock market to develop reporting regulations in line with international standards, such as the GRI, and integrated reports under ESG (Environmental, Social, Governance) criteria that include long-term strategic corporate sustainability objectives and that do not result in simple regulatory reporting standards. The participation of regulatory bodies and business and professional associations should exert greater influence on the regulation of corporate sustainability disclosure, both to shareholders and other stakeholders, as suggested by Costa & Agostini (2016) and Larrinaga et al. (2002).

The implications of this study suggest the need for future qualitative content analysis research to assess not only compliance with requirements, but also the quality of sustainability information in separate or integrated sustainability reports based on international standards and ESG criteria, as well as to analyze the influence of sustainability reporting on financial performance and value of companies in the stock market.

References

Abdul Rahman, R. & Alsayegh, M. F. (2021). Determinants of corporate environment, social and governance (ESG) reporting among Asian firms. *Journal of Risk and Financial Management*, 14(4), 167. https://doi.org/10.3390/jrfm14040167

Aboagye-Otchere, F. K., Simpson, S. N. Y. & Kusi, J. A. (2020). The influence of environmental performance on environmental disclosures: An empirical study in Ghana. *Business Strategy & Development*, 3(1), 98-111. https://doi.org/10.1002/bsd2.81

Agostini, M., Costa, E. & Korca, B. (2022). Non-financial disclosure and corporate financial performance under directive 2014/95/EU: evidence from Italian listed companies. *Accounting in Europe*, 19(1), 78-109. https://doi.org/10.1080/17449480.2021.1979610

Aragón-Correa, J. A., Marcus, A. A. & Vogel, D. (2020). The Effects of Mandatory and Voluntary Regulatory Pressures on Firms' Environmental Strategies: A Review and Recommendations for Future Research. *Academy of Management Annals*, 14(1), 339-365. https://doi.org/10.5465/annals.2018.0014

Aranguren, N. & Maldonado, S. (2019). Governance and Type of Industry as Determinants of Corporate Social Responsibility Disclosures in Latin America. *Latin American Business Review*, 21(1) 1-35. https://doi.org/10.1080/10978526.2019.1697185

Aubert, D., & Venegas C. (2022). Factores influyentes en grado de cumplimiento GRI de reportes de sostenibilidad de empresas peruanas cotizadas. *CAPIC REVIEW*, 20, 1-16. https://doi.org/10.35928/cr.vol20.2022.173

Balluchi, F., Lazzini, A. & Torelli, R. (2021). Credibility of environmental issues in non-financial mandatory disclosure: Measurement and determinants. *Journal of Cleaner Production*, 288, article 125744. https://doi.org/10.1016/j.jclepro.2020.125744

Bergmann, A. & Posch, P. (2018). Mandatory Sustainability Reporting in Germany: Does Size Matter? *Sustainability*, 10(11), 3904. https://doi.org/10.3390/su10113904

Braam, G., De Weerd, L., Hauck, M. & Huijbregts, M. (2016). Determinants of corporate environmental reporting: The importance of environmental performance and assurance. *Journal of Cleaner Production*, 129, 724-734. https://doi.org/10.1016/j.jclepro.2016.03.039

Bunclark, L. & Barcellos-Paula, L. (2021). Sustainability reporting for sustainable supply chain management in Peru. *Sustainable Production and Consumption*, 27, 1458-1472. https://doi.org/10.1016/j.spc.2021.03.013

Caballero, A., Valdivieso, L., Álvarez, A. L., Beaumont, M., Vásquez, D., Llaxacondor, A., ... & Rouillón, E. (2019). *La calidad de la información de los reportes de sostenibilidad corporativa en el mercado de valores peruano: un enfoque en el sector* financiero. Programa de Inversión Responsable (PIR). Pontificia Universidad Católica del Perú (PUCP). https://pir.pe/wp-content/uploads/Reporte-PIR-2018.pdf

Carini, C., Rocca, L., Veneziani, M. & Teodori, C. (2018). Ex-ante impact assessment of sustainability information–the directive 2014/95. *Sustainability*, 10(2), 560. https://doi.org/10.3390/su10020560

Caputo, F., Leopizzi, R., Pizzi, S. & Milone, V. (2020). The non-financial reporting harmonization in Europe: Evolutionary pathways related to the transposition of the Directive 95/2014/EU within the Italian context. *Sustainability*, 12(1), 92. https://doi.org/10.3390/su12010092

Carrasco, F., Correa, C. & Larrinaga, C. (2022). Evolución de la contabilidad social y medioambiental: reflexiones sobre el papel constitutivo de la contabilidad. *Contaduría Universidad de Antioquia*, 81, 165-186. https://doi.org/10.17533/udea.rc.n81a07

Carungu, J., Di Pietra, R. & Molinari, M. (2021). Mandatory vs Voluntary Exercise on Non-Financial Reporting: Does a Normative/coercive Isomorphism Facilitate an Increase in Quality? *Meditari Accountancy Research*, 29(3), 449-476. https://doi.org/10.1108/MEDAR-08-2019-0540

Costa, E. & Agostini, M. (2016). Mandatory disclosure about environmental and employee matters in the reports of Italian-listed corporate groups. *Social and Environmental Accountability Journal*, 36(1), 10-33. https://doi.org/10.1080/0969160X.2016.1144519

Christensen, H. B., Hail, L. & Leuz, C. (2021). Mandatory CSR and sustainability reporting: economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176-1248. https://doi.org/10.1007/s11142-021-09609-5

De Villiers, C. & Alexander, D. (2014). The institutionalisation of corporate social responsibility reporting. *The British Accounting Review*, 46(2), 198-212. https://doi.org/10.1016/j.bar.2014.03.001

Díaz-Becerra, O. A., Leon-Chavarri, C. & Ampuero-Alfaro, B. G. (2021). An analysis of the content and quality of corporate sustainability reports according to GRI standards in Peruvian mining companies supervised by the SMV in 2018: deficiencies and opportunities. *Revista Contemporânea de Contabilidade*, 18(47), 140-154. https://doi.org/10.5007/2175-8069.2021.e77331

Powell, W.W. & DiMaggio, P.J. (1991). *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press. https://archive.org/details/isbn_9780226677095

Doni, F., Bianchi Martini, S., Corvino, A. & Mazzoni, M. (2020). Voluntary versus mandatory non-financial disclosure: EU Directive 95/2014 and sustainability reporting practices based on empirical evidence from Italy. *Meditari Accountancy Research*, 28 (5), 781-802. https://doi.org/10.1108/MEDAR-12-2018-0423

Fortanier, F., Kolk A. & Pinkse, J. (2011). Harmonization in CSR reporting. *Management International Review*, 51(5), 665. https://doi.org/10.1007/s11575-011-0089-9

Gerwing, T., Kajüter, P., & Wirth, M. (2022). The role of sustainable corporate governance in mandatory sustainability reporting quality. *Journal of Business Economics*, 92(3), 517-555. https://doi.org/10.1007/s11573-022-01092-x

Gulenko, M. (2018). Mandatory CSR reporting—literature review and future developments in Germany. *Sustainability Management Forum*, 26, 3-17. https://doi.org/10.1007/s00550-018-0476-9

Hernández-Pajares, J. & Pocomucha, K. (2021). Analysis of compliance with the mandatory sustainability report in Peruvian listed companies. *Studies in Business & Economics*, 16(2), 84-100. https://doi.org/10.2478/sbe-2021-0026

Hernández-Pajares, J. (2023). Exploring the research on sustainability reporting: a comprehensive bibliometric and literature review in the Latin American context. *Revista de Gestão Ambiental e Sustentabilidade*, 12(1), 1-44, article 2801. https://doi.org/10.5585/2023.22801

Ivic, A., Saviolidis, N. M. & Johannsdottir, L. (2021). Drivers of sustainability practices and contributions to sustainable development evident in sustainability reports of European mining companies. *Discover Sustainability*, 2(1), 1-20. https://doi.org/10.1007/s43621-021-00025-y

Ioannou, I. & Serafeim, G. (2017). The consequences of mandatory corporate sustainability reporting (Research working paper N.° 11-100). Harvard Business School. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1799589

Jackson, G., Bartosch, J., Avetisyan, E., Kinderman, D. & Knudsen, J. (2019). Mandatory non-financial disclosure and its influence on CSR: An international comparison. *Journal of Business Ethics*, 162, 323-342. https://doi.org/10.1007/s10551-019-04200-0

Korca, B., Costa, E. & Farneti, F. (2021). From voluntary to mandatory non-financial disclosure following Directive 2014/95/EU: An Italian case study. *Accounting in Europe*, 18(3), 353-377. https://doi.org/10.1080/17449480.2021.1933113

Larrinaga, C., F. Carrasco, C. Correa, F. Llena & J. Moneva. (2002). Accountability and Accounting Regulation: The Case of the Spanish Environmental Disclosure Standard. *European Accounting Review*, 11(4), 723–740. https://doi.org/10.1080/0963818022000001000

Loza-Adaui, C. R. (2020). Sustainability Reporting Quality of Peruvian Listed Companies and the Impact of Regulatory Requirements of Sustainability Disclosures. *Sustainability*, 12(3), 1135. https://doi.org/10.3390/su12031135

Luque-Vílchez, M. & Larrinaga, C. (2016). Reporting models do not translate well: Failing to regulate CSR reporting in Spain. *Social and Environmental Accountability Journal*, 36(1), 56-75. https://doi.org/10.1080/0969160X.2016.1149301

Macias, H. A. & Ficco, C. (2022). ¿Inició la era de reportes de sostenibilidad obligatorios? Transición desde lineamientos GRI hacia estándares de ISSB. *Revista*

Activos, 20(2), 9-17. https://revistas.usantotomas.edu.co/index.php/activos/article/view/8619

Matuszak, L. & Rozanska, E. (2017). CSR disclosure in Polish-listed companies in the light of Directive 2014/95/EU requirements: Empirical evidence. *Sustainability*, 9, 2304. https://doi.org/10.3390/su9122304

Masoud, N. & Vij, A. (2021). The effect of mandatory CSR disclosure on firms: empirical evidence from UAE. *International Journal of Sustainable Engineering*, 14(3), 378-389. https://doi.org/10.1080/19397038.2020.1821405

Mio, C., Fasan, M., Marcon, C. & Panfilo, S. (2020). The predictive ability of legitimacy and agency theory after the implementation of the EU directive on non-financial information. *Corporate Social Responsibility and Environmental Management*, 27(6), 2465-2476. https://doi.org/10.1002/csr.1968

Mion, G. & Loza, C. R. (2019). Mandatory Nonfinancial Disclosure and Its Consequences on the Sustainability Reporting Quality of Italian and German Companies. *Sustainability*, 11(17), 4612. https://doi.org/10.3390/su11174612

Moneva, J. M., Jara-Sarrúa, L., Hernández-Pajares, J. & Del Barco, J. P. (2019). The disclosure of social issues in Latin American sustainability reports: an exploration of their disclosure in Argentina, Chile, and Peru. In I. García-Sánchez, & J. Martínez (Eds.), *The Disclosure and Assurance of Corporate Social Responsibility: A Growing Market,* 135-174. https://www.cambridgescholars.com/product/978-1-5275-2323-4

Neu, D., Warsame, H. & Pedwell, K. (1998). Managing public impressions: environmental disclosures in annual reports. *Accounting, organizations and society*, 23(3), 265-282. https://doi.org/10.1016/S0361-3682(97)00008-1

Ottenstein, P., Erben, S., Jost, S., Weuster, C. W. & Zülch, H. (2022). From voluntarism to regulation: effects of Directive 2014/95/EU on sustainability reporting in the EU. *Journal of Applied Accounting Research*, 23(1), 55-98. https://doi.org/10.1108/JAAR-03-2021-0075

Posadas, S. C. & Tarquinio, L. (2021). Assessing the Effects of Directive 2014/95/EU on Nonfinancial Information Reporting: Evidence from Italian and Spanish Listed Companies. *Administrative Sciences*, 11(3), 89. https://doi.org/10.3390/admsci11030089

Pocomucha, K. & Venegas, C. (2021). Nivel de cumplimiento de información medioambiental en los reportes de sostenibilidad en sector minería y energía de empresas peruanas y chilenas. *Contabilidad y Negocios*, 16(31), 71-94. https://doi.org/10.18800/contabilidad.202101.005

Radu, O. M., Dragomir, V. D. & Hao, N. (2023). Company-Level Factors of Non-Financial Reporting Quality under a Mandatory Regime: A Systematic Review of Empirical Evidence in the European Union. *Sustainability*, 15(23), 16265. https://doi.org/10.3390/su152316265

Rosales, C. A., Álvarez, M. S. & Lam, I. C. (2021). Análisis del contexto económico en el Perú a partir del COVID-19. *Revista Scientific*, 6(21), 237-251. https://doi.org/10.29394/Scientific.issn.2542-2987.2021.6.21.12.237-251

Roy, P. P., Rao, S. & Zhu, M. (2022). Mandatory CSR expenditure and stock market liquidity. *Journal of Corporate Finance*, 72, 1-29. https://doi.org/10.1016/j.jcorpfin.2022.102158

Schröder, P. (2022). Mandatory non-financial reporting in the banking industry: assessing reporting quality and determinants. *Cogent Business & Management*, 9(1), article 2073628. https://doi.org/10.1080/23311975.2022.2073628

Superintendencia del Mercado de Valores (SMV). (2020). Resolución SMV 018-2020-SMV/02. (2020). Diario Oficial El Peruano, 8 de febrero. https://www.gob.pe/institucion/smv/normas-legales/3948354-018-2020

Superintendencia del Mercado de Valores (SMV). (2016). Resolución SMV 033-2015-SMV/01. Diario Oficial El Peruano, 15 de diciembre. https://www.gob.pe/institucion/smv/normas-legales/3947586-033-2015

Scott, W. Richard (1987). The Adolescence of Institutional Theory. *Administrative Science Quarterly* 32(4) 493-511. https://doi.org/10.2307/2392880

Sepúlveda-Alzate, Y. M., Garcia-Benau, M. A. & Gómez-Villegas, M. (2022). Materiality assessment: the case of Latin American listed companies. *Sustainability Accounting, Management and Policy Journal*, 13(1), 88-113. https://doi.org/10.1108/SAMPJ-10-2020-0358

Venturelli, A., Caputo, F., Cosma, S., Leopizzi, R. & Pizzi, S. (2017). Directive 2014/95/EU: Are Italian Companies Already Compliant? *Sustainability*, 9(8), 1385. https://doi.org/10.3390/su9081385

Wang, K. T. & Li, D. (2016). Market reactions to the first-time disclosure of corporate social responsibility reports: Evidence from China. *Journal of Business Ethics*, 138(4), 661-682. https://doi.org/10.3390/su9081385

Wang, X., Cao, F. & Ye, K. (2018). Mandatory corporate social responsibility (CSR) reporting and financial reporting quality: Evidence from a quasi-natural experiment. *Journal of Business Ethics*, 152(1), 253-274. https://doi.org/10.1007/s10551-016-3296-2

Wang, M. C. (2017). The relationship between firm characteristics and the disclosure of sustainability reporting. *Sustainability*, *9* (4), 624. https://doi.org/10.3390/su9040624

Zucker, L.G. (1987). Institutional theories of organization. *Annual Review of Sociology*, 13, 443-464. http://dx.doi.org/10.1146/annurev.so.13.080187.002303

About the authors

Julio César Hernández Pajares has a PhD in Accounting and Finance and a master's in International Business Management from the Universidad de Zaragoza, and is a Public Accountant from the Pontificia Universidad Católica del Perú. He is a senior lecturer, researcher, and author of publications in indexed scientific journals in the areas of financial accounting and IFRS, corporate governance, social responsibility and sustainability at the University of Piura. He was Director of the Accounting and Auditing Program and Vice-Dean of the Faculty of Economics and Business Administration at the Universidad de Piura. Formerly auditor at Deloitte & Touche, he is a consultant and advisor to companies.

julio.hernandez@udep.edu.pe
https://orcid.org/0000-0002-7481-2912

Yulliana Marised Llauce Ontaneda has a master's degree in Business Administration from PAD Business School and is a Public Accountant from Universidad de Piura. She has a CCA Certificate from the Global Chartered Controller Institute. She is the director of the Accounting and Auditing Program at Universidad de Piura and a professor in the academic area of accounting for undergraduate courses and graduate programs in financial accounting, cost and management control courses. She has worked in the accounting area of hydrocarbon companies, as well as in financial auditing, and provides advice to companies on issues related to costs and management control.

yulliana.llauce@udep.edu.pe
https://orcid.org/0000-0002-0192-0154

Macarena Ayleen Mansilla Mahmud holds a bachelor's degree in economics from the Universidad de Piura. He is a research and teaching assistant at Universidad de Piura, Lima Campus, Peru. Her research interests include organizational behavior, human talent management, and corporate sustainability.

macarena.mansilla.m@udep.edu.pe
https://orcid.org/0000-0003-4345-6254

Does US Interest Rate Sentiment Impact Latin American ETFs?

¿Impacta el sentimiento estadounidense de las tasas de interés en los fondos latinoamericanos negociados en bolsa (ETF)?

Humberto Valencia Herrera

EGADE Business School, Tecnológico de Monterrey, Mexico

Received: February 27, 2024.
Approved: May 13, 2024.

Abstract

This article examines the dependence of Exchange Traded Fund (ETF) returns in six Latin American countries on interest rate and the Federal Reserve (FED) sentiment in the United States (US) news, during the period 2022 to 2023. For each country, robust regressions with zero to two lags for positive and negative sentiments, and previous returns were used. It was found that sentiment is statistically significant for some lags of ETF returns in Brazil, Chile, and Peru, in both, local currency and US dollar. The Latin American 40 ETF also depends on sentiment in US currency. Furthermore, a moment effect on returns in US currency and a mean reverting effect in local currency was identified. A panel data model for the considered countries' ETFs with random effects and zero to two lags in the change of sentiment shows that all considered changes in sentiment are statistically significant for returns, except for the change in positive sentiment without lags.

Keywords: *sentiment, ETF, returns, interest rate, Latin America.*
JEL Classification: *G11, G15.*

Resumen

En este artículo se analizó la dependencia de los rendimientos de fondos cotizados en bolsa (ETF) de seis países latinoamericanos respecto al sentimiento en relación con las tasas de interés y la reserva federal (FED) en las noticias de Estados Unidos (EE.UU.) durante el período de 2022 a 2023. Para cada uno de los fondos se usaron regresiones robustas con cero a dos rezagos para sentimientos positivos y negativos, y las rentabilidades previas. Se encontró que el sentimiento es estadísticamente significativo para algunos rezagos en los retornos de los ETF de Brasil, Chile y Perú, tanto en la moneda local como en el dólar estadounidense. El ETF Latin American 40 depende asimismo del sentimiento respecto a la moneda estadounidense. También hay un efecto de momento sobre los rendimientos en moneda estadounidense y un efecto de reversión media en moneda local para todos los ETF considerados. El modelo de datos de panel para los ETF de los países considerados con efectos aleatorios y dos rezagos muestra que todos los cambios en el sentimiento considerados son estadísticamente significativos para los rendimientos, excepto el cambio en el sentimiento positivo sin rezagos.

Palabras clave: *sentimiento, fondos negociados en bolsa, retornos, tasa de interés, Latinoamérica.*
Clasificación JEL: *G11, G15.*

1. Introduction

The use of sentiment information in financial markets is an expanding field that has captured the attention of both practitioners and academics. In this article, we delve into a novel area of research, exploring how shifts in US interest rate sentiment, as reflected in US news, can impact Latin American stocks ETF returns on a daily basis. This investigation into the relationship between changes in sentiment from news and social media and future asset returns is a recent development. We are not aware of any existing studies that have examined how changes in sentiment in US news about the US Fed policy can influence the asset prices of other countries on a daily basis. This paper aims to fill this gap by examining these dynamics for Latin American stock ETFs.

The dynamic may be understood as follows: Changes in economic expectations modify the information available in the markets. Changes in US inflation expectations and economic activity determine the actions of the FED. Analysts and financial commentators consider this information, and this changes the tone of published news. News sentiment and other information influence decision-makers in financial markets and result in changes in asset prices. Although many economic and financial articles consider that changes in monetary policy directly affect treasuries and bond prices, these also influence stock prices. In the case of foreign shares, this effect has two facets: directly on companies and indirectly through the exchange rate. The direct effect is due to changes in the economic potential of companies, which translates into changes in their share price. An interest rate increase in the United States will increase the financing cost of companies that fund themselves there, including foreign companies. As a result, their share prices will fall. Countries that have governments financed in US currency will have higher financial costs. Their taxpayers, both companies and consumers, will have to pay for the increased interest rate. The indirect effect through the exchange rate is mixed. Because most Latin American countries produce raw materials and other goods, an increase in interest rates in the United States will result in a fall in their international economic activity and the depreciation of the local currency. Such a situation is evident in countries with high export activity, such as Chile, Brazil, and Peru. In addition, if the monetary policy of the country in question consists of offering high interest rates in local currency to attract capital, as could be the case of Mexico, an increase in the US funding rate may signal that the local funding rate will remain high. As a result, the local currency will appreciate against the US dollar. The local interest rate policy can counteract the restrictive effect of a policy of increasing rates by the FED. Other factors may also be at play. For example, in the event of a policy to increase investments and favor export activity, as in the case of Mexico, changes in economic activity resulting from increased interest rates in the United States will not necessarily result in the depreciation of the local currency. As a result, there can be differences in the impact of sentiment news and other information on returns in US and local currencies.

The rest of the paper is organized as follows: the next section describes the theoretical framework, followed by a discussion of the employed methodology. The third section includes the main results, and the last section contains the conclusions and recommendations.

2. Theoretical Framework

The literature that uses data mining to extract sentiment information from news and social media is relatively recent. Investors' sentiment, captured by social media in communications related to companies, is relevant to investment returns. However, most of the research has been done for the US market. In this respect, Wu and Gu (2023) propose a simplified mechanism for capturing market sentiment using tweets from Twitter (now X) for stocks in the US market. Mendoza-Urdiales et al. (2022) show that Twitter sentiment is important for explaining changes in returns concerning a moving average for some stocks in the US market, while Cristescu et al. (2023) show a lagged relationship between news and some US stocks.

While many studies address the importance of news for US stock markets, few papers consider news sentiment and its impact on stocks in developing countries. For example, Wu et al. (2022) found that mixing information with sentiment derived from news and stock market reactions offers predictive power for stocks in Vietnam. Chari et al. (2023) propose an aggregate indicator of market sentiment based on the news in India and analyze its relationship with the aggregate Indian stock market on a daily basis.

Ample related economic literature measures market sentiment using proxy variables. Many of these studies have centered on US market sentiment and its relevance for the US market using variables that attempt to mimic market sentiment. For example, Kabiri et al. (2023) document its importance in explaining phenomena during the Great Depression of 1920 to 1934. Han et al. (2022) found that market sentiment impacts US stock returns. Nakhli et al. (2022) observed a bidirectional relationship between market sentiment and stock market performance in the US. However, Ur et al. (2023), who used the Baker and Wurgler (2006) sentiment index to predict US stock behavior, found that the index has weak predictive power. Nevertheless, when combined with other indicators, the overall predictive power of the model improves. Dumiter et al. (2023) use the daily sentiment index downloaded from the Federal Reserve Bank of San Francisco and show that it related strongly to the US stock market from 2004 to 2022. Using another economy, Chen et al. (2022) found an asymmetric effect of investor sentiment on stock excess returns using the Shanghai Stock Exchange 50 Index Stocks. Lv et al. (2022) analyze the impact of investor sentiment on stock returns for more than 35 years and found that the relationship has changed over time. While it used to be trending, more recently, it has become mean reverting.

Some papers have used sentiment analysis based on communications to consider aspects of monetary policy. For example, Tadle (2022) analyzes the Federal Open Market Committee (FONC) communications using sentiment analysis, showing that these have an impact on the future rates of Fed funds. The minutes were also shown to have an impact on the US dollar exchange markets, with the tone of the FONC minutes modifying financial market expectations.

The economic relationship between monetary policy and asset markets has been well addressed for closed economies using data in relatively long intervals of one month or more. Many papers consider treasuries and bonds as well as stocks for a national economy. The FED target rate affects short- and long-term interest rates because it modifies expectations on interest rate markets (Ehrmann & Fratzscher, 2007). As a result, prices for bonds and treasuries change. However, they also affect stocks. For example, Bernanke and Kuttner (2005) show that the unexpected component of the target rate influences equity prices in the US. They extend Kuttner's seminal work using Fed funds futures and evaluate the impact of the unexpected component of the target rate changes on equity prices. Labadie and Giovannini (1991) show an inverse contemporary relationship between nominal stock returns and nominal interest rates, although this is not statistically significant in the US. For Brazilian stocks, De Pontes and Rêgo (2022) show that domestic interest rates have less influence on stock returns than Gross Domestic Product, Risk Brazil, and Ibovespa points. For ASEAN-5 countries, Juhro et al. (2021) find a negative relationship between interest rate movements in a contractionary monetary policy shock and real stock prices, although the movement lags for two months.

Other papers have analyzed how changes in the US FED monetary policy affect prices and returns of foreign bonds, treasuries, and stocks over time. There is ample economic literature that documents its effect in the long run for many countries using monthly, quarterly, or yearly data. For example, for Asian markets, Yang and Hamori (2014) analyze the economic effect of the US monetary policy on the Singaporean, Thai, and Indonesian stock markets, using monthly returns and US Treasury bill rates from 1990 to 2012. They found that during the boom period, there was an inverse relationship, which disappeared during the recession period. Hindrayani et al. (2019) found that the US monetary policy had a negative effect on ten ASEAN stock markets during economic expansion periods, using yearly data for the period 2008-2016. Using monthly returns, Zubair Muntaz and Smith (2019) analyze the economic relationship between US interest rates and European stock markets. They found a positive relationship for many of the analyzed European stock markets, except for the Finnish, Swedish, UK, Slovenian, and Ukrainian markets. The relationship is more robust in developed markets than in developing ones, and changed during the crisis period. Lakdawala et al. (2021) analyze how surprises in the change of the Fed policy rate affect international bond yields. They found that for advanced economics, the transmission is through the term premium in yields. For emerging markets, the expected component of yields reacts to uncertainty. For Latin American stocks, Cabezón (2012) used quarterly data to show that changes in the US interest rate affect Chilean stocks.

3. Methodology

In this section we discuss the data and estimated variables, together with the models and equations employed throughout the paper.

We used ETFs denominated in US dollars from six Latin American countries: iShares MSCI Brazil ETF (EWZ), iShares MSCI Mexico ETF (EWW), iShares MSCI Chile ETF (ECH), Global X MSCI Argentina ETF (ARGT), iShares MSCI Peru and Global Exposure ETF (EPU), and Global X MSCI Colombia ETF (GXG). We also considered the iShares Latin America 40 ETF (ILF). Table 1 shows the considered ETFs, the total asset value, and the average volume as of December 20, 2023 (see Table 1).

Table 1. Latin America ETFs, General Information

Symbol	ETF Name	Country	Assets ($MM)	Average Volume
EWZ	iShares MSCI Brazil ETF	Brazil	$5,933	23,871,064
EWW	iShares MSCI Mexico ETF	Mexico	$1,995	2,460,109
ECH	iShares MSCI Chile ETF	Chile	$612	498,994
ARGT	Global X MSCI Argentina ETF	Argentina	$116	60,338
EPU	iShares MSCI Peru and Global Exposure ETF	Peru	$97	16,537
GXG	Global X MSCI Colombia ETF	Colombia	$37	19,811
ILF	iShares Latin America 40 ETF	Broad Latin America	$1799	1,208,117

* Assets and average volume reported on December 20, 2023.

Source: VettaFi, https://etfdb.com/

Using ETF prices, we calculated the return and other statistics for the period 2022 to 2023, as shown in Table 2. Argentina ETF (ARGT) has the highest standard deviation (7.23) of all the considered ETFs in the period of analysis. All ETF funds are skewed to the right, except Peru ETF (EPU), which is skewed to the left, and Chile ETF (ECH), which has almost no skew (0.02). Furthermore, all ETF funds show positive kurtosis. The most platykurtic fund is Mexico ETF (EWW), with kurtosis of 1.77, as shown in Table 2 (see Table 2).

Table 2. Summary Statistics for Prices and Sentiment Indicators, 2022-2023

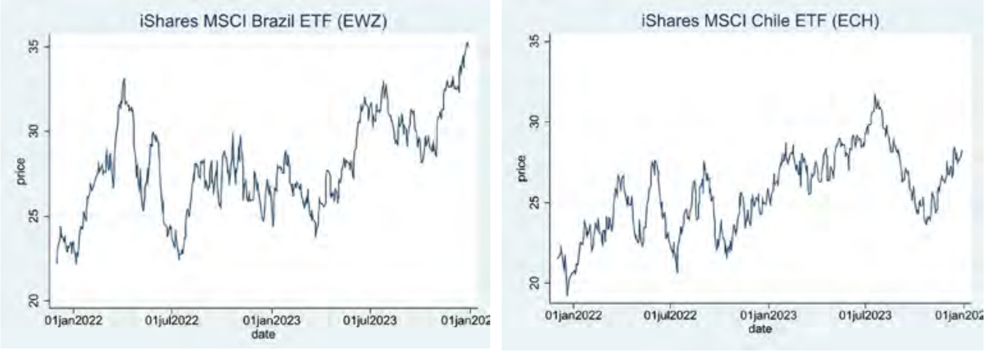
	Closing Price					
Stock Names	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**Minimum**	**Maximum**
Latin America ETF(ILF)	23.31	2.25	0.26	2.55	18.54	29.33
Brazil ETF (EWZ)	27.92	2.76	0.20	2.53	22.23	35.27
Mexico ETF (EWW)	52.63	6.90	0.25	1.77	41.98	68.46

	Closing Price					
Stock Names	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**Minimum**	**Maximum**
Chile ETF (ECH)	25.52	2.45	0.02	2.44	19.20	31.77
Argentina ETF (ARGT)	35.43	7.23	0.42	2.13	22.84	52.25
Peru ETF (EPU)	28.51	2.86	-0.17	2.40	22.04	34.60
Colombia ETF (GXG)	21.22	3.39	0.88	2.65	16.52	29.44
Sentiment Indicator						
Positive Sentiment	0.24	0.04	-1.72	13.00	0.00	0.36
Negative Sentiment	0.08	0.05	1.52	8.13	0.00	0.39

Source: Prepared by the author.

Except for Chile, ETF prices in these countries demonstrated a peak during the first semester of 2022. All countries, except Columbia, exhibit prices that increased during the second semester of 2022 and the first semester of 2023. The same increasing tendency can be observed at the end of 2023, as shown in Figure 1 (see Figure 1).

Figure 1. Prices of Latin America ETFs



https://doi.org/10.36105/theanahuacjour.2024v24n1.04

Source: Prepared by the author.

For sentiment analysis, we used news from Google News as an RSS feed, and searched for news using the terms: interest rate, FED, or Federal Reserve in the

United States. We downloaded the available news, up to one hundred articles, for each calendar day during 2022 and 2023.

We used the VADER (Valence Aware Dictionary for Sentiment Reasoning) algorithm (Hutto & Gilbert, 2014) to estimate a positive and a negative sentiment indicator from the first line of each news item. VADER uses a given lexicon of words, labeled according to their semantic meaning as positive or negative. Using a sample of microblog-like texts (tweets), the lexicon was curated and validated. Lexicon tokens were classified and assigned a sentiment intensity score from -4 to 4, from very negative to very positive. These lexical features were then combined with five rules that considered syntactical and grammatical conventions that emphasize sentiment intensity: the use of an exclamation point (!) increases the intensity of the sentiment; while a word in all caps, among non-caps words, emphasizes the sentiment implied by the word. The use of degree modifiers also modifies the intensity of a word, for example, “the market is extremely optimistic”. Based on the tri-gram preceding a sentiment-laden lexicon expression, we were able to capture 90% of events that involved a negation flipping the meaning of a text. Based on the algorithm, a text received a positive, negative, and neutral lexicon rating, on a scale of zero to one. We named the VADER positive (negative) lexicon, rating the news positive (negative) sentiment indicator. In addition, before submitting the text to the VADER algorithm, we filtered the text for the stop words suggested in the VADER corpus. Stop words refer to words that are so common that they have very little meaning, such as “the”, “a”, or “is”. For these purposes, we used the VADER open source code available at nltk (nltk/nltk/sentiment/vader.py in github.com), version 3.3.2.

For each business date, we calculated a positive daily sentiment indicator and a negative daily sentiment indicator. The daily polarity indicators refer to these daily indicators.

The positive (negative) daily sentiment indicator on business day t is an average of the positive (negative) news sentiment in the news on that day, only considering the news with positive (negative) indicators greater than 0.1 if the following day is a business day. In the case that one or more non-business days follow day t, we calculated the daily positive or negative sentiment indicator of business day t as an average of the positive or negative sentiment in the news on that day and all the following non-business days before a business day, excluding news with polarity indicators less than or equal to 0.1.

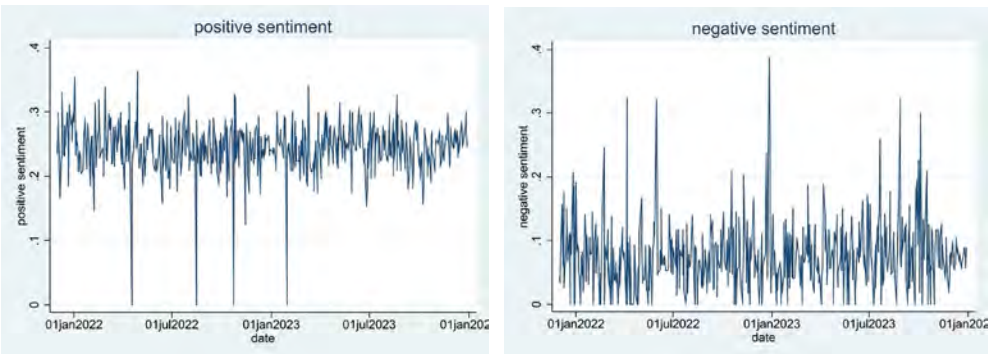
The mechanics of calculating the positive and negative daily sentiment indicators consider that at closing time on day t, an investor can take a position on the ETFs

https://doi.org/10.36105/theanahuacjour.2024v24n1.04

based on the news sentiment of that day and the previous day. When there is a holiday or a weekend on the previous day, information from the news during the holiday or the weekend can be used, in addition to the news of the day.

Regarding interest rate sentiment, a positive sentiment indicator of two or three is most frequent, with few periods at the zero level. However, a negative sentiment is less pronounced, normally falling between zero and two during the period of analysis, with some peaks above two, as shown in Figure 2 (see Figure 2).

Figure 2. Positive and Negative US Interest Sentiment Indicators



Source: Prepared by the author.

For each ETF *i*, we also calculated the forward returns for one business day using the equation:

$$f_{i,t,T} = ln(P_{i,t+1}) - ln(P_{i,t}), \tag{1}$$

where $P_{i,t}$ is the closing price of fund i in business day t.

Our research employed robust regressions for each ETF, considering its returns as the dependent variable, with the independent variables comprising changes in the daily positive and negative polarity indicators, Δp_t, = $p_t - p_{t-1}$ and Δn_t, = $n_t - n_{t-1}$ for the day under consideration, the day before, and the day before that, that is m =2. Therefore, we captured the changes on the polarity indicators for the day, as well as for the one and two previous days. To consider the momentum or the mean reverting effects in returns, we considered the previous day's return, $r_{i,t\text{-}1}$ in the equation for each asset *i*.

$$r_{i,t} = a_i + b_{i,1} * r_{i,t-1} + \sum_{T=0}^{m} b_{i,2,T} * \Delta p_{t-T} + \sum_{T=0}^{m} b_{i,3,T} * \Delta n_{t-T} + \varepsilon_{i,t} \quad (2)$$

We repeated this exercise, considering the returns in local currency, using the exchange rate provided by yahoo.com for each date.

Following this, we conducted panel data regressions for both fixed and random effects. We then employed the Hausman (1978) test to determine the validity of these models. A panel model allows for an explanation of variations in the returns of the considered assets through time as a group, as well as for the combination of temporal and cross-sectional information in the same model: variations through time and between the different assets. In this case, changes in sentiment regarding the different assets and within each asset over time explain the returns.

A panel model also allows for the simultaneous modeling of the effect of changes in the independent factors for all assets while allowing for different constants for each asset. A different constant for each asset allows for a constant difference in returns among different assets. That is, it allows for constant factors that explain these differences, as shown in equation (3),

$$r_{i,t} = a_i + b_1 * r_{i,t-1} + \sum_{T=0}^{m} b_{2,T} * \Delta p_{t-T} + \sum_{T=0}^{m} b_{3,T} * \Delta n_{t-T} + \varepsilon_{i,t} \quad (3)$$

Sometimes, an equation that models the constant as drawn randomly for the sample of assets explains these differences. In this case, a random panel model can be used. The advantage of a random panel is that the model results are more generalizable than a fixed panel. If the above is not the case, the difference in magnitude of the returns of the assets is specific to each asset, and we must independently estimate the constant for each asset. That is, we must prefer a fixed panel model.

We used the Hausman (1978) test to examine whether the constant factor explaining differential returns of assets can be a random draw from a distribution. In this case, the test recommends a random panel model.

4. Analysis of Results

We begin by discussing the results of each individual regression considering the ETF returns in USD dollars, followed by the results of each individual ETF with returns in local currency. We then discuss the observed results using a panel model with random effects, given that the Hausman (1978) test showed this to be better than a fixed effects model.

The returns in US dollars of all considered ETFs, except EWW from Mexico, show a positive effect from changes in the negative sentiment in US interest rates and a negative effect from changes in the positive sentiment. However, the effects are only statistically significant in the following cases: EWZ, the ETF from Brazil, shows a statistically significant effect from changes in the negative sentiment with no and one lag, and coefficients of 0.0379 and 0.0189, respectively. The ETFs from Chile and Peru, ECH, and EPU, respectively, show statistically significant negative effects from changes in the positive sentiment with one lag, and regression coefficients of -0.0616 and -0.0347, respectively. In the case of the ETF ILF, its regression also has a statistically significant positive coefficient for changes in the negative sentiment with no lags (0.0318). It is worth noting that all regressions show a moment effect on returns from the previous day, although this was only statistically significant for EPU, GSG, and ILF ETFs, with coefficients of 0.1141, 0.1300, and 0.0943, respectively, as can be observed in Table 3 (see Table 3).

When considering ETF returns in local currency, the pattern of a positive effect on returns from a change in negative sentiment and a negative effect on returns from a change in positive sentiment is maintained for the ETFs of the considered countries, except in the cases of Mexico and Peru. In the latter exceptions, the coefficient does not differ from zero, with a confidence level of 95%. For EWZ Brazil, the coefficients for the change in negative sentiment without lags and with one lag are positive and statistically significant, at 0.0380, and 0.0442, respectively. For ECH Chile, coefficients for the change in positive sentiment without lags and with one lag are negative and statistically significant, at -0.0482 and 0.0821, respectively. For EPU Peru, coefficients for a change in positive sentiment with one lag and with two lags are negative and statistically significant, that is, -0.0614 and -0.0388, respectively, as can be appreciated in Table 4 (see Table 4).

Contrary to the results using US dollars, in local currency, the ETFs of all countries show mean reversion. The coefficients with respect to the previous return are negative, more than minus one, and statistically significant. That is, -0.2939, -02464, -02798, -0.0127, -0.1643, and -0.1359, for the EWZ, EWW, ECH, ARGT, EPU, and GXG ETFs, respectively.

Table 3. Latin America ETF Returns in USD as a Function of US Interest Rate Sentiment and Previous Returns

Country		**Brazil**	**Mexico**	**Chile**	**Argentina**	**Peru**	**Colombia**	**LatinAmerica40**
ETF		EWZ	EWW	ECH	ARGT	EPU	GXG	ILF
Return	L1	0.0612	0.0491	0.0002	0.0643	0.1141 *	0.1300 *	0.0943 *
		1.42	1.09	0	1.53	2.73	3.14	2.15
Δ positive	.	-0.0178	-0.0175	-0.0297	-0.0270	-0.0060	-0.0056	-0.0255
		-0.98	-1.32	-1.63	-1.46	-0.44	-0.38	-1.66
	L1	-0.0322	-0.0158	-0.0616 *	-0.0349	-0.0347 *	-0.0230	-0.0345
		-1.52	-1.01	-2.89	-1.61	-2.19	-1.33	-1.93
	L2	-0.0280	0.0069	-0.0286	-0.0135	-0.0201	-0.0080	-0.0246
		-1.56	0.52	-1.57	-0.73	-1.49	-0.54	-1.61
Δ negative	.	0.0365 *	0.0064	0.0108	0.0256	0.0081	0.0199	0.0318 *
		2.63	0.63	0.77	1.8	0.78	1.74	2.71
	L1	0.0379 *	-0.0155	0.0072	0.0209	0.0058	0.0206	0.0233
		2.31	-1.28	0.43	1.24	0.47	1.52	1.67
	L2	0.0189	-0.0088	0.0178	0.0034	0.0106	0.0072	0.0121
		1.36	-0.86	1.26	0.24	1.01	0.63	1.03
constant		0.0012	0.0010	0.0006	0.0018 *	0.0005	0.0001	0.0011
		1.51	1.57	0.67	2.09	0.83	0.06	1.42

* Ninety-five percent confidence level. For each independent variable, the first row is the coefficient; the second row is the t value.
Source: Prepared by the author.

https://doi.org/10.36105/theanahuacjour.2024v24n1.04

Table 4. Latin America ETFs Returns in Local Currency as a Function of US Interest Rate Sentiment and Previous Returns

		Brazil	Mexico	Chile	Argentina	Peru	Colombia
		EWZ	EWW	ECH	ARGT	EPU	GXG
return	L1	-0.2939 *	-0.2464 *	-0.2798 *	-0.0127	-0.1643 *	-0.1359 *
		-7.07	-5.67	-6.88	-0.59	-3.66	-3.24
Δ positive	.	-0.0169	-0.0174	-0.0482 *	-0.0297	-0.0255	-0.0139
		-0.89	-1.22	-2.29	-1.56	-1.36	-0.80
	L1	-0.0257	-0.0237	-0.0821 *	-0.0361	-0.0614 *	-0.0356
		-1.15	-1.42	-3.32	-1.62	-2.81	-1.75
	L2	-0.0222	0.0018	-0.0296	-0.0109	-0.0388 *	-0.0140
		-1.17	0.12	-1.4	-0.57	-2.08	-0.81
Δ negative	.	0.0380 *	0.0169	0.0084	0.0258	0.0053	0.0254
		2.59	1.54	0.52	1.76	0.37	1.89
	L1	0.0442 *	-0.0093	0.0128	0.0172	-0.0015	0.0157
		2.55	-0.71	0.67	0.99	-0.09	0.99
	L2	0.0274	-0.0103	0.0244	0.0042	0.0151	0.0210
		1.87	-0.93	1.5	0.29	1.05	1.56
constant		0.0011	0.0005	0.0008	0.0040 **	0.0004	-0.0001
		1.25	0.73	0.81	4.58	0.46	-0.1

* Ninety-five percent confidence level. For each independent variable, the first row is the coefficient, the second row is the t value.

Source: Prepared by the author.

The Hausman (1978) test suggests the use of a random effects model. The test cannot reject the hypothesis that the results from the fixed effects model are equal to those of the random effects model with a 95% confidence level. The p value is 99.9%. Therefore, we consider the results from the random effects model reported in Table 5 (see Table 5).

The panel data results of the random effects model suggest that, on aggregate, there is dependence of returns in Latin America on changes in interest rate sentiment in the US, and the effect is persistent over various days. The coefficients from a positive change with zero, one, and two lags are all negative, -0.0133, -0.0362, and -0.0153, respectively. They are statistically significant with one and two lags. The coefficients from a positive change with zero, one, and two lags are all negative, -0.0133, -0.0362, and -0.0153, respectively. They are statistically significant with one and two lags. The coefficients from a negative change are all positive from none to two lags, 0.0198, 0.0155, and 0.0150, respectively, and statistically significant. It is also worth noting that the effect of the changes has the opposite sign: ETF returns are lower with a positive change in interest rate sentiment in the US, while if a change in negative sentiment occurs, the ETF returns are higher. In addition, there is a persistent moment effect from the previous day in the returns. The coefficient is 0.0599 and statistically significant (see Table 5).

Table 5. ETF Returns in USD in Latin America as a Function of US Interest Rate Sentiment, Panel Data Results

		Latin America Countries ETFs			
		RE		FE	
Return	L1	0.0599	*	0.0594	*
		3.35		3.32	
Δ positive	.	-0.0133		-0.0132	
		-1.93		-1.93	
	L1	-0.0362	*	-0.0362	*
		-4.51		-4.51	
	L2	-0.0153	*	-0.0153	*
		-2.23		-2.23	

		Latin America Countries ETFs			
		RE		FE	
Δ negative	.	0.0198	*	0.0198	*
		3.74		3.74	
	L1	0.0155	*	0.0155	*
		2.48		2.48	
	L2	0.0150	*	0.0150	*
		2.83		2.83	
Constant		0.0006		0.0006	
		1.86		1.86	

RE: random effects, FE, fixed effects, * 95% confidence.

ETFs from Brazil, Mexico, Chile, Argentina, Peru, and Colombia.

Source: Prepared by the author.

5. Conclusions and Recommendations

In the case of the ETFs from Brazil, Chile, and Peru, EWZ, ECH, and EPU, there is dependence on interest rate sentiment from US news both in local currency and in US dollars. The dependence is also observed for ITF, the Latin American 40 ETF. The effect is more clearly observed using a panel data model with random effects, where all coefficients for changes in sentiment are statistically significant, except for the change in positive sentiment without lags.

This dependency is likely related to the importance of exports and dependence on US funding of the companies in the considered ETF. This is an area for further research.

The dependency on previous ETF returns depends on the considered currency. In the US currency, we observed a moment effect. In local currency, we observed a mean reverting effect. We recommend further analysis with emphasis on the changes in exchange rates.

For Mexico, neither positive nor negative sentiment indicators helped predict ETF returns. This is probably due to the high integration of the Mexican market with the US market, which makes the news available promptly to all investors on an intraday basis. This finding supports the efficient market hypothesis.

For further areas of research, an extension of the analysis to other markets and instruments is recommended, as well as the consideration of news related to other economic and financial policies, such as those of the European Central Bank. Another promising area of research is the impact of sentiment news on asset prices on an intraday basis and the sentiment captured in social media communications.

References

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61, 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Bernanke, B. S., & Kuttner, K. N. (2005). What explains the stock market's reaction to federal reserve policy? *The Journal of Finance*, 60(3), 1221–1257. https://doi.org/10.1111/j.1540-6261.2005.00760.x

Cabezón, F. (2012). Assessing the effects of foreign financial shocks in the Chilean economy. *Revista de Análisis Económico*, 27(2), 121–143. http://dx.doi.org/10.4067/S0718-88702012000200004

Chari, S., Desai, P. H., Borde, N., & George, B. (2023). Aggregate News Sentiment and Stock Market Returns in India. *Journal of Risk & Financial Management*, 16(8), 376. https://doi.org/10.3390/jrfm16080376

Chen, M., Guo, Z., Abbass, K., & Huang, W. (2022). Analysis of the impact of investor sentiment on stock price using the latent Dirichlet allocation topic model. *Frontiers of Environmental Science*, 10, article 1068398. https://doi.org/10.3389/fenvs.2022.1068398

Cristescu, M. P., Mara, D. A., Nerișanu, R. A., Culda, L. C., & Maniu, I. (2023). Analyzing the Impact of Financial News Sentiments on Stock Prices—A Wavelet Correlation. *Mathematics*, 11 (23), article 4830. https://doi.org/10.3390/math11234830

De Pontes, L. S. & Rêgo, L. C. (2022). Impact of macroeconomic variables on the topological structure of the Brazilian stock market: A complex network approach. *Physica A: Statistical Mechanics and Its Applications*, 604, article 127660. https://doi.org/10.1016/j.physa.2022.127660

Dumiter, F. C., Turcaș, F., Nicoară, Ștefania A., Bențe, C., & Boiță, M. (2023). The Impact of Sentiment Indices on the Stock Exchange—The Connections between Quantitative Sentiment Indicators, Technical Analysis, and Stock Market. *Mathematics* (2227-7390), 11(14), article 3128. https://doi.org/10.3390/math11143128

Ehrmann, M., & Fratzscher, M. (2007). Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness? *Journal of Money, Credit and Banking*, 39(2–3), 509–541. https://doi.org/10.1111/j.0022-2879.2007.00034.x

Han, Z., Sakkas, N., Danbolt, J., & Eshraghi, A. (2022). Persistence of investor sentiment and market mispricing. *Financial Review*, 57, 617–640. https://doi.org/10.1111/fire.12301

Hausman, J. A. (1978). Specification test in econometrics. *Econometrica* 46 (6), 1251–1271. https://doi.org/10.2307/1913827

Hindrayani, A., Putri, F. K. & Puspitasari, I. F. (2019). Spillover effect of US monetary policy to ASEAN stock market. *Journal Economia. Review of Business and Economic Studies*, 15(2) 213–242. https://doi.org/10.21831/economia.v15i2.26314

Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious rule-based model for sentiment analysis of social media text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor. https://doi.org/10.1609/icwsm.v8i1.14550

Juhro, S. M., Iyke, B. N., & Narayan, P. K. (2021). Interdependence between monetary policy and asset prices in ASEAN-5 countries. *Journal of International Financial Markets, Institutions and Money*, 75, article 101448. https://doi.org/10.1016/j.intfin.2021.101448

Kabiri, A., James, H., Landon-Lane, J., Tuckett, S., & Nyman, R. (2023). The role of sentiment in the US economy: 1920 to 1934. *Economic History Review*, 76, 3–30. https://doi.org/10.1111/ehr.13160

Labadie, A. G. P. & Giovannini, P. (1991). Asset prices and interest rates in cash-in-advance models. *Journal of Political Economy*, 99(6), 1215–1251. https://doi.org/10.1086/261798

Lakdawala, A., Moreland, T., & Schaffer, M. (2021). The international spillover effects of US monetary policy uncertainty. *Journal of International Economics*, 133, article 103525. https://doi.org/10.1016/j.jinteco.2021.103525

Lv, Y., Piao, J., Li, B., & Yang, M. (2022). Does online investor sentiment impact stock returns? Evidence from the Chinese stock market. *Applied Economics Letters*, 29(15), 1434–1438. https://doi.org/10.1080/13504851.2021.1937490

Mendoza-Urdiales, R. A., Núñez-Mora, J. A., Santillán-Salgado, R. J., & Valencia-Herrera, H. (2022). Twitter Sentiment Analysis and Influence on Stock Performance Using Transfer Entropy and EGARCH Methods. *Entropy*; 24(7), 874. https://doi.org/10.3390/e24070874

Nakhli, M. S., Dhaoui, A., & Chevallier, J. (2022). Bootstrap rolling-window Granger causality dynamics between momentum and sentiment: Implications for investors. *Annals of Finance*, 18, 267–283. https://doi.org/10.1007/s10436-021-00399-z

Tadle, R. C. (2022). FOMC minutes sentiments and their Impact on financial markets. *Journal of Economics and Business*, 118, article 106021.https://doi.org/10.1016/j.jeconbus.2021.106021

Ur Rehman, M., Raheem, I. D., Al Rababa'a, A. R., Ahmad, N., & Vo, X. V. (2023). Reassessing the predictability of the investor sentiments on US stocks: The role of uncertainty and risks. *Journal of Behavioral Finance*, *24* (4), 450–465. https://doi.org/10.1080/15427560.2022.2037598

Wu, S. & Gu, F. (2023). Lightweight scheme to capture stock market sentiment on social media using sparse attention mechanism: a case study on Twitter. *Journal of Risk and Financial Management*,16 (10), 440. https://doi.org/10.3390/jrfm16100440

Wu, S., Liu, Y., Zou, Z., & Weng, T. H. (2022). S_I_LSTM: Stock price prediction based on multiple data sources and sentiment analysis. *Connection Science*, 34, 44–62. https://doi.org/10.1080/09540091.2021.1940101

Yang, L. & Hamori, S. (2014). Spillover effect of US monetary policy to ASEAN stock markets: Evidence from Indonesia, Singapore1 and Thailand. *Pacific-Basic Finance Journal*, 26, 145-155. https://doi.org/10.1016/j.pacfin.2013.12.003

Zubair Mumtaz, M. & Smith, Z. A. (2019). Examining spillover effect of US monetary policy to European stock markets: A Markov-Switching approach. *Estudios de Economia*, 46(1), 89-124. http://dx.doi.org/10.4067/S0718-52862019000100089

About the author

Dr. Humberto Valencia Herrera has an extensive career as a researcher, teacher, and professional in finance, economics, and management. He holds a PhD and a Master's Degree in Economics and Decision Sciences from the Engineering Economic Systems Department, now part of the Management Science and Engineering Department at Stanford University, United States. He has been a finance, economics, management, and industrial engineering professor and researcher, and has collaborated with educational institutions in Mexico and the United States, such as the Tecnológico de Monterrey, the Universidad Iberoamericana, and Stanford University. He has numerous publications in international and national academic journals. He is currently researching the use of artificial intelligence for investment decision-making, financial economics, energy economics, and technological development economics. He believes in the efficient and responsible management of businesses and economies, balancing environmental and social considerations with human needs.

humberto.valencia@tec.mx
https://orcid.org/0000-0002-4843-5965

Large Exposures: Implicit Credit Risk Concentration add-ons and the Basel Framework

Grandes exposiciones: add-ons *por concentración de riesgo de crédito implícita y el marco de Basilea*

José Juan Chávez

EGADE Business School, Tecnológico de Monterrey, Mexico

Received: March 6, 2024.
Approved: May 24, 2024.

Abstract

The Basel large exposures standard, addressed to banks, is in the process of being implemented, and although the new rule is reducing the limit for the large credit exposures in banks, and is geared to better control the portfolio risk parameters that make a portfolio more prone to losses due to credit concentration, it is important to know the share of risk and capital implicit to this new rule. In this work, the implicit add-ons for credit risk concentration are determined through a Monte Carlo credit risk model, and the results are compared with current capital requirements. The author also analyzed the complete Basel framework to understand how the concentration risk is addressed in an integrated approach rather than with a specific capital supplement.

Keywords: *large exposures, LEX, credit risk concentration add-ons, granularity adjustment, HHI, ICAAP, CVaR.*
JEL Classification: *F38, G21, G28.*

Resumen

La regulación de Basilea sobre grandes exposiciones está en proceso de implementación y, aunque es una regulación que reduce el límite del tamaño para las exposiciones de gran tamaño en los bancos y está orientada a un mejor control de los parámetros de riesgo que ocasiona que los portafolios se inclinen más a tener pérdidas grandes por concentración de crédito, es importante conocer la porción de riesgo de crédito y capital que lleva implícita. En este trabajo se determinan, mediante una simulación Montecarlo a través de un modelo de riesgo de crédito, los add-ons *implícitos (o añadidos implícitos) debidos a la concentración de crédito, y se comparan con los requisitos actuales de capital. Al mismo tiempo, se analiza el marco completo de Basilea para entender cómo se aborda el riesgo de crédito de concentración, esto es, de forma integral, más que mediante un suplemento particular de capital.*

Palabras clave: *grandes exposiciones, LEX,* add-ons *por concentración de riesgo de crédito, ajuste de granularidad, HHI, ICAAP, CVaR.*
Clasificación JEL: *F38, G21, G28.*

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.05

1. Introduction

A regulation on bank's large exposures was issued in 2019 by the Bank for International Standards (Basel Committee) and is being implemented by countries within the Basel Accord following local implementation schedules. In Mexico, the regulation entered into effect in October 2023. The large exposures regulation is a tool to deal with credit risk concentration, specifically "name concentration". However, questions on the effectiveness of this regulation were raised regarding the adequacy of the capital requirement. Thus, there is a need to analyze if the current capital requirement is aligned with the charge of capital implicit in this regulation regarding credit risk concentration. Due notice is given to the fact that the credit risk due to concentration accounts for an important share or capital at risk, measured through the computed add-ons.

Concentration risk arises basically from two kinds of concentration of credit risk: The so-called "name risk" referred to as the one resulting from a single counterparty or a group of connected counterparties, and the concentration in economic sectors. This work addresses the "name concentration" problem since this is the focus of the large exposures regulation.

The Basel regulation placed its initial focus on credit risk capital requirements given that it is the main component of capital for most banks. The regulatory framework included two approaches:

I. The standard model, based on agencies' external ratings, where the risk weight of a loan for corporate loans depends on the credit rating of the obligor, the risk weight of other credit transactions was fixed by the regulator according to the type of transaction and parameters like the loan-to-value ratio in the case of residential loans.

II. The internal ratings-based approach (IRB), with models developed internally by the banks and assessed and approved by the regulator. The models determined the three basic risk parameters for credit: probability of default, loss given default, and exposure at default (PD, LGD, and EAD). At the same time, the IRB had two approaches: foundation IRB and advanced IRB (FIRB and AIRB). In the foundation approach, the bank develops an internal model to estimate the probability of default but has to use the regulatory parameters for the LGD and EAD. Under this approach, banks compute the capital requirement with a closed formula that provides the comparison capital requirement used by the author in this work.

At the beginning (2000–2004) there was a proposal to include a component of the capital requirement resulting from the concentration level of the bank's portfolio. Eventually, the Basel Committee did not include this capital component, and the discussion on how to deal with credit risk concentration remained open. After years of discussion, the proposal for a different approach emerged in 2014 and concluded in 2019. This was called the large exposures approach or LEX. This work looks retrospectively and links the initial discussion with the final proposal.

2. Objective

This work seeks to determine the add-on implicit in the large exposures regulation (LEX, following the Bank for International Settlements—BIS—practice) and to assess its importance in the regulatory capital. The regulation has followed a different path from computing and assigning a specific supplement of capital due to credit risk concentration, relying more upon the complete Basel framework, as well as on current and in-process capital supplements. Therefore, it is important to quantify the risk implicit in the new LEX rule. The difference between the capital requirement without concentration, measured by the capital requirement through the Basel IRB formula, and the requirement including concentration in the credit portfolio will be the adjustment add-on—the share of risk not included in the regulatory requirement. To do so we will go from a capital requirement with only a systemic risk to a capital requirement including an idiosyncratic risk derived from the risk concentrated in a single counterparty or a group of connected counterparties. We hypothesize that the concentration credit risk has an important share of capital at risk since real-world portfolios are heavily concentrated.

3. Previous Studies on Credit Risk Concentration and add-on Computation

The problem of credit risk concentrations has been extensively addressed in the past. Works can be classified according to the problems addressed from 2000 to the present.

3.1 Stage 1: Works Preceding the 2004 BIS Document

Assessment of the granularity approach or the ASFR framework adopted by the Basel Committee. Studies and literature collected by the BIS's Committee on Banking Supervision (BCBS) itself helped define the June 2004 *International Convergence of Capital Measurement and Capital Standards* document.

Works included in the BIS Selected Literature on Concentration Risk in Credit Portfolios—published between 2001 and 2004. (BCBS, 2005).

a) Basel II and its asymptotic single-risk-factor model foundation (3 works): Gordy (2003), Wilde (2001a), Basel Committee on Banking Supervision (2004), represents the research, presentation, and publishing of the asymptotic IRB formula that excludes all concentration risk in the capital requirement.

b) Granularity adjustment for single name concentrations (six works). They address the 2001 first BIS proposal to adjust the capital requirement for single-name concentration. Wilde (2001b) summarizes the main problem with the 2001 adjustment: "[T]he granularity adjustment as presented in Basel II is inaccurate and [so is] the belief expressed in its derivation." As we know, this adjustment was discarded.

c) VaR adjustment for sector concentration (two works).

d) Estimation of default dependence (nine works).

e) Contagion in credit portfolios (four works).

We can consider this selection of literature as an initial departing point for concentration risk as presented in the large exposures' regulation. These works were developed before the 2004 Basel document. They explain the IRB formula, and the initial proposal for adjusting such formula and show problems aside from name concentration. Works on correlation were useful to determine the correlation formula that accompanies the IRB capital formula. Apart from correlation in defaults, those works address dependence regarding sectors. Contagion, as well as concentration and dependence in economic sectors, is out of the scope of this work.

3.2 Stage 2. Works After the Publishing of the BIS Committee 2004 Convergence Document

This stage is summarized in the BIS's Committee on Banking Supervision 2006 Working Paper No. 15, *Studies on credit risk concentration* (BCBS, 2006) that addresses the main issues around credit risk concentration, some of them finally present in the BIS regulation and especially in the LEX:

a) The economic capital due to concentration and the technical difficulties in computing this capital, particularly in sector concentration.

b) The use of the Hirschman-Herfindahl Index (HHI).

c) The use of credit limits to manage concentration risk.

d) Business interconnectedness and its impact on contagion.

e) Stress that testing is a tool to identify the effect of concentration risk in capital requirements.

f) Issues around data to handle and properly address the concentration risk, including those to consolidate the total exposure.

The topics related to economic capital were particularly useful for this document; the rest helped us understand the complete Basel framework as an integrated tool.

3.3 Stage 3. Works Published Before the Final LEX Regulation was Released

The IRB Basel formula follows the Asymptotic Single Risk Factor Framework (ASRF) that assumes infinity granular portfolios, which does not consider the credit concentration that accompanies most corporate loan portfolios. Therefore, the natural attempt was to compute a granularity adjustment based on the Hirschman– Herfindahl Index, whose inverse provides the number of loans for a given level of concentration, this was done by Gordy and Lutkebohmert in 2013, as quoted by Nokkala (2022, p. 380): "The granularity adjustment of Gordy and Lutkebohmert (2013) uses portfolio exposure distribution and aligns the fully diversified IRB unexpected loss with non-diversified portfolios' corresponding unexpected loss."

Regarding the portfolio size to study concentration risk, Nokkala states the following:

> The literature on credit portfolios does give some guidance on how to construct realistic portfolios in terms of exposure distribution with a given portfolio size n. Heterogenous credit sizes are practically observed in research and Galaasen et al. (2020) presents [sic] an "80% to 20%" rule, stating that 20% of the largest credits constitute 80% of a portfolio's exposure. (Nokkala, 2022, p. 382).

In our proposal, the concentration level is the one implicit in the LEX regulation.

Martin Hibbeln published an extensive book on the matter (Hibbeln, 2010). Following different approaches, he determined add-ons using parametric models as well as Montecarlo simulations.

The BIS finally followed a comprehensive approach including different components of its regulation, but research continues extending the lines we have mentioned. One interesting line is the use of complex systems to analyze contagion risk due to credit concentration (Relim et al., 2019).

4. The Large Exposures (LEX) Regulation Components and their Implications

In this work, we will use the global regulation (Financial Stability Institute, 2022) and the specific implementation in Mexico for examples, parameters, and precise implications.

The BIS LEX framework was concluded and released by the BIS to enter into effect as of January 1st, 2023. It was implemented in Mexico in 2023 and became effective in October 2023 for systemic banks. The BIS assessed Mexico as LEX regulatory-compliant in December 2023 (BCBS, 2023, p. 7).

Main components of the Financial Stability Institute, 2022:

- The LEX regulation defines Tier 1 capital as the capital reference to determine the LEX limits. This is to ensure that banks consider only high-quality capital to absorb losses derived from high credit risk concentration.
- LEX requires banks to consolidate their credit exposures at name (single counterparty) or group of interconnected counterparties. The banks must conduct an assessment of economic interdependencies to define connectedness due to economic interdependency.

- The regulation defines a large exposure as a consolidated exposure equal to or higher than 10% of the bank's Tier 1 capital.
- Exposure limits set by LEX:
    - 25% of Tier 1 capital for any counterparty or group of interconnected counterparties.
    - 15% of Tier 1 capital for Global Systemic Important Banks (G-SIBs) and connected counterparties.
- Connections include not only control relationships among counterparties but any relevant economic interdependency (due to concentration on sales, suppliers, loans, guarantees, or another important dependency.)
- Banks monitor their LEX and report them to regulators. In case of any limit breach, the banks must remediate immediately.
- Exposures in LEX include both banking and trading books and in-balance and off-balance elements. The target is to consolidate all credit risk derived from the relation with the counterparties or group of connected counterparties. This approach differs from the previous one, which focused on loans.
- Mitigation. The LEX permits the use of mitigants used for regulatory capital computation purposes to reduce exposures—such as collaterals, guarantees, credit protections, etc.

The LEX implementation in Mexico contains all BIS components:

- LEX permits banks to conduct the assessment on counterparties or groups of interconnected counterparties only when the exposure is equal to or higher than 5% of Tier 1 capital.
- The exposure of the four main counterparties or group of interconnected counterparties must be lower than the Tier 1 capital.
- Besides de G-SIBs, Local Systemic Banks (D-SIBs) are included in the 15% of Tier 1 capital limit.

Let us rethink the LEX regulation and its implications in the management of the credit concentration risk and the metrics to manage the credit risk.

The LEX regulatory approach does not include a specific capital requirement for credit risk concentration but requires the development of a framework to manage such risk. Nonetheless, the LEX regulation's components have important implications for the parameters used to compute the capital requirement under the IRB framework. As shown in Table 1, we can align the components with the credit risk key parameters that are important to determine the capital requirement (see Table 1).

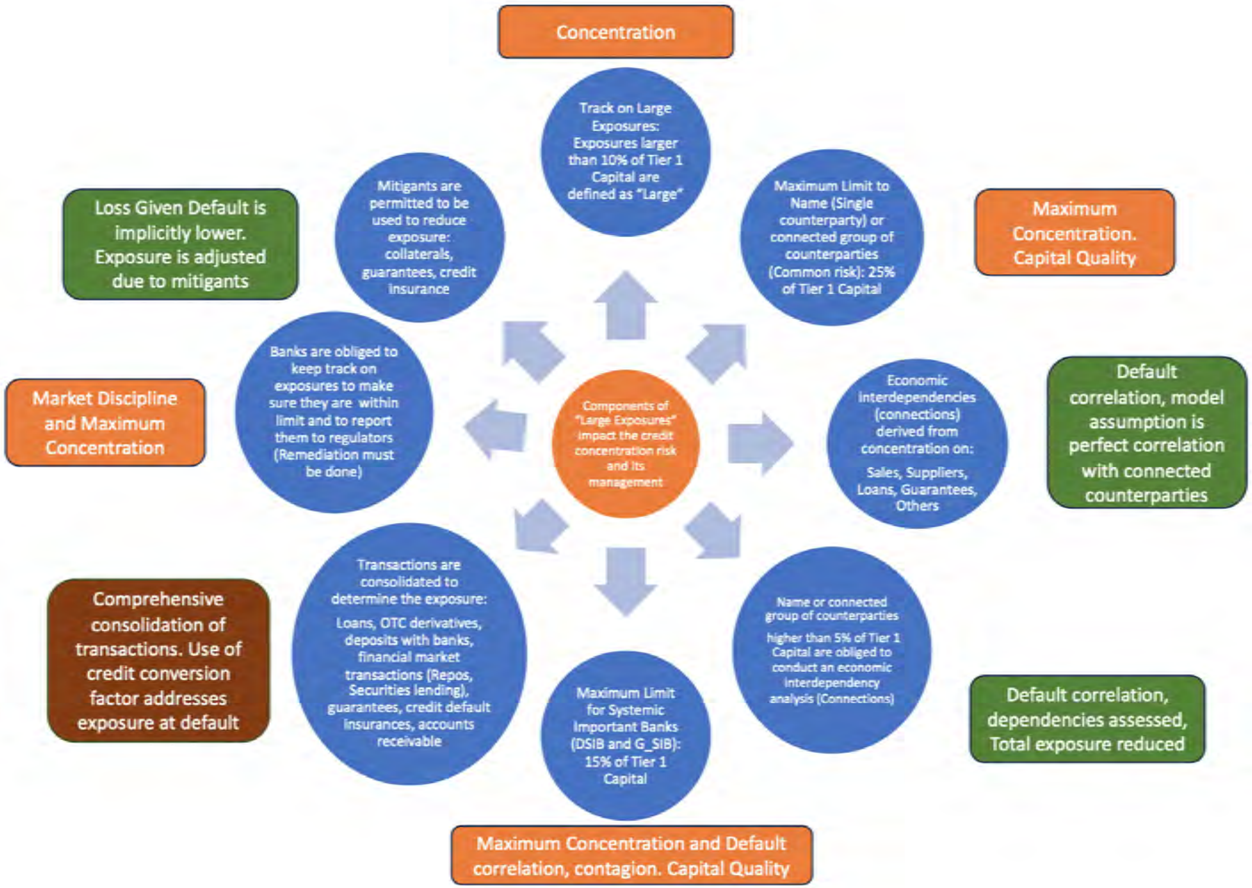
Table 1. LEX Components and Related Credit Risk Parameters

LEX Component	Related Credit Risk Parameter
Large Exposure equal to or higher than 10% of Tier 1 capital Limit of 25% of Tier 1 capital Limit of 15% of Tier 1 capital	**Concentration** vs infinite granular approach in the IRB Basel formula. Specific levels of concentration can be derived from LEX rules.
Exposures include: • Banking book (in- and off-balance) • Trading book Determination of exposure includes the use of credit conversion factors	**EAD (Exposure at Default).** Size of the exposure through consolidation of all exposures and the use of credit conversion factors.
Use of mitigants (collaterals, guarantees, credit insurance, etc.)	**LGD (Loss Diven Default)**. The use of permitted mitigants implies a lower LGD.
Economic interdependency analysis Grouping connected counterparties (Due to both control and economic interdependencies)	**Correlation**. Connected counterparties are considered by the LEX as a single exposure for concentration purposes, so a perfect correlation is assumed for connected counterparties.
Limits to G-SIBs and D-SIBs	**Systemic risk and diversification**. Limiting exposures with G-SIBs and D-SIBs forces the system to diversify the funding of banks that may have systemic impact and contagion.

Source: Prepared by the author.

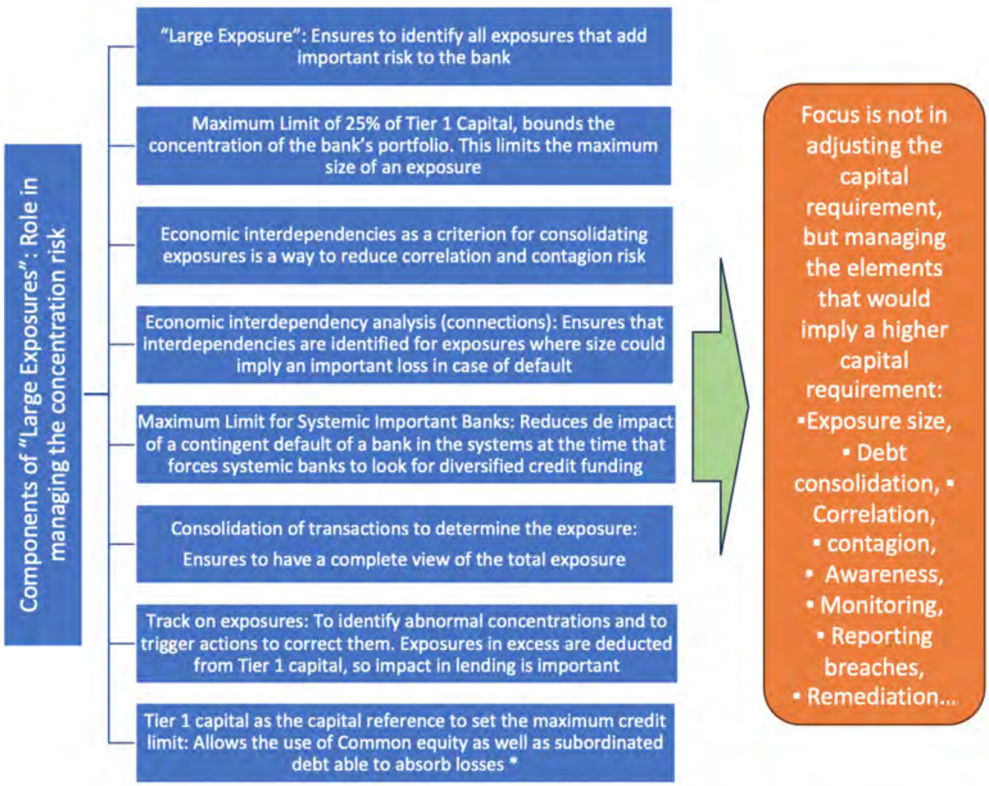
In theory, all those elements provide a framework to manage the concentration risk. In Figure 1 all the elements of the framework are joined and show at the center the management of the concentration risk (see Figure 1). We can trace how every component affects the credit risk parameters, as shown in Figure 2 (see Figure 2). Our objective is to model portfolios derived from the LEX regulation and to compute the concentration component that is completely absent from the Basel IRB formula.

Figure 1. Components of Large Exposures Regulation and Effect on Risk Management and Risk Metrics Parameters



Source: Prepared by the author.

Figure 2. Components of Large Exposures Regulation and Effect on Risk Management and Risk Metrics Parameters



* In the case of Mexico, only subordinated debt convertible to capital is allowed. The law must be modified to permit the use of other absorbing losses debt to be used.

Source: Prepared by the author.

Before addressing the task of computing capital requirements, it is worth pointing out that the LEX regulation is not the only component of the Basel regulation dealing with handling credit risk concentration. The complete framework is important to deal with this risk, but of course, not only with this risk. The next section addresses this issue.

5. The Basel Regulation Framework and its Connection with the Concentration Risk

At the end of the day, the LEX regulation is looking to avoid impacts on the financial system derived from large losses in banks due to large exposures in loans granted and other credit exposures.

As explained in the *Large exposures standard: executive summary* (Financial Stability Institute, 2022) document, the LEX standard is part of the Basel III reform package that complements the Basel Committee on Banking Supervision's risk-based capital framework to achieve a:

- Microprudential objective of serving as a backstop to the risk-based capital regime by protecting banks from incurring large losses from the default of a single counterparty or group of connected counterparties.
- Macroprudential objective of supporting efforts to manage systemic risks by reducing the interconnectedness between systemically important banks.

It is important to highlight the following:

- The systemic focus of the standard.
- The seeking of reduction of the interconnectedness.
- To bound exposures to limit large losses.
- The risk-based capital regime.

Thus, it is not only important to study the complete framework but also to compute the impacts on capital according to the implicit risks in the LEX.

First, let's address the BIS framework's components and their connection with the credit concentration risk. For our purpose, the relevant components are as follows:

I. The Basel IRB capital requirement. It determines the capital requirement for the IRB approach. The base assumption is a capital requirement over an infinite granular portfolio where concentration is absent. For non-IRB banks, there is a standard approach. In theory, this requirement is higher than the one for the IRB. Nonetheless, current rules and the IRB capital floors implemented recently may come close to both requirements.

II. The LEX standard as previously explained.

III. Conservation capital supplement. This supplement applies to all banks and may help them face any losses, including large ones (concentration).

IV. Capital Buffer for Systemic Important Banks (G-SIBs and D-SIBs). Applies only to Systemic Banks. In Mexico it ranges from 0.6% to 2.5%, representing 6.5% of the Risk-Weighted Assets (RWAs).

V. Total Loss-Absorbing Capacity (TLAC), applies only to Systemic Banks and is the maximum between 6.5% of the RWA's and 3.75% of the adjusted assets for leverage ratio computation (Mexico rule).

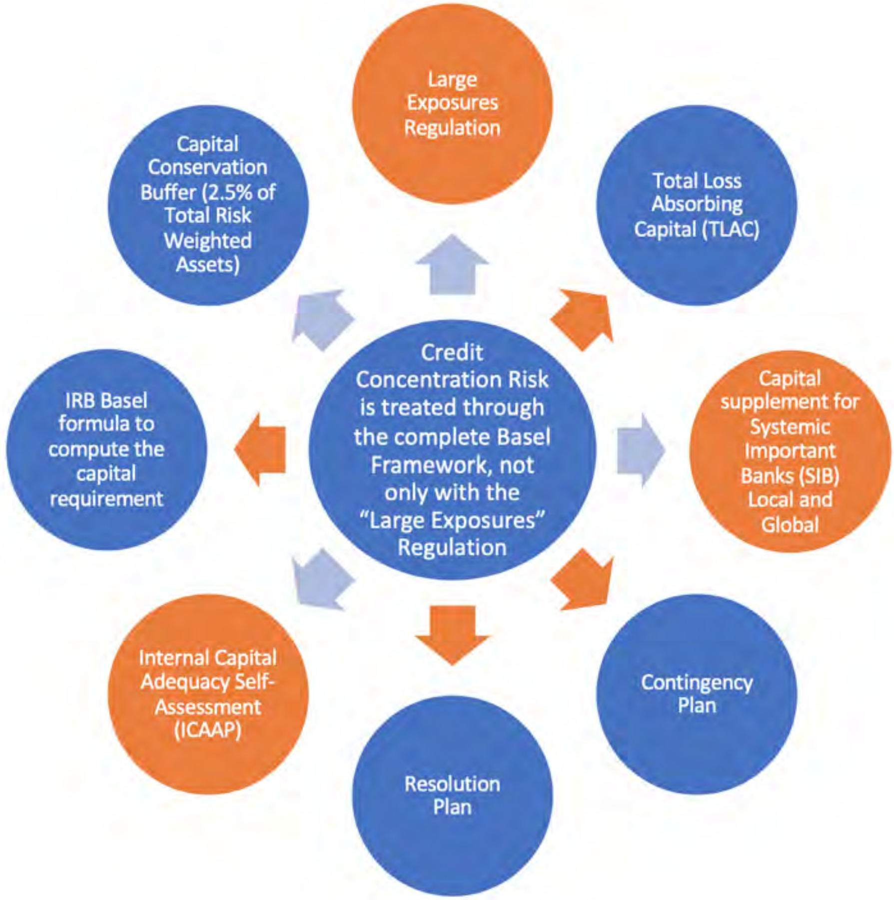
VI. Internal Capital Adequacy Assessment Process (ICAAP). It is an annual regulatory exercise that seeks to assess if the total capital that the bank has is enough to absorb the losses that the bank may face under different scenarios, including those of adverse economic conditions. Banks must demonstrate capital adequacy in all scenarios; otherwise, they must present a preventive action plan. The ICAAP is linked to other Basel regulation components to fulfill the complete capital regulation: the TLAC supplement, systemic capital supplements, conservation capital supplements, liquidity requirements, contingency plans, resolution plans, etc. Since the assessment must show that the bank is fulfilling all capital supplements and requirements in any scenario.

VII. Contingency plan. This plan is a detailed document that has all ordered feasible actions that the bank can execute to bring back the bank's capital ratio to compliant levels and to ensure continuity in its operations. It is a confidential plan, updated annually.

VIII. Resolution plan: This plan is confidential and entails a detailed and ordered process in case the bank's capital ratio falls below regulatory limits without the possibility of recovery. This plan is for the financial regulatory authority to take control and execute needed actions to protect public deposits due to the bank's financial problems, for example, insolvency derived from large losses due to credit risk concentration.

Figure 3 shows the integration of the pieces into a complete framework (see Figure 3). For our purposes, we place the credit concentration risk in the center but is a complementary piece of the framework.

Figure 3. Concentration Risk and its Integration into the BIS Regulation Framework



Source: Prepared by the author.

The author presents this framework in articulated form to understand the role that every piece has in managing concentration risk.

The 2008 banking crisis triggered important initiatives that will be reflected in specific regulations later. The focus of those initiatives was oriented to the resilience and stability of the financial system and of course to protect the economy, as explained by the Basel Committee on Banking Supervision:

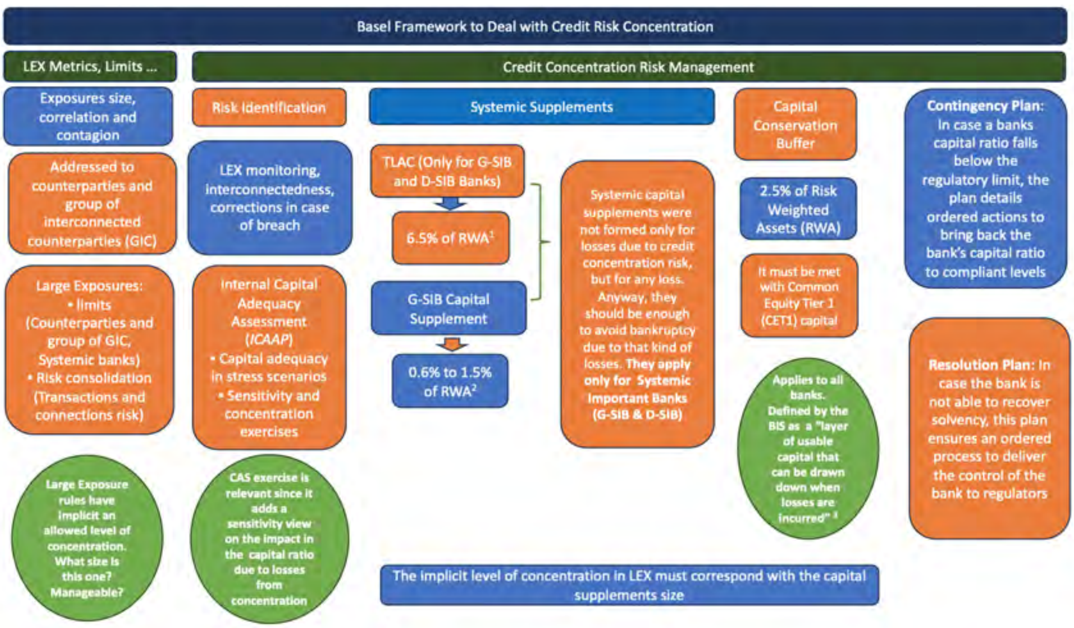
The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.05

> This consultative document presents the Basel Committee's proposals to strengthen global capital and liquidity regulations with the goal of promoting a more resilient banking sector. The objective of the Basel Committee's reform package is to improve the banking sector's ability to absorb shocks arising from financial and economic stress, whatever the source, thus reducing the risk of spillover from the financial sector to the real economy. (BCBS, 2009, p. 1).

In that sense, we must interpret the articulation of the regulation pieces explained before. The regulation addresses the main systemic components of the financial system—the important part is the system, not a specific group of banks.

Let us use Figure 4 to complete the articulation of the pieces (see Figure 4).

Figure 4. Role of the Pieces of the BIS Regulation Framework in Managing Credit Concentration Risk



[1] The rule in Mexico is a maximum between 6.5% of the risk-weighted assets or 3.75% of the risk-adjusted assets for leverage ratio computation purposes.

[2] In Mexico the supplement is assigned by the regulator to every bank designated as locally systemically important or D-SIB. The size of the supplement depends on the systemic impact of the bank.

[3] Financial Stability Institute (2019).

Source: Prepared by the author.

We can notice, on the one hand, that systemic banks have two heavy capital supplements to deal with losses: G-SIB and D-SIB losses (from 0.6% to 1.5%) and the TLAC supplement of approximately 6.5% of the bank's risk-weighted assets. On the other hand, the rest of the banks are only obliged to form the capital conservation supplement, not the TLAC supplement, as shown in Table 2 (see Table 2).

Table 2. Summary of Banks' Capital Supplements

Supplement (as % of Risk-Weighted Assets or RWAs)	Mexican banks
Capital conservation 2.5% of RWAs	All Banks
Domestic Systemic Important Banks: 0.6% to 1.5% of RWAs	Domestic Systemic Important Banks (6)
Total loss-absorbing capacity (TLAC): 6.5% of RWAs or 6.75% of risk adjusted assets used for leverage ratio computation.	Domestic Systemic Important Banks (6)

Source: Prepared by the author, based on information from several sections of CNBV (2024a).

There are minimum regulatory capital requirements—including the supplements. Regulatory reporting ensures that those minimums are complied with. To assess any potential risk, banks must conduct a regulatory Internal Capital Adequacy Assessment Process (ICAAP) annually. In the manual and template to conduct such assessment, the Comisión Nacional Bancaria y de Valores (CNBV, 2023, pp. 32–33), the Mexican regulator, includes in Section 3.4.15 Sensitivity Analysis, the following: a) Simultaneous write-offs of the ten main counterparties or group of connected counterparties, (write-offs adjusted by loss given default). Sensitivity must include Public Sector Entities (PSEs such as Petróleos Mexicanos [Pemex] and Comisión Federal de Electricidad [CFE]).

The effect of this sensitivity on the regulatory capital (through the capital ratio) must be computed and disclosed to the regulator.

We need to point out several issues about this sensitivity.

The simultaneous write-off assumption is very heavy for any portfolio since a Bank that uses the maximum concentration limits, by definition, among the 10 main counterparties must include the four maximum large exposures, that in its limit would sum the 100% of the Tier 1 capital, and the next six large exposures below those four.

That implies a perfect correlation among the ten main counterparties or group of connected counterparties regardless of sectorial connections and activity, including Public Sector Entities (PSEs).

The reader can easily guess that a loss of this size would be an important share of any bank's capital.

As an important reference to assess the relevance of concentration risk in the Mexican financial system, let us review the information presented in Table 3. This shows the vulnerabilities identified by regulators in the ICAAP for the years 2017 to 2023 (see Table 3). On average, fifteen banks showed vulnerabilities due to credit concentration. In 2022 ICAAP showed a vulnerability in sixteen banks involving 12.1% of the assets of the banking system (see notes included in Table 3). Note as well that, on average, eight banks presented capital shortfall in the ICAAP exercise.

Table 3. Vulnerabilities Identified by Mexican Regulators in the Internal Capital Adequacy Assessment Process (ICAAP) for the Years 2017–2023

		Vulnerable Banks in (ICAAP) due to:				
CESF Report as of March:	Period of the ICAAP	Loan Concentration		Capital Shortfall	Liquidiry Coverage Ratio	Depositors
2018	2017 - 2019*	19		8	8	15
2019	2018 - 2020*	15		5	3	16
2020	2019 - 2021*	12		6	3	14
2021	2020 - 2022**	12	0.4% ***	10	2	20
2022	2021 - 2023	12	0.5% ***	7	1	13
2023	2022 - 2024	16	12.1% ***	10	0	14
2024	2023 - 2025	18	0.45% ***	9	0	16
	Average	15		8	2	15

* At least one bank presented a risk of loan concentration due to sensitivity to accumulated write-offs of the ten largest counterparties that drove its capital ratio below 10.5% (minimum regulatory level).

** A least one bank presented a risk of loan concentration due to sensitivity to accumulated write-offs of the ten largest counterparties that drove its capital ratio to the minimum regulatory level.

*** Assets of vulnerable banks due to loan concentration to system total assets (percentage).

Source: Prepared by the author with information from Consejo de Estabilidad del Sistema Financiero (2018–2024).

Although the result of this sensitivity does not imply a failed (ICAAP) regulatory exercise, the bank may eventually face the question of how it would recover from a loss of this size. In such a case the action plan that the bank must outline in its contingency plan (CP) should be enough to face such losses. A bank with an important risk in this sensitivity at least will be on a watch list. Therefore, results from one exercise (ICAAP) can be linked to another (CP) straightforwardly, which is one of the benefits of the integrated framework from a regulatory point of view.

Obviously, there is an offset of the write-offs. The risk-weighted assets will also be reduced, but in any case, the net effect is important.

That risk is present in all banks, but the supplements for systemic banks are enough to cover any loss of this size. A regular (non-systemic) bank has only the capital conservation supplement of 2.5%. We will address this issue later when we compute the capital concentration add-on.

The LEX regulation is oriented to all banks including those that use the standard approach to compute their regulatory capital. For this reason, instead of including a specific capital requirement in the IRB formula, the approach is to control important parameters that have to do with the capital requirement, regardless of the bank being systemic or not, or uses the IRB or the standard approach to compute its capital requirement. The parameters are as follows:

- Total exposure consolidation.
- Limit to the size exposure, being more acid to systemic bank's exposure.
- Correlation.
- Connections and contagion.

The process that we will follow to determine the add-on is the following:

I. Assess the LEX regulation released by both the BIS and Mexico and its implications for credit portfolio concentration to design the portfolios.

II. Compute the regular capital requirement with the IRB formula approach as a comparison yardstick.

III. Determine feasible levels of concentration implicit in the LEX regulation.

IV. Compute the capital requirements through value at risk and the conditional value at risk metrics for every portfolio designed.

V. Use the Hirschman-Herfindahl Index to make portfolio size homogeneous and comparable.

VI. Using a CreditMetrics-like model, obtain the credit risk metrics (VaR and CvaR) for diverse levels of concentration through a Montecarlo simulation.

VII. Compute the concentration add-on for diverse levels of concentration.

VIII. Assess the resulting add-on regarding capital requirements and its importance.

6. Levels of Concentration Implicit in the LEX Regulation

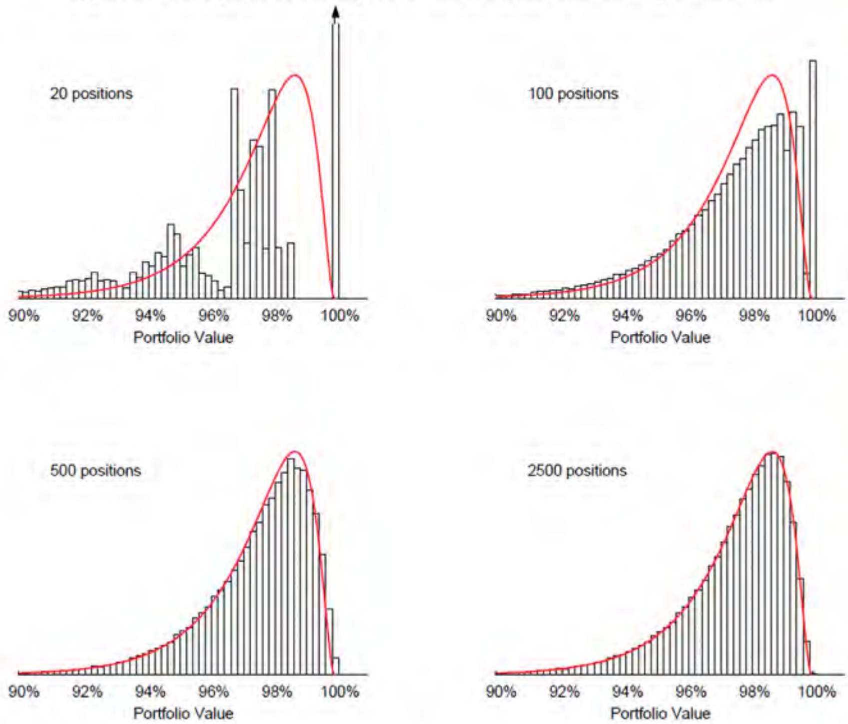
LEX regulations released by the BIS and CNBV in Mexico have implications for credit portfolio concentration, regardless of whether the bank computes the capital requirement measured with the IRB formula or determined with the standard model.

The IRB formula assumes infinite granularity in the portfolio. For practical purposes, we reduce the problem dimension based on Yi Xiao and Finger's research.

The problem we face is the same one that Yi Xiao and Finger (2002) addressed since, in both cases, we are parting from homogeneous and fine-grained portfolios. In our case, a corporate one under the Basel IRB Basel formula, that assumes infinite granularity and homogeneous risk, homogeneity and fine-grained exposures in the case of retail exposures, in both cases the dimension of the problem is reduced to analyze the problem in a more tractable size.

> Consider a portfolio with an average default probability of 5%, an average recovery rate of 50%, and an average correlation of 10%. Assuming there are 20, 100, 500, or 2500 exposures in this portfolio, each to a distinct obligor, we carry out full-blown Monte Carlo simulations with 100,000 scenarios for each case. The distribution of the portfolio is calculated from the simulation and shown as a histogram in Figure 1. With a small number of exposures in the portfolio, the distribution is rather discrete, dominated by the properties of individual positions.
>
> With more and more exposures in the portfolio, the default of a single exposure has less and less impact on the total pool, and the distribution becomes progressively smoother. The distribution eventually converges to the homogeneous and fine-grained limit shown as solid lines in Figure 1. (Yi Xiao & Finger, 2002, p. 3. Please note that the following Figure 1 is part of the quote).

Figure 1: Convergence of a loan pool to the homogeneous and fine-grained limit.



Obligor correlation $\rho = 10\%$, default probability $p = 5\%$, recovery rate $R = 50\%$. Exposure sizes are normally distributed with standard deviation equal to 1/3 of mean.

Source: Yi Xiao and Finger (2022, p.3).

On the one hand, in the quotation above note that in Xiao and Finger convergence was at 100,000 trials in a Monte Carlo simulation exercise.

On the other hand, we will use the one presented by the Committee on Banking Supervision (2006) as a reference yardstick.

> Credit concentration is sometimes known as lack of granularity. This section discusses how to extend the ASRF model to incorporate the effect of granularity.
>
> To fix ideas, consider how economic capital (credit VaR) varies over a sequence of loan portfolios with the following structure: they all contain a number of exposures

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.05

> to similar credits which are all of the same size with the exception of one that is ten times that size. Table 1 depicts the tail of the simulated loss distribution for seven such portfolios of different sizes ranging from 10 to 3000 credits. As the number of credits increases the importance in the portfolio of the single large exposure declines and the economic capital converges to the one corresponding to the infinitely granular case (BCBS, 2006, p. 9).

The following Table 4 (Table 1 of the BCBS, 2006, p. 9, and Table 4 in this text) appears on page 9 of the BCBS document (see Table 4).

Table 4. Scenarios Present in BSI's Basel Committee on Banking Supervision (BSCS, 2006) Working Paper 15

Table 1

A stylised example of the effect of granularity on portfolio risk

Number of loans	10	50	100	500	1,000	2,000	3,000
VaR(95%)	.0526	.0508	.0459	.0393	.0386	.0378	.0389
VaR(99%)	.5263	.1695	.1009	.0786	.0773	.0762	.0758
VaR(99.9%)	.5263	.1864	.1284	.0982	.0971	.0950	.0947

Note: Credit VaR at the specified level of confidence expressed as a fraction of total portfolio exposure. The calculations assume PD=1% and asset correlation of 20%.

Source: BCBS (2006, p. 9).

With this information, we can infer a simple add-on, if we know the increase in concentration, assuming that the 3000 exposures portfolio mimics the granular portfolio case. For instance, at a 99.9% level of confidence passing from three thousand loans to one hundred implies an add-on of 3.37% using the VaR risk metric.

In this same document, the following question is asked: "How important is the effect of name concentration on economic capital?" (BCBS, 2006, p. 9) The answer is important for our purposes:

- For large credit portfolios of over 4000 exposures, the effect is 1.5% to 4%.
- For smaller portfolios (with 1000 to 4000 loans) the effect ranges from 4% to 8%.

The 3000 exposures portfolio size is similar to the one proposed by Xiao and Finger (2002).

In our exercise, we assumed that we had a portfolio of 3000 granular exposures (same size and risk), and we derived alternative concentrated portfolios as follows, according to the main sizes found in Basel Committee on Banking Supervision (2006):

I. A portfolio of 3000 exposures but including the four allowed largest exposures (each one representing 25% of the Tier 1 capital, assuming a capitalization of 10.5% of the portfolio), the rest of the exposures remain the same size.

II. A portfolio of 2000 exposures but including the four allowed largest exposures (each one representing 25% of the Tier 1 capital, assuming a capitalization of 10.5% of the portfolio), the rest of the exposures remain the same size.

III. Portfolios of 1000, 500, 100, and 50 exposures were built in the same way: four allowed the largest exposures (each one of 25% of the Tier 1 capital, assuming a capitalization of 10.5% of the portfolio) the rest of the exposures remain the same size. As we reduce the number of loans, the money size of each exposure grows (since the rest of the portfolio, apart from the four main exposures, is divided into a lower number of loans), increasing the concentration effect.

The monetary amount of the portfolio and loans was selected as follows:

- Portfolio of loans $ 3347.00
- Regulatory capital $ 351.40 (10.5% of portfolio loans, as Tier 1 capital)
- Size of each of the four permitted large exposures: $ 87.9.

If we were to consider the Tier 1 capital of systemic banks, the level of concentration would be higher, but we can calculate the add-on with a rule we can derive from our results.

We also worked on the 3000-size portfolio with equal-size exposures in Table 5 (see Table 5).

Table 5. Concentration Effect Assuming Scenarios Derived from the Large Exposure Regulation

Original Number of Loans	3,000		2,000		1,000		500		100		50	
	LOANS	HHI	LOANS	HHI	LOANS	HHI	LOANS	HHI	LOANS	HHI	LOANS	HHI
4 Maximums (25% each)	331	0.00302	317	0.00316	281	0.00356	229	0.00437	90	0.01110	50	0.02017
4 Maximums + 1 SIB (15%)	307	0.00326	295	0.00339	264	0.00378	219	0.00457	90	0.01114	50	0.02018
4 Maximums + 2 SIB (15% each)	286	0.00350	276	0.00363	250	0.00400	210	0.00476	89	0.01118	50	0.02020
4 Maximums + 3 SIB (15% each)	267	0.00374	259	0.00386	237	0.00422	202	0.00496	89	0.01123	49	0.02021
4 Maximums + 3 SIB + 10 Large Loans (10% each)	209	0.00479	205	0.00488	194	0.00516	174	0.00575	89	0.01125	47	0.02132

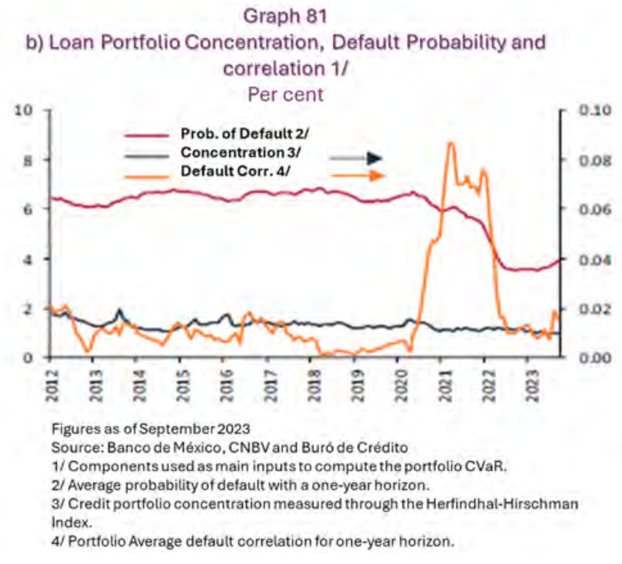
Source: Prepared by the author.

In the first line, we will compute the add-ons for all scenarios, and based on the results, we will infer the rest of the cases.

In the BCBS's framework, we find the following reference regarding the "Maximum permissible concentration under EU large exposures rules. Such calculations give estimates of 13% to 21% higher portfolio value-at-risk for this highly concentrated portfolio versus a perfectly granular one that is comparable in all other dimensions." (BCBS, 2004). Also, note 8 states the following: "Directive 93/6/EEC of 15 March 1993. An estimate of the HHI for such a portfolio would be about 0.0156"; following the HHI inverse rule (Márquez Diaz-Canedo, 2003, pp. 198-199) that number would be around 64 loans, so that estimate falls within our scenarios.

According to the Banco de México (the Mexican Central Bank) in its second semester report of 2023 on financial stability (Banco de México, 2023) the concentration (measured by the Herfindahl-Hirschman Index HHI) of Mexican banks ranged from 0.01 to 0.02 during the 2012–2023 period (see Graph 1). This is a significant level of concentration. Using the metric 1/HHI, we find that it is equivalent to portfolios with a number of loans between 50 and 100. As stated before, those numbers are consistent with the scenarios designed for this work.

Graph 1. Banco de México Graph on Loan Portfolio Concentration, Default Probability, and Correlation



Source: Banco de México (2023, p. 67).

Thus, for our purposes, the relevant HHI and associated loan portfolios are between 100 and 50 loans.

The parameters and other assumptions for our exercise are the following:

- Portfolio correlation: 20%.
- 1-year probability of default: 1%.
- Loss Given Default: 100%.
- Simulations by scenario: 500,000 (Recall that Xiao and Finger [2002] used 100,000 simulations, Martin Hibbeln 3,000,000, and other authors 400,000). The Montecarlo Simulation is a methodology capable of providing very accurate results for specific levels of concentration, as is our case. Nonetheless, the computation time burden is enormous.

The model used is a CreditMetrics-like model developed by the author. The Cholesky decomposition is used to obtain credit-correlated default scenarios. The original methodology is disclosed in JP Morgan and Reuters (1996).

The metrics to be used to compute the add-on are the Value at Risk of the portfolios and The Conditional Value at Risk (VaR and CVaR). Nonetheless, although we will compare the results from both, CvaR will be the final chosen metric to compute the add-on. We know from Artzner et al (1999) that VaR fails as a coherent measure of risk, specifically in the subadditivity property at high levels of confidence, as is the case in our research. The CVaR has become the dominant risk metric for many standards.

7. Capital Requirements and VaR Using the Basel IRB Formula

As a comparison yardstick, we computed the capital requirement using the Basel IRB formula. Is important to point out that this formula considers no concentration at all, which is why there is a need to compute the add-on resulting from passing to the real world, where concentration is an important risk factor to be considered.

The capital requirement for a loan under the selected assumptions is computed with the Basel IRB formula as follows.

Capital requirement:

$$(K) = \left[LGD * N\left[(1-R)^{-0.5} * G(PD) + \left(\frac{R}{1-R}\right)^{0.5} * G(0.999)\right] - PD * LGD\right] * (1-1.5*b)^{-1} * (1+(M-2.5)*b) \quad \text{(Equation 1)}$$

Where the correlation is:

$$(R) = 0.12 * \frac{1-e^{-50*PD}}{1-e^{-50}} + 0.24 * \frac{1-(1-e^{-50*PD})}{1-e^{-50}} \quad \text{(Equation 2)}$$

Maturity adjustment:

$$(b) = \left(0.11852 - 0.05478 * Ln(PD)\right)^2 \quad \text{(Equation 3)}$$

Ln = natural logarithm

N(x) = standard normal cumulative distribution function

G(x) = standard normal inverse cumulative distribution function

Source: BCBS (2004, para. 272).

Notice that the confidence level of this requirement is 99.9%. To align the formula with our exercise, we fix the term as one year.

The reference capital requirement using this formula and our assumptions is 13.03%.

8. Results Computing the VaR and CVaR of Portfolios with Different Concentration Levels

8.1 Introduction to Formulas and Calculations

Simulation algorithm:

- We obtain a vector of n independent, identical standard normal distributed random variables N(0,1); where n represents the size of the portfolio or portfolio replica. Let us call this vector Z.
- Since we are using a single correlation value of 20% for all elements in the portfolio, we part from an N x N size correlation matrix with a value of 1 in the diagonal and 0.2 value in all the rest elements of the matrix. The correlation matrix is C.

- We obtain the Cholesky decomposition of the correlation matrix (inferior triangular) and design this result as the M Matrix. To do so we used the algorithm described in JP Morgan and Reuters (1996, p.254, Appendix E): "Routines to simulate correlated normal random variables", and section E.2, "Applying the Cholesky decomposition." Following this algorithm, beginning from the correlation matrix C, and considering that we have standard normal random variables where the standard deviation is equal to 1, then C is equal, in terms of M, to:

$$C = M^T * M \qquad \text{Equation 4}$$

Let *i* and *j* index be the row and the column of an N x N matrix.

The diagonal elements of M are computed with:

$$a_{ii} = \left(s_{ii} - \sum_{k=1}^{n-1} a_{ik}^2\right)^{1/2} \qquad \text{Equation 5}$$

s_{ii} represents any element in the diagonal of C.

The rest of the elements of M are computed with:

$$a_{ij} = \frac{1}{a_{ii}}\left(s_{ij} - \sum_{k=1}^{n-1} a_{ik}a_{jk}\right)^{1/2} \qquad \text{Equation 6}$$

s_{ij} represents any element out of the diagonal of C.

j=i+1, i+2,...,N

- Once having M, we obtained correlated vectors (Z_c) of the normal distributed random variables by applying the following formula.

$$\boldsymbol{Z_c} = (\boldsymbol{M} * \boldsymbol{Z^T})^T \qquad \text{Equation 7}$$

- Every random variable represents a loan. We decide if a loan is paid or defaulted if the random variable is equal to or lower than the fix threshold of -2.3263, since we set the probability of default of the exercise at 1%.
- If the loan is paid, its value is equal to the original exposure, if the loan is defaulted, the value of the loan is equal to:

$$Loan = Expsoure * (1\text{-}LGD) \qquad \text{Equation 8}$$

- In every trial we obtained loans paid and loans defaulted, the sum of all values gives the value of the portfolio in that trial.
- We repeated that process 500,000 times, recording every result.

8.2 Computing the Risk Metrics

Risk metrics have evolved over time. The first risk metric used was exclusively volatility (standard deviation). Value at risk (VaR) was set as a new risk metric standard paradigm in the 1990's. Artzner et al (1999) pointed out that the value at risk was not a coherent metric of risk, given that it does not fulfill the subadditivity property that ensures that the risk in a portfolio is lower than considering every element of the portfolio in a separated way and adding the individual risks. This happens especially in credit portfolios with a very low probability of default and which compute the VaR using elevated levels of confidence.

According to Venegas (2008, p. 694):

> The value at risk of X at a level (of confidence) of 1-q denoted by – VaR, is defined as the worst value of the portfolio, in a given period, [t,T], for a confidence interval of (1-q)100%. In a more accurate way:

$$\mathbb{P}_\theta\left\{-VaR_{1-q}^{X} \leq X\right\} = 1 - q \qquad \text{Equation 9}$$

Since we are using a Montecarlo method to compute the VaR, we will use this alternate expression presented by Venegas (2008, p. 694):

$$VaR_{1-q}^{X} = -Inf\{X \in \mathbb{R} | \mathbb{P}_\theta\{X > x\} \leq 1 - q\} \qquad \text{Equation 10}$$

In our work we computed the value of the portfolio for every one of the 500,000 scenarios and obtained first the average value of the portfolio. In the credit risk the total loss is divided into two components: the average loss is called expected loss and this constitutes the provision for credit losses (credit allowance). Losses that go far from the average losses to the VaR are equal to the economic capital or capital requirements at the confidence level that the VaR was computed. Using losses with positive sign we have:

$$Capital = VaR_{1-q}^{X} - Expected\ Loss \qquad \text{Equation 11}$$

To find the VaR of the portfolio:

1. All scenarios are arranged from worst to best portfolio values.
2. To find the VaR at a confidence level of 99.9%, for 500,000 scenarios we compute 500000*(1-99.9%) = 500. Therefore, to find the VaR, the value of the portfolio in the 500 scenario will be the value at risk.
3. To obtain the capital requirement we subtract the expected loss from the value, this is called the "unexpected loss."

The conditional VaR (CVaR), in turn, is a metric that fulfills all properties of the coherent risk framework, including sub-additivity. This metric works with the losses once we have overpassed the value at risk loss—it is computed as the average of all losses exceeding the Value at Risk and includes all losses conditional to exceed the value at risk. Venegas (2008, p. 706) defines CvaR as follows:

$$\mathcal{E}_{1-q}^{X} = VaR_{1-q}^{X} - E\left[X + VaR_{1-q}^{X} \middle| VaR_{q}^{X} + X < 0\right] \qquad \text{Equation 12}$$

To obtain the CVaR we will compute the average of losses that are higher than the VaR, that is straightforward since we already have ordered the complete set of simulated losses, we have to include in the computation of the average all excluded scenarios from the VaR computation.

Next, we compute the Capital requirement through the CvaR as follows:

$$Capital = \mathcal{E}_{1-q}^{X} - Expected\ Loss \qquad \text{Equation 13}$$

8.3 Portfolio Composition

To assess the effect of credit concentration in the portfolios we designed the following portfolios. Concentration in a portfolio increases as the number of loans (exposures) decreases.

1) Portfolio A 3000 has all exposures of the same size. In our exercise, this case is the most similar to an IRB granular portfolio.

2) Portfolio B 3000 has four exposures with an individual limit of 25% of Tier 1 Capital. The rest are exposures of the same size. This portfolio includes in a granular portfolio the effect of the maximum credit limit to a counterparty or group of

connected counterparties: 25%. We assume that four counterparties or groups of connected counterparties use this limit, so the limit of 100% of Tier 1 capital allocated in four counterparties or groups of connected counterparties is reached.

3) The rest of the portfolios (with sizes of 2000, 1000, 500, 100, 50, 40) have four exposures with an individual limit of 25% of Tier 1 Capital. The rest exposures are homogeneous in size. This follows the same logic of Portfolio B 3000.

Once we defined the portfolio size and composition, we executed the simulation process and computed the risk metrics. The results were as presented in Table 6 (see Table 6).

Table 6. Results of VaR and CVaR for Built Portfolios

	Value at Risk				Conditional Value at Risk			
	VaR Confidence Level				CVaR Confidence Level			
Portfolio Size	99.90%	99.50%	99.00%	95.00%	99.71%	98.56%	97.10%	85.40%
3,000 Equal	13.60%	8.53%	6.57%	2.80%	13.59%	8.53%	6.59%	2.86%
3,000 Portfolio	13.90%	8.88%	6.91%	3.06%	13.90%	8.87%	6.91%	3.06%
2,000 Portfolio	14.14%	8.95%	6.98%	3.04%	14.01%	8.93%	6.94%	3.07%
1,000 Portfolio	14.28%	8.98%	6.93%	3.05%	14.13%	8.93%	6.93%	3.07%
500 Portfolio	14.34%	9.02%	7.00%	3.07%	14.19%	9.01%	7.01%	3.12%
100 Portfolio	14.85%	10.01%	7.40%	3.66%	15.08%	9.87%	7.30%	3.58%
50 Portfolio	16.51%	10.67%	8.72%	3.56%	16.60%	10.70%	8.67%	3.57%
40 Portfolio	16.68%	11.43%	8.95%	3.97%	17.42%	11.26%	9.05%	3.71%

Source: Prepared by the author.

Recall that our comparison reference value is the VaR implicit in the IRB formula capital requirement—that is, 13.03%.

When a metric changes from VaR to CvaR, it is good practice to select the latter's confidence level to replicate the risk of the VaR. One example is to use 97.25% for CVaR and 99.9% for VaR, as shown in Graph 2 (see Graph 2). In this work, we determined the CVaR confidence level as 99.71% to align both metrics, as in Graph 3 (see Graph 3). Other works identify a confidence level of 99.72%.

https://doi.org/10.36105/theanahuacjour.2024v24n1.05

Graph 2. Value at Risk for Every Portfolio at a 99.9% Confidence Level

Source: Prepared by the author.

Graph 3. Conditional Value at Risk for Every Portfolio at 99.71% Confidence Level

Source: Prepared by the author.

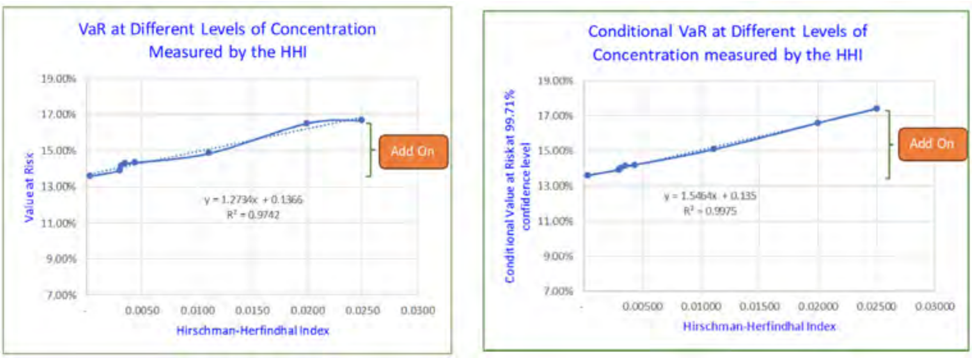
The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.05

9. Computing the Concentration add-on and Assessment of Results

We obtained the add-ons by comparing the risk metric with a given level of concentration minus the capital requirement with the IRB formula. Not all banks are IRB, but the idea is to get a comparison parameter, and this is a sound one.

Graph 4 shows the resulting VaR and CVaR. It is easy to notice that the CVaR metric is much more stable. This has to do with its coherence, and so we used it to determine the add-one (see Graph 4).

Graph 4. Concentration add-on with VaR and CVaR



Source: Prepared by the author.

Despite the size of the simulations, it is advisable to correct any deviation coming from the method used and the slow convergence that we noted. For this reason, we obtained a fitted add-on from a linear regression between the HHI and the obtained add-on. Results are shown in Tables 7 and 8 and Graph 5 (see Table 7, Table 8 and Graph 5).

Table 7. Fitted Concentration add-on with the VaR Metric

Portfolio Size	1/HHI	HHI	Obtained VaR	Fitted VaR	VaR Add On
3,000 Equal	3,000	0.00033	13.60%	13.76%	0.73%
3,000 Portfolio	331	0.00302	13.90%	14.11%	1.08%
2,000 Portfolio	317	0.00316	14.14%	14.13%	1.10%
1,000 Portfolio	281	0.00356	14.28%	14.18%	1.15%
500 Portfolio	229	0.00437	14.34%	14.29%	1.26%
100 Portfolio	90	0.01111	14.85%	15.16%	2.13%
50 Portfolio	50	0.02000	16.51%	16.31%	3.28%
40 Portfolio	40	0.02500	16.68%	16.96%	3.93%

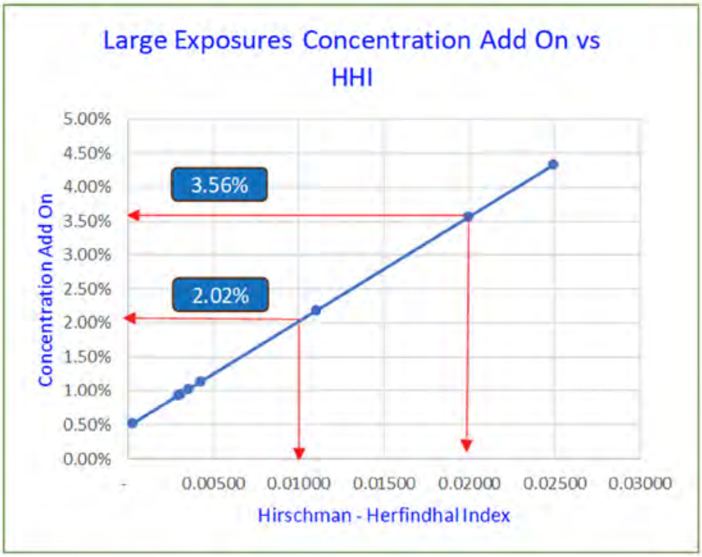
Source: Prepared by the author.

Table 8. Fitted Concentration add-on with the CVaR Metric

Portfolio Size	1/HHI	HHI	Obtained CVaR	Fitted CVaR	CVaR Add On
3,000 Equal	3,000	0.00033	13.59%	13.55%	0.52%
3,000 Portfolio	331	0.00302	13.90%	13.97%	0.94%
2,000 Portfolio	317	0.00316	14.01%	13.99%	0.96%
1,000 Portfolio	281	0.00356	14.13%	14.05%	1.02%
500 Portfolio	229	0.00437	14.19%	14.18%	1.15%
100 Portfolio	90	0.01111	15.08%	15.22%	2.19%
50 Portfolio	50	0.02000	16.60%	16.59%	3.56%
40 Portfolio	40	0.02500	17.42%	17.37%	4.34%

Source: Prepared by the author.

Graph 5. Fitted Concentration add-on with CVaR



Source: Prepared by the author.

As expected, there is a linear relation between the HHI and the add-on. If we use this relation to obtain the add-on for concentrations of 0.01 and 0.02 (the range of concentrations for the Mexican Bank system as of December 2023) we find add-ons of 2.02% and 3.56%, as shown in the previous graph.

The results are similar to those in other works. We have mentioned the Basel Committee on Banking Supervision (2004) and the EU large exposure rules quoted in that paper.

- Large Exposures of more than 4000 are 1.5% to 4%.
- Smaller portfolios (1000 to 4000 Loans) range from 4% to 8%.

One extensive study on concentration is Hibbeln (2010). This author uses several approaches to compute the add-on. He estimated the add-ons with the following results (see Table 9).

Table 9. Estimated add-ons from Hibbeln (2010)

	Add On	
HHI	HQMC	LQMC
0.010	1.56%	1.82%
0.020	2.93%	3.64%

HQMC: High-Quality Montecarlo (3 million simulations)

LQMC: Low Quality Montecarlo

Source: Estimated by the author based on Figure 4.3. Granularity add-on for heterogeneous portfolios, in Hibbeln (2010).

The resulting add-ons for the scenarios initially proposed are presented in Table 10 (see Table 10). Recall that the add-on for a completely grained portfolio of 3000 loans is 0.52. In the next case, it passes from a granular to a concentrated portfolio having the four maximum exposures permitted by the LEX regulation—the add-on is 0.94% for the first 3000-size portfolio scenario, and 1.21% for the last.

The add-on changes little for portfolio sizes of 2000 and 1000, but for sizes of 100 and 50 (with HHI equal to 0.01 and 0.02), in our relevant area, the add-on increases considerably: ranging from 2.19% to 2.21% for the former and 3.59% to 3.77% for the latter.

Table 10. Add-ons for the Proposed Scenarios

Original Number of Loans	3,000			2,000			1,000		
Scenario	LOANS	HHI	Add On	LOANS	HHI	Add On	LOANS	HHI	Add On
4 Maximums (25% each)	331	0.0030	0.94%	317	0.0032	0.96%	281	0.0036	1.02%
4 Maximums + 1 SIB (15%)	307	0.0033	0.97%	295	0.0034	0.99%	264	0.0038	1.05%
4 Maximums + 2 SIB (15% each)	286	0.0035	1.01%	276	0.0036	1.03%	250	0.0040	1.09%
4 Maximums + 3 SIB (15% each)	267	0.0037	1.05%	259	0.0039	1.07%	237	0.0042	1.12%
4 Maximums + 3 SIB + 10 Large Loans (10% each)	209	0.0048	1.21%	205	0.0049	1.22%	194	0.0052	1.27%

Original Number of Loans	500			100			50		
Scenario	LOANS	HHI	Add On	LOANS	HHI	Add On	LOANS	HHI	Add On
4 Maximums (25% each)	229	0.0044	1.15%	90	0.0111	2.19%	50	0.0202	3.59%
4 Maximums + 1 SIB (15%)	219	0.0046	1.18%	90	0.0111	2.19%	50	0.0202	3.59%
4 Maximums + 2 SIB (15% each)	210	0.0048	1.21%	89	0.0112	2.20%	50	0.0202	3.59%
4 Maximums + 3 SIB (15% each)	202	0.0050	1.24%	89	0.0112	2.21%	49	0.0202	3.60%
4 Maximums + 3 SIB + 10 Large Loans (10% each)	174	0.0057	1.36%	89	0.0112	2.21%	47	0.0213	3.77%

Source: Prepared by the author.

It is also important to note that "vertically" the increase of the add-on from the best (only 4 maximum large exposures permitted) to the worst scenario (4 maximum LE + 3 D-SIB Exposures +10 Loans larger Than 10% of Tier 1 capital) is moderated (see Table 10, and for complementary support, see Table 11). Therefore, we conclude that the main contribution to concentration is by one of the four largest loans.

"Horizontally" the average increase of the add-on is 2.59%, so the most critical component is the implicit reduction in the number of loans according to the inverse of the HHI. We can take as a concise result of this work the one in Table 10 for concentration levels of HHI= 0.01 and HHI=0.02, which is an add-on ranging from 2.2% to 3.6% in higher concentrations, as in Table 10 (see Table 10).

So, in summary, the vertical behavior (increasing base concentration) is the key driver. Adding more concentration to base scenario does not contribute heavily to capital requirement, as shown in Table 11 (See Table 11).

Table 11. Horizontal View: Increase in add-ons from "Best" to "Worst" Concentration Scenarios

Potfolio Size	3,000	2,000	1,000	500	100	50
Increase in Add On From scenario 1 to 5	0.27%	0.27%	0.25%	0.21%	0.02%	0.18%

Source: Prepared by the author.

Please note that the add-on is 4.34% for a forty-loan equivalent portfolio size, and concentration escalates the capital requirement heavily.

Assigned to a specific supplement of capital or not, the conclusion is that credit portfolio concentration implicit in LEX accounts for an important share of the capital at risk. In perspective, it is important to know how important that number is. Consider the following for some Mexican banks: In Table 12 we can see that the amount of required capital due to concentration in LEX regulation is especially important, accounting for more than 50% of the total TLAC supplement, and on average, it represents more than 20% of the regulatory capital (see Table 12).

Table 12. Capital Share Due to Concentration According to the LEX

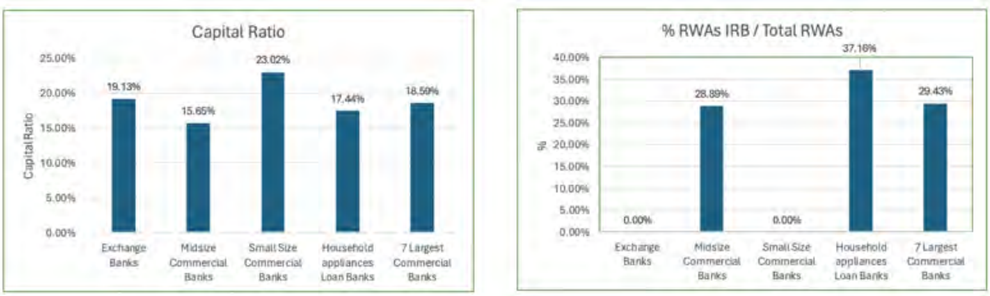
	Banamex	Banorte	BBVA México	HSBC	Santander	Scotiabank
Capital due to concentration (LEX)	28,298	33,905	70,331	20,050	32,329	17,084
% of Regulatory Capital	19%	17%	20%	23%	20%	23%
IRB Bank		x	x	x	x	

Source: Prepared by the author.

We explained before that there is not a specific capital requirement for concentration, but the pieces of the whole regulatory framework work together to handle the concentration risk. We also explained that systemic banks have two additional capital buffers apart from the capital conservation supplement and that the last one is a supplement for all banks.

It is true that capitalization ratios are, in general, higher than minimum regulatory ones, but it is important to know the marginal contribution of risk components to total risk and to capital at risk. In Graph 6 we can find the total capital ratio and by type of bank, compare such data with the one in Table 13, which shows the capital requirement for Mexican systemic banks, meaning that those have higher requirements. Recall that for the rest of the banks, the minimum ratio is 10.5% (see Graph 6 and Table 13).

Graph 6. Mexico, Total Capital Ratio by Type of Bank as of December 2023



Source: Prepared by the author with data from CNBV (2024b).

Table 13. Mexico Systemic Banks: Minimum Total Capital Ratio and Supplements as of December 2023

Capital Ratio & Suplements / Bank	Banamex	Banorte	BBVA México	HSBC	Santander	Scotiabank
Minimum Total Capital Ratio	8.00%	8.00%	8.00%	8.00%	8.00%	8.00%
Capital Conservation Supplement	2.50%	2.50%	2.50%	2.50%	2.50%	2.50%
D-SIB Supplement	1.20%	0.90%	1.50%	0.60%	1.20%	0.60%
Capital requirement without TLAC supplement	11.70%	11.40%	12.00%	11.10%	11.70%	11.10%
Capital requirement with 50% of TLAC supplement*	**14.95%**	**14.65%**	**15.25%**	**14.35%**	**14.95%**	**14.35%**
Total Capital Ratio as of December 2023	**18.82%**	**20.72%**	**18.27%**	**15.78%**	**17.54%**	**15.82%**
Ratio excess as od December 2023	**3.87%**	**6.07%**	**3.02%**	**1.43%**	**2.59%**	**1.47%**
Capital requirement with TLAC supplement fully formed	22.07%	23.97%	21.52%	19.03%	20.79%	19.07%

*Starting in December 2022, the TLAC supplement is being formed by one-fourth of the total requirement every year. Banks must complete this supplement by December 2025.

The author uses the 6.5% of the RWA in this table as a reference since some banks are using the 3.75% of the adjusted assets for the Leverage Ratio (LR) computation.

Source: Prepared by the author with data from CNBV (2024c).

What is capitalization like in the world? In Graph 7 we can notice that world capitalization levels are much higher than 10.5%. Only three countries have ratios lower than 12% and only one with an average capital ratio of 6%. In red labels, we have the ratios of other selected countries: the United States 16.3%; Mexico 17.7%, and Switzerland 19.7% (see Graph 7, in next page). Consider the minimum concentration component that we have computed and keep in mind that those ratios must support several types of losses including concentration implicit in LEX.

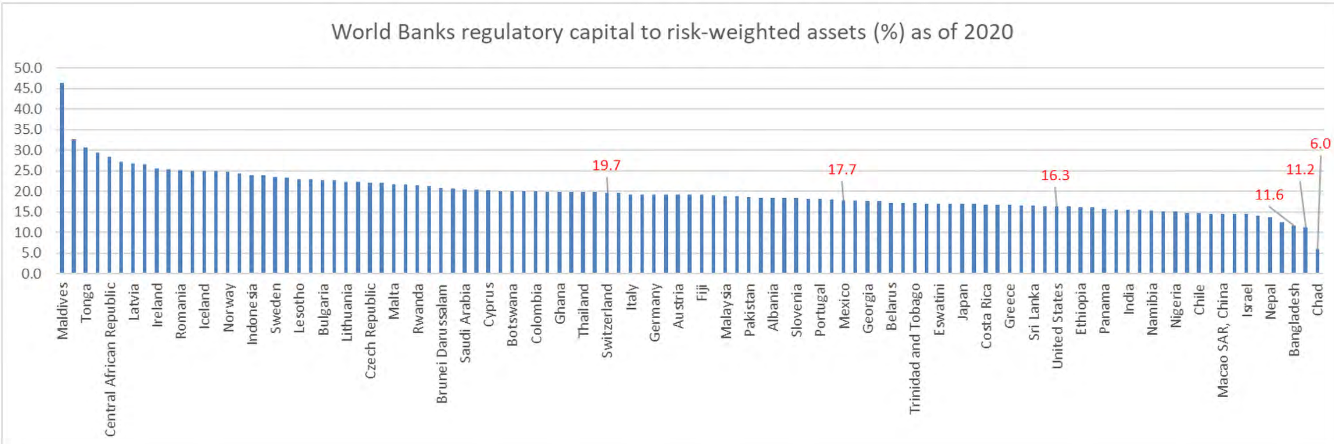
10. Conclusions

We have explained that the regulation did not follow the approach of computing a specific granularity adjustment for the bank's capital requirement due to the credit concentration risk. However, a holistic approach through the complete regulatory framework, including the LEX regulation, is a way to limit the concentration risk, and account for consolidated exposures, including not only all those with credit risk but also adding the exposures of all connected counterparties. At the same time, this approach addresses correlation and contagion and is supported by other pieces of the Basel regulation. Nonetheless, the LEX has an implicit minimum concentration and, therefore, a share of capital at risk. We have computed the add-ons and demonstrated that such a share is important. Since there is not a specific capital supplement for concentration, the existing supplements must be enough to absorb any loss, including the losses eventually coming from the concentration risk. We have also pointed out that the whole focus of the Basel regulation is systemic. Finally, we showed that although capital levels are aligned with the add-on, those levels must face any kind of loss. Losses implicit in the LEX due to concentration are particularly important and account for more than half of the TLAC supplement (6.5% of RWAs). This confirms the relevance of the size of the add-on.

The BIS implemented the LEX regulation to address the concentration risk, but once banks adopt the new regulation completely, the residual risk will remain relevant.

In Table 3, the author presents a summary of data from seven years of ICAAP in Mexico (2017-2023) showing that, on average, an important number of banks (15) present a vulnerability due to credit risk concentration. At least in one year, this involves an important share of the total system assets. LEX regulation started in Q4 2023 in Mexico.

Graph 7. World Banks' Regulatory Capital as of 2020



Disclaimer of data provider: Deposit takers' capital adequacy is a ratio of total regulatory capital to their assets held, weighted according to the risk of those assets.

Source: World Bank (n.d.).

If credit concentration would materialize, the results of this work show that for non-systemic banks, the capital conservation supplement would not be enough (3.6% vs 2.5%). For systemic banks this implies that the add-on represents 55% of the complete TLAC supplement (3.6% vs 6.5%) still in the formation process. A current offset for this risk is the levels of capital shown by Mexican banks. Nonetheless, the stress test scenarios in the ICAAP show vulnerabilities due to credit concentration, meaning that under stress conditions, situations may change considerably.

References

Artzner, P., Delbaen, F., Eber, J. & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *(9)*3, 203–228. https://doi.org/10.1111/1467-9965.00068

Banco de México. (2023). *Reporte de estabilidad financiera: segundo semestre 2023*. https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-el-sistema-financiero/%7B6B881BAE-59D7-05D2-0B38-1E72528030BA%7D.pdf

Basel Committee on Banking Supervision (BCBS). (2004). *International convergence of capital measurement and capital standards: A revised framework*. Bank for International Settlements. https://www.bis.org/publ/bcbs107.htm

Basel Committee on Banking Supervision (BCBS). (2005). *Selected literature on concentration risk in credit portfolios*. Bank for International Settlements. https://www.bis.org/bcbs/events/rtf05biblio.htm

Basel Committee on Banking Supervision (BCBS). (2006). *Studies on credit risk concentration: An overview of the issues and a synopsis of the results from the Research Task Force project*. (Working Paper No. 15). Bank for International Settlements. https://www.bis.org/publ/bcbs_wp15.htm

Basel Committee on Banking Supervision (BCBS). (2009). *Strengthening the resilience of the banking sector*. Bank for International Settlements. https://www.bis.org/publ/bcbs164.htm

Basel Committee on Banking (BCBS). (2023). Regulatory Consistency Assessment Programme (RCAP). Assessment of Basel large exposures regulations – Mexico. https://www.bis.org/bcbs/publ/d564.htm

Comisión Nacional Bancaria y de Valores. (CNBV). (2023). *Manual para la elaboración de la evaluación de suficiencia de capital bajo escenarios supervisores e internos 2023-2025*. https://portafolioinfo.cnbv.gob.mx/PortafolioInformacion/Manual_BM.pdf

Comisión Nacional Bancaria y de Valores. (CNBV). (2024a). *Disposiciones de carácter general aplicables a las instituciones de crédito*. https://shorturl.at/aIMS1

Comisión Nacional Bancaria y de Valores. (CNBV). (2024b). Bancos de Importancia Sistémica Local. https://www.gob.mx/cnbv/acciones-y-programas/bancos-de-importancia-sistemica

Comisión Nacional Bancaria y de Valores. (CNBV). (2024c). *Portafolio de Información*. https://portafolioinfo.cnbv.gob.mx/Paginas/Inicio.aspx

Consejo de Estabilidad del Sistema Financiero. (2018–2024). *Informe anual sobre el estado que guarda la estabilidad del sistema financiero en México y sobre las activi-*

dades realizadas por el Consejo de Estabilidad del Sistema Financiero. https://www.cesf.gob.mx/en/CESF/Publicaciones_e_informes

Financial Stability Institute. (FSI). (2019). *The capital buffers in Basel III: Executive summary*. Bank for International Settlements. https://www.bis.org/fsi/fsisummaries/b3_capital.htm

Financial Stability Institute (FSI). (2022). *Large exposures standard: Executive summary*. Bank for International Settlements. https://www.bis.org/fsi/fsisummaries/lex.pdf

Gordy, M. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, *12*(3), 199-232. https://doi.org/10.1016/S1042-9573(03)00040-8

Gordy, B. & Lutkebohmert, E. (2013). Granularity Adjustment for Regulatory Capital Assessment. *International Journal of Central Banking*, *9*(3), 33-71. https://www.ijcb.org/journal/ijcb13q3a2.pdf

Hibbeln, M. (2010). *Risk Management in Credit Portfolios: Concentrations Risk and Basel II*. (Contributions to Economics). Physica Heidelberg. https://doi.org/10.1007/978-3-7908-2607-4

JP Morgan & Reuters. (eds.) (1996). *RiskMetrics-Technical Document*. Morgan Guaranty Trust Company of New York. https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a

Márquez Diaz-Canedo, J. (2003). Suficiencia de capital y riesgo de crédito en portafolios de préstamos bancarios. In Elizondo, A. (Ed.). *Medición integral del riesgo de crédito*. Limusa.

Nokkala, Jan. (2022). Are large credit exposures a source of concentration risk? *Bank i Kredyt*, *53* (4), 375-398. https://bankikredyt.nbp.pl/content/2022/04/BIK_04_2022_02.pdf

Relim, T., Mariano, A., Monteiro, S., Grubisic, V. & Oliveira, E. (2019). Credit Risk Concentration in the Banking Industry: A Systematic Literature Review. 34th IBIMA Conference, November 13–14, Madrid. https://www.researchgate.net/publication/339314572_Credit_Risk_Concentration_in_the_Banking_Industry_A_Systematic_Literature_Review

Venegas Martínez, F. (2008). *Riesgos financieros y económicos*. Cengage Learning.

Wilde, T. (2001a). The IRB approach explained. *Risk Magazine*, *14*(5), 87-90. https://www.risk.net/infrastructure/1530304/irbapproach-explained

Wilde, T. (2001b). Probing granularity. *Risk Magazine*, *14*(8), 103-106. https://www.risk.net/infrastructure/1530360/probing-granularity

World Bank. (n.d.). *Global Financial Development Database. (DataBank)* https://shorturl.at/aflPT

Yi Xiao, J. & Finger, C. (2002). *Treatment of Retail and Small Business Exposures*. (RiskMetrics Group Product Technical Note). RiskMetrics Group.

About the author

José Juan Chávez obtained the following degrees: a Bachelor's in Economics from the School of Economics at the Universidad Nacional Autónoma de México (UNAM), a Master's in Finance, and a PhD in Financial Sciences at the Instituto Tecnológico de Monterrey (ITESM; EGADE Business School). He has been a professor at the ITESM, where he has taught courses at the undergraduate and graduate levels—both master's and PhD—for over 15 years, especially in risk management, corporate finance, and credit financing. Chávez has wide experience in the financial sector in Mexico. He worked for 34 years in Grupo Financiero Scotiabank Inverlat (GFSI) in corporate credit financing, credit policy, and credit analysis, credit risk modelling and management, and liquidity risk management. He successfully led the Basel Accord implementation in GFSI for ten years, and for the last eight years has worked in optimizing regulatory and economic capital with excellent results. He was head of the Risk Committee of the AMIB (Mexican Brokerage Houses and Investment Funds Association) from late 2020 to early 2024, where he led the group that determined the collateral haircuts for repo and securities lending markets, applying state-of-the-art market risk and credit risk standards. Currently, he is devoted to research and education activities.

jjchavezgud@gmail.com
https://orcid.org/0009-0002-9593-7789

Comparing the Performance of Long Short-Term Memory Architectures (LSTM) in Equity Price Forecasting: A Research on the Mexican Stock Market

Comparación del desempeño de arquitecturas de memoria a corto y largo plazo (LSTM) en el pronóstico de precios de acciones: una investigación sobre el mercado bursátil mexicano

Samuel García

EGADE Business School, Tecnológico de Monterrey, Mexico

Received: March 1, 2024.
Approved: May 17, 2024.

Abstract

This study compares the performance of univariate and multivariate Long Short-Term Memory (LSTM) to predict next-day closing prices on four stocks in the consumer retail sector of the Mexican Stock Exchange. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), and Root Mean Squared Error (RMSE) are used to test the networks' performance. Results show a better performance on multivariate price forecasts when using 20-day and 15-day length sequences, generating consistent results for the sample, including illiquid and liquid stocks. On the other hand, univariate LSTM discloses lower forecast performance when predicting the price of illiquid stocks.

Keywords: *forecast, stocks, univariate, multivariate, LSTM.*
JEL Classification: *G1, G15, G20, C6.*

Resumen

Este trabajo compara el desempeño de la memoria de corto y largo plazo (LSTM, por sus siglas en inglés) univariada y multivariada en la predicción de los precios de cierre del día siguiente de cuatro acciones del sector de consumo minorista en la Bolsa Mexicana de Valores. El error absoluto medio (MAE, por sus siglas en inglés), el error porcentual absoluto medio (MAPE, por sus siglas en inglés), la mediana del error porcentual absoluto (MdAPE, por sus siglas en inglés) y la raíz del error cuadrático medio (RMSE, por sus siglas en inglés) se utilizan para probar el desempeño de las redes. Por un lado, los resultados muestran un mejor desempeño en el pronóstico multivariado de precios cuando se utilizan secuencias de 20 y 15 días de duración, generando resultados coherentes para la muestra, incluidas las acciones líquidas e ilíquidas. Por otro lado, la LSTM univariada revela un desempeño de pronóstico menor para la predicción del precio de acciones ilíquidas.

Palabras clave: *predicción, acciones, univariada, multivariada, LSTM.*
Clasificación JEL: *G1, G15, G20, C6.*

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.06

1. Introduction

Given the growing complexity of the global financial industry and the unstable nature of the financial markets, pricing analysis of financial assets—like stocks—and predicting future prices and returns in the financial market is a complex and challenging activity, highly valued in the financial sector. Since noise and non-parametric and non-linear dynamics are characteristic of the stock market, its traditional statistical tools to analyze historical data—where past events have great importance in predicting future states (e.g., price and returns) and trends—may struggle to model those dynamics on stock prices over time (Pramod & Mallikarjuna, 2020; Bhandari et al., 2022).

In recent years, developments in ML, AI, and Deep Learning (DL) have played a central role in enhancing stock price prediction. A case in point is that academics have noticed the advantages of DL models when capturing non-linear features of data sequences through Recurrent Neural Networks (RNN) and LSTM networks (Tianxiang & Zihan, 2020). ML is a sub-field of AI, which tries to emulate some human cognitive features like the learning process to identify patterns and/or classify specific sets of objects and is currently used in the financial sector because of its analytical capabilities to analyze and manage big data (Lu, 2017; Liebergen, 2017).

Artificial Neural Networks (ANN) are part of deep learning, which attempt to recreate the logic of the human brain to perform cognitive tasks. These models are mainly based on the interconnection of individual neurons, which creates a network (Nielsen, 2015; Krenker et al., 2011; Tirozzi et al, 2007); RNNs and LSTMs are a subset of Neural Networks, mainly designed to capture information on historical data.

According to the literature reviewed, there is no extensive research on the price forecasting capabilities of LSTM in Latin American markets. Given the importance of AI and DL techniques in the analysis and prediction of financial assets' prices, as well as the growing presence of Latin American financial markets on the global financial landscape, this research is focused on comparing the performance of univariate and multivariate LSTM when predicting next day closing price of four Mexican stocks from the consumer retail sector in the Mexican Stock Exchange (BMV, in Spanish), as well as analyzing the impact of the size of the sequence length used for prediction accuracy. As mentioned, the sample used for this work includes four stocks, two of them liquid when comparing the 3-month and 10-day average traded volume with the other two stocks in the sample. The contribution of this research is to test the performance of LSTM when predicting stock prices, using different historical timeframes, to predict the prices of four Mexican stocks from the consumer retail sector.

As part of the results, it can be observed that the size of the rolling window impacts the performance of the univariate model when predicting the price of illiquid stocks, assessed through four performance metrics to measure the magnitude of errors: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MDAPE), and Root Mean Squared Error (RMSE), whereas the performance of the multivariate output shows consistency for both illiquid and liquid stocks.

This paper is ordered as follows: Section 2 discloses works related to LSTMs prediction capabilities for a stock process; Section 3 provides an overview of the tested architectures and their features; Section 4 contains methodology and model implementation; Section 5 includes performance metrics; Section 6 details preliminary results; and Section 7 contains conclusions.

2. Related Works

Pricing prediction of financial products is a significant issue in the financial sector and the academy. Currently, several ML and AI models are used to enhance price prediction accuracy; most are based on RNNs, LSTM, and other DL. As mentioned in the previous section, RNNs were mainly designed to capture information on historical data. Focusing on RNNs and LSTM DL models, academic research is centered on analyzing the prediction capabilities of the DL through plain and mixed deep learning models, testing different variables, architectures, and levels of model parameters to obtain a better analysis and prediction accuracy. For instance, Nourbakhsh and Habibi (2023) combined Convolutional Neural Network and LSTM as well as specific variables used in fundamental analysis, to enhance the model's accuracy measured through MAE and MAPE. Also, Zaheer et al. (2023) explored the capabilities of a hybrid deep-learning model based on single and mixed RNN, LSTM, and CNN architectures to predict closing and high prices on the next trading day of the Shanghai Composite Index, where they found that a single layer RNN outperforms the other tested models showing the lowest MAE and RMSE metrics.

Another interesting option to enhance prediction capabilities is found in Tianxiang and Zihan (2020), who proposed a method to predict the West Texas Intermediate oil price, from January 1986 to January 2020, with the LSTM and GM (1,1) model, based on a multi-step prediction method. The model showed significant prediction accuracy measured through MAPE and RMSE. The model effectively captured long-term effects with lower frequency and also price trends. The work performed by these authors can show how mixing different DL models enriches the existing literature on price prediction of financial assets.

The model proposed in this paper analyses how the length of historical data would contribute to better price forecasting, as mentioned by Bhandari et al (2022) who used 15 years of market and macroeconomic data, as well as technical indicators to predict the closing price for the S&P 500 index through a multivariable LSTM. They found the best performance results were based on RSME, MAPE, and a correlation coefficient obtained through a single-layer model.

Although this research evaluates price prediction for the next trading day, other studies examined the prediction accuracy for longer periods. For example, Ghosh et al. (2019) employed LSTM techniques on historical stock price data of five companies from some pre-decided sectors in the Indian market, to infer future trends. The authors proposed a framework based on LSTM models to calculate the best time length to forecast the future share price of a company from a particular sector, as well as predict the future growth of a company for periods of 3 and 6 months, and 1 or 3 years. They found a decrease in the error level when using test data for longer periods, dependencies, and the same growth rate in companies from a certain sector.

Published studies on price prediction have also explored the impact of transformed variables, the number of layers in models, parameter levels, and the length of historical data used, for better model learning, and enhancing prediction capabilities on DL models when compared to other statistical tools. For instance, Andi (2021) normalized variables on a data set to compare the performance of LSTM with other prediction models, like linear regression and the Lasso algorithm, concluding that the first model obtained the most accurate forecast on the bitcoin price based on accuracy, precision, recall, and sensitivity because using common variation ranges on the variables allow capturing trends.

Finally, Pramod and Mallikarjuna (2020) explored predicting Tata Motors Limited’s stock price using LSTM. The output produced a low loss and low error rate. They also found that increases in layers and epoch batch rates had a positive impact on the performance.

Overall, there is an extensive body of research focused on measuring the accuracy of DL models to enhance forecasting capabilities. For example, LSTM architectures combined with other DL models have been used, as well as other techniques like transformed variables to improve model performance, all these evaluated under different financial markets in Asia, Europe, and North America. However, there is no extensive body of research assessing the performance of LSTM models in predicting stock prices in the Latin American financial markets. The contribution of this research is to test the performance of LSTM, using different timeframes to predict the prices of four stocks issued by Mexican firms, with different liquidity attributes, in the Mexican Stock Exchange.

3. A Brief on LSTM

Recurring Neural Networks (RNNs) have loops to feedback other neurons in the architecture, hence the output of a neuron in the network impacts the input of another neuron, resulting in closed paths for the transmission of information in the network (Haykin, 2010). LSTMs are a type of RNN architecture, used to find patterns in data, where the occurrence of events of interest is uncommon in time and frequently mixed with other events (Bhandari et al., 2022; Pramod & Mallikarjuna, 2020).

LSTMs deal with the problem of “long-term dependencies”, present in RNNs, by retaining information from past inputs contained in a variable number of time steps, so they can manage to learn and allow facts of interest to persist over time while overcoming the vanishing and exploding gradient problem. As mentioned previously, this network can find relationships in historical data where the existence of the event of interest is scarce in the data set (Benchaji et al, 2021a; Yu et al, 2019; Benchaji et al., 2021b).

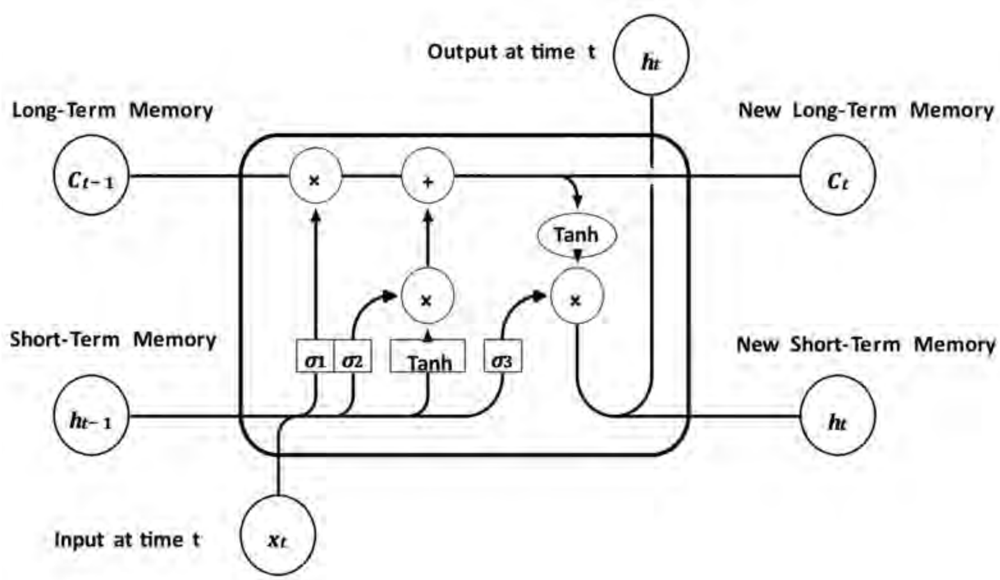
In general, a LSTM architecture (see Figure 1) has explicit memory blocks containing different states: a hidden state (h) and a cell state (C), which allow to store and manage both, short and long-term information through three gates (stages), each one performing an individual function:

1. Forget gate, which chooses, through a sigmoid function (σ_1), whether information coming from h_{t-1} and current input (x_t) needs to be remembered (values near to 1) or is irrelevant and can be forgotten (values near to 0).

2. Input (update) gate, this allows learning from the input x and h_{t-1} to update C, which contains the long-term information; the layer includes two parts: first, a sigmoid layer it will decide which new values will be stored in the cell state and second, a tangent layer creates a vector of new candidate data with values between -1 and 1 to rate relevant data. Then, the output of the input gate is obtained through multiplying the values of sigmoid layer and the tangent layer.

3. Output Layer determines the new hidden state (h_t), based on h_{t-1}, x_t and the tanh of the current cell state (C_t).

Figure 1. General Representation of a LSTM Cell



Source: Prepared by the author.

Where:

σ = Sigmoid Function

Tanh = Hyperbolic Tangent Function

4. Methodology and Model Implementation

4.1 Coding & Data Overview

The analysis for the Research was done using Scikit-learn, a Python-related library to create and implement ML models and perform statistical analysis and modelling; TensorFlow, a high-level open-sourced end-to-end platform to create DL and AI models, and Keras is a high-level open-sourced library which takes the underlying operations provided by other platforms like TensorFlow.

For this research, I used daily market data obtained from Yahoo Finance. The dataset contains stock transactions executed in the Mexican stock market, in the consumer retail sector from January 1, 2020, to February 9, 2024 (1036 workdays). This sector is important not only because it includes companies selling several retail products related to the basic needs of the Mexican population, distributed across Mexico, but the sector was resilient during the pandemic, presenting the smallest drop in value in the Mexican financial market. This sector had the speediest recovery in comparison to other sectors (Landazuri Aguilera & Ruíz Pérez, 2021).

The analyzed stocks were the following:

Grupo Comercial Chedraui, S.A.B. de C.V. (ticker: CHDRAUIB.MX)

- La Comer, S.A.B. de C.V. (ticker: LACOMERUBC.MX)
- Organización Soriana, S. A. B. de C. V. (ticker: SORIANAB.MX)
- Wal-Mart de México, S.A.B. de C.V. (ticker: WALMEX.MX)

The four stocks were selected, since all of them are nationwide supermarkets, selling comparable retail products with similar target markets, making them comparable in terms of business models.

Table 1 shows some market data for the four stocks used for the research (see Table 1). WALMEX and LACOMERUBC would be considered the most liquid stocks in the sample because both have the greater number of shares outstanding and average traded volume on a three-month and ten-day timeframes, which allow the stocks to be easily traded in the stock market at the current fair market price (Armitage et al., 2014).

Table 1. Market Statistics for the Analyzed Stocks

Statistics	Walmex	Soriana	Chedraui	La Comer
Average Volume on a 3-Month timeframe	15.08 M	65.88 k	311.35 k	645.09 k
Average Volume on a 10-Day timeframe	15.27 M	2.56 k	295.92 k	224.8 k
Shares Outstanding	17.46 B	1.8 B	959.82 M	1.09 B
Implied Shares Outstanding	17.46 B	1.85 B	959.82 M	N/A
Intraday Market Cap	1.19 T	62.76 B	120.30 B	N/A
Enterprise Value	1.21 T	81.19 B	163.16 B	N/A

Source: Prepared by the author with data from Yahoo Finance as of February 20, 2024.

The research was based on six variables extracted from the data set, including open, high, low, closing, and adjusted closing prices, as well as volume. During a trading day, open and close are prices at which the stock began and ended trading in the stock market, high and low prices are the highest and lowest traded prices for that stock, during a trading day. Adjusted (Adj) close price is the closing price after considering any splits and dividend distributions. Finally, volume indicates the total quantity of stocks traded during a day.

All variables in the dataset were normalized considering a 0 to 1 range to maintain a common scale and to contribute to the model's accuracy.

4.2 Model Implementation and Training

As mentioned in Section 1, this research aims to compare the prediction accuracy of univariate *vs*. multivariate LSTMs on stocks related to the consumer retail sector in Mexico. The research was performed through the following LSTM core architectures:

- Two hidden layers with 50 units each. For every network, the output h_t (see Figure 1), is an input for $X_{t\,,}$ at time *t*, as shown before in Figure 1.
- A Dense layer with five neurons, to convert the output of the final layer into a vector.
- Finally, the vector flows to a linear activation, used to predict the next day's closing price.

Comparison between univariate and multivariate LSTM networks is performed using sequence lengths of 20, 15, and 10 historical stock prices and volume (described in Section 4.1). For example, Figure 2 discloses an architecture with a sequence length of 20 daily data (see Figure 2).

Figure 2. Architecture of a LSTM with a Sequence Length of 20 and 50 units

```
Layer (type)                 Output Shape
=================================================
lstm_2 (LSTM)                (None, 20, 50)

dropout_2 (Dropout)          (None, 20, 50)

lstm_3 (LSTM)                (None, 50)

dropout_3 (Dropout)          (None, 50)

dense_2 (Dense)              (None, 5)

dense_3 (Dense)              (None, 1)

=================================================
```

Source: Prepared by the author.

Both LSTMs predicted the close price for the next trading day. Multivariate, forecasting was based on the six features mentioned in section 4.1. Univariate forecasting was run using the close price.

4.3 Hyperparameters

Following Wiese and Omlin (2009), several test runs were executed to find the best combination of hyperparameters before executing the runs of the research. The model was compiled using the following hyperparameters:

- Dropout technique of 30% to avoid overfitting and to allow the network to get a better generalization.
- Learning rate to adjust the weights in response to changes in the gradient. For this research, the learning rate is 0.0018.
- The m square error (MSE) Loss Function is commonly used on regression tasks. It calculates the magnitude of the average error between the model's prediction ($\hat{y}_i$) and the target value (y_i) by taking the average of the squared

https://doi.org/10.36105/theanahuacjour.2024v24n1.06

difference between these two values. Squaring differences results in a higher penalty for material deviations from the target value.

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{1}$$

Where:

- n is the total sample size.
- $\hat{y}_i$ is model's prediction.
- y_i is the target value.

5. Performance Metrics

To compare both architectures, the prediction accuracy was evaluated through four different metrics:

- MAE shows the arithmetic mean over the absolute difference between $\hat{y}_t$ and y_t (residuals) at time t in the analyzed timeframe.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{2}$$

  Where "n" is the total sample size.

- MAPE. This indicator measures prediction accuracy as a percentage based on the average of the ratios of individual absolute errors, at each point in time. Defining the error between $\hat{y}_t$ and y_t at time t as a ratio, as follows:

$$e_t = \frac{y_t - \hat{y}_t}{y_t} \tag{3}$$

MAPE is represented as:

$$MAPE = \frac{\sum_{t=1}^{n}|e_t|}{n} * 100 \tag{4}$$

- MDAPE. It is a performance metric used to evaluate the accuracy of forecasts in time series analysis. Unlike MAPE, MDAPE uses the median of the absolute percentage errors. This property enables MDAPE to be less sensitive to outliers than MAPE. Mathematically, MDAPE is represented as follows:

$$\text{MDAPE} = median\left(\frac{|e_t|}{y_t}\right) * 100 \tag{5}$$

- RMSE measures the difference between $\hat{y}_t$ and targets y_t at time t, through squaring the errors, taking the mean, and finally calculating the square root. RMSE is used to quantify the error on $\hat{y}_t$, when y_t is a continuous number and gives a friendly view of the model's performance, since it shows data on the same scale/units as the Target variable. RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{6}$$

6. Preliminary Results

Both architectures described in Section 4.2 were tested using the hyperparameters described in Section 4.3; early stopping was used to prevent overfitting. All tests were run considering a dataset from January 1st, 2020, to February 9th, 2024, encompassing 1036 trading days.

6.1 First Test

The test was run using a historical timeframe (sequence length) of 20 days. Table 2 shows the results for both architectures, after replicating 10 times the test over the same stock to provide model reliability (see Table 2).

Results suggest that the multivariate architecture has more consistent performance results (i.e., MAE, MAPE, MDAPE, and RMSE) on the four stocks than the univariate, where results for WALMEX and LA COMER differ significantly from those for SORIANA and CHEDRAHUI.

When comparing performance results between univariate and multivariate models, it can be observed that the multivariate model outperforms when forecasting prices of less liquid stocks: the univariate MAPE for SORIANA and CHEDRAUI is 357% and 1021% higher than the same multivariate metric (see Table 2). Additionally, the univariate MDAPE for the two cases is 457% and 1319% higher than the same metric obtained through the multivariate model. Similar differences are observed for RMSE where the univariate results are 92% and 1252% higher.

https://doi.org/10.36105/theanahuacjour.2024v24n1.06

When comparing MAPE and MDAPE metrics obtained from the two models for WALMEX and LA COMER, the univariate results outperform the multivariate in almost all indicators (see Table 2). By way of illustration, the univariate MAE for WALMEX is 0.8500 and the multivariate is 1.1400. The results may imply that stock liquidity impacts the forecast capability of the univariate LSTM. Finally, the multivariable average MAE (1.0325) is -80.77% compared to the univariate (5.3700), MAPE -77.42%, MDAPE -81.14 %, and RSME -76.67% respectively.

6.2 Second Test

The sequence length was changed from 20 to 15 days. Table 3 shows the performance results (see Table 3).

Observing performance metrics, a shortened sequence length shows a positive impact on the univariate architecture when compared with the first test, lowering differences in performance metrics: as Table 3 shows, the average MAE is 1.7650, MAPE 3.0450%, MDAPE 2.5375% and RMSE 2.5305 among the four stocks, however the results are higher than those obtained with the multivariate.

Multivariate LSTM discloses more consistent and accurate results, showing small differences in the four indicators, on average MAE is 1.64, MAPE 2.42%, MDAPE 2.24%, and RMSE 1.93. Additionally, when comparing performance metrics between both architectures for the analyzed stocks, the multivariable average results are more accurate than the univariate: MAE is -6.94%, MAPE -20.36 %, MDAPE -11.53 %, and RSME -23.61% respectively.

Table 2. Long Short-Term Memory (LSTM), 20 days

Timesteps	20	**Architecture**	2 HL / 50 HU
Batch size	16	**Data**	From Jan 1, 2020 to Feb 9, 2024
Early stopping	yes	**# Days**	1036
Learning rate	0.0018	**Frequency**	Daily

		Univariate				Multivariate			
Issuer	**Ticker**	**Median Absolute Error (MAE)**	**Mean Absolute Percentage Error (MAPE) %**	**Median Absolute Percentage Error (MDAPE) %**	**Difference between actual and predicted values (RMSE)**	**MAE**	**MAPE (%)**	**MDAPE (%)**	**RMSE**
WALMEX	WALMEX	0.8500	1.2600	0.8900	1.1796	1.1400	1.6800	1.4000	1.4259
SORIANA	SORIANA	2.1700	6.8600	6.3500	2.5707	1.0100	1.5000	1.1400	1.3365
CHEDRAUI	CHDRAUI	17.8800	17.4900	17.3100	18.2680	1.0500	1.5600	1.2200	1.3514
LA COMER	LACOMER	0.5800	1.4900	1.0600	0.7953	0.9300	1.3800	1.0700	1.2085
Average		**5.3700**	**6.7750**	**6.4025**	**5.7034**	**1.0325**	**1.5300**	**1.2075**	**1.3306**
Δ Univariate vs multivariate						-80.77%	-77.42%	-81.14%	-76.67%

Source: Prepared by the author.

Table 3. Long Short-Term Memory (LSTM), 15 days

Timesteps	15	**Architecture**	2 HL / 50 HU
Batch size	16	**Data**	From Jan 1, 2010 to Feb 9, 2024
Early stopping	yes	**# Days**	1036
Learning rate	0.0018	**Frequency**	Daily

		Univariate				Multivariate			
Issuer	**Ticker**	**Median Absolute Error (MAE)**	**Mean Absolute Percentage Error (MAPE) %**	**Median Absolute Percentage Error (MDAPE) %**	**Difference between actual and predicted values (RMSE)**	**MAE**	**MAPE (%)**	**MDAPE (%)**	**RMSE**
WALMEX	WALMEX	2.6000	4.0500	3.1900	3.4742	1.6000	2.3600	2.1400	1.9060
SORIANA	SORIANA	1.4600	4.5500	4.8600	1.7166	1.5400	2.2800	2.0900	1.8417
CHEDRAUI	CHDRAUI	2.4700	2.2400	1.1200	4.1867	1.7600	2.6000	2.5300	2.0231
LA COMER	LACOMER	0.5300	1.3400	0.9800	0.7444	1.6700	2.4600	2.2200	1.9611
Average		**1.7650**	**3.0450**	**2.5375**	**2.5305**	**1.6425**	**2.4250**	**2.2450**	**1.9330**
Δ Univariate vs multivariate						-6.94%	-20.36%	-11.53%	-23.61%

Source: Prepared by the author.

Table 4. Long Short-Term Memory (LSTM), 10 days

Timesteps	10	**Architecture**	2 HL / 50 HU
Batch size	16	**Data**	From Jan 1, 2010 to Feb 9, 2024
Early stopping	yes	**# Days**	1036
Learning rate	0.0018	**Frequency**	Daily

		Univariate				**Multivariate**			
Issuer	**Ticker**	**Median Absolute Error (MAE)**	**Mean Absolute Percentage Error (MAPE) %**	**Median Absolute Percentage Error (MDAPE) %**	**Difference between actual and predicted values (RMSE)**	**MAE**	**MAPE (%)**	**MDAPE (%)**	**RMSE**
WALMEX	WALMEX	0.8700	1.2800	0.9400	1.1899	1.2900	1.8800	1.5500	1.6678
SORIANA	SORIANA	2.5600	8.1100	7.6200	2.9482	1.5700	5.0200	4.4100	1.8782
CHEDRAUI	CHDRAUI	27.2200	26.6300	26.8500	27.7901	13.0100	12.8700	12.5600	13.6385
LA COMER	LACOMER	0.5500	1.4000	1.0200	0.7522	0.7100	1.7900	1.3700	0.9799
Average		**7.8000**	**9.3550**	**9.1075**	**8.1701**	**4.1450**	**5.3900**	**4.9725**	**4.5411**
Δ Univariate vs multivariate						-46.86%	-42.38%	-45.40%	-44.42%

Source: Prepared by the author.

6.3 Third Test

The third test was performed using the same database used for the first and second tests, the sequence length was changed to 10 trading days.

Table 4 shows the performance test results for both architectures (see Table 4). The average performance metrics for the univariate deteriorated when compared with the results in the second test (see Table 3), mainly for less liquid stocks (SORIANA and CHEDRAUI, In particular, SORIANA MAE varied from 1.45 in the second test to 7.8; MAPE from 4.55% to 8.11%, RMSE from 1.7166 to 2.9482, and MDAPE from 4.86% to 7.62%. Additionally, the performance metrics for the multivariate show the worst results when compared with tests one and two.

Although the multivariable performance results deteriorate when compared to those obtained for the same model in the first and second tests, these numbers are better than those in the univariate model. On average, multivariate (4.1450) MAE is -46.86% than univariate (7.8000), MAPE -42.38 %, MDAPE -45.40 % and RSME -44.42% respectively.

7. Conclusions

Stock price prediction is a very researched and complex area because all variables involved in trading activities have a nonlinear behavior. Thus, there is an interest in developing models that will allow more accurate and consistent forecasts. This study focuses on comparing the performance of univariate and multivariate LSTM in predicting prices for stocks in the consumer retail sector in Mexico, as well as the impact of the size of the sequence length on the models. The performance results under different sequence lengths were analyzed in Section 6.

In general, results show that the univariate LSTM works better when predicting prices over liquid stocks, although the performance in this model was less consistent among the four stocks in the sample than the multivariate. Multivariate LSTM shows accurate and consistent performance metrics when predicting prices for liquid and illiquid stocks, producing minor errors, measured through the performance metrics.

Sequence length impacts the accuracy of price prediction on both tested models. For instance, the univariate model disclosed a better performance with a sequence

https://doi.org/10.36105/theanahuacjour.2024v24n1.06

length of 15 trading days, whereas the multivariate shows a better performance with a sequence length of 20 days and 15 days.

Finally, it is worthwhile to continue exploring in future works the impact of price volatility and trends on predicting prices of illiquid stocks traded in developing economies.

References

Andi, H. (2021). An accurate bitcoin price prediction using logistic regression with LSTM machine learning model. *Journal of Soft Computing Paradigm*, *3*(3), 205–217. https://doi.org/10.36548/jscp.2021.3.006

Armitage, S., Brzeszczyński, J. & Serdyuk, A. (2014). Liquidity measures and cost of trading in an illiquid market. *Journal of Emerging Market Finance, 13*(2), 15–5196. https://doi.org/10.1177/0972652714541340

Benchaji, I., Douzi, S. & Ouahidi, B. E. (2021a). Credit card fraud detection model based on lstm recurrent neural networks. *Journal of Advances in Information Technology*, *12*(2), 113–118. https://doi.org/10.12720/jait.12.2.113-118

Benchaji, I., Douzi, S., Ouahidi, B. E. & Jaafari, J. (2021b). Enhanced credit card fraud detection based on attention mechanism and lstm deep model. *Journal of Big Data*, 8(1), 1–21. https://doi.org/10.1186/s40537-021-00541-8

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R. & Khatri, R. K. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*, article 100320. https://doi.org/10.1016/j.mlwa.2022.100320

Ghosh, A., Bose, S., Maji, G., Debnath, N. & Sen, S. (2019). Stock Price Prediction Using LSTM on Indian Market Share. *International Conference on Computer Applications in Industry and Engineering, 63*, 101–110. https://doi.org/10.29007/qgcz

Haykin, S. (2010). *Neural Networks and Learning Machines.* Pearson Education India.

Krenker, A., Bešter, J. & Kos, A. (2011). Introduction to the Artificial Neural Networks. In K. Suzuki (Ed.), *Artificial Neural Networks: Methodological Advances and Biomedical Applications.* InTech. https://doi.org/10.5772/15751

Landazuri Aguilera, Y. & Ruíz Pérez, R. (2021). El desempeño del sector productos de consumo frecuente en época de pandemia. *Memorias del Coloquio Nacional de Investigación en las Ciencias Económicas y Administrativas*, 114–132. https://shorturl.at/eEX78

Liebergen, B. V. (2017). Machine learning: a revolution in risk management and compliance? *Journal of Financial Transformation*, *45,* 60–67. https://www.capco.com/Capco-Institute/Journal-45-Transformation

Lu, Y. (2017). *Deep neural networks and fraud detection* [Master's thesis], Uppsala University. https://shorturl.at/ozEGI

Nielsen, M. (2015). Neural networks and deep learning. *Determination Press*, 25, 15–24. http://neuralnetworksanddeeplearning.com/

Nourbakhsh, Z., & Habibi, N. (2023). Combining LSTM and CNN methods and fundamental analysis for stock price trend prediction. *Multimedia Tools and Applications*, *82*,17769–17799. https://doi.org/10.1007/s11042-022-13963-0

Pramod, B. & Mallikarjuna, S. P. (2020). Stock Price Prediction Using LSTM. *Test Engineering and Management*, *83*, 5246-5251. https://shorturl.at/bjtA2

Tianxiang, Y. & Zihan, W. (2020). Crude oil price prediction based on LSTM network and GM (1,1) model. *Grey Systems: Theory and Application*, *11*(1), 80–94. https://doi.org/10.1108/GS-03-2020-0031

Tirozzi, B., Puca, S., Pittalis, S., Bruschi, A., Morucci, S., Ferraro, E. & Corsini, S. (2007). *Neural networks and sea time series: reconstruction and extreme-event analysis.* Springer Science & Business Media. https://doi.org/10.1007/0-8176-4459-8

Wiese, B. & Omlin, C. (2009). Credit card transactions, fraud detection, and machine learning: Modelling time with LSTM recurrent neural networks. In Bianchini, M., Maggini, M., Scarselli, F. & Jain, L.C. (Eds.), *Innovations in neural information paradigms and applications.* Springer, 231-238. https://doi.org/10.1007/978-3-642-04003-0_10

Yu, Y., X. S., Hu, C. & Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural Computation*, *31*(7),1235–1270. https://doi.org/10.1162/neco_a_01199

Zaheer, S., Anjum, N., Hussain, S., Algarni, A., Iqbal, J., Bourouis, S. & Ullah, S. (2023). Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. *Mathematics*, *11*(3), 590. https://doi.org/10.3390/math11030590

About the author

Samuel García is Mexican and has earned two master's degrees: the first in Economics and Technological Change and the second in Finance, and a Bachelor's in Business Administration. He has 25 years of experience in national and international financial institutions. He is a senior executive with knowledge of market surveillance, compliance, risk management, structured finance, financial products, and strategic management. Samuel worked at Citi for 16 years, where he was responsible for building and providing strategic direction for the ICRM Surveillance program, covering sales and trading as well as banking businesses across the Latin American Region—with more than 20 countries, including Mexico.

a1018966@tec.mx
https://orcid.org/0000-0001-7366-1406

Machine Learning Analysis of Consolidated Purchasing: A Case Study of Antiretroviral Medication 2019 Pricing Trends in Mexico

Análisis de aprendizaje automático de compras consolidadas: un estudio de caso sobre las tendencias de precios de medicamentos antirretrovirales en México en 2019

Blanca Iveth Mayorga Basurto

EGADE Business School, Tecnológico de Monterrey, Mexico

Galo Moncada Freire

Business School, Tecnológico de Monterrey, Mexico

Received: February 27, 2024.
Approved: May 15, 2024.

Abstract

This paper investigates trends in antiretroviral medication prices and their impact on public health in Mexico during 2019. Using three machine learning models developed in Python (logistic regression, random forest, and K-Nearest Neighbors or KNN), this study discerns increasing or decreasing patterns in antiretroviral (ARV) drug price fluctuations using a dataset comprising 15,220 observations of ARV drugs acquired between 2016 and 2019. Results indicate that random forests exhibited the highest precision in predicting price changes, followed by KNN and logistic regression. Significant factors affecting acquisition prices, such as drug type and duration of procurement strategy, were identified. In addition to analyzing price trends, the paper explores the budgetary considerations associated with these fluctuations, providing insights into the financial implications for healthcare systems and stakeholders. It is important to note that this paper focuses on a specific ARV pharmaceutical purchasing scheme. Moreover, the study emphasizes the creation of a unified and detailed medication price database, highlighting the significant effort invested in compiling complete and comprehensive information from various sources. This study's findings underscore the effectiveness of initiatives such as consolidated purchasing approaches and the integration of newer, cost-effective medications into treatment protocols. These initiatives have led to significant cost savings in antiretroviral medication procurement, contributing to improved access for individuals living with HIV/AIDS. Overall, the research highlights the importance of data-driven approaches and strategic planning in optimizing pharmaceutical purchasing processes and ensuring sustainable access to essential medications for public health interventions.

Keywords: *purchasing strategies, public health, Mexico, machine learning.*
JEL Classification: *I11, I18, H57.*

Resumen

Este trabajo examina las tendencias en los precios de medicamentos antirretrovirales y su impacto en la salud pública, centrándose en el contexto mexicano en 2019. Por medio de técnicas de aprendizaje automático, el estudio analiza las fluctuaciones en los precios de los medicamentos antirretrovirales, con base en un conjunto de datos que comprende 15,220 observaciones de medicamentos antirretrovirales (ARV) adquiridos entre 2016 y 2019, con énfasis particular en el año 2019. Los resultados indican que los "bosques aleatorios" demostraron la mayor precisión en la predicción de cambios de precios, seguidos por K-Nearest Neighbors *o KNN (algoritmo de k vecinos más próximos) y la regresión logística. Se identificaron factores significativos que afectan los precios de adquisición, como el tipo de medicamento y la duración de la estrategia de adquisición. Además, el estudio profundiza en las consideraciones presupuestarias, evaluando las implicaciones financieras de estas fluctuaciones de precios. Estos hallazgos destacan la efectividad de iniciativas como los enfoques de adquisición consolidada y la integración de medicamentos más nuevos y rentables en los protocolos de tratamiento, lo que conduce a ahorros significativos y mejor acceso para las personas que viven con VIH/sida. Es importante destacar que este artículo se centra en analizar un esquema específico de adquisición de medicamentos antirretrovirales. Los estudios futuros ampliarán este análisis para abarcar un espectro más amplio de esquemas de adquisición de medicamentos, proporcionando así una comprensión más completa de la dinámica de fijación de precios y sus implicaciones para la salud pública. Además, la investigación en curso perfeccionará la metodología e integrará fuentes de datos adicionales para mejorar la precisión y profundidad del análisis.*

Palabras clave: *ARV, tendencias de precios, compras consolidadas, México.*
Clasificación JEL: *I11, I18, H57.*

https://doi.org/10.36105/theanahuacjour.2024v24n1.07

1. Introduction[1]

In April 2019[2], a citizen petition was submitted to the Secretaría de Hacienda y Crédito Público (SHCP, Ministry of Finance and Public Credit) to address an extremely urgent case related to the shortage of medications intended for the treatment of over 100,000 patients infected with the Human Immunodeficiency Virus (HIV) in public hospitals. Dr. Carlos Urzúa, then Minister of SHCP, pointed out that pharmaceutical purchasing had became complicated due to many HIV specialists' insistence on acquiring over thirty different HIV inhibitors, in order to perform various therapeutic combinations. These factors, coupled with restrictions imposed by antiretroviral (ARV) drug providers and associated cost overruns, posed a risk to the supply of ARV medications and the care of all HIV patients in the public sector (Urzúa, 2020).

In response to the situation, the SHCP summoned various multidisciplinary groups including medical experts, scientists, and administration and finance professionals with experience in consolidated purchasing. These groups met to analyze the problem's root causes and seek possible solutions, focusing primarily on the public health perspective. As a result of this analysis, a drastic change in the consolidated purchasing model was implemented in 2019. This change also involved the modification of the purchasing procedure, previously coordinated by the Mexican Social Security Institute (IMSS, Instituto Mexicano del Seguro Social).

In this context, interest in conducting a comprehensive analysis of this new approach to consolidated purchasing arose in order to assess its effectiveness. Efficient resource management in the healthcare sector is a crucial concern today, and one of the most important fields in this area is the procurement and distribution of ARV medications for HIV treatment. While it cannot eradicate HIV infection, "Antiretroviral

[1] This study was possible thanks to the support and collaboration of various government entities in Mexico. We especially thank the Secretaría de Hacienda y Crédito Público (Ministry of Finance and Public Credit), Secretaría de Salud (Health Ministry), CENSIDA (Centro Nacional para la Prevención y Control del VIH y el sida), IMSS (Mexican Social Security Institute), ISSSTE (Institute of Security and Social Services for State Workers), Pemex (Mexican Petroleum), and INAI (National Institute for Transparency, Access to Information, and Protection of Personal Data) for providing us with the necessary data to conduct this research. Their contribution was fundamental to the success and significance of the results obtained.

[2] In preparing this paper, we consulted and interviewed officials from the Mexican government, primarily from the SHCP, the Secretaría de Salud (SS), and the Centro Nacional para la Prevención y Control del VIH y el sida (CENSIDA).

therapy (ART) suppresses viral replication, increases $CD4^+$ cell-count, decreases AIDS-related mortality and morbidity and comorbidities, improves the quality of life of HIV-infected patients, and prevents sexual transmission of HIV" (Lozano & Domingo, 2011). Within this framework, there is a need to explore the results of a radical ARV procurement strategy implemented by the Mexican government, and how this strategy has affected both the economic aspect and the quality of care for patients.

This research takes shape as part of a collaborative effort that brings experts from various disciplines together, including finance, information technology, and ARV medication administration. The main objective is to assess whether the paradigm shift in the ARV medication procurement process—through the implementation of consolidated purchasing—has achieved the expected benefits both in terms of economic efficiency and improvement of healthcare provided to HIV patients.

Consolidated purchasing has been presented as a promising strategy to achieve significant economic savings and ensure the constant availability of essential medications. In this regard, this article thoroughly analyzes whether this approach has led to substantial economic savings and improved the quality of care for HIV patients at the same time.

A crucial aspect of the methodology is the application of advanced data analysis and machine learning techniques, which play a key role in evaluating this strategy's results. These techniques allow for a rigorous and comprehensive analysis of ARV medication procurement data, and the exploration of patterns and relationships that help better understand the consolidated purchasing strategy's underlying dynamics and assess its real impact.

The essential contribution of this research lies in three interrelated dimensions. First, it examines whether consolidated purchasing strategies for ARV medications have had a positive impact on the costs and financial sustainability of HIV treatment programs. Pattern and trend identification through machine learning techniques offers valuable information for informed decision-making regarding resource allocation.

Second, it addresses a gap in the literature by exploring how the implementation of advanced data analysis and machine learning techniques can optimize the financial administration of ARV procurement. These techniques provide a predictive insight into future trends in pharmaceutical purchasing, enabling efficient resource planning and allocation.

Lastly, this research highlights the importance of a multidisciplinary approach to addressing the complex challenges in ARV procurement management. Collaboration among experts in different fields provides a unique and enriching perspective in analyzing the challenges linked to the purchasing and efficient distribution of ARV medications.

2. Methodology

The method adopted is designed to comprehensively evaluate the efficiency and effectiveness of the ARV medication consolidated purchasing scheme in Mexico during 2019. The methodological approach includes the following steps:

- *Demand and Coverage Analysis*: It starts with a detailed analysis of statistical data sourced from international organizations and the Government of Mexico. (WHO, 2021a; WHO, 2021b). This analysis encompasses the demand for ART among individuals living with HIV/AIDS, internationally, in Latin America, and specifically in Mexico, including relevant information regarding ART coverage worldwide (33%), in Latin America (38%), and in Mexico, where 172,221 people have access to it. Of these, 99.2% receive it from the public sector and 0.8% from the private sector (SS, 2018; UNAIDS, 2020; UNAIDS, 2021).
- *Budgetary Analysis and Savings Evaluation:* Budgets are highlighted as a key factor in determining savings derived from variations in ARV medication prices. Fluctuations in average prices are quantified, which were around 55% in this study.
- *Medication Price Analysis:* We conducted a detailed analysis of ARV medication prices using the set of prices of all 15,220 medications procured by the National Center for HIV Prevention and Control of HIV and AIDS (Centro Nacional para la Prevención y Control del VIH y el sida, CENSIDA) from 2017 to 2020 as a basis. Additionally, we have a database of medications procured in fiscal years prior to 2019, when the remaining agencies and entities procured ARV medications outside the consolidated purchasing scheme. In this way, the results of different contracting schemes can be contrasted using these databases.
- *Analysis with Machine Learning Algorithms:* In order to project ARV medication prices, we employed machine learning algorithms in R and Python statistical environments (McKinney, 2013). We identified relevant data sets, including

variables such as purchasing agencies and entities, procured ARV regimens, and we classified all 15,220 data items used in the model construction. Additionally, we carried out analyses on the suppliers with allocations in 2017, 2018, and 2019, in response to the relevance identified through information requests to the Mexican government.

The methodology describes in detail the implementation of logistic regression, random forest, and K-Nearest Neighbors (KNN) models, including a theoretical explanation of the functioning of each. Research results are presented at the end of the development section, providing a thorough data-supported analysis of the consolidated purchasing scheme's effectiveness in 2019.

The adopted methodology seeks to ensure comprehensiveness and rigor in evaluating the ARV medication consolidated purchasing scheme. It relies on robust analytical approaches and machine learning algorithms to draw objective conclusions based on relevant empirical evidence.

3. Development

We conducted a comprehensive analysis to evaluate the effectiveness of the 2019 consolidated purchasing program, which ensures the supply of ARV medications to HIV patients.

We refer to the 2019 consolidated purchasing, which comprises the 2019-2020 fiscal cycle, from April 1, 2019, to March 31, 2020.

In order to ensure result accuracy and ensure that the decrease in ARV medication costs was not due to insufficient use of budgetary resources nor to a reduction in the quality of regimens provided by the Secretaría de Salud (SS, Health Ministry), we applied rigorous evaluation criteria:

- *Quality:* We examined improvements in HIV patient treatment.
- *Cost (budget):* We analyzed the reduction in ARV medication procurement prices.

The analysis was based on identifying and segmenting various relevant variables, including state, age, and agency. Table 1 shows the amounts of ARV medication procured, the quantity of ARV procured, and various key indicators at different times (see Table 1).

Table 1. ARV Medication Procured from 2017 to 2020

Cutoff Date	Amount of ARV Procured (in Mexican pesos, $MX)	Quantity of ARV Procured	Individuals on ART	Individuals with CD4 <200	Awarded Providers
31/03/17	537,812,507.78	339,566			7
30/06/17	1,276,375,861.82	608,514			3
30/09/17	488,984,118.03	232,862			6
31/12/17	960,073,004.33	485,409	87,026	13,101	8
31/03/18	393,950,549.44	251,783	89,315	12,813	8
30/06/18	985,625,837.20	446,229	91,194	13,259	3
30/09/18	815,559,854.53	438,252	93,037	13,762	8
31/12/18	871,150,963.21	467,876	95,732	14,004	8
31/03/19	350,896,989.49	221,428	97,245	3,498	8
30/06/19	811,635,301.73	521,787	98,765	3,498	14
30/09/19	685,147,693.28	537,324	99,531	6,921	14
31/12/19	630,004,095.86	425,418	100,409	11,273	15
31/03/20	553,478,872.57	346,809	102,288	2,147	15

Source: Compiled by the authors with data from CENSIDA (2017a), CENSIDA CENSIDA (2017b), CENSIDA (2017c), CENSIDA (2017d), CONASIDA (2018), CENSIDA (2018a), CENSIDA (2018b), CENSIDA (2018c), CENSIDA (2018d), CENSIDA (2019e), CENSIDA (2019f), CENSIDA (2019g), CENSIDA (2019h), CENSIDA (2019i), CENSIDA (2020a), CENSIDA (2021a) y CENSIDA (2021b).

The study relies on data collection and examination to provide a comprehensive understanding of the consolidated purchasing scheme's effectiveness in terms of its influence on treatment quality and cost reduction in the procurement of ARV medication for HIV patients. Segmentation based on key variables allows for a complete and enriching view of the results obtained, enabling a precise and detailed evaluation of the implemented initiative.

3.1. Demand and Coverage Analysis

The analysis focuses on the comprehensive evaluation of demand and coverage of ARV treatment regimens provided to HIV patients in Mexico. The aim is to understand the distribution, effectiveness, and characteristics of care programs in the public sector. The data used came from various official sources, including CENSIDA reports, which is the primary national agency for the prevention and control of HIV/AIDS in Mexico, as well as responses to information requests submitted to the Plataforma Nacional de Transparencia (PNT, National Transparency Platform) of the Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI, National Institute for Transparency, Access to Information, and Protection of Personal Data), Mexico's principal body responsible for ensuring transparency and access to public information.

According to the National Report on Monitoring Commitments and Expanded Goals to End AIDS, the distribution of ART in Mexico is largely concentrated in public sector institutions, with 99.2% of individuals receiving treatment in these entities, while 0.8% receive it in the private sector (SS, 2018).

Within the public sector, government institutions that provide ART to HIV patients include Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE, Institute of Security and Social Services for State Workers), IMSS, Petróleos Mexicanos (Pemex, Mexican Petroleum), Secretaría de la Defensa Nacional (Sedena, National Defense Ministry), Secretaría de Marina (Semar, Ministry of the Navy), and Secretaría de Salud (SS) through CENSIDA.

Regarding the specific patient distribution by institution and cutoff date, we can observe in Table 2 that the SS treats the largest number of patients, with a total of 98,100 patients under ART as of June 2019. Likewise, IMSS plays an important role, treating 55,818 patients as of December 2018, while Pemex treats 881 patients on the same date. However, it is worth noting that data provided by Sedena, Semar, and ISSSTE do not specify the number of patients treated (see Table 2).

Table 2. Patients Receiving ART

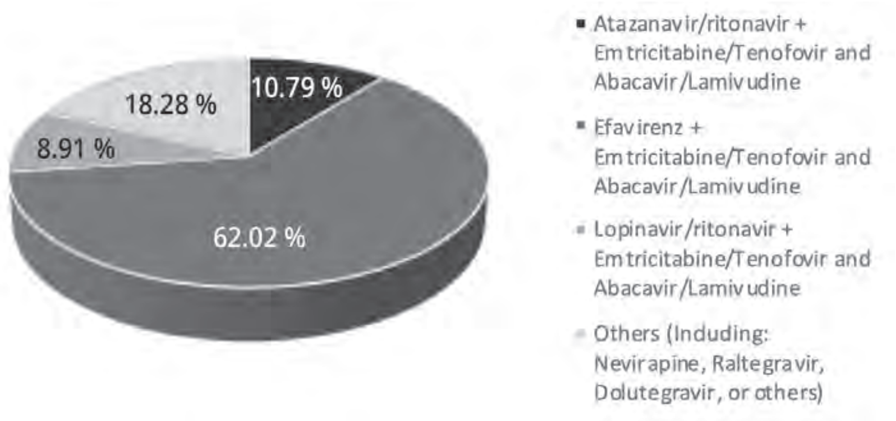
Institution	Number of Patients on ART	Cutoff Date
SS	98,100	June 10, 2019
IMSS	55,818	December 1, 2018
Pemex	881	June 25, 2019
Semar	No data available	No data available
Sedena	No data available	No data available
ISSSTE	No data available	No data available
Total	**154,799**	

Source: Compiled by the authors with data from information requests made to INAI in 2021.

3.2. ARV Treatment Regimens in Mexico

Mexico's ARV medicine market situation reflects concentration on first-line regimens, as revealed by the study "Antiretroviral Purchasing and Prescription Practices in Mexico: Constraints, Challenges and Opportunities" and shown in Figure 1 (Chaumont et al., 2015) (see Figure 1).

Figure 1. Antiretroviral Regimens in Mexico



Source: Compiled by the authors with data from information requests made to INAI in 2021 and with information from the basic framework and catalogue of medications issued by the General Health Council (Consejo de Salubridad General, 2017).

According to records up to June 2019 shown in Figure 1, regimens commonly used by government institutions include Efavirenz plus Emtricitabine/Tenofovir and Abacavir/Lamivudine, administered to 96,013 patients; Atazanavir/Ritonavir plus Emtricitabine/Tenofovir and Abacavir/Lamivudine, administered to 16,697 patients; Lopinavir/Ritonavir plus Emtricitabine/Tenofovir and Abacavir/Lamivudine, administered to 13,795 patients, and other regimens that incorporate agents such as Nevirapine, Raltegravir or others, administered to approximately 28,294 patients on ART (CENSIDA, 2019e).

Moreover, treatment regimens used adhere to CENSIDA's guidelines for administering ARV medications, which predominantly consist of groupings of Nucleoside Reverse Transcriptase Inhibitors (NRTIs) and/or Protease Inhibitors (PIs). These groupings vary, with Emtricitabine/Tenofovir plus a third agent such as Efavirenz, Atazanavir/Ritonavir, or Lopinavir/Ritonavir being the most common. Additional agents such as Nevirapine, Raltegravir, or Dolutegravir are also used, albeit in smaller proportions, particularly in cases of viral resistance.

Treatment regimen selection is guided by expert recommendations and customized to address the specific conditions and mutations present in each patient (CENSIDA, 2021c).

Table 3 provides a summary of the amount each of the regimens has been prescribed, commonly supplied by government agencies and entities in the Mexican Government before consolidated purchasing (see Table 3).

Table 3. Most Used ARV Treatment Regimens

ART Regimen	CENSIDA[3]	IMSS[4]	Pemex[5]	Total
Efavirenz + Emtricitabine/Tenofovir and Abacavir/Lamivudine	66,880	28,888	245	96,013
Others (Including: Nevirapine, Raltegravir, or Dolutegravir, among others)	21,166	6,938	190	28,294

[3] Response to request 1200800005619. Numbers as of June 10, 2019.
[4] Response to request 64101707619. Numbers as of December 2018.
[5] Response to request 1857200281519. Numbers as of June 25, 2019.

ART Regimen	CENSIDA[3]	IMSS[4]	Pemex[5]	Total
Atazanavir/Ritonavir + Emtricitabine/ Tenofovir and Abacavir/Lamivudine	4,504	12,053	140	16,697
Lopinavir/Ritonavir + Emtricitabine/Tenofovir and Abacavir/Lamivudine	5,550	7,939	306	13,795
Total	**98,100**	**55,818**	**881**	**154,799**

Source: Compiled by the authors with data as of June 2019 from information requests made to INAI in 2021.

3.3 Budgetary Analysis and Savings Evaluation

This analysis focuses on examining the financial aspect of acquiring ARV medications for HIV/AIDS treatment in Mexico, using different sources of information for a comprehensive assessment.

We initially analyzed projected and published savings in 2019 by CENSIDA. During the 2019 to 2020 analysis period, we can observe an allocation of resources from the Catastrophic Expenses Fund of the Seguro Popular (SP, People's Insurance). CENSIDA obtained the authorization from the Technical Committee of the Fideicomiso del Sistema de Protección Social en Salud (Trust for the Social Protection System in Health), resulting in an allocation of 3,207,277,135.17 Mexican pesos. Of this amount, 2,820,476,482.87 were allocated for ARV medication procurement, while the remaining 386,800,652.30 were used to finance viral load, CD4, and genotype tests. This allocation highlights a significant component of estimated savings, totaling 1,373,101,058 Mexican pesos. According to this report, the estimated savings at that time would represent an approximately 55% reduction compared to purchases from the previous year, and they originate from the procurement of single-source and generic ARV medications (CENSIDA, 2019a).

Based on the response provided by CENSIDA, derived from information access request No. 0001200157620 submitted through the National Transparency Platform, the evaluation focuses on the purchasing periods spanning from 2015 to 2020, with a special emphasis on 2019 and 2020, as shown in Table 4 (see Table 4).

Table 4. Expended Budget on HIV/AIDS Care with Resources from the Fideicomiso del Sistema de Protección Social en Salud (Trust for the Social Protection System in Health), in Mexican pesos

Fiscal Year	Period[6]	ARV Medication	Laboratory Tests (viral load, CD4, and genotype)	Total
2015	2015-2016	2,525,542,291.89	348,666,843.00	2,874,209,134.89
2016	2016-2017	2,826,011,136.95	325,624,483.00	3,151,635,619.95
2017	2017-2018	3,097,064,812.92	329,155,024.44	3,426,219,837.36
2018	2018-2019	3,023,233,644.43	348,368,940.43	3,371,602,584.86
2019	2019-2020	2,471,152,477.44	363,581,350.00	2,834,733,827.44

Source: Compiled by the authors with data from information requests made to INAI in 2020.

In the 2018 fiscal year, spanning from April 1, 2018, to March 31, 2019, considerable resources were allocated to ARV medication procurement, reaching an expended budget of 3,023,233,644.43 Mexican pesos. However, a significant decrease is evident in the purchases corresponding to the 2019 fiscal year, with a reduction in the expended budget of 552,081,166.99, representing an 18.26% contraction compared to the previous year and marking it out as the lowest expended budget since 2015.

These numbers reflect the budget executed with resources from the Fideicomiso del Sistema de Protección Social en Salud (Trust for the Social Protection System in Health), although it is relevant to mention that the amounts do not include credit notes, which are incorporated into the estimated savings prior to the consolidation of ARV purchasing.

Specifically, the allocation of resources in different periods for ARV pharmaceutical purchasing and laboratory tests can be observed. Data illustrate an evolution in the expended budget over the years studied, with numbers ranging from 2,525,542,291.89 to 3,097,064,812.92 Mexican pesos for ARV medications and between 325,624,483.00 and 363,581,350.00 for laboratory tests, between 2015 and 2020.

The financial analysis provides a comprehensive insight into resource allocation and changes in the budget allocated for pharmaceutical purchasing.

[6] The periods cover from April 1, 2015, to March 31, 2020.

3.4 Medication Price Analysis

This study embarked on a comprehensive analysis of the numbers disclosed by CENSIDA regarding procured medications, focusing on the effects following the implementation of a renewed therapeutic approach, derived from a collaboration between SS and SHCP. We identified the medications that experienced the most significant reductions in unit prices in this context.

Table 5 shows the medications that reflected notable changes in their unit prices after the adoption of the new therapeutic approach (see Table 5).

It is important to note that this representation is a selection based on the numbers published by CENSIDA up to December 31, 2019. It is worth mentioning that data present in CompraNet, the official platform for government procurement registration, does not provide a standardized breakdown of the details of each medication procured.

This analysis is based on all medications reported by CENSIDA. Table 6 summarizes the data used, considering breakdowns by state, institute, and hospital, throughout each month from January 2017 to July 2020 (see Table 6). The results provide a holistic view of the evolution of ARV medication prices and procurement in Mexico (CENSIDA, 2021b).

In conclusion, the decrease in the budget allocated for purchasing ARV medications in 2019 compared to 2018 was largely offset by the estimated savings in their procurement, allowing for greater efficiency in the utilization of resources allocated for this purpose. Additionally, the implementation of the new therapeutic approach contributed to a reduction in the unit price of some ARV medications (Fernando & Pere, 2011; European Medicines Agency, 2018; NIH, n.d.).

Table 5. Price Comparison of Medications Procured in 2018 vs. Medications Procured with Consolidated Purchasing in 2019 (Mexican pesos)

ARV	Monthly Average Quantity Procured	2018 Unit Price (as of March 2019)	2018 Monthly Average	2019 Unit Price after Consolidated Purchasing	2019 Monthly Average	Decrease %
ABACAVIR, 300 mg: 60 tablets per package	1,477	MX$543	MX$801,849		MX$664,650	21%
ABACAVIR/LAMIVUDINE, 600/300 mg: 30 tablets per package	10,800	MX$990	MX$10,692,000	MX$282	MX$3,045,600	251%
DARUNAVIR, 400 mg: 60 tablets per package	3,500	MX$3,286	MX$11,500,545	MX$1,600	MX$5,600,000	105%
DARUNAVIR, 600 mg: 60 tablets per package	1,300	MX$4,481	MX$5,824,949	MX$1,834	MX$2,384,200	144%
EFAVIRENZ 600 mg, EMTRICITABINE 200 mg, TENOFOVIR disoproxil succinate 300.6 mg: 30 tablets per package	100,000	MX$1,300	MX$130,000,000	MX$800	MX$80,000,000	63%
EFVIRENZ, 600 mg: 3 0 tablets per package	11,900	MX$162	MX$1,927,800	MX$99	MX$1,178,100	64%
EMTRICITABINE-TENOFOVIR, 200/245 mg: 30 tablets per package	26,000	MX$2,061	MX$53,582,880	MX$710	MX$18,460,000	190%
LOPINAVIR/RITONAVIR, 200/50 mg: 120 tablets per package	900	MX$1,730	MX$1,557,000	MX$988	MX$888,750	75%
TENOFOVIR DISOPROXIL FUMARATE, 300 mg: 30 tablets per package	2,050	MX$2,000	MX$4,100,246	MX$1,040	MX$2,132,246	92%

Source: Compiled by the authors with data from CENSIDA, 2019e.

Table 6. Summary of Medications Used in the Development of Machine Learning Models

ARV	2016	2017	2018	2019	2020
ABACAVIR, 2 g Bottle 240 ml	578	578	578	578	577
ABACAVIR, 300 mg. 60 tablets	500	475	543	450	309
ABACAVIR/LAMIVUDINE, 600/300 mg. 30 tablets	1379	1379	990	282	399
ATAZANAVIR, 300 mg. 30 capsules	2765	2668	2641		
BICTEGRAVIR/EMTRICITABINE/TENOFOVIR ALA-FENAMIDE, 50/200/25 mg. Box with 30 tablets				1720	1720
DARUNAVIR 150 mg. 240 tablets	4979	4979	4979	4979	4979
DARUNAVIR 400 mg. 60 tablets	3319	3286	3286	1600	1521
DARUNAVIR 600 mg. 60 tablets	4979	4979	4481	1834	2049
DARUNAVIR 75 mg. 480 tablets		4979	4979	4979	
DARUNAVIR/COBICISTAT, 800/150 mg. 30 tablets			2960	2915	2915
DIDANOSINE, 250 mg. 30 capsules	657	657			
DIDANOSINE, 400 mg. 30 capsules	1057	1057			
DOLUTEGRAVIR, 50 mg. 30 tablets	4077	4077	3335	3269	3000
DOLUTEGRAVIR/ABACAVIR/LAMIVUDINE, 50/600/300 mg. 30 tablets			4665	4135	3000
EFAVIRENZ, 600 mg. 30 tablets	259	221	162	99	85
EFAVIRENZ/EMTRICITABINE/TENOFOVIR DISO-PROXIL, 600/200/245 30 tablets	2404	2332	2332	805	800
ELVITEGRAVIR/COBICISTAT/EMTRICITABIN A/ TENOFOVIR, 150/150/200/245 mg. 30 tablets			2000	2000	
EMTRICITABINE/TENOFOVIR ALAFENAMIDA, 200/10 gr. 30 tablets				1720	1720
EMTRICITABINE-TENOFOVIR, 200/245 mg. 30 tablets			2134	710	710
EMTRICITABINE, 200 mg. 30 capsules	490	343	601		
EMTRICITABINE/TENOFOVIR ALAFENAMIDA, 200/25 mg. 30 tablets					1720
EMTRICITABINE/TENOFOVIR DISOPROXIL FUMARATE, 200/245 mg. 30 coated tablets	2125	2061			

ARV	2016	2017	2018	2019	2020
ENFUVIRTIDE, INJECTABLE SOLUTION. 108 mg. 60 3 ml syringes, 60 1 ml syringes, and 180 alcohol wipes.	22450	22450	22608	20330	20330
ETRAVIRINE, 100 mg. 120 tablets	5429	5429	5429		
ETRAVIRINE, 200 mg. 60 tablets			5429	5429	5429
GLECAPREVIR/PIBRENTASVIR, 100/40 mg. 4 boxes, with 7 strips of 3 tablets each					68250
LAMIVUDINE, 150 mg. 60 tablets	584	394	394	365	335
LAMIVUDINE. 1 g per 100 ml. 240 ml and pipette	835	835	894	891	890
LAMIVUDINE/ZIDOVUDINE, 150/300 mg. 60 tablets	772	772	595	425	234
LOPIANVIR/RITONAVIR, 100/25 mg. 60 tablets	1100	1100	1100	1100	1100
LOPIANVIR/RITONAVIR, 200/50 mg. 120 tablets	2656	2063	988	1730	2010
LOPIANVIR/RITONAVIR, 8.0/2.0 g per 100 ml. Amber glass jar with 160 ml and dosing cup	1714	1714	1714	1714	1714
MARAVIROC 150 mg. 60 tablets	6622	6622	6622	6622	6612
MARAVIROC 300 mg. 60 tablets	6622	6622	6622	6622	6612
NEVIRAPINE, 1.0 g per 100 ml. 240 ml and dosing cup	333	328	313	313	313
NEVIRAPINE, 200 mg. 60 or 100 tablets	384	380	377		
RALTEGRAVIR, 400 mg. 60 tablets	5309	4247	3610	3395	3574
RITONAVIR, 100 mg. 30 tablets	348	348	348	348	348
SOFOSBUVIR-VELPATASVIR, 400/100 mg. 28 tablets.					49896
TENOFOVIR, 300 mg or 245 mg. 30 tablets	2000	2000	2000	1040	510
TIPRANAVIR, 250 mg. 120 capsules	3229	3229			
ZIDOVUDINE INJECTABLE SOLUTION, 200 mg. 5 vials (200 mg/20 ml)				995	993
ZIDOVUDINE ORAL SOLUTION, 1 g per 100 ml. 240 ml	515	500	680	485	460
LOPIANVIR/RITONAVIR, 200/50 mg. 60 tablets		1100		1100	
ETRAVIRINE, 200 mg. 120 tablets		5429	5429		

Source: Compiled by the authors with data from CENSIDA (2019e).

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.07

3.5 Analysis Using Machine Learning Algorithms

In this phase of the study, we proceeded with the creation and evaluation of machine learning models using logistic regression, random forest, and KNN techniques. The analysis procedure consisted of the stages shown in Figure 2 (see Figure 2).

Figure 2. Diagram of the Machine Learning Algorithm Analysis

Source: Compiled by the authors.

a) *Data collection and preparation:* We obtained data from various sources, including the National Transparency Platform and CENSIDA. Subsequently, we carried out data cleaning and transformation with the aim of using it in the analysis.

b) *Exploratory data analysis:* We employed various visualization techniques and descriptive statistics to better understand the data, as well as to detect patterns and relationships between variables.

c) *Variable selection and application of Machine Learning Models:* We conducted the identification of the most pertinent variables for the analysis, followed by the implementation of diverse machine learning models, such as logistic regression, random forest, and KNN. We employed these models to forecast potential decreases in ARV medication prices.

d) *Training and validation of each model:* We divided the data into training and validation sets and adjusted model parameters to achieve the best performance. We then evaluated the trained models using metrics such as accuracy, precision, and recall assessing their performance.

e) *Model comparison:* The performance of different algorithms is compared to determine which one provides the best results for the given dataset, predicting whether there was a variation in ARV medication price.

In this section, we elaborated in more detail on the key components of our data analysis and machine learning modeling. We addressed the independent and dependent variables used in the analysis of ARV medication procurement, as well as the machine learning methods applied, based on logistic regression, random forest, and KNN techniques.

3.5.1 Independent Variables (Medication Characteristics)

- ARV Type: We used this variable to categorize medications in terms of their commercial and market identification, allowing for the consideration of differences between generic, branded, or single-source medications.
- Short description: Brief description of the 45 different medications analyzed.
- Detailed description: Detailed description of the 45 different medications analyzed.
- Unit price: Amount in Mexican pesos paid per unit procured corresponding to the purchase year.
- Price from previous period: Amount in Mexican pesos paid per unit procured corresponding to the immediate previous purchase year.
- Quantity procured: Number of units procured corresponding to the purchase year.
- Total amount: The total amount in Mexican pesos paid for all units procured of this medication corresponding to the purchase year.
- Active components: This variable includes the chemical composition of the medication and its potential therapeutic effect, enhancing our understanding of the analyzed ARVs.
- Brief presentation: An indicator that includes a summarized presentation of the medication: capsule, tablet, solution, suspension, pill.
- Detailed presentation: An indicator that includes an extensive presentation of the medication: capsule, tablet, solution, suspension, pill.

- Dosage: A quantitative variable that plays a vital role in quantifying the amount of active ingredients in a treatment unit, including the unit of measurement.
- Content per package: Linked to treatment duration, this variable provides insights into medication availability and continuity.
- Supplier/Manufacturer: A factor that adds an important nuance by identifying the supply and production source of ARV medications, whether it be the name of the individual or legal entity from which the medication was purchased.
- Client: Government department or entity in Mexico that receives the purchased medications.
- Procurement Year: An indicator that contextualizes transactions over time, allowing for analysis of temporal patterns in medication purchasing.

We analyzed and selected relevant columns for the study exploratory data analysis (EDA), correlations, and feature importance with random forest.

3.5.2 Dependent Variable: Price Change

The dependent variable "price change" captures variations in ARV medication prices. This binary variable, encoded as "1" for price increase and "0" for decrease, allows us to analyze trends in procurement prices over time.

3.5.3 Data Analysis Methods and Machine Learning Modeling

3.5.3.1 Logistic regression

In this phase of the study, logistic regression was employed as one of the primary machine learning techniques for analyzing medication prices. Logistic regression remains a cornerstone in predictive modeling, particularly in domains where the outcome of interest is binary or categorical. Its flexibility, interpretability, and ability to handle complex relationships make it indispensable in the analysis of healthcare data, including pharmaceutical pricing dynamics Hastie, Tibshirani, and Friedman (2021a) and Hastie, Tibshirani, James, and Witten (2021b).

According to the logistic regression approach described by Hastie et al. (2021a) in their book *The Elements of Statistical Learning*, the procedure for analyzing medication prices using logistic regression is as follows:

- *Data preprocessing*. Missing values should be removed, and categorical variables encoded. Additionally, the dataset should be split into training and testing sets. We created a database containing the prices and characteristics of ARV medications procured by the Government of Mexico from 2015 to 2020. We selected relevant predictor variables that may affect medication prices, such as: medication type, short description, quantity procured, total amount, active components, summarized presentation, dosage, content per container, supplier/manufacturer, client, and procurement year. Finally, we partitioned the dataset into training and testing subsets, with an 80% training and 20% testing ratio.
- *Model training*. The logistic regression model is the training set. The goal is to find the optimal coefficient values that maximize the probability of the model correctly predicting each observation's class. The selected predictive variables are medication type, short description, quantity procured, total amount, active components, summarized presentation, dosage, content per container, supplier/manufacturer, client, and procurement year.
- *Model evaluation*. The model's accuracy is evaluated on the test data using measures such as the error rate and confusion matrix. Logistic regression modeled the probability of price change based on the independent variables.

Through the logistic function, we examine the relationships between these variables and their impact on the dependent variable.

We formulated a logistic regression model in Python with the aim of anticipating whether medication prices exhibit an increase or a decrease. To accomplish this, a function is employed to model $p(X)$, producing values that are exclusively binary, zero or one for all X values. In logistic regression, the logistic function is introduced within this context.

The logistic regression formula is expressed as:

$$log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Where $p(X)$ is the probability that $Y = 1$ given $X = x$, β_0 is the intercept term, and $\beta_1, \ldots, \beta_p$ are the coefficients of the predictor variables $X_1, \ldots, X_p$, also known as the regression coefficients.

The logistic regression formula is:

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}}$$

This function converts the weighted sum of the input variables and coefficients into a probability value, which falls within the range of 0 and 1. If this probability surpasses the threshold of 0.5, the model indicates a prediction for the positive category; otherwise, it predicts the negative category.

In the analysis of medication prices, logistic regression offers a robust framework for modeling the likelihood of price changes based on various predictor variables. Its application in forecasting price fluctuations in pharmaceutical markets has yielded valuable insights for policymakers and healthcare administrators (Hosmer Jr et al., 2013).

With the increasing availability of healthcare data, logistic regression serves as a powerful tool for analyzing trends in medication pricing and procurement. Its ability to handle large datasets and incorporate diverse predictor variables makes it indispensable for understanding the drivers of price variations in the pharmaceutical sector (Agresti, 2015).

3.5.3.2 Random forest

The implementation of random forest allowed for a more sophisticated and accurate analysis. This technique combines multiple decision trees to identify subtle patterns and nonlinear relationships in the data.

Aligned with Hastie et al.'s (2021b) description of the random forest methodology, this approach involves the creation of decision trees using a set of training data. However, a distinct feature of this technique is that during the development of each tree, a random subset of predictors is chosen from the complete set of available predictors when evaluating a potential split. Among these selected predictors, only one is used for the actual split. Moreover, with each split, a new subset of predictors is randomly sampled, typically with the number of predictors being close to the square root of the total number of predictors ($m \approx \sqrt{p}$).

Random forest is recognized for its versatility and capability to handle complex datasets in various domains (Fernández-Delgado et al., 2014). In healthcare analytics, it has been extensively utilized for predicting clinical outcomes, disease diagnoses, and medication responses.

Opting for a smaller *m* value when constructing random forest proves advantageous, particularly in scenarios where there is substantial correlation among numerous predictors. This strategy aids in predicting the decrease in the price of ARV medications by effectively handling the complexities and interrelationships among various factors involved.

According to Marsland (2014), the random forest algorithm leverages ensemble learning by aggregating multiple decision trees, each trained on a bootstrap sample and a subset of features, thereby mitigating overfitting and enhancing generalization performance.

Moreover, in financial domains, random forest has proven effective for tasks such as credit risk assessment, stock market prediction, and fraud detection (Liaw & Wiener, 2002). Its ability to handle large volumes of data, nonlinear relationships, and high-dimensional feature spaces makes it a valuable tool for analyzing pharmaceutical pricing dynamics and forecasting changes in medication costs.

3.5.3.3 K-Nearest Neighbors Method

The K-nearest neighbors (KNN) method, as described by Hastie et al. (2021b), plays a crucial role in our predictive approach. This method relies on the idea that a new data point's category can be predicted by examining the categories of nearby data points. It assumes that data points with similar characteristics tend to belong to the same category, providing useful insights for classification and prediction tasks and improving our understanding of price changes.

Implementation of the KNN technique entails the following procedural steps:

- *Selection of K* value: A suitable value for *K*, denoting the number of nearest neighbors considered during prediction, is determined.
- *Distance computation*: Utilizing a Euclidean metric, the distance between the data point to be predicted and all other points in the training set is computed.
- *Neighbor identification*: The *K* nearest points to the data point under prediction are identified based on the calculated distances.
- *Category prediction*: Among these *K* neighbors, the category that appears most frequently is designated as the predicted category for the data point in question.

By following these steps, the KNN method facilitates informed predictions regarding price changes, thereby enhancing decision-making processes within our predictive framework.

KNN demonstrates its effectiveness particularly in situations where there is no clear functional relationship between the features and categories, as well as when the data exhibits nonlinear structures. However, it is important to consider that KNN may present significant computational consumption, especially in extensive datasets, due to the distance calculation required for all points in the training set.

In the context of this research, the KNN method was employed in conjunction with the two predictive approaches mentioned earlier: logistic regression and random forest, in order to explore its predictive potential and complement the comprehensive analysis of ARV medication prices, following the methodology proposed by Hastie, Tibshirani, and Friedman (2021a), and Hastie, Tibshirani, James, and Witten (2021b) and Badal and Sungkur (2023).

We separated data into training and testing sets with 80% of the data for training and 20% for testing. Prior to this separation, the data underwent preprocessing that normalized and standardized the variables, ensuring the consistency and accuracy of the results.

Given the significance of KNN in pharmaceutical price prediction, it is noteworthy to mention the seminal work by Cover and Hart (1967) which underscore KNN's versatility, efficacy, and simplicity in pharmaceutical price forecasting by leveraging similarities among medication attributes and purchase patterns.

4. Scope and Limitations

The scope of this study is based on the evaluation of consolidated purchasing of ARV medications in Mexico, carried out by the Federal Government through the SS. Specifically, the analysis focuses on the procurements made by CENSIDA with the purpose of meeting the therapeutic demands of HIV patients during the 2019 fiscal year. This interval spans from April 1, 2019, to March 31, 2020, covering the transactions aimed at addressing the medical needs of 98,100 patients undergoing ARV treatment. This patient cohort constitutes approximately 63.37% of the total population of individuals identified in this category. It is important to highlight that this analysis is exclusively limited to procurements made by the Federal Government, excluding any information regarding purchases made by private entities, which represent only 0.8% of the total population receiving ARV treatment in Mexico.

To better understand patterns in ARV medication procurement by the Mexican government in 2019, it has been decided to adopt an initial approach focused on the most representative medication within our dataset. This decision is based on the need to delve into the analysis of a specific medication to identify and understand the factors influencing its price and variability. The choice of this approach is justified by the wide gap observed between the minimum and maximum prices of the various medications procured. By focusing on a specific medication in this initial study, we aim to establish a solid foundation for understanding the determinants of price and procurement patterns before expanding our analysis to the entirety of ARV medications.

It is essential to recognize that this study faces certain intrinsic limitations due to the nature and source of the data used for the analysis:

- *Inconsistencies in information:* Discrepancies have been identified in the total expenditure amounts present in various official sources such as information access requests addressed to the INAI Institute, compared to the numbers published on the official CENSIDA website. Despite this inconsistency, the analysis was based on the data provided directly by the official CENSIDA portal.
- *Heterogeneous data source:* Since we relied upon data obtained from multiple official government sources and international organizations, there is a possibility of introducing a certain degree of bias due to inherent variations in the sources of information.
- *Absence of data in international statistics:* The lack of specific data for Mexico in statistics published by international organizations, such as the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO), led to the need to complement these statistics with information provided by Mexican government agencies responsible for the official data generation. Obtaining this information often involved submitting requests to the INAI.

Despite these limitations, however, we expect this study to provide a comprehensive and objective analysis of the consolidated purchasing of ARV medications in Mexico during the 2019 fiscal year, providing valuable insights for the evaluation of the effectiveness and efficiency of this procurement system in the realm of HIV patient treatment.

5. Results

This study aimed to evaluate the consolidated purchasing of ARV medications in Mexico in 2019. It primarily employed machine learning techniques to predict changes in drug prices and analyze their economic impact. Below, we present the results obtained from the analyses conducted.

In order to better understand ARV medication procurement patterns, we applied exploratory data analysis techniques. Table 7 shows a statistical summary of the numerical variables of the medications considered in this study (see Table 7), while Table 8 displays the statistics of the categorical variables (see Table 8).

The analysis reveals a variety in the characteristics and prices of the procured medications, suggesting a wide range of therapies and treatment approaches.

Table 7. Summary of Numerical Variables

Characteristic	Count	Mean	Std	Min	25%	50%	75%	Max	Unique Ratio
Year	1653	2019.40	0.49	2019	2019	2019	2020	2020	0.00
Purchase Year	1653	2019	0.00	2019	2019	2019	2019	2019	0.00
Month Number	1653	5.06	2.10	1	3	6	6.00	12.00	0.01
Unit Price	1653	2099.11	2621.30	99	578	1714	2915	20330	0.02
Quantity Procured	1653	1107.89	5992.21	1	18	75	467	104000	0.40
Total Amount	1653	1621455.51	8032435.54	99	20568	98462.83	613440	145750000	0.81
Previous Period Price	1365	2656.09	3055.69	162	680	2000.12	3610.07	22608.35	0.02

Source: Compiled by the authors.

Table 8. Summary of Categorical Variables

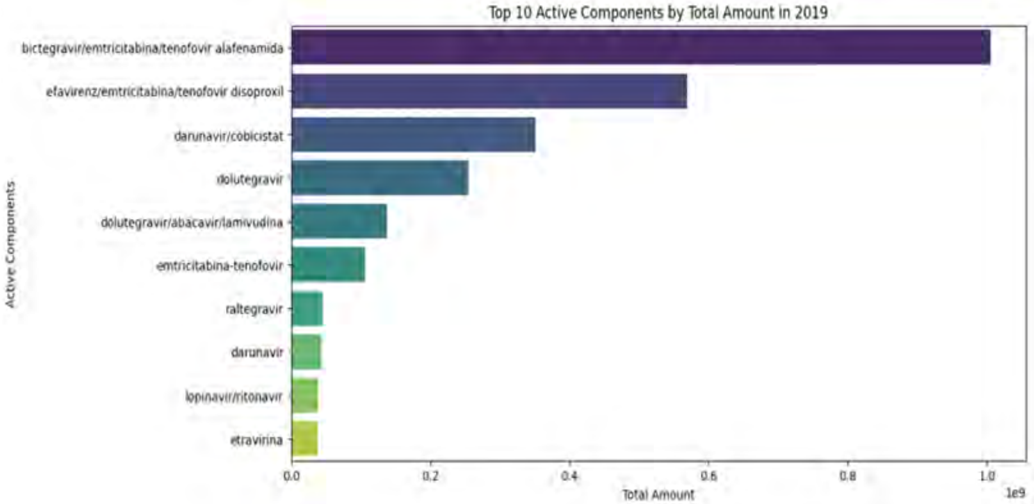
Characteristic	Count	Unique	Top	Freq.	Unique Ratio
Supplier	1653	15	Glaxosmithkline México, S.A. de C.V.	283	0.01
Client	1580	41	Mexico City	50	0.02
Presentation Brief	1653	4	Tablet	1301	0.00
Presentation Detailed	1653	9	Tablet	1213	0.01
Container Content	1653	13	30 tablets	589	0.01
Dosage	1653	25	50/200/25 mg	158	0.02
Active Components	1653	23	Lopinavir/Ritonavir	161	0.01
Short Description	1653	33	Bictegravir/Emtricitabina/Tenofovir Alafenamida	158	0.02
ARV Type	1642	2	Single source	890	0.00

Source: Compiled by the authors.

5.1 Results of the Medication Price Analysis

The comparative analysis of ARV medication prices revealed a downward trend during the period of consolidated purchasing. On average, we observed a 58.6% reduction in the prices of key medications, i.e. medications representing 80% of the budget, as shown in Figure 3. This significant decrease in prices confirms the effectiveness of the consolidated purchasing strategy in achieving substantial economic savings (see Figure 3).

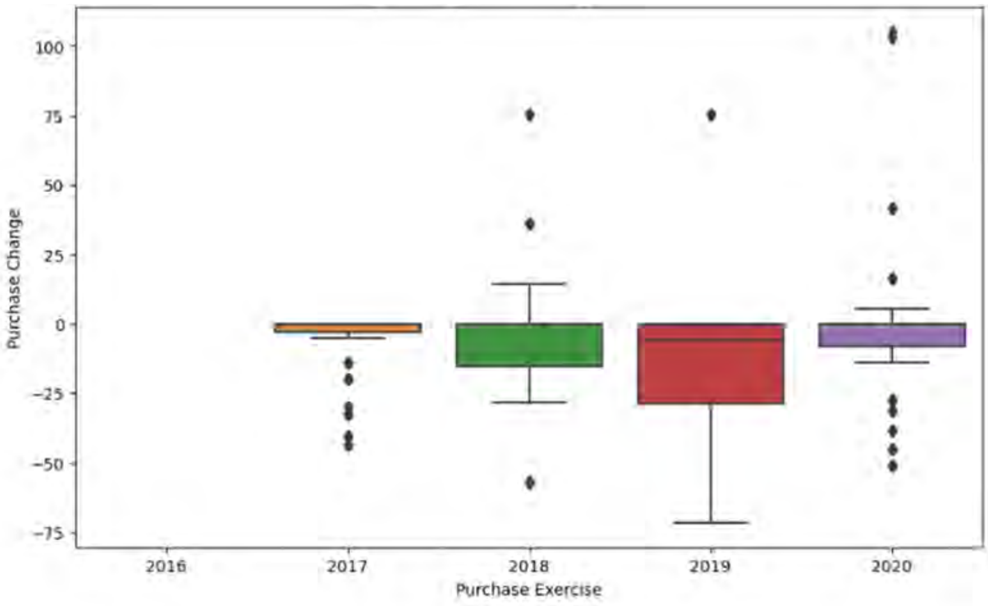
Figure 3. Top 10 Active Components by Total Amount in 2019



Source: Compiled by the authors.

Globally, considering all medications procured, the decrease in ARV medication prices for the 2019 procurement period is 26.1%, as shown in Figure 4 (see Figure 4).

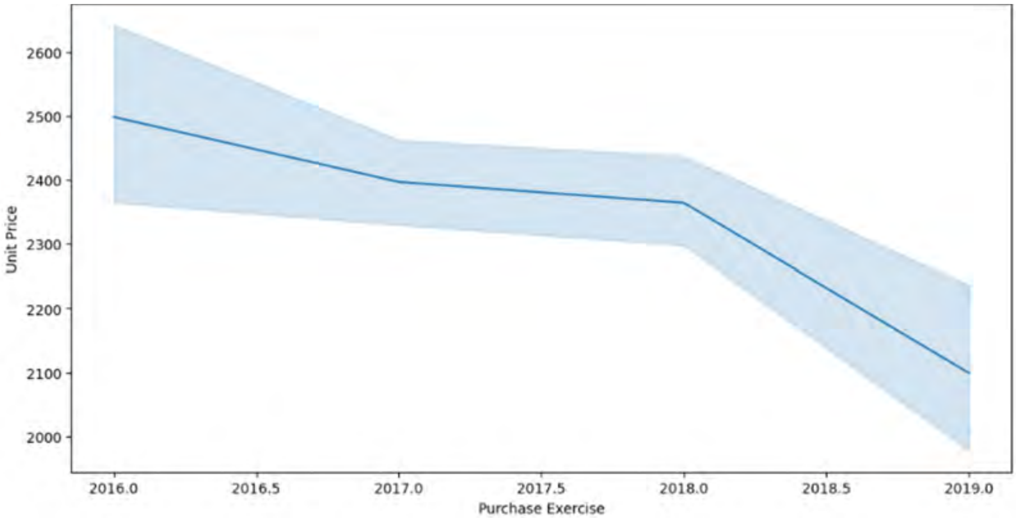
Figure 4. Percentage Change in Prices by Purchase Exercise



Source: Compiled by the authors.

According to numbers provided by the Mexican government, total economic savings during the analyzed period ranged from 1,070 million Mexican pesos to 1,559 million in total, considering laboratory tests, credit notes, and changes in ARV regimens, depending on the consulted source and specific criteria considered. This represents an average reduction of 55% of the budget allocated to CENSIDA for ARV medication procurement, amounting to 2,820,476,482.87 Mexican pesos. A subsequent study will analyze the database to contrast the results with these numbers; however, the detailed analysis of medications' unit prices during the 2019 procurement period reveals a clear downward trend, highlighting a marked decrease compared to previous periods, as shown in Figure 5. This result is evident in the visualization of temporal trends, where unit prices show a descending trajectory throughout different months of the year. This reduction in unit prices translates into a positive impact on total procurement costs, generating substantial savings for ARV medication procurement during the analyzed period (see Figure 5).

Figure 5. Temporal Price Trends



Source: Compiled by the authors.

5.2 Budget Analysis Results and Savings Evaluation

Information provided by CENSIDA highlights the allocation of resources from the Catastrophic Expenses Fund of the People's Insurance for the period from 2019 to 2020. Out of the 3,207,277,135.17 Mexican pesos in the authorized budget, 2,820,476,482.87 was allocated to ARV medication procurement. This report emphasized a crucial component of estimated savings, amounting to 1,373,101,058 Mexican pesos, representing approximately a 55% reduction compared to purchases from the previous year. These savings were primarily derived from the procurement of single-source and generic ARV medications (CENSIDA, 2019a; CENSIDA, 2019c).

The analysis also encompasses the evolution of the expended budget over the years analyzed, ranging from 2,525,542,291.89 Mexican pesos to 3,097,064,812.92 for ARV medications and from 325,624,483.00 to 363,581,350.00 Mexican pesos for laboratory tests, from 2015 to 2020. It is relevant to mention that the reported amounts do not include credit notes, which are incorporated into the estimated savings prior to the consolidation of ARV purchasing.

The budget analysis exhibited a downward trend in resource allocation for ARV medication procurement in Mexico. In the 2018 fiscal year, an expended budget of 3,023,233,644.43 Mexican pesos was reached. However, we observed a significant decrease in purchases corresponding to the 2019 fiscal year, with a reduction in the expended budget by 552,081,166.99, representing a 18.26% contraction compared to the previous year. This decrease marks the lowest budget since 2015.

It is important to highlight that the results of the analysis of our database align with the savings trend. This is due to the evident reduction in the unit price of several medications, which coincides with the overall cost decrease. However, it is crucial to highlight that this study did not assess the overall savings through the calculation of treatment costs for patients treated throughout the 2019 period. We will conduct this detailed analysis in a subsequent study.

5.3 Machine Learning Modeling and Price Change Prediction

For the three models generated in this analysis, we used a database composed of 15,220 observations of ARV medications procured between 2016 and 2019. The focus was on 2019 procurements, limited to medications with the active components Efavirenz, Emtricitabine, and Tenofovir. Each observation contained detailed information about various relevant attributes, such as medication type, quantity procured, total amount, active components, supplier/manufacturer, client, and procurement year. We selected attributes that could influence price changes.

5.3.1 Analysis with Logistic Regression

During training, we adjusted the model coefficients using the gradient descent algorithm. This involved iteratively updating the coefficients to minimize the logarithmic loss function, which measures the difference between the predicted probabilities and the actual outcomes. Feature selection techniques were employed to identify the most influential variables affecting price changes in ARV medications. Additionally, cross-validation was performed to assess the model's generalization ability and to mitigate overfitting.

We thoroughly evaluated the logistic regression model using data collected on ARV medications and their price changes. In addition to traditional evaluation metrics like precision, sensitivity, and specificity, supplementary techniques such

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.07

as cross-validation and bootstrapping were utilized to validate model robustness. Calibration plots were generated to scrutinize the calibration of predicted probabilities across diverse thresholds.

We generated a confusion matrix, shown in Table 9, in order to analyze the correct and incorrect classification of the model's predictions. The confusion matrix offers a comprehensive overview of the outcomes, encompassing true positives, false positives, true negatives and false negatives (see Table 9).

Table 9. Confusion Matrix for the Logistic Regression Model

	No Decrease Prediction	**Decrease Prediction**
No Decrease	56	8
Decrease	6	64

Source: Compiled by the authors.

We calculated precision, sensitivity, and specificity metrics to assess the overall performance of the model in classification. We obtained the following metrics:

- Precision: 0.8750
- Sensitivity: 0.9143
- Specificity: 0.8750

Although the logistic regression model showed reasonable ability to predict price changes in ARV medications, it revealed certain limitations in its predictive capacity in some instances. Despite an overall accuracy of 89.29%, we identified areas for improvement in sensitivity and specificity.

To enhance the performance of the logistic regression model, we will consider incorporating additional data to enrich the model.

5.3.2 Analysis with Random Forest

Throughout the training phase, we made adjustments to fine-tune the model's settings using the Grid Search method. Parameters such as the number of trees in the forest, maximum tree depth, and minimum number of samples required to split a node were fine-tuned to enhance predictive accuracy. Feature importance analysis was conducted to identify significant predictors of ARV medication price changes.

We thoroughly evaluated the random forest model using standard evaluation metrics to understand its performance and predictive capability. In addition to traditional evaluation metrics, such as precision, sensitivity, and specificity, techniques such as out-of-bag error estimation and permutation feature importance were employed to evaluate the model's robustness and generalization ability. Moreover, visualization techniques, such as partial dependence plots and feature interaction analysis, were utilized to gain insights into the complex relationships between predictors and price changes.

The confusion matrix offers a comprehensive breakdown of true positives, correct positives, correct negatives, false positives, and false negatives, providing valuable insights into the model's performance, as shown in Table 10 (see Table 10).

Table 10. Confusion Matrix for the Random Forest Model

	No Decrease Prediction	**Decrease Prediction**
No Decrease	63	1
Decrease	3	67

Source: Compiled by the authors.

We assessed the random forest model's discriminative ability between positive and negative instances using the ROC curve. We quantified the quality of the model by calculating the AUC.

The precision of the random forest model, 98.53%, suggests that the model has a high level of confidence in identifying instances where prices are likely to increase. With a sensitivity that identifies 95.74% of actual price increases in ARV medications, this implies that the model effectively captures the majority of instances where prices are indeed observed to increase. The specificity, 98.53% of the time, the model accurately predicts instances where prices remain stable or decrease.

The overall accuracy of the random forest model indicates that the model's predictions, both positive and negative, are correct 97.32% of the time. It demonstrates the model's effectiveness in providing accurate forecasts of price changes in ARV medications.

To address any potential overfitting of the model, we will consider further optimization of hyperparameters and inclusion of more training data to improve

https://doi.org/10.36105/theanahuacjour.2024v24n1.07

the model's generalization. Additionally, feature selection techniques may be explored to identify the most relevant attributes contributing to the model's predictions.

5.3.3 K-Nearest Neighbors (KNN) Method

The KNN method, leveraging similarity between nearby data points, was employed to predict price changes of ARV medications based on drug characteristics. Prior to model training, dataset preprocessing was performed to ensure compatibility with the KNN algorithm. Techniques such as partitioning, normalization, and standardization of independent variables were employed.

For analysis purposes, we partitioned the dataset into training sets comprising 80% of the data and test sets comprising the remaining 20%.

For each data point in the test set, we calculated the Euclidean distance between that point and all points in the training set. The Euclidean distance is calculated using the formula:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Where n is the number of features, x_i and y_i are the values of feature i in the test point and training point respectively.

In our analysis, we chose a K value which signifies the number of nearest neighbors used to predict the label of a data point within the test set. We set the K value to 5. We determined the most common label among the K nearest neighbors and assigned that label to the test point. If the majority of neighbors experienced a price increase, the test point was labeled "1" (price increase), and if the majority experienced a decrease, it was labeled "0" (price decrease).

Once we assigned labels to all test points, we constructed the confusion matrix to evaluate the KNN model's performance on the test set. The confusion matrix compares the model's predictions against the actual outcomes, providing insight into the accuracy of the classifications.

The KNN model demonstrated a precision of 97.67%, a sensitivity of 94.33%, and a specificity of 97.67% in predicting price changes in ARV medications. The confusion matrix associated with these results is shown in Table 11 (see Table 11).

Table 11. Confusion Matrix Associated with KNN Analysis Results

	No Decrease Prediction	Decrease Prediction
No Decrease	58	2
Decrease	4	70

Source: Compiled by the authors.

Below is a comparative analysis of the logistic regression, random forest, and K-nearest neighbors (KNN) models in terms of precision, sensitivity, and specificity for predicting price changes in ARV.

Although the KNN model provided reasonable capability to predict price changes in ARV medications, it also revealed certain limitations in its predictive capacity in some instances. The precision results for each method are detailed in Table 12 (see Table 12).

Table 12. Precision Results for the Three Methods Used

Method	Precision	Sensitivity	Specificity	Accuracy
Logistic regression	87.50%	91.43%	87.50%	89.29%
Random forest	98.53%	95.74%	98.53%	97.32%
K-nearest neighbors	97.67%	94.33%	97.67%	96.30%

Source: Compiled by the authors.

The results show that the random forest demonstrated 98.53% precision in predicting price changes, followed by KNN and logistic regression. This suggests that the random forest is the most suitable method for analyzing and predicting price fluctuations in ARV medications.

Additionally, we observed that the models yielded consistent results with conventional data analysis. We identified influential factors in procurement prices, such as medication type, quantity procured, and duration of consolidated purchasing strategy. We detected complex interactions among these variables, highlighting the importance of considering multiple dimensions when analyzing ARV drug prices.

Specifically, the logistic regression model indicated that the duration of consolidated purchasing strategy had a significant effect on reducing procurement prices. On the other hand, the random forest model underscored the importance of medication type in price determination. The KNN model identified clustering patterns among different ARV medications, suggesting the existence of complex relationships between them in terms of prices and characteristics.

6. Discussion

The implementation of a new consolidated purchasing scheme by the Mexican government for ARV medication procurement has shown significant results. Data reveal that this approach succeeded in generating substantial savings in the cost of these medications, contributing to an optimization in public health investment. These results are consistent with government reports reporting savings of 1.7 billion pesos by implementing the new procurement scheme (Ramírez, 2019).

A key factor in these savings relates to the inclusion of the drug Bictegravir, which has demonstrated greater efficacy and a broader resistance profile for viruses, as well as fewer adverse effects (DHHS, 2021). This inclusion allowed the Mexican government to achieve a more efficient and simplified ARV regimen for over 36,000 patients, resulting in increased coverage and optimization of ARV treatment in the country. This is supported by statements from those involved in the procurement process, who highlight the improvement in care and the ability to effectively meet demand.

Substantial savings in purchases, representing nearly a 60% decrease compared to previous prices, enabled the Mexican government to carry out an update to the basic care package and make state-of-the-art medications available at reduced costs. Furthermore, the implementation of the consolidated purchasing scheme has set a precedent that aligns with international best practices and positions Mexico as a leader in medication provision globally.

When comparing the results obtained in this research with government reports, we can observe that savings in the purchasing of ARV medications were consistent with price decreases recorded during the period when the new procurement scheme was implemented. However, it is important to highlight that the cause of these savings has not yet been precisely determined. While the integration of the drug Bictegravir was a key factor, further research is needed to quantify its coverage and evaluate its impact compared to medications procured in periods prior to 2019.

We will make an attempt to include comparisons with previous studies related to savings in ARV medications and procurement policies.

7. Conclusions

It should be noted that, despite the positive results, this study faced limitations related to the nature of the data and information sources. Discrepancies in expenditure numbers and the heterogeneity of data sources may have introduced some degree of bias into the results.

Given the complexity and variability observed in the prices of ARV drugs procured by the Mexican government, the decision has been made to postpone the construction of predictive models for other medications in this initial paper. This decision is based on the need to focus our efforts on understanding in depth the factors affecting the most representative medication before generalizing our findings to the rest of the medications. Additionally, this approach will allow us to validate our models and methodologies in a more controlled context before applying them to a wider range of drugs. Once we have established a solid framework and identified key patterns in the most representative medication, we can expand our analysis to other ARV medications, allowing us to obtain a more comprehensive understanding of procurement patterns and price determinants in the context of pharmaceutical purchasing by the Mexican government.

In summary, this research provides a deep understanding of consolidated purchasing of ARV medications in Mexico. Machine learning models proved valuable in predicting price changes, with significant economic savings observed through this procurement approach.

However, it is imperative to acknowledge the constraints when interpreting the outcomes owing to variations in data and information origins. These insights can facilitate well-informed decision-making in ARV medication management and the enhancement of HIV patient care.

References

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons.

Badal, Y.T. & Sungkur, R.K. Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and Information Technologies* 28, 3027–3057 (2023). https://doi.org/10.1007/s10639-022-11299-8

Chaumont, C. G. V., Bautista-Arredondo, S., José Calva, J., Isaac Bahena-González, R., Hitz Sánchez-Juárez, G., en Der, L., González de Araujo-Muriel, A., Magis-Rodríguez, C. & Hernández-Ávila, M.(2015). Antiretroviral purchasing and prescription practices in Mexico: constraints, challenges and opportunities. Salud Pública de M*éxico* 57 (2) 171–82. https://doi.org/10.21149/spm.v57s2.7606

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory,* 13(1), 21-27. https://doi.org/10.1109/TIT.1967.1053964

Department of Health and Human Services (DHHS). (2021). *Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV*. https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/adult-adolescent-arv/guidelines-adult-adolescent-arv.pdf

European Medicines Agency. (2018, June 22). *Biktarvy*. *Summary of Product Characteristics*. https://www.ema.europa.eu/en/medicines/human/EPAR/biktarvy

Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*(1), 3133–3181.

Fernando, L., & Pere, D. (2011). Tratamiento antirretroviral de la infección por el VIH. *Enfermedades Infecciosas y Microbiología Clínica*, *29*, 455–465. https://doi.org/10.1016/j.eimc.2011.02.009

Gobierno de México, Consejo de Salubridad General. (2017). Cuadro básico y catálogo de medicamentos. Issue 2017, 162–245. https://www.csg.gob.mx/descargas/pdf/priorizacion/cuadro-basico/med/catalogo/2017/EDICION_2017_MEDICAMENTOS-FINAL.pdf

Gobierno de México, Secretaría de Salud (SS). (2018). Informe nacional del monitoreo de compromisos y objetivos ampliados para poner fin al SIDA. (Informe GAM). https://www.gob.mx/censida/documentos/informe-nacional-del-monitoreo-de-compromisos-y-objetivos-ampliados-para-poner-fin-al-sida-informe-gam-mexico-2018?%20idiom=es

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017a). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas* *que Viven con VIH,* 3(1), 11-18. https://www.gob.mx/cms/uploads/attachment/file/226743/Boletin_Nal_CENSIDA_AT_IN_ene_mar_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017b). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 3(2), 9-16. https://www.gob.mx/cms/uploads/attachment/file/254117/Boletin_Nal_CENSIDA_AT_IN_abr_jun_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017c). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 3(3), 9-14. https://www.gob.mx/cms/uploads/attachment/file/299954/Boletin_Nal_CENSIDA_AT_IN_jul_sep_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017d). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 3(4), 11-18. https://www.gob.mx/cms/uploads/attachment/file/332699/Boletin_Nal_CENSIDA_AT_IN_oct_dic_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018a). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(1), 15-22. https://www.gob.mx/cms/uploads/attachment/file/332700/Boletin_Nal_CENSIDA_AT_IN_ene-mzo2018.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018b). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(2), 9-16. https://www.gob.mx/cms/uploads/attachment/file/370474/Boletin_Nal_CENSIDA_AT_IN_abr_jun_2018_1.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018c). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(3), 9-16. https://www.gob.mx/cms/uploads/attachment/file/420856/Boletin_Nal_CENSIDA_AT_IN_jul_sep_2018.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018d). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(4), 5-12. https://www.gob.mx/cms/uploads/attachment/file/441126/Boletin_Nal_CENSIDA_AT_IN_oct_dic_2018.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019a, July 19). *Ahorros producto de la estrategia de adquisición de antirretrovirales 2019*. https://www.gob.mx/cms/uploads/attachment/file/478447/ahorros_medicamentos.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019b). Datos Abiertos de la Compra Consolidada de Medicamentos ARV 2019-2020. Gobierno de México.

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019c). *Los nuevos mecanismos de optimización y mejora de los tratamientos antirretrovirales en México*. *5*(3), 4–7.

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019d). *Vigilancia Epidemiológica de casos de VIH/SIDA en México. Registro Nacional de Casos de SIDA Actualización al Cierre de 2019*.

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019e, November 11). *Antirretrovirales*. 2019. http://www.censida_2015.salud.gob.mx/contenidos/censida/antirretrovirales.html

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019f). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(1), 15-22. https://www.gob.mx/cms/uploads/attachment/file/473713/Bolet_n_DAI_1ER_TRIM_ENE_MZO2019_VF1.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019g). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(2), 12-19. https://www.gob.mx/cms/uploads/attachment/file/500787/BOLET_N_DAI_2DO_TRIM_ABRIL_JUNIO2019_VF_2.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019h). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(3), 13-21. https://www.gob.mx/cms/uploads/attachment/file/525270/Bolet_n_DAI_3er_trim._Vol.5_N.3_Jul-Sept_VF2.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019i). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(4), 16-22. https://www.gob.mx/cms/uploads/attachment/file/545616/Bolet_n_DAI_4to_trimestre_VF.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2020a). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 6(1), 20-26. https://www.gob.mx/cms/uploads/attachment/file/563143/Bolet_n_DAI_1er_trim_2020_V090720_f.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2021a, January 20). *Adquisición de Antirretrovirales Diciembre 2018*. 2021. http://www.censida_2015.salud.gob.mx/descargas/transparencia/ARV_diciembre_18.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2021b, January 20). *Compra de Antirretrovirales CENSIDA 2019*. 2021. http://www.censida_2015.salud.gob.mx/contenidos/censida/antirretrovirales.html

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2021c, March 31). *Boletín de Atención Integral de Personas con VIH/Censida*. 2021. https://www.gob.mx/censida/es/articulos/boletin-de-diagnostico-y-tratamiento-antirretroviral-censida?idiom=es

Gobierno de México, Centro Nacional para la Prevención y el Control del VIH y el sida (CONASIDA). (2018). *Boletín del Grupo de Información Sectorial en VIH, sida e ITS*. SS, Semar, Sedena, IMSS, ISSSTE, Pemex.

Hastie, T., Tibshirani, R. & Friedman, J. (2021a). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hastie, T., Tibshirani, R., James, G. & Witten, D. (2021b). An introduction to statistical learning. *Springer Texts in Statistics* (Vol. 102). https://doi.org/10.1007/978-1-4614-7138-7

Hosmer Jr, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied Logistic Regression.* John Wiley & Sons.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2020). *GLOBAL HIV STATISTICS*. https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2021). *HIV treatment*. https://www.unaids.org/es/topic/treatment

Liaw, A. & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

Lozano, F. & Domingo, P. (2011). Tratamiento antirretroviral de la infección por el VIHAntiretroviral therapy for HIV infection [Antiretroviral therapy for HIV infection]. *Enfermedades Infecciosas y Microbiología Clínica*, *29*(6), 455–465. doi.org/10.1016/j.eimc.2011.02.009

Marsland, S. (2014). *Random Forests. In Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC.

McKinney, W. (2013). Python for Data Analysis. O'Reilly Media, Inc.

National Institutes of Health (NIH). (n.d.). Aprobación de los medicamentos contra el VIH por la FDA. https://hivinfo.nih.gov/es/understanding-hiv/infographics/aprobacion-de-los-medicamentos-contra-el-vih-por-la-fda

Ramírez, M. (2019, July 23). Presume México su cambio de estrategia en VIH; #IAS2019. *El Economista*. https://www.eleconomista.com.mx/opinion/Presume-Mexico-su-cambio-de-estrategia-en-VIH-IAS2019-20190724-0007.html

Urzúa, C. (2020, January 20). El episodio de los antirretrovirales. *El Universal*.

World Health Organization (WHO). (2021a). *Global Price Reporting Mechanism*. https://apps.who.int/hiv/amds/price/hdd/

World Health Organization (WHO). (2021b). *HIV/AIDS*. https://www.who.int/health-topics/hiv-aids#tab=tab_1

About the authors

Blanca Iveth Mayorga Basurto is a professional in the field of information technology and data analysis, with a master's degree in Information Technology Management. She is currently in the process of completing her Doctorate in Financial Sciences at EGADE Business School, demonstrating a profound commitment to research and the application of advanced machine learning techniques in financial and public health contexts.

biveth@gmail.com
https://orcid.org/0000-0002-3643-3743?lang=en

Galo Moncada Freire works as a professor of finance and economics at the Tecnológico de Monterrey (ITESM) and at the Universidad Nacional Autónoma de México (UNAM). He holds a PhD in Economic Sciences from the Universidad Autónoma Metropolitana (UAM–Mexico), a master's in Business Economics from Tecnológico de Monterrey, and a degree in Economics from the Universidad Católica de Guayaquil. He also has consultation experience in administration, finance and economics and has worked in consultancy, banking and the commercial sector.

galomf@yahoo.com
https://orcid.org/0009-0007-4831-6599

The Impact of the Social Media Sentiment Index on S&P500 Returns

El índice de sentimiento en las redes sociales y su impacto en los rendimientos del S&P 500

Lizeth Gordillo Martínez

EGADE Business School, Tecnológico de Monterrey, Mexico

Received: April 1st, 2024.
Approved: May 30, 2024.

Abstract

The study of the development and analysis of sentiment indexes through social media is a recent technique that has captured interest because it can identify stock price tendencies. Also, using artificial intelligence to quickly analyze large volumes of data from various information sources has created a new way of evaluating massive amounts of information from social media. Natural language processing (NLP) is the preferred method for this research. Originating in the 1950s, NLP emerged at the junction of artificial intelligence and linguistics. Initially, it was used to retrieve information in text. It uses methods based on statistics to effectively index and search large sections of text.

Keywords: *Twitter, social media, natural language processing, sentiment analysis, artificial intelligence, news.*
JEL Classification: *G1, G02, C01.*

Resumen

El estudio de la construcción y el análisis de índices de sentimiento en redes sociales es una técnica reciente que ha captado interés por su capacidad para identificar tendencias en los precios de las acciones. Además, la aplicación de inteligencia artificial para analizar rápidamente grandes volúmenes de datos de diversas fuentes de información ha creado una nueva forma de evaluar información masiva de redes sociales. El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) es el método preferido que se sigue en la investigación. Originado en los años cincuenta, el NLP surgió de la intersección entre la inteligencia artificial y la lingüística. En un comienzo se empleó para recuperar información textual, con métodos basados en estadísticas para indexar y buscar de manera eficaz en grandes secciones de texto.

Palabras clave: *Twitter, redes sociales, procesamiento del lenguaje natural, análisis de sentimiento, inteligencia artificial, noticias.*
Clasificación JEL: *G1, G02, C01.*

1. Introduction

Stock markets operate on sentiment, where investors' herd mentality often results in emotional overreactions and investment choices. Contemporary social media plays a significant role in affecting sentiments and decisions in the stock market. Access to news and information is crucial for successful investment strategies, with social media serving as an auxiliary resource for decision-making, facilitating the rapid dissemination of financial news, sometimes even faster than traditional news tickers (Lugmayr, 2012).

The rapid expansion of digital data creation has ushered in the big data era, made increasingly significant as around 70% of the world's collected data originates from social media, sentiment analysis, identifying sentiments and emotions expressed about societal issues on social media (Lee et al., 2020).

We are quoting several cases where social media spread the word of relevant events that will affect financial markets significantly. Our financial markets are described as the S&P 500 Index and their returns according to these events. Social media and artificial intelligence have made this evaluation process more efficient, automated, and data transparent tracking to process bid data coming from Twitter.

In a recent paper, the Federal Reserve Board of Governors revealed that the Twitter financial sentiment index, which applies natural language processing to social media data, can assist in forecasting stock market returns for the following day. The study observed that sentiment declines following unexpected policy tightening, indicating that the data can also somewhat predict the extent of interest rate increases by the central bank (Smith, 2023).

Throughout 2021 and 2022, the Silicon Valley Bank (SVB) found itself with an abundance of bank deposits and not enough loans to disburse. Consequently, the bank was compelled to invest these customer deposits into long-term securities, primarily bonds. Typically, banks holding such securities aim to retain them until maturity to optimize returns. U.S. Treasury bonds, backed by the U.S. Government, carry a very low risk of default. However, Silicon Valley Bank's inadequate risk management practices left it unprepared for the Federal Reserve's rapid interest rate increases from under 1% to 4.75% within a year. This abrupt surge in interest rates resulted in a sharp decline in bond prices, highlighting the inverse relationship between interest rates and bond values. The hikes in interest rates unveiled mismatches in assets and liabilities at various regional banks across the U.S. (Manda, 2023).

Existing studies claim that investors and portfolio managers insight into how the Silicon Valley Bank collapse influenced various equity sectors, while regulators can gain an understanding of its broader effects. The role of social media in quickly triggering bank runs remains under investigation, as the U.S. has not seen a bank failure since 2008. Research works show that investors did not foresee the swift bank run, and also that the run did not uniformly affect all sectors of the U.S. market. This information is crucial for regulators attempting to address the risks posed by social media-driven “flash mob” reactions to the stability of banks. It is imperative for policymakers to closely watch social media channels to proactively identify and mitigate the risk of contagious financial panics (Yousaf & Goodell, 2023).

On April 14th, 2022, Elon Musk made his well-known bid to acquire Twitter for 43 billion dollars in a private offer. In response to this takeover bid, Twitter announced that it would conduct a thorough review. Through interviews and social media, Elon Musk expressed his intention to transform Twitter into a global free speech platform, emphasizing that the acquisition was not for business purposes or to boost visibility.

On April 25th, Twitter’s board unanimously and publicly agreed to the acquisition for 44 billion dollars. Following the completion of this transaction, Twitter would officially transition into a private entity (Jia & Xu, 2022).

Primarily, Twitter aimed to leverage financial backing from Elon Musk to enhance its platform, software development, and IT maintenance, thereby gaining a competitive edge over rivals like Facebook and Instagram. Additionally, Twitter’s founders viewed Musk’s acquisition as an opportunity to implement an exit strategy, allowing them to divest their shares to Musk and disengage from the company.

Elon Musk, on the other hand, envisioned transforming Twitter by making its algorithms open source, meaning the rationale behind tweet visibility to users would be publicly accessible. This approach contrasted with the current norms in social media, characterized by content restrictions and pre-validation requirements. Musk aspires to create a more open environment that champions free speech (Jia & Xu, 2022).

The background of this study begins with Tetlock (2007) who analyzes the interactions of the media, which in this case is the print media of *The Wall Street Journal* and the influence that these interactions have on the share price. It is identified that a low perception of the company expressed in the news precedes a high trading volume and a high perception of the company precedes a low trading volume. For this article, we will use autoregressive vectors.

For the methodology and the extraction of data and its analysis, we will be using daily data that was extracted with the Twitter API focused on the 2225 American companies corresponding to the SPX (Standard & Poors 500 Index). Once the information has been extracted, we will go on to identify the influence of negative news concerning positive news. An automated process was used for data extraction, calculating sentiment, creating time series, and estimating VAR. To determine a specific period, a rolling widow will be applied on day 1, day 2, day 3... until day 18 taking Elon Musk's event. In the case of 18 days before and after the event, we found a statistically significant impact of the compound index on the returns of the S&P 500.

Given that we want to identify if there is a bidirectional relationship between the sentiment Index and the S&P 500, our hypotheses will be the following:

H0: The composite sentiment has a negative effect and a statistical effect on the S&P 500 returns.

H1: The composite sentiment has a non-negative effect on the S&P 500 returns.

The study concludes by assessing the practical and theoretical implications of incorporating a sentiment index using artificial intelligence into the analysis of financial markets described by their main stock indices, as well as its limitations and potential challenges. This analysis contributes to understanding the complexity and dynamics of the markets through current technology and new media self-creation by society, providing investors, analysts, and regulators with a valuable tool for informed and strategic decision-making which includes a methodology that can be replicated in different media outlets.

2. Theoretical Framework

Before the advent of social media, the question of how the news affected the share price was already appearing in various sources of financial literature. Short-term stock prices could be predicted by financial news (Gidófalvi, 2001). Advances in technology developed a dynamic environment for creating sharing and collaborating among Internet users, allowing investors to access information on the performance of companies quickly (Li et al., 2014). News was available in real-time through various sources of information on the Internet and social media. Thomson Reuters News API was used to construct a daily series for the DJI Index's sentiment scores (Tikkanen, 2021). Unstructured data, such as news, could be analyzed with machine learning to identify causality through entropy transfer. Multiple kernel learning identified distinct types of information for the Korean market (Nam & Seong, 2019). Several disciplines document the impact of news on stock prices—computer science, statistics, economics, and finance. The reliability of Big Data's computational models in price predictions is very sensitive and can lead to profit losses in the industry analyzed (Shah et al., 2018). Various studies have questioned the hypothesis of efficient markets in different sectors and over different periods. All phenomena that have not been explained by this theory can be described by the adaptive markets' hypothesis and the fractal market's hypothesis (Núñez-Mora & Mendoza-Urdiales, 2023).

Text analysis is relatively new in the study of financial market behavior. Tetlock (2007) analyzed political forecasts and the relationship between analyst behavior and market outcomes. Schindler (2013) researched market psychology and the development of the Case-Shiller Index, used to measure house prices. Thaler (2015) developed behavioral economics and behavioral finance theory. O'Shaughnessy (2006) worked on quantitative investment strategies, including the consideration of behavioral factors.

Feuerriegel, Heitzmann, and Neumann (2015) analyzed oil prices to determine the relationship between these prices and terrorist attacks using the methodology of autoregressive vectors. Ruan et al. (2018) set out to determine the level of trust of users and identify whether there is a correlation with the results of financial markets. Derakhshan and Beigy (2019) used econometrics and artificial intelligence to demonstrate the relationship between news and price returns. Other people's opinions are an essential piece of information for decision-making. The use of the Internet has generated the availability of these opinions in large volumes on different topics, creating complexity due to contrasting opinions.

Broadstock and Zhang (2019) evaluated how Twitter influences share price. They assessed the stock's intraday returns and how they reacted to sentiment about the stock. This demonstrated how the share price is susceptible to price factors involving social media sentiment. Mendoza-Urdiales et al. (2022) used a database of more than 50 million tweets using artificial intelligence to obtain the information and analyze the results. In the same study, the polarity of news is highlighted, whether in a positive or a negative tone, giving greater weight to negative news in the reaction of the share price. It was observed that negative comments have a more relevant impact on the price. This study aims to leverage all accessible information sources that mention companies with recognizable tickers which can be automatically extracted to evaluate the sentiment of each text.

In another study**,** the authors indicate that sentiment has an impact on prices and there is an asymmetric effect—for example, negative news tends to have a stronger impact than positive news (Mendoza-Urdiales et al., 2022). One other problem mentioned is the growing volume managing these large databases and the challenge this involves. Investors must analyze the information within an extremely brief time frame. Several market players used Twitter to follow decision-makers and official company channels, and it became a popular source of information (Heston & Sinha, 2016). Steinert and Herff (2018) elaborated on the prediction of stock returns from the news.

There are also documented instances of false information affecting a stock's price. As a historical example, one of the most notorious instances of false information impacting the financial market was the cyberattack on the Associated Press (AP) news agency in 2013. In this incident, hackers compromised the AP's Twitter account, posted a fabricated tweet alleging explosions at the White House, and claimed that President Barack Obama had been injured. Although the tweet was quickly deleted and confirmed to be false, it caused a temporary dip in financial markets. The Dow Jones Industrial Average lost several points in a matter of minutes.

Silicon Valley Bank (SVB) was a pivotal financial institution in the startup and technology ecosystem, particularly in Silicon Valley. Founded in 1983, SVB's specialty was providing banking services to startups, technology companies, and the venture capital industry. Its unique focus on this niche allowed it to better understand the needs of its customers and offer financial products and services tailored to fast-growing, high-risk businesses.

SVB not only provided traditional banking services such as account management and loans, but also offered equity financing, investment services, and advice to companies in various phases of growth, ranging from nascent startups to more mature, publicly traded entities. In addition, his deep involvement with the community of investors and entrepreneurs allowed him to be a key player in the networking and business development network within the technology and innovation sector (Akhtaruzzaman et al., 2023).

2.1 The Bankruptcy of SVB in California, United States, in March 2023

SVB was an American commercial bank that failed due to a lack of deposits to cope with rising interest rates. Its strategy focused on financing technology startups that pose a greater risk because they do not yet have enough income. The authorities took control of them and initiated measures to cover depositors (Van Vo & Le, 2023). It was the biggest U.S. bank failure since 2008. The following was posted on Twitter: "Europe calls for calm in the face of the failure of Silicon Valley Bank. The Commission recalled that its presence in Europe is very limited, but they assured that they will be very vigilant (Radio 5, 2013, March 13)."

The news was not well-received by the media and the share price fell by 60%, the adverse effect on the share price corroborates the hypothesis that the negative effects of a news story have a greater effect than positive news. Pandey et al. (2023) also analyzed this.

3. Methodology

In this paper, we will use the methodology developed by Mendoza-Urdiales et al. (2022). Our sampling plan is as follows: The sentiment index was created by obtaining the historical tweets referring to each stock of 2,557 companies in the S&P 500 index. This index represents 85% of US equity markets. We extracted the English comments mentioning each company ticker in Twitter ($+Ticker) (Núñez-Mora & Mendoza-Urdiales, 2023). A daily time frame was set up. The time interval ran from October 3, 2022, to April 27, 2023. The Twitter API was used through Rstudio.

3.1 Data Extraction

The information was expressed in text, and the next step was to convert the text into variables. We describe the process sequentially:

a) Request Type. The *rtweet* package offers various request options based on the level of Twitter access granted. The most frequently used function is "search_tweets" which allows for the retrieval of tweets from up to seven days before extraction, with a limit of up to 500,000 tweets.

b) The identification of the ticker is done with '$+Ticker'.

c) Identify the maximum number of tweets—that is 500,000 tweets.

d) We determined that English was the language of extraction.

e) The way to check if we were extracting the entire tweet is through the code tweet_mode= 'extended' to make sure that we were getting all the information of the tweet which was up to 280 characters. Refinitiv Eikon API was used to extract the financial information of the companies. Refinitiv software uses its RIC code.

f) Selection of variables. We selected 2,557 American companies representing 85% of the capital markets and extracted all the comments in English identified by a ticker on Twitter ($+Ticker).

To create sentiment analysis, we needed a dependent variable, in this case, stock returns. The independent variable was the composite sentiment index created from tweets. With all the above, we ran several statistical analyses and converted the information into qualitative and quantitative data. The NLP algorithm was executed using Python, utilizing the TextBlob library, which fragments each text into words and assigns a numerical value to each word according to the categories of the libraries, one value for polarity and another for subjectivity. Here we used the composite sentiment index as an independent variable.

3.2 The Econometric Model

Vector autoregressive models (VAR) gained prominence in the field of econometrics through Sims's work in 1980. They serve as an extension of univariate autoregressive models to a broader context. VARs function as multi-equation regression models, incorporating multiple dependent variables, thus positioning themselves as a midpoint between singular time series models and concurrent equations models.

They have been frequently presented as a preferable option to the expansive simultaneous equation models used in structural analysis. In its most basic form, a VAR model can be bivariate, involving just two variables, y1t and y2t.

The present value of each variable is influenced by a unique mix of the past k values of both variables.

$$Y_{1t\,=}\beta_{10+}\beta_{11}\,Y_{1t-1+}\ldots{+}{+}\ldots{+}{+}\;\beta_{1k}\,Y_{1t-k}+\propto_{1t}Y_{2t-1}\,\alpha_{1k}\,Y_{2t-k}\,U_{1t}$$

$Y_{2t} = \beta_{20+}\beta_{21}\ Y_{2t-1+}\ \ldots{+}{+}\ldots{+}{+}\ \beta_{2k}\ Y_{2t-k} + \propto_{2t}\ Y_{2t-1}\ \alpha_{2k}\ Y_{1t-k}\ U_{2t}$

Another advantage of VAR models lies in the compactness of their notation. For instance, take the previously mentioned scenario where k equals 1, meaning each variable's current value relies solely on the past values of y1t and y2t, accompanied by an error term. This could be expressed as:

$Y_{1t} = \beta_{10+}\beta_{11}\ Y_{1t-1+} \propto_{1t} Y_{2t-1} + U_{1t}$

$Y_{2t} = \beta_{20+}\beta_{21}\ Y_{2t-1+} \propto_{2t} Y_{1t-1} + U_{2t}$

Autoregressive vectors (VARs) are used to identify relationships between variables and the bias that emerges from simultaneous equations if it is overlooked. We select autoregressive vectors to find the bidirectional relationship between two variables. We want to find the explanation for the possible existence of two-way causality between stock prices and the composite sentiment index.

The Granger representation theorem states that if there exists a dynamic linear model with stationary disturbances and the data are I(1), then the variables must be cointegrated of order (1,1) (Brooks, 2019).

3.3 The Granger Causality Test

This test is a statistical hypothesis test for determining whether a one-time series is useful in forecasting another. While the Granger causality test is not a test of true causality, it checks whether past values of one variable help predict the future value of another variable better than using past values of the target variable alone.

The test is typically applied to a pair of time series X and Y. Here is how the test is mathematically structured using a simple bivariate model.

Step 1: Model only with past values of the target variable.

First, we fit an autoregressive model (AR) to the target variable Y:

$$Y_t = \propto + \sum_{i=1}^{p} \beta_i\, \gamma_{t-i} + \epsilon_t$$

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.08

Where:

- α is a constant.
- β_i are the coefficients of lagged values of \(Y\).
- ε_t is the error term.
- ρ is the number of lags.

Step 2: Augmented model with past values of both variables.

Next, we fit an augmented model that includes past values of both \(Y\) and the potential causal variable X:

$$Y_t = \propto ' + \sum_{i=1}^{p} \beta'_i \gamma_{t-i} + \sum_{j=1}^{q} y_j x_{t-j} + \epsilon'_t$$

Where:

- $\alpha ' \beta '_i$ and y_i are coefficients.
- ε_t is the new error term.
- q can be equal to p or different, depending on the model selection.
- ρ is the number of lags.

3.4 Granger Causality Test

To determine if **X** Granger causes **Y**, we performed an F-test to compare these two models. The null hypothesis is that **X** does not Granger-cause **Y**, which implies that the coefficients y_j for all lags **j** are zero:

$$H_0: \gamma 1 = \gamma 2 = \ldots = \gamma q = 0$$

If the null hypothesis is rejected (meaning at least one of the y_i coefficients is significantly different from zero), then we conclude that **X** Granger causes **Y**.

This method checks if the inclusion of the past values of **X** significantly improves the model's ability to predict **Y**, compared to the model that only includes past values of **Y**. It is essential to note that if **X** Granger causes **Y**, it does not imply a causal relationship in a traditional sense, but rather a predictive directional relationship (Brooks, 2019).

4. Results

The Autoregressive Vector results are presented in Table 1, showing that the sentiment variable (named compound) has an almost immediate negative effect on the SPX index performance. Considering a 95% confidence interval with p-value ≤ 0.05, the statistically significant coefficient is only present within a day (Lag 1), and the effect on performance is negative (as expected).

Moreover, neither variable influences the sentiment equation, meaning that there is no feedback from the sentiment itself or from its performance.

As for the result of the influence of the SPX index performance on itself, the results show intermittence (considering a 95% confidence interval with p-value ≤ 0.05), the statistically significant coefficients are present within a day (Lag 1) and again on day 8th (Lag 8), and the effect on the performance is negative (see Table 1).

This result concludes that there is a negative sentiment in the market around the date of Twitter's acquisition. As we know, this acquisition involved highly criticized actions from Elon Musk, generating uncertainty and, consequently, a negative sentiment.

5. The Model's Results

As a key date, we selected October 28, 2022, when Elon Musk bought Twitter. Our results came out positive with the variation of the composite index. When you see a positive return, it already hits the composite index positively. The result we got is that sentiment is positively affected, and lagging positive sentiment is observed. In data analysis, statistics, and financial modeling, a rolling window is a method where computations are conducted on a set number of sequential data points. After each calculation, the window moves forward one or more periods to incorporate new data. This technique is employed to evaluate time series data, allowing analysts to detect trends, assess variability, and conduct various other analyses on a dataset that is continually updated. We use 18 days on a rolling window testing daily, day 1, day 2, day 3, day 4, until day 18 (Brooks, 2019).

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.08

Table 1. Regression Results for SPX Returns with the statistically Significant Variables Highlighted (p-value ≤ 0.05)

Summary of Regression Results

Model:	VAR
Method:	OLS
Date:	Wed, 06, Mar, 2024
Time:	13:18:58

No. of Equations:	2.00000	BIC:	-7.52723
Nobs:	30.0000	HQIC:	-8.60723
Log likelihood:	85.5925	FPE:	0.000149006
AIC:	-9.11525	Det(Omega_mle):	6.07088e-05

Results for equation SPX Returns

	coefficient	std. error	t-stat	prob
const	0.003549	0.001573	2.256	0.024
L1.SPX Returns	-0.432398	0.210097	-2.058	0.040
L1.Compound	-0.003140	0.001554	-2.021	0.043
L2.SPX Returns	-0.104791	0.236009	-0.444	0.657
L2.Compound	0.000025	0.001650	0.015	0.988
L3.SPX Returns	-0.084670	0.223500	-0.379	0.705
L3.Compound	-0.000549	0.001663	-0.330	0.741
L4.SPX Returns	0.117132	0.217736	0.538	0.591
L4.Compound	-0.002552	0.001694	-1.507	0.132
L5.SPX Returns	0.244943	0.209261	1.171	0.242
L5.Compound	0.001060	0.001642	0.646	0.519
L6.SPX Returns	-0.140113	0.206786	-0.678	0.498
L6.Compound	0.001319	0.001746	0.756	0.450
L7.SPX Returns	-0.341930	0.210242	-1.626	0.104
L7.Compound	0.001889	0.001757	1.075	0.282
L8.SPX Returns	-0.418846	0.199433	-2.100	0.036
L8.Compound	0.001559	0.001517	1.027	0.304

Results for equation Compound

	coefficient	std. error	t-stat	prob
const	0.054744	0.279006	0.196	0.844
L1.SPX Returns	44.207016	37.263753	1.186	0.235
L1.Compound	0.141436	0.275564	0.513	0.608
L2.SPX Returns	-13.429302	41.859507	-0.321	0.748
L2.Compound	-0.114001	0.292632	-0.390	0.697
L3.SPX Returns	33.155597	39.640905	0.836	0.403
L3.Compound	-0.008319	0.294945	-0.028	0.977
L4.SPX Returns	8.745297	38.618612	0.226	0.821
L4.Compound	0.353053	0.300419	1.175	0.240
L5.SPX Returns	8.416371	37.115352	0.227	0.821
L5.Compound	0.096010	0.291294	0.330	0.742
L6.SPX Returns	8.994378	36.676488	0.245	0.806
L6.Compound	-0.036604	0.309656	-0.118	0.906
L7.SPX Returns	-36.473068	37.289423	-0.978	0.328
L7.Compound	0.300948	0.311576	0.966	0.334
L8.SPX Returns	-24.754087	35.372264	-0.700	0.484
L8.Compound	-0.218066	0.269132	-0.810	0.418

Correlation matrix of residuals

	SPX Returns	Compound
SPX Returns	1.000000	0.059417
Compound	0.059417	1.000000

Source: Prepared by the autor.

Granger causality Wald-test. H_0: Compound does not Granger-cause SPX Returns.

Conclusion: fail to reject H_0 at a 5% significance level.

The composite sentiment index is significant at 5% and negatively affects the performance of the Standard & Poor's Index with parameter -0.003140. What we are saying is that around the event on October 28, 2022, we selected 18 days before and 18 days after. Remember that the composite index is the sum of the positive, the negative, and the neutral sentiment. What dominated this period was negative sentiment and that hit the performance negatively.

In the S&P 500 performance equation, the influential lag is L1 which is the lag of the composite sentiment index at -0.003140.

Performance is negatively influenced by the composite sentiment index which in turn is influenced by lags in the performance of the S&P 500.

Granger causality simply indicates a correlation between the current value of one variable and the past values of others; it does not imply that changes in one variable directly cause changes in another (Brooks, 2019).

There is no Granger causality in the sense of the composite index to the performance of the S&P 500, i.e., the composite does not cause the performance of the S&P 500 in the Granger sense.

5.1 Impact of the Composite Sentiment Index on S&P 500 Returns

The findings validate the distinct connection between the composite sentiment index and the S&P 500 returns index, showing significant influence at the 5% level and a negative impact on the S&P 500 index's performance. The composite index affects the S&P stock index's performance with a -.003 impact.

The test concludes that you cannot reject the null hypothesis. Given that we cannot reject it, then the composite index does not, in Granger's sense, cause the return.

6. Discussion

X (formerly Twitter), as a frontrunner among social media platforms, is a constant presence in people's daily lives and has become a vital component of digital communication, but the emergence of short video-centric platforms, like TikTok, changed the game rules. This global social media boasted a sizeable user base of approximately 319 million people in 2022. It facilitates the widespread distribution of information, underpinning the value of enabling instant and unrestricted creation and sharing of ideas and information by all users. Additionally, it provides access to trending topics and has a considerable influence (Jia & Xu, 2022).

X (formerly Twitter) primarily generates its revenue through advertising, operating as an advertising-centric company. It is likely that Elon Musk recognized the potential in this network advertising sector and aimed at enhancing his presence in the digital advertising realm, seeing an opportunity to profit from acquiring the platform. Additionally, Musk might have viewed the ownership of Twitter as a means to shape public opinion more effectively, furthering his business and political agendas (Jia et al., 2023).

Observing X (formerly Twitter), it is evident that the emergence of short video-centric social media platforms has disrupted the growth trajectory of traditional social media. The rate of increase in this network daily active users has been on a downward trend in recent years, plummeting from a peak of 34% to 11%. Conversely, new social media platforms, with TikTok at the forefront, are steadily chipping away at the finite market share. X (formerly Twitter) has reported a profit decrease since 2019 (Jia et al., 2023).

Following Elon Musk's announcement of his intention to buy Twitter, the move garnered significant interest from various societal sectors. Musk, known for his unconventional approach, is a key opinion leader and, with this acquisition, he places himself at the forefront of the social media sector that moves the financial market. He has political and personal branding motivations behind this acquisition.

The event we determined as relevant was the purchase of Twitter by Elon Musk, the CEO of Tesla and SpaceX. This was one of the most controversial acquisitions in recent memory. In April 2022, Musk revealed that he had acquired a significant stake in Twitter, becoming one of its largest individual shareholders. On April 14, 2022, Musk offered to buy Twitter for $44 billion, proposing a price of $54.20 per share. Musk argued that his interest in buying it was his desire to ensure free speech on the platform to ensure the functioning of a democracy.

According to Musk, Twitter had the potential to be the quintessential free speech platform, but he believed it did not live up to that potential under its current administration. Initially, Twitter's board of directors responded to Musk's offer by implementing a measure to make it harder for the company to a hostile takeover by allowing other shareholders to buy more shares at a reduced price. After further negotiations, the board accepted Musk's offer on April 25, 2022.

The acquisition faced several hurdles and controversies. Over the next few months, Musk expressed concerns about the number of fake and spam accounts on the platform, leading to a temporary pause on the deal as he sought more information.

After a legal battle and the possibility of a trial, Elon Musk proceeded with the purchase of Twitter, closing the deal in October 2022.

The S&P 500 captures the wider market sentiment concerning the leading publicly traded companies sourced from the world's most extensive focus group. The connection between S&P 500 returns and Twitter acquisition can be reflected in our sentiment index: it shows a relevant impact on negative news compared with positive comments. This opens a new discussion: Why does negative news have a greater relevance on price stock than positive comments?

An increasing amount of research highlights the human inclination to favor negative news over positive. But what drives this preference? Stuart Soroka proposes that this bias might stem from neurological or physiological predispositions, as the potential risks associated with negative information greatly exceed the potential advantages of positive information (Soroka, 2015).

However, sharing news content on social media differs fundamentally from choosing and reading articles. A study mentioned by Arianna Huffington, cited in Soroka (2015), acknowledges this distinction. While we might be more inclined to share positive content on social media, our tendencies when it comes to reading news still lean towards prioritizing negative information. Ultimately, there is an expanding pool of evidence that underscores humans' tendency to favor negative over positive news content (Soroka, 2015).

7. Conclusions

This study underscores the innovative use of social media sentiment, particularly X (formerly Twitter) data, to assess its influence on stock prices, highlighting the transformative role of artificial intelligence in analyzing vast volumes of data from diverse sources rapidly. This approach enables investors to identify positive or negative market signals more efficiently. The research pivots around Twitter's role in the current information ecosystem, emphasizing the evolving nature of news analysis in the digital age, influenced by the Internet and social media. By leveraging Twitter data to construct a sentiment index and applying an autoregressive vector (VAR) model, the study explores the bidirectional relationship between this sentiment index and the S&P 500 index returns, through the Granger causality.

The introduction of news sentiment from social media into the analysis is novel in two main respects: showcasing the impact of social commentary on stock performance and employing natural processing language (NPL) for the automated text analysis. The specific case of Elon Musk's acquisition of Twitter and its impact on the S&P 500 is examined, aiming to reveal the relationship between market returns and aggregate sentiment. This research contributes to the understanding of market dynamics. An example is the announcement of Silicon Valley Bank's financial problems. This bank went bankrupt due to poor performance of its risk management strategy. In this sense, the methodology in this paper offers valuable insights for investors and regulators by presenting a replicable methodology for sentiment analysis across different media outlets. This can be used by investors analyzing market sentiment to detect movements in the value of assets that can be generated from a specific event. An application for public policy can be made when an abnormal event appears and the bank needs to be rescued, regulations should be put in place. It can be a part of bankruptcy prediction tools (Kliestik et al., 2018).

Key findings include the identification of a significant, negative correlation between the composite sentiment index and the S&P 500's performance, highlighting the predominance of negative sentiment in influencing market behavior around specific events. The absence of Granger causality between the sentiment index and S&P 500 returns suggests complex dynamics in how sentiment reflects or predicts market movements.

After applying the methodology, we accept H0—the composite sentiment has a negative effect on S&P 500 returns.

The discussion reflects on Twitter's changing role amid the rise of short video platforms like TikTok, noting Twitter's decline in user growth and its implications for the platform's advertising revenue and market position. Musk's acquisition is contextualized within this shifting landscape, suggesting his motives may extend beyond financial gain to include influencing public opinion and leveraging Twitter for political and personal branding purposes.

On the one hand, a limitation of this research is that the autoregressive vector technique remains a linear technique and does not capture complexity when there is no linearity. On the other hand, the absence of the Granger causality, which is a concept that refers to linearity, allows us to consider other types of models for future studies regarding prediction.

The study also contemplates the broader market sentiment captured by the S&P 500 regarding Twitter's acquisition, opening discussions on the disproportionate impact of negative versus positive news on stock prices. This observation aligns with broader psychological research indicating a human predisposition towards negative information, a trend that persists in social media engagement and news consumption patterns.

References

Akhtaruzzaman, M., Boubaker, S., & Goodell, J. W. (2023). Did the collapse of Silicon Valley Bank catalyze financial contagion? *Finance Research Letters*, 56, 104082. https://doi.org/10.1016/j.frl.2023.104082

Broadstock, D. C., & Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters*, *30*, 116-123. https://www.sciencedirect.com/science/article/pii/S1544612318307888

Brooks, C. (2019). *Introductory econometrics for finance*. Cambridge University Press.

Derakhshan, A., & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, 85, 569-578. https://www.sciencedirect.com/science/article/pii/S0952197619301666

Feuerriegel, S., Heitzmann, S. F., & Neumann, D. (2015, January). Do investors read too much into news? How news sentiment causes price formation. In the *48th Hawaii International Conference on System Sciences*, IEEExplore, 4803-4812. https://ieeexplore.ieee.org/abstract/document/7070391

Gidófalvi, G. (2001). Using news articles to predict stock price movements. Department of Computer Science and Engineering, University of California, San Diego. https://people.kth.se/~gyozo/docs/financial-prediction.pdf

Heston, S. L., & Sinha, N. R. (2016). News versus Sentiment: Predicting Stock Returns from News Stories. (FEDS Working Paper No. 2016-48). *Finance and Economics Discussion Series*, 1–35. https://doi.org/10.17016/feds.2016.048

Jia, J., Pan, H., & Su, J. (2023). Analysis of Motivations, Process, and Implications of Elon Musk's Acquisition of Twitter. *BCP Business & Management*, *47*, 145-153. https://doi.org/10.54691/bcpbm.v47i.5185

Jia, Q., & Xu, S. (2022). An Overall Analysis of Twitter and Elon Musk M&A Deal. *Highlights in Business, Economics and Management*, *2*, 436-441. https://doi.org/10.54097/hbem.v2i.2399

Kliestik, T., Misankova, M., Valaskova, K., & Svabova, L. (2018). Bankruptcy prevention: New effort to reflect on legal and social changes. *Science and Engineering Ethics*, *24*, 791-803. https://doi.org/10.1007/s11948-017-9912-4

Lee, H., Lee, N., Seo, H., & Song, M. (2020). Developing a supervised learning-based social media business sentiment index. *The Journal of Supercomputing*, 76, 3882-3897. https://doi.org/10.1007/s11227-018-02737-x

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826-840. https://doi.org/10.1016/j.ins.2014.03.096

Lugmayr, A. (2012). Predicting the future of investor sentiment with social media in stock exchange investments: A basic framework for the DAX performance index. In M. Friedrichsen, & W. Mühl-Benninghaus (Eds.), *Handbook of social media management: value chain and business models in changing media markets* (pp. 565-589). Springer. https://doi.org/10.1007/978-3-642-28897-5_33

Manda, V. K. (2023). The Collapse of Silicon Valley Bank. *MAR-Ekonomi: Jurnal Manajemen, Akuntansi Dan Rumpun Ilmu Ekonomi*, *2*(1), 59-70. https://jurnal.seaninstitute.or.id/index.php/marekonomi/article/view/232

Mendoza-Urdiales, R. A., Núñez-Mora, J. A., Santillán-Salgado, R. J., & Valencia-Herrera, H. (2022). Twitter sentiment analysis and influence on stock performance using transfer entropy and EGARCH methods. *Entropy*, *24*(7), 874. https://doi.org/10.3390/e24070874

Nam, K., & Seong, N. (2019). Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market. *Decision Support Systems*, *117*, 100-112. https://doi.org/10.1016/j.dss.2018.11.004

Núñez-Mora, J. A., & Mendoza-Urdiales, R. A. (2023). Social sentiment and impact in US equity market: an automated approach. *Social Network Analysis and Mining*, *13*(1), 111. https://doi.org/10.1007/s13278-023-01116-6

O'Shaughnessy, J. P. (2006). *Predicting the markets of tomorrow: A contrarian investment strategy for the next twenty years*. Penguin. https://shorturl.at/hxHIV

Pandey, D. K., Hassan, M. K., Kumari, V., & Hasan, R. (2023). Repercussions of the Silicon Valley Bank collapse on global stock markets. *Finance Research Letters*, *55*, 104013. https://doi.org/10.1016/j.frl.2023.104013

Radio 5 [@radio5_rne]. (2023, March 13). "Europa llama a la tranquilidad frente a la quiebra del Silicon Valley Bank. La Comisión recuerda que su presencia en Europa es muy limitada, pero aseguran que estarán muy pendientes." [Tweet] X (formerly Tweeter). https://twitter.com/radio5_rne/status/1635251865624674304

Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). Using Twitter trust network for stock market analysis. *Knowledge-Based Systems*, 145, 207-218. https://www.sciencedirect.com/science/article/pii/S0950705118300248

Schindler, F. (2013). Predictability and persistence of the price movements of the S&P/Case-Shiller house price indices. *The Journal of Real Estate Finance and Economics*, *46*(1), 44-90. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2192128

Shah, D., Isah, H., & Zulkernine, F. (2018, December 10–13). Predicting the effects of news sentiments on the stock market. In *2018 IEEE International Conference on Big Data*, Seattle, WA, United States. https://doi.org/10.1109/BigData.2018.8621884

Smith, Molly (2023; May 23) Fed Tracks Market Sentiment With New Index Built from 4.4 Million Tweets. *Bloomberg*. https://www.bloomberg.com/news/articles/2023-05-23/fed-index-built-from-4-4-million-tweets-predicts-size-of-hikes?embedded-checkout=true

Soroka, S. (2015; May 25). *Why do we pay more attention to negative news than to positive news? British Politics and Policy at LSE* [blog entry]. https://eprints.lse.ac.uk/62222/

Steinert, L., & Herff, C. (2018). Predicting altcoin returns using social media. *Plos one*, 13(12), article 0208119. https://doi.org/10.1371/journal.pone.0208119

Tikkanen, M. (2021). Predicting the FTSE All-Share index daily close-to-close price direction using sentiment analysis on tweets from UK. LUT University. https://lutpub.lut.fi/handle/10024/163481

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139-1168. https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2007.01232.x

Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. W.W. Norton & Company. https://shorturl.at/adC17

Van Vo, L., & Le, H. T. T. (2023). From Hero to Zero: The Case of Silicon Valley Bank. *Journal of Economics and Business*, *127*, article 106138. https://doi.org/10.1016/j.jeconbus.2023.106138

Yousaf, I., & Goodell, J. W. (2023). Responses of US equity market sectors to the Silicon Valley Bank implosion. *Finance Research Letters*, *55*, article 103934. https://doi.org/10.1016/j.frl.2023.103934

About the author

Lizeth Gordillo Martínez began her professional career as coordinator for the Center for Technology and Financial Innovation while a BA student in international trade at the Tecnológico de Monterrey, Mexico City Campus. Gordillo Martínez conducted training workshops on databases specializing in finance and technology in various companies such as Bloomberg, Reuters, Economática, SaS, Eviews, and Numerix, and taught financial administration to 14 generations of students at the undergraduate level. Starting in 2008, she worked full-time leading negotiations with clients from the financial sector in the Latin America region at multinational companies with a technological and financial profile like Bloomberg, Thomson Reuters, Numerix, and Identy. Among her professional achievements are the creation of financial labs for private universities with Bloomberg and Reuters terminals, and the implementation of a Thomson Reuters ticker with financial indicators at the Mexican Stock Exchange (BMV) building. Gordillo Martínez secured Grupo Bancolombia as the first client—and first international bank—for Numerix in Latin America. There, she implemented a CVA calculation module for the bank's treasury. At the same time, she continued with her graduate studies, earning a master's degree in Finance, and in 2021 she started a PhD in financial sciences.

gordillo.lizeth@gmail.com
https://orcid.org/0009-0000-9210-4955

Causality Study on Financial Inclusion Issues with Data Science Techniques: The Mexican Case

Estudio de causalidad sobre problemas de inclusión financiera con técnicas de ciencia de datos: el caso de México

Itzel Coquis Rioja

EGADE Business School, Tecnológico de Monterrey, Mexico

Mario Iván Contreras Valdez

Tecnológico de Monterrey, Mexico

Received: March 1, 2024.
Approved: May 10, 2024.

Abstract

The current article explores the causes of financial inclusion among the Mexican population. It leverages data from the Encuesta Nacional de Inclusión Financiera (ENIF) (INEGI, 2021) to develop two machine learning models aimed at identifying individuals who are part of the financial system. These models are assessed using both artificial intelligence methodologies and traditional statistical significance tests. The findings suggest that factors such as education level, monthly income, future-oriented behavioral preferences over present ones, saving capacity, and access to smartphones are significant drivers that enhance the likelihood of financial inclusion. Consequently, there is a potential for implementing public policies to incentivize individuals to voluntarily adopt formal financial services.

Keywords: *financial inclusion, artificial intelligence, machine learning.*
JEL Classification: *C13, C54, D14.*

Resumen

El presente artículo explora las causas de la inclusión financiera entre la población mexicana. Con datos de la Encuesta Nacional de Inclusión Financiera (ENIF) (INEGI, 2021) desarrolla dos modelos de aprendizaje automático con el objetivo de identificar a individuos que forman parte del sistema financiero. Estos modelos son evaluados valiéndose tanto de metodologías de inteligencia artificial como de pruebas estadísticas de significancia tradicionales. Los hallazgos sugieren que factores como nivel educativo, ingreso mensual, preferencias orientadas hacia el futuro sobre las presentes, capacidad de ahorro y acceso a teléfonos inteligentes son impulsores significativos que aumentan la probabilidad de inclusión financiera. En consecuencia, existe un potencial para la implementación de políticas públicas dirigidas a incentivar a los individuos para que adopten voluntariamente servicios financieros formales.

Palabras clave: *inclusión financiera, inteligencia artificial, aprendizaje automático.*
Clasificación JEL: *C13, C54, D14.*

1. Introduction

Throughout human history, technological advancements have catalyzed numerous social changes. Initially, these innovations often impacted only small groups of people; however, as democratization progresses, their benefits become increasingly accessible to broader populations. In the realm of economic and financial interactions, such advancements have led to new modes of transactional engagement. One such innovation is financial inclusion. According to basic economic theory (Varian, 1992), uncertainty detrimentally affects individuals' utility. In such circumstances, the ability to safely transfer wealth across time periods becomes a valuable tool for enhancing societal welfare. Financial markets and institutions facilitate this temporal decision-making process. Despite these advantages and available technologies, many countries have yet to fully exploit them. Demirguc-Kunt and Klapper (2012) define financial inclusion as the capacity of individuals to access financial products and services that meet their needs sustainably. Building upon this definition, Tram et al. (2023) highlight three primary dimensions that determine financial inclusion: penetration, availability, and utilization of financial services.

Given the significance of financial services and markets in modern economies, financial inclusion has emerged as a primary focus for governments worldwide (Sun & Wang, 2023). Consequently, there has been a surge in statistical research aimed at addressing this imperative. The abundance of information on this issue has facilitated the development of various methodologies; many of these operate under the premise that accurate measurement is essential for assessing the efficacy of public policies aimed at improvement. However, it is worth noting that the type of information gathered through surveys may have varying potential in different scenarios.

On this note, the causality behind an individual's consideration for financial inclusion has predominantly been approached from a theoretical and qualitative standpoint. Many discussions on enhancing financial inclusion primarily focus on customer preferences and income levels. However, given the multidimensional nature of this issue, it is important to consider other factors or features that may influence individuals' decisions regarding whether to access the financial system or not.

The present document outlines a machine learning procedure conducted on the Encuesta Nacional de Inclusión Financiera (ENIF, or National Financial Inclusion Survey) (INEGI, 2021) dataset, focusing on financial inclusion in Mexico during

2021. The aim is to characterize individuals falling within the broader definition of financial inclusion. By identifying the a priori conditions for inclusion in this dataset, policymakers can devise more effective public policies to enhance the likelihood of individuals belonging to the financial system. This research endeavors to transition from qualitative and theoretical perspectives to statistical and quantitative methodologies; through these, the impact of different variables can be evaluated to incentivize financial inclusion effectively. The importance of this research lies in its utilization of classification artificial intelligence models. These models yield two crucial outcomes: firstly, they identify the variables that contribute to the conditional probability of individuals being considered in financial inclusion. Secondly, they uncover nonlinear relationships between features and the response variable. Furthermore, the results of this research can be utilized by governments to implement public policies aimed at increasing the likelihood of individuals accessing the financial system.

To structure the present paper, Section 2 develops the literature review; Section 3 presents the data and methodology to evaluate the relationship; Section 4 presents the main results of these models and develops the interpretation and analysis of them; finally, Section 5 sets out the conclusions, and the last section includes the references.

2. Literature Review

Financial inclusion is a term that can be traced back several decades. The Community Reinvestment Act (1977) was enacted by the US to regulate banks, preventing them from focusing only on rich districts and obliging them to provide financial services, regardless of the population's income level. In this regard, Leyshon and Thrift (1995) understand financial exclusion to be all those mechanisms that serve to prevent certain social groups from entering the financial system through the availability of financial instruments. Further refinements to the definition encompass the specification that it prevents poor or unprotected groups from gaining access to the formal financial system or to regulated financial instruments in a low-cost and safe environment, where credit might be the instrument most required (Conroy, 2005; Mohan, 2006; Rangarajan Committee, 2008; Marín & Schwabe, 2018).

Its importance in incentivizing other economic variables is closely connected with financial inclusion. One of the earliest works in this regard was done by the economist

Schumpeter (1911), in a text where the author shows that the financial sector has the potential to boost economic growth and economic development. The primary function by which it operates is the ability of the banking system to facilitate capital accumulation, thereby enabling firms to utilize the capital in the production process. Consequently, this leads to an increase in the output and productivity of productive factors, ultimately enhancing the purchasing power of households. (King & Levine, 1993; Beck et al., 2007). However, those initial analyses only addressed innovation within the financial system, which is also known as financial development. In this regard, Johnson and Arnold (2014) note that public policy has undergone a shift in its priorities, from Financial Development to Financial Inclusion. In this case, focusing efforts not only on the strength of the banking system and the free flow of capital is a first step, but the goal nowadays has also turned towards enabling the economically vulnerable population to have access to this financial system.

In the case of Mexico, Segovia and Cepeda (2024) use panel data to measure the level of financial inclusion in different states from the periods of 2005 to 2018. Their findings show that the increase in banking credit for non-financial firms has the potential to increase GDP per capita. In this instance, the Region variable seems to be a proper indicator for capturing the effect described in this document. In this same vein, De la Cruz and Alcántara (2011) provide statistical evidence for the positive relationship between the development of the private banking system and economic development. A particular experience of this relationship is described in Bruhn and Love (2014), where the incursion of Banco Azteca in the banking system via its Elektra stores shows a shift in poverty reduction because of the incorporation of financial and commercial infrastructure, making clear the importance of this element in the financial inclusion of Mexican society.

The growing importance of financial inclusion has generated another issue in relation to the definition and subsequent measurement of the variable. Huong et al. (2023) expose this problem as a global phenomenon, where the debate is centered on a universally accepted term; however, the shared understanding that can be reached is that financial inclusion entails access to formal financial services at an affordable cost. In this regard, the authors also note that currently there is a lack of an adequate method for assessing financial inclusion. The main problem arises from the fact that proposed indexes have arbitrary weights for the characteristics of people who are considered to have financial inclusion.

In view of this, Sarma (2015) developed a financial inclusion index for 81 countries using a number of economic indicators that worked as a reference point; further research has been extended using the same procedure for 98 more countries (Arora, 2014). For a local case, Dircio-Palacios Macedo et al. (2023) proposed the construction of an index for the municipalities of Mexico. Their initial results indicate that there are variations among municipalities, making it challenging to implement a universal index.

Another perspective on this same logic refers to the concept of digital financial inclusion (DFI). Xi and Wang (2023) provide a number of outcomes concerning the influence of DFI on economic growth. Their conclusions highlight certain attributes that could serve as a more concise definition since technological applications might expose unsuspecting individuals to scams and fraud in electronic mediums (Yue et al., 2022). Because of this, Schuetz and Venkatesh (2020) argue that one of the main obstacles to reaching financial inclusion, along with geographical access and the high relative cost of certain services, is the presence of financial illiteracy. All of these sources remark on the need not only for infrastructure or technology but also for the understanding of personal finance.

So far, the impacts of financial inclusion and some of the problems in measuring them can be analyzed; however, from a complete perspective, the factors contributing to people's involvement should also be considered. In that respect, Cruz-García et al. (2020) developed a model to identify the probability of a given municipality being considered to offer financial inclusion. Their results show that geography might play a major role in the determination of financial inclusion, as certain statistics seem to have proven their ability to predict the probability. Jamil et al. (2024) present a complementary analysis of the impact that age has on this topic. Their hypothesis starts from the fact that the older population who have retired from the labor market might need to have a strong understanding of their financial situation in order to live mostly from their savings. Therefore, this characteristic within the population could serve as an indicator of whether people have knowledge about this topic or not.

3. Data and Methodology

The present study uses the Encuesta Nacional de Inclusión Financiera (ENIF) (INEGI, 2021), which has been conducted in Mexico every three years since 2012, when the Instituto Nacional de Estadística y Geografía (INEGI, Statistics and Geography National Institute) and the Comisión Nacional Bancaria y de Valores (CNBV, National Baking and Stock Commission) joined forces to obtain relevant statistics on financial inclusion. The purpose of these exercises is to identify the main challenges to financial inclusion in the country; on this basis, public policies that are the responsibility of the Consejo Nacional de Inclusión Financiera (CONAIF, National Council for Financial Inclusion) could be implemented in future proposals.

The current data comes from the latest survey conducted in 2021, with information compiled from June 28 to August 13. Using a confidence level of 90%, the sample was defined to be local and nationally representative. The sample size consisted of 13,554 people, representing 90,328,320 people over 18 years of age. The questionnaire was separated and compiled into four modules with different aggregate levels of the households. The current study is developed based on the TModulo file, which contains the answers from the individual questionnaires completed in the survey. The data consists of a rectangular grid of 13,554 rows and 382 columns with the variables and identifiers.

To conduct causality modeling, two artificial intelligence models are introduced to analyze the categorical variable indicating whether an individual demonstrates financial inclusion or not. The aim of this approach is to discern the directional impact of individual features that influence the likelihood of belonging to the financial system. The outcomes of these models will enable the identification of exogenous variables that drive individuals to be part of the financial system. Such insights can then be leveraged by policymakers to enhance financial inclusion indicators.

The first model we propose is logistic regression with categorical exogenous variables. Originally introduced by Berkson (1944) to investigate the impact of certain drugs on patient survival, logistic regression has evolved to become a cornerstone of artificial intelligence models. It serves as a bridge between black-box classification models and traditional inference statistics. The logistic function, which is central to this model, can be analytically expressed as follows:

$$F(z_i) = \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{1 + e^{-z_i}}$$

Where *e* is the exponential number and *F* represents the cumulative logistic function. In sum, using this model allows us to compute the probability P_i for the observation $Y_i = 1$ where the value of Z_i represents a linear function; that is:

$$P_i = \frac{1}{1 + e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}}}$$

Moreover, since some variables in the survey are categorical, with responses being either Boolean or categorical, we incorporate transformations into dummy variables. This allows for comparisons with individual responses from the questionnaire. Conceptually, the model constructed resembles parallel lines containing several categories in the result. In consequence, the estimates derived from logistic regression can be interpreted as the probability differences between individual features.

To further enhance this analysis, a tree-based model is proposed to complement the causality analysis. Such models aim to segment the response space of observations, enabling a discriminant function to allocate observations based on their features. In this context, the decision tree will be a classification model, given that the predictor consists of dummy variables.

The method for adjusting such a tree involves binary splitting with the features of the observation. Each time a bifurcation occurs, a node is created with two branches that can further expand using other variables. Since the objective of the model is to maximize the probability of the observation belonging to a specific region or class, the setup is defined as follows:

$$E = 1 - \max_{k}(\hat{p}_{mk})$$

Where $\hat{p}_{mk}$ is the proportion of observations in the *m* region belonging to the *k* class. To make the empirical fit, however, another measure is preferred, for tree-based models the Gini Index can be defined as:

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

This expression allows for values of the Bernoulli Density Function to operate in a mean approach to maximize the likelihood of the observation belonging to a category or not.

The rationale behind selecting the previous models over more advanced machine learning ones is their interpretability. In many classification problems tackled with artificial intelligence, the primary objective is to achieve the best possible prediction for the data. However, this often results in the creation of black-box models, where the features or exogenous variables are not directly interpretable. While these models serve as powerful forecasting tools, they lack transparency in understanding the underlying behavior of the phenomenon.

The logistic regression and decision tree models belong to a subset of machine learning models that accommodate non-linear data behavior, in contrast to traditional multivariate linear models. However, they retain the necessary ability to explain and understand the relationship between the response and exogenous variables. Using more complex and flexible models may lead to difficulties in measuring the direction and impact of variables, as traditional linear models are unable to capture non-linear impacts (James et al., 2017). Additionally, compared to Probit models, logistic regression allows for the use of a distribution with heavier tails than the Gaussian Density. Therefore, in the presence of possible outliers, logistic regression proves to be more robust (Gujarati & Porter, 2009).

As mentioned earlier, the dataset comprises all observations from the survey. To assess the a priori variables that may contribute to financial inclusion, the selected variables mainly pertain to household financial inclusion characteristics. The following variables were chosen to identify the features influencing inclusion in the financial system. Table 1 presents the variable names along with brief descriptions of their meanings and categorical levels (see Table 1).

Table 1. Variable Description for Exogenous Features

Variable	**Description**
Location size	Number of inhabitants of the location; it consists of four categories "More than 100,000," "Between 15,000 and 99,999," "Between 2,500 and 14,999" and "Less than 2,500."
Region	Six regions in which INEGI divides the territory; "Northwest," "Northeast," "West," "CDMX," "Southwest" and "Southeast."
Sex	Sex: "Male" or "Female."
Age	Age, with valid values in the range of 18 to 96.
Education	Completed schooling level, ranging from "None" to "Postgraduate."
Civil	Marital status: "Single," "Free Union," "Separated," "Divorced," "Widow," "Married."
Indigenous	Proxy variable to characterize if the person belongs to an indigenous group; the question only asks if the person speaks an indigenous language.
Government Program	If he or she is a beneficiary of a government welfare program.

Variable	**Description**
Occupation	Main occupation: “worker,” “unpaid worker,” “unemployed,” “student,” “homemaker,” “retired,” “disabled” and “not economically active.”
Labor Class	“Worker,” “unpaid worker,” “laborer,” “self-employed,” “employer.”
Income	Monetary amount of income measured per month.
Income Frequency	Frequency of payment: “weekly,” “fortnight,” “monthly” and “yearly.”
Stable	If the perceived income is regular in frequency or not.
Smartphone	If the person owns a smartphone.
Budget	If the person has a budget control over its monthly expenditure.
Annotations	If the person uses a system to keep track of his or her own expenditure.
App	If the person uses a digital tool to keep control over their expenditure.
Resources	If the person, in any of the last three years, experienced a shortage of money by the end of the month.
Considerations	How often does the person consider if there is enough money to make a purchase: “always,” “sometimes,” “never,” “not answered,” “don’t know.”
Future	How often does the person prefer to consume rather than save for the future: “always,” “sometimes,” “never,” “not answered,” “don’t know.”
Goals	How often the person set a financial or economic goal and work to achieve it: “always,” “sometimes,” “never,” “not answered,” “don’t know.”
Driver	How often does income and expenditure take control over the life of the person: “always,” “sometimes,” “never,” “not answered,” “don’t know.”
Present	How much does the person agree with the statement: they prefer thinking about the present rather than the future: “agree,” “neither agree nor disagree,” “disagree” “not answered,” “don’t know.”
No Income	How long would the person be able to maintain their current level of expenditure: “don’t know,” “not answered,” “no savings,” “1-3 weeks,” “1-3 months,” “3-6 months” and “6+ months.”
Savings	In the last 12 months, how much money was the person capable of saving in reference to their income: “1 week,” “1 fortnight,” “1 month,” “1+ month,” and “don’t know.”
Financial Education	If the person has ever taken a course in financial education.

Source: Prepared by the authors using the variable descriptor of ENIF survey referring to the module data subset.

From another research perspective, it is imperative to establish the definition of financial inclusion. According to INEGI (2021), it entails "the access to and use of formal financial services under proper regulation that guarantees consumer protection schemes and promotes financial education." While this definition encapsulates the multidimensional nature of financial inclusion, it is essential to devise a measurable way to assess it. In this study, following the logic of the ENIF design, we have chosen the so-called "filter" variables to represent the various categories of financial inclusion. Three levels or questions have been considered, and their interpretations as variables are presented in Table 2 (see Table 2).

Table 2. Variable Description for Predicted Variable

Variable	Description
Savings	Answer to the question: Did you save any money in a financial account?
Payroll Card	Answer to the question: Do you have a payroll card?
Debit Card	Answer to the question: Do you have a debit card?
Inclusion	Boolean variable if all previous answers are "Yes." This will be used as the dependent variable in the classification

Source: Prepared by the authors using the variable descriptor of ENIF survey referring to the Module data subset.

To conduct this study, two models with the dependent variable in Table 2 will be conducted separately. The variables in Table 1 will be used as features in the machine learning models previously presented. To perform the evaluation, the confusion matrix approach will be implemented. In this sense, 60% of the persons will be randomly selected as part of the training dataset, and the remaining 40% will correspond to the testing dataset. The main statistics to evaluate are the accuracy ratio, McNemar's test, and Cohen's Kappa. Furthermore, the interpretability of the models will allow the evaluation of the significance and impact of the variables included to obtain the best model to predict the presence, or not, of financial inclusion.

3. Results

To implement the previous models in the datasets we begin by gathering and transforming the variables into dummy ones. By doing so, the model can be executed as a parallel line's Logit model and decision tree. At the same time, a balanced data transformation is proposed as the data contains different amounts of observations in the classifications that were previously defined. The first step is to identify the number of persons belonging to each side of the financial inclusion, and then a random sampling of the same length is taken for both sides. This provides balanced data that can be analyzed with the prior variables without having a size skewness in the over or under-identification of the features. Table 3 contains the counting for each of the classifications in the variable presented in Table 2 (see Table 3).

Table 3. Counting Statistics for Response Variables

Variable	Yes	No
Inclusion	1632	11922

Source: Prepared by the authors using the variable descriptor of ENIF survey referring to the Module data subset.

In such case, the number of observations per model will be the value for the "Yes" response. The remaining observations will be randomly sampled to fit the same value. A backward-looking selection algorithm will be implemented to present only the best models in the model selection for the main features to be included.

The first model to consider is the one with the "inclusion" variable in the Logit specification. Table 4 contains the results for the regression estimates and p-value for t-test significance values (see Table 4) and Table 5 presents the confusion matrix with the main statistics to conduct the evaluation (see Table 5). It is important to mention that in the model selection process, the multicollinearity phenomena were controlled using the variance inflation factor. The presented model was the best in calibration.

Table 4. Logit Results for Inclusion as Response Variable

Coefficient	Estimate	Standard Error	p-value
Intercept	-5.47	1.55	4.27×10^{-4}
Income	1.14×10^{-4}	2.22×10^{-5}	2.52×10^{-7}
Education (Elementary)	4.43×10^{-1}	1.32	7.38×10^{-1}
Education (Middle School)	1.38	1.30	2.88×10^{-1}
Education (Middle Technical)	1.89	1.45	1.95×10^{-1}
Education (High School)	2.12	1.30	1.04×10^{-1}

Coefficient	**Estimate**	**Standard Error**	**p-value**
Education (High Technical)	3.45	1.43	1.58×10^{-5}
Education (Bachelor)	2.90	1.31	2.69×10^{-2}
Education (Post Graduate)	4.84	1.69	4.12×10^{-3}
Government Program (Yes)	1.01	5.17×10^{-1}	4.97×10^{-2}
Smartphone (Yes)	1.15	3.20×10^{-1}	3.34×10^{-4}
Resources (Not Enough money)	-5.25×10^{-1}	1.92×10^{-1}	6.18×10^{-3}
No Income (Not answered)	5.50×10^{-1}	1.29	6.71×10^{-1}
No Income (No savings)	-3.14×10^{-2}	7.69×10^{-1}	9.67×10^{-1}
No Income (1-3 weeks)	4.33×10^{-1}	7.68×10^{-1}	5.73×10^{-1}
No Income (1-3 months)	1.29	7.57×10^{-1}	8.74
No Income (3-6 months)	1.56	7.81×10^{-1}	4.56×10^{-2}
No Income (6+ months)	2.45	7.95×10^{-1}	2.03×10^{-3}
Labour Stability (Fixed)	1.27	1.88×10^{-1}	1.15×10^{-11}
Location Size (100000+)	6.48×10^{-1}	2.84×10^{-1}	2.23×10^{-2}
Location Size (15M - 99M)	5.36×10^{-1}	3.31×10^{-1}	1.06×10^{-1}
Location Size (2.5M - 15M)	6.23×10^{-1}	3.57×10^{-1}	8.14×10^{-2}
Financial Education (Yes)	1.26	3.93×10^{-1}	1.35×10^{-3}

Source: Prepared by the authors with information from ENIF. The results are displayed as the maximum likelihood estimator for the coefficients. The p-value refers to the significance t-Test.

Table 5. Confusion Matrix for Inclusion as Response Variable

	No	Yes
No	256	61
Yes	66	361
Accuracy	0.83	
Kappa	0.65	
p-value McNemar's Test	0.72	

Source: Prepared by the authors. The confusion matrix in the first part of the table represents the empirical category the observation belongs to in the rows, while the columns are for the forecast. The main diagonal counts for the observations are correctly identified, while the secondary diagonal is badly allocated. The accuracy ratio is the proportion of correctly identified observations relative to the whole sample. Kappa stated for a cleaner accuracy interpretation, while McNemar's Test captures the bias error, the null hypothesis states that the incorrectly identified observations are the same.

The original model included all the previously mentioned variables; however, the significance test and evaluation using the confusion matrix did not provide evidence of their significance over the response variable. Additionally, we considered variables that may cause multicollinearity, and in this instance, the VIF criterion was employed to refine the model, retaining only the variables without severe multicollinearity. Therefore, only the best model is presented here. In this case, considerations were made to assume significance under a 95% confidence level, although in some cases, we also allowed for a 90% confidence level if the interpretation of such a variable aligned consistently with results obtained by previous authors. The selection criteria also considered the highest value for Accuracy and Cohen's Kappa, along with the non-rejection of the McNemar's Test of bias in the error.

To interpret the model, each variable will be explained based on the increment or decrement of the probability of having financial inclusion. It is worth mentioning that the estimated values presented in Table 4 correspond to the log-odds of the model, so only the sign of the estimate can be read as an increment or decrement; still, to have a proper reading, it is necessary to transform such values into traditional probabilities.

The first variable to be significant in the model is the observation's monthly income, t a numeric variable measured in Mexican pesos, so the marginal increment in probability is 0.0025% per extra monthly peso of income. This can be understood as a positive relationship between income and the likelihood of having this type of financial inclusion.

The second variable incorporated in the model is Education. In this case, we have a categorical variable with eight levels, so the estimates are compared with respect to the lowest level of education which, in this case, is labeled as "None." In essence, we are presenting the results for dummy variables in the education category. The first elementary and middle school levels are not statistically significant compared to those with no education at all, meaning that there is no statistical difference between those three groups. However, from the level of middle technical, high school, bachelor, and postgraduate, a significant alteration can be perceived. Compared to those without any education, middle technical individuals have an increase in probability of being in financial inclusion of 25.82%, the high school increments by 35.13% the chances, high technical tends to have a marginal increment of 32.42%, bachelors have a difference of 46.94% and postgraduates of 36.37%. As may be appreciated, the increase in financial inclusion is significant when we isolate this category as a comparison variable. The interpretation of this result can be explained by the capabilities of people reaching these educational levels; as their income tends to be higher and more regular than those in jobs with lower education levels, the incentive and ability to save money for the future increases as well. Another interesting aspect to contemplate is the fact that in those educational levels, extra information in financial education is provided, allowing individuals to be conscious about such topics.

The next variable is being the beneficiary of any government program. In this sense, the Mexican Government has developed several social programs intended to aid those with a higher economic vulnerability. In doing so, many programs lean toward the digital transfer of resources. The result is that the population in such conditions was incentivized to enter the financial system in order to benefit from these programs. When isolating for the marginal impact of this variable compared to those without government aid, the probability increases by 17.9%, allowing for a positive relationship.

The Boolean variable of owning a smartphone follows. This technology enables various activities in daily life, one of which is accessing financial services or applications to understand personal finance better. According to WallStreetZen (2023), gen Z individuals learn about money using YouTube and TikTok digital platforms. Although this can potentially enhance financial inclusion, the survey also mentions the incentives for content creators to provide false information to generate more engagement. In this sense, it is worth noticing and questioning not only the quantity but the quality of financial inclusion, as greater access to financial services also has the potential to be used to scam or to provide incorrect information to financial consumers. In this respect, possessing a smartphone provides an extra 27.14% probability of attaining financial inclusion.

Next comes a variable about the income and expenditure level of the individual. The question refers to whether or not the person arrives at the end of the month with enough money. The reference category is that the person has “enough money” by the end of the month. For the estimated value of the coefficient, the difference is significant and has an impact on the probability of -11.58%. The previous result explains that a person with insufficient money is less likely to have financial inclusion. This aspect, which might seem trivial, is not so in an economic context where households present high expenditure levels that are not met by their current income. This inability to save money and fear to ask for consumer credit might explain some of the low levels of financial inclusion in Mexico. In this case, creating instruments suited for those types of individuals, either in saving or allowing for credit, may be more productive than simply expanding the current infrastructure.

The next variable to be incorporated into the model delves into another aspect of purchasing power and future preparedness. It aims to determine how long an individual could sustain their current expenses if they suddenly ceased to receive income. The reference category for this question is “Don’t Know.” The results indicate that there is no statistically significant difference among the responses “No Savings,” “Not Answered,” and “1-3 weeks.” However, the real impact surfaces among those capable of maintaining their expenses for one month or more. For individuals able to sustain their lifestyle for “1-3 months,” the probability of financial inclusion increases by 24.96%. This likelihood rises to 26.1% for those able to manage for “3–6 months,” and for those who can sustain themselves for more than six months, the chances of financial inclusion increase by a substantial 35.85%. This underscores the importance of the income-expenditure relationship, highlighting the need for financial instruments that facilitate savings and enhance financial security.

Another variable that was shown to be statistically significant was the stability of income perception. People with a fixed income present an increase in probability compared to variable income of 27.3%. In this case, the Mexican labor structure works against financial inclusion as a considerable portion of the labor market belongs to the informal sector. In this instance, the inability to access to a financial instrument may have caused an exacerbated impact during the pandemic because of the contingency and reduction in consumption.

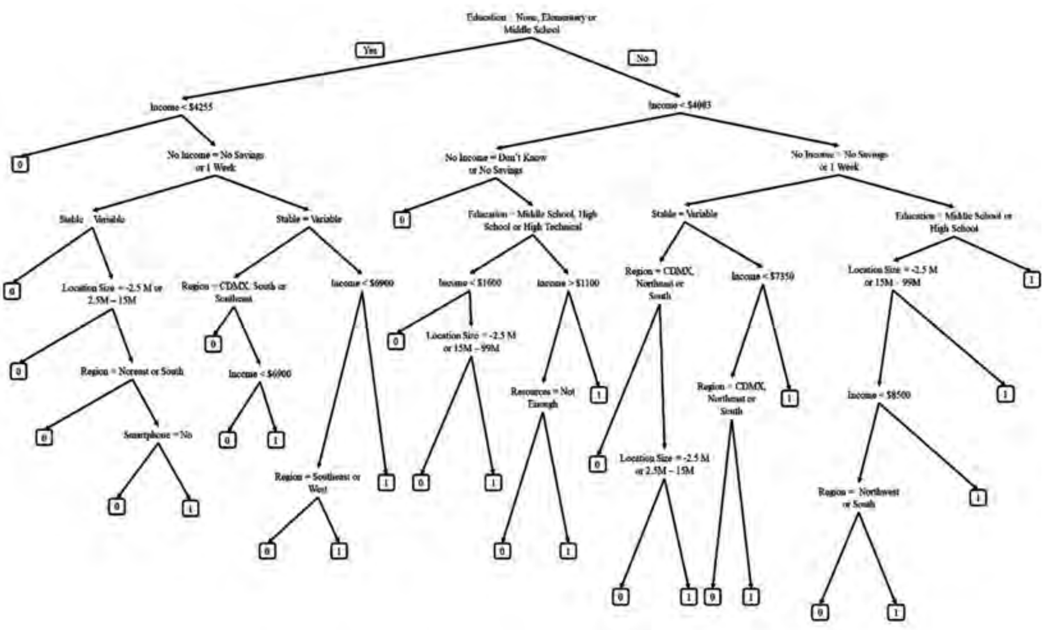
The location size may be used as a proxy variable for the infrastructure availability of the individual. In Mexico, INEGI measures this variable as “less than 2500,” “between 2500 and 15,000,” “between 15,000 and 99,000” and “more than 100,000” people. The first two categories can be described as rural areas, while the latter are urban or semi-urban locations. For the interpretation of this estimate, the reference point is for communities with fewer than 2500. Therefore, when the individual belongs to a community between 2500 and 15,000, the chances to be part of financial inclusion increases by 12.27%; when it has between in 15,000 and 99,000 persons, the probability increases by 10.82%. Despite that, it is worth mentioning that these results are valid only under a 90% confidence level, so it might be argued that there is no difference in these three categories. However, in the urban area where the population reaches more than 100,000 people, the probability increases by 14.27% with a confidence level of 95%. As such, the result of this variable can be interpreted as a positive impact of infrastructure on financial inclusion.

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.09

The final variable to be included is the response to the question of whether the person has ever taken a "personal finance" course. As such, the increment in probability for financial inclusion due to this fact is 22.1%. This strengthens the notion that having access to this type of information on the benefits and types of financial instruments, as well as how to make best use of them, has a positive and significant impact.

For the second model, the results of the estimation are better presented graphically. Figure 1 presents the branches and conditions for the identification of the observation belonging to the financial inclusion (1) or not (0) (see Figure 1). Table 6 presents the confusion matrix for the decision tree (see Table 6).

Figure 1. Tree Identifying Financial Inclusion



Source: Prepared by the authors to represent the variables as nodes in the decision tree. The left branches are for when the condition is true and the right branches when the statement is false. The 0 represents the observation without financial inclusion and the 1 represents those who have it.

Table 6. Confusion Matrix for Decision Tree Model

	No	Yes
No	245	69
Yes	79	348
Accuracy	0.80	
Kappa	0.59	
p-value McNemar's Test	0.12	

Source: Prepared by the authors. The confusion matrix in the first part of the table represents the empirical category the individual belongs to, in the rows, while the columns are for the forecast. The main diagonal counts for the individuals correctly identified while the secondary diagonal those badly allocated. The accuracy ratio is the proportion of correctly identified observations relative to the whole sample. Kappa stated for a cleaner interpretation of the accuracy while McNemar's Test captures the bias error; the null-hypothesis states that the incorrectly identified observations are the same.

For the interpretation of this tree, the paths leading to the observation belonging to the financial inclusion group were gathered in Table 7 (see Table 7). As a result, 14 sub-groups were identified by the algorithm as the combination of features that allow a person to belong to the financial inclusion group. In each row the description of the characteristics is included.

Table 7. Features Identified by the Decision Tree that Allows a Person to Belong to the Financial Inclusion Category

Sub-group	Characteristics
1	No education, elementary or middle school Income > $4255 No savings or for less than a week Stable income Location size larger than 15,000 Not from northeast or south region Owns smartphone
2	No education, elementary or middle school Income > $6900 Savings for more than a week Variable Income From northeast or northwest region

https://doi.org/10.36105/theanahuacjour.2024v24n1.09

Sub-group	Characteristics
3	No education, elementary or middle school Income between $4,255 and $6,900 Savings for more than a week Fixed income Not from southeast or northwest region
4	No education, elementary or middle school Income > $6,900 Savings for more than a week Fixed income
5	Middle school, high school or high technical Income < $1,083 Savings for at least 1 week Location with more than 100,000 people
6	Bachelor or postgraduate education Income > $1,083 Savings for at least 1 week Enough resources to reach the end of the month
7	Bachelor or postgraduate education Income < $1,083 Savings for at least 1 week
8	Education above middle school Income > $4,083 No savings or less than a week Variable income Not from CDMX, northeast or south Location size above 100,000 people
9	Education above middle school Income between $4,083 and $7,350 No savings or less than a week Fixed income Not from CDMX, northeast or south
10	Education above middle school Income > $7,350 No savings or less than a week Fixed income
11	Education above High School Income > $4,083 Savings for more than a week

Sub-group	Characteristics
12	Middle or high school Income > $4,083 Savings for more than a week Location with more than 100,000 people
13	Middle or high school Income > $8,500 Savings for more than a week Location with more than 100,000 people
14	Middle or high school Income between $4,083 and $8,500 Savings for more than a week Location with more than 100,000 people Not from northwest nor south

Source: prepared by the authors using information from the decision tree model.

It is noteworthy that the most influential variables in identifying a person's level of financial inclusion once again center around education levels, with middle and high school education serving as pivotal benchmarks that define the disparity. Following closely is the individual's region, where sectors in the north exhibit a higher propensity to utilize financial instruments. Additionally, the size of the location plays a significant role in differentiating observations, enabling individuals with lower economic standing to access financial instruments, particularly in areas with larger population sizes. When interpreted as a proxy variable for infrastructure, this underscores the critical importance of such features in fostering financial inclusion.

When comparing the two models, it's important to highlight that those who opted to incorporate all the previously mentioned variables, driven by a principle of simplicity and effectiveness in achieving predictability, refined the models to showcase the most optimal results attainable. In both cases, some variables are incorporated and represent a major influence on the discriminatory analysis. For instance, the education level, location size, income, and income stability are present. On the other hand, the Logit model includes, explicitly and with a huge impact in probability, the feature of the individual having any type of personal financial education. The second important different variable is the access to a smartphone, which the Logit model uses as a positive relationship with financial inclusion; however, the decision tree does not use it.

Concerning the results in the machine learning evaluation, the accuracy level of both models is around 80%; however, logistic regressions seem to have a better performance when using the Kappa criteria. Also, both algorithms present an adequate value for the McNemar skewness test, allowing for a proper fit and unbiased errors. It can be resolved that artificial intelligence algorithms under the classification basis are a proper way to make the study of financial inclusion drivers. Furthermore, the main advantage of these models is their capability to detect the non-linear effect of independent variables over the response without having to sacrifice the interpretability of the results, as it may happen with more sophisticated models like neural networks or support vector machines.

In this instance and using the Kappa and accuracy criteria, the Logit models seem to have a better performance in the detection of the variables belonging to financial inclusion. Nevertheless, the decision trees also work as a reinforcement of the importance of variables in the features that drive the financial inclusion phenomena.

4. Conclusions

The current document highlights the significance of financial inclusion as an objective for any developing nation. The international attention this variable is receiving allows for multiple studies to be performed and a plethora of causes and effects to be proposed. To contribute to this, a machine learning approach is suggested. As part of this, Logit and decision Tree models are used as a technological tool that makes it possible to capture non-linear relationships for features of the observations and the probability of belonging to the financial inclusion sector. By using analysis related to statistical and artificial intelligence evaluation, the best models are computed with information from the ENIF survey. The main contribution of this paper is to quantify the impact of some characteristics of the observations on the probability of a person having financial inclusion. As such, it provides a better understanding of the phenomena that can be implemented in public policies to improve the indicator.

The main findings of the models provide information about the importance of education levels over financial inclusion. Individuals reaching middle school have a higher and significant improvement in the probability of being part of the financial system. This situation can be attributed to education programs where topics on personal finance are introduced too late in the curriculum. Consequently, individuals who only attain a basic level of education may never receive formal instruction on financial matters. To control for this condition, the explicit question about receiving a course in personal finance is included. With it, it can be noticed that having this extra knowledge is a significant driver to increase the probability.

Education is one of the main studied and supported variables to be included in the financial inclusion documents; however, the other features that have a major impact on it are the amount of wealth, the consistency of a perceived income, and the size of the location. In terms of the amount of wealth, the model found that the marginal effect of having enough money to survive for at least one month without any income is almost equal to the educational level.

References

Arora, R. U. (2014). Access to finance: An empirical analysis. *The European Journal of Development Research*, *26*(5), 798–814. https://doi.org/10.1057/ejdr.2013.50

Beck, T., Demirguc-Kunt A., & Martinez Peria, M. S. (2007). Reaching out: Access to and use of banking services across countries. *Journal of Financial Economics*, *85*(1), 234–266. https://doi.org/10.1016/j.jfineco.2006.07.002

Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, *39*(227). https://doi.org/10.2307/2280041

Bruhn, M. & Love, I. (2014). The real impact of improved access to finance: Evidence from Mexico. *The Journal of Finance*, *69*(3), 1347–1376. https://doi.org/10.1111/jofi.12091

Community Reinvestment Act (1977). Federal Reserve.

Conroy, J. (2005). APEC and financial exclusion: missed opportunities for collective action? *Asia Pacific Development Journal*, *12*(1), 53–79.

Cruz-García, P., Dircio-Palacios Macedo, M. del C. & Tortosa-Aucina, E. (2020). Financial inclusion and exclusion across Mexican municipalities. *Regional Science Policy & Practice*, *13*(5), 1496–1526. https://doi.org/10.1111/rsp3.12388

De la Cruz Gallegos, J. L. & Alcántara Lizárraga, J.A. (2011). Crecimiento económico y el crédito bancario: un análisis de causalidad para México . *Revista de Economía*, *28*(77), 9–38.

Dircio-Palacios Macedo, M. del C., Cruz-García, P., Hernández-Trillo, F. & Tortosa-Ausina, E. (2023). Constructing a financial inclusion index for Mexican municipalities. *Finance Research Letters*, *52*(C), article 103368. https://doi.org/10.1016/j.frl.2022.103368

Demirguc-Kunt, A. & Klapper, L. (2012). Measuring financial inclusion: The global findex database (No. 6025). The World Bank.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2017). *An Introduction to Statistical Learning with Application in R*. Springer.

Jamil, A. R. M., Law, S.H., Khair-Afham, M.S. & Trinugroho, I. (2024). Financial inclusion and income inequality in developing countries: The role of aging populations. *Research in International Business and Finance*, *67*(PA). https://doi.org/10.1016/j.ribaf.2023.1

Gujarati, D. & Porter, D. (2009). *Basic Econometrics*. McGraw-Hill Series Economics.

Huong Tram, Thi Xuan, Lai, Tien Dinh, Huong Nguyen, Thi Truc. (2023). Constructing a composite financial inclusion index for developing economies. *The Quarterly*

Review of Economics and Finance, 87 (257–265). https://doi.org/10.1016/j.qref.2021.01.003

Instituto Nacional de Estadística Geografía e Informática (INEGI). (2021). Encuesta Nacional de Inclusión Financiera (ENIF) 2021.

Johnson, S. & Arnold, S. (2014) Inclusive financial markets: is transformation under way in Kenya? *Development Policy Review*, 32 (5), 639–642. https://doi.org/10.1111/j.1467-7679.2012.00596.x

King, R. G. & Levine, R. (1993) Finance and growth: Schumpeter might be right. *The Quarterly Journal of Economics, 108*(3), 713–737. https://doi.org/10.2307/2118406

Leyshon, A. & Thrift, N. (1995). Geographies of financial exclusion: financial abandonment in Britain and the United States. *Transactions of the Institute of British Geographers*, *New Serie*s *20*(3), 312–341. https://doi.org/10.2307/622654

Marín, A. G. & Schwabe, R. (2018). Bank competition and financial inclusion: Evidence from Mexico. *Review of Industrial Organization*, *55*(2), 257–285. https://doi.org/10.1007/s11151-018-9673-5

Mohan, R. (2006). Economic growth, financial deepening, and financial inclusion. Address at the Annual Bankers' Conference 2006, Hyderabad on November 3, 2006. http:// rbidocs.rbi.org.in/rdocs/Speeches/PDFs/73697.pdf

Rangarajan Committee. (2008). Report of the Committee on Financial Inclusion. Government of India.

Sarma, M. (2015). Measuring Financial Inclusion. *Economic Bulletin*. *35*(1), 604–611.

Schuetz, S. & Venkatesh, V. (2020). Blockchain, adoption, and financial inclusion in India: Research opportunities. *International Journal of Information Management*, *52*. https://doi.org/10.1016/j.ijinfomgt.2019.04.009

Schumpeter, J. (1911). *The theory of economic development*. Harvard University Press.

Segovia, M. A. F. & Cepeda, L. E. T. (2024). Financial development and economic growth: New evidence from Mexican States. *Regional Science Policy & Practice*, article 100028. https://doi.org/10.1016/j.rspp.2024.100028

Sun, T. & Wang, X. (2023). Adoption of financial inclusion in a world of depleting natural resources: the importance of information and communication technology in emerging economies. *Resources Policy,* 85. https://doi.org/10.1016/j.resourpol.2023.103901

Tram, T. X. H., Lai, T. D. & Nguyen, T. T. H. (2023). Constructing a composite financial inclusion index for developing economies. *Quarterly Review of Economics and Finance*, *87*, 257–265.https://doi.org/10.1016/j.qref.2021.01.003

Varian, Hal R. (1992). *Microeconomic Analysis*. Norton.

WallStreetZen. (2023). Where Did Gen Z Learn About Money. https://www.wallstreetzen.com/blog/genz-money-social-media-survey/

Xi, W. & Wang, Y. (2023). Digital financial inclusion and quality of economic growth. *Heliyon, 9*. https://doi.org/10.1016/j.heliyon.2023.e19731

Yue, P., Korkmaz, A.G., Yin, Z. & Zhou, H. (2022). The rise of digital finance: financial inclusion or debt trap? *Finance Research Letters,* 47(A). https://doi.org/10.1016/j.frl.2021.102604

About the authors

Itzel Coquis Rioja graduated from Tecnológico de Monterrey (ITESM) Campus Mexico City with a degree in Business Administration and holds a Master's in Energy Administration and Renewable Resources. She is currently a doctoral student in Financial Sciences. After completing her studies, she began her professional career at Coca-Cola Femsa, in Human Resources, in 2011. She worked in the area of market intelligence at Millward Brown as an Account Coordinator. Years later, she returned to ITESM with the goal of pursuing a postgraduate degree. Today she is the Bachelor Director of Marketing at Tecnológico de Monterrey Campus Mexico City. Since 2017 she has been a professor at the Business School. Students acknowledge her balanced vision of the business world, as well as the care she takes in the professional and personal development of students.

itzelcoquis@tec.mx
https://orcid.org/0000-0001-6284-8320

Mario Iván Contreras Valdez studied for a bachelor's degree in economics and finance at Tecnológico de Monterrey (ITESM) and graduated with honors. Subsequently, he entered EGADE Business School to study for a Ph.D. in Finance, where he obtained recognition for an outstanding thesis. He is currently the Bachelor Director of Finance at Tecnológico de Monterrey Campus Mexico City. Additionally, he is a consultant to firms on topics of M&A and Risk Management. His research areas include cryptocurrencies, quantitative finance, and risk management.

marioivan.contrerasv@tec.mx
https://orcid.org/0000-0003-4870-4376

Prioritizing the Net Sentiment Score: A Banking Industry Case Study

Por qué el índice de sentimiento neto debería ser una prioridad: un estudio de caso de la industria bancaria

José Guadalupe Mendoza-Macías

Quantum Analytics, Sinaloa, Mexico

Román Alejandro Mendoza-Urdiales

Quantum Analytics, Sinaloa, Mexico

Received: April 3, 2024.
Approved: June 4, 2024.

Abstract

This study analyzes the impact of social media comments on the stock performance of banks registered on the United States stock market. We use artificial intelligence to monitor and extract comments in real time, together with natural language processing, to identify the sentiment of each comment. Comments were classified as positive or negative and were added by the hour for each bank during the observed period. Our results showed that both positive and negative comments have a significant effect on the stock performance, with negative comments having a more pronounced, asymmetrical impact. This study contributes to understanding how social media interactions influence the market value of companies, highlighting the importance of monitoring and managing the online perception of the companies listed on the stock market.

Keywords: *stock performance, net sentiment score (NSS), robust linear model regression, brand value impact.*
JEL Classification: *G41.*

Resumen

El artículo analiza el impacto de los comentarios en redes sociales sobre el rendimiento de las acciones de los bancos en el mercado de valores de EE.UU. Se empleó inteligencia artificial para monitorear y extraer comentarios en tiempo real, y se utilizó el procesamiento de lenguaje natural para calcular el sentimiento de cada comentario. Los comentarios se clasificaron como positivos o negativos y se agregaron, por hora, para cada banco durante el período observado. Los resultados mostraron que tanto los comentarios positivos como los negativos tienen un efecto significativo en el rendimiento de las acciones, con un impacto asimétrico más pronunciado en el caso de los comentarios negativos. Este estudio contribuye a la comprensión de cómo la interacción en redes sociales puede influir en el valor de mercado de las empresas y destaca la importancia para las compañías que cotizan en bolsa de monitorear y gestionar la percepción en línea.

Palabras clave: *rendimiento accionario, índice de sentimiento neto (ISN), regresión de modelo lineal robusto, impacto en el valor de la marca.*
Clasificación JEL: *G41.*

1. Introduction

The practice of manipulating information for personal financial gain predates academic literature. During the Napoleonic wars, a man claiming to be Colonel du Bourg stated that Napoleon I was murdered in the battle against the Bourbons (Evans,1859). Within a very short period, this claim greatly benefited the securities price backed by the British government. It was later revealed that the news was false, and a subsequent investigation by the Committee of the Stock Exchange concluded that Lord Cochrane had orchestrated the manipulation, selling over £1 million in assets during the frenzy.

The above example is but one of the multiple documented instances of the stock market being impacted significantly and swiftly through news manipulation, with the information source seldom verified. This is primarily due to market players seeking premiums, and being exposed to (and accepting) the risk. In modern times, more contemporary methods are employed to induce similar market effects in order to garner substantial yields within a brief period.

Social media serves as a common medium through which influence on publicly traded companies can be exerted. These platforms, either consciously or unconsciously, have virtually eliminated the gap between consumers and firms. In recent years, consumers have discovered that they wield significant power (at their fingertips) (Plourde, 2023). Situations arise where individuals, feeling unheard by the brands managed by firms, utilize the power of social media to demand attention not only from the brands themselves, but also from other customers, regulators, government agencies, and so forth. This phenomenon has informally and positively established social media as the "official" channel for communication between brands and end users, removing any intermediaries.

Brands have noticed this and are leveraging social media to successfully promote their products, as well as monitoring their reputation on the platforms. Reichheld (2011) describes the introduction of the term "net promoter score (NPS)" by the management consulting company, Bain and Company, in 2003. This metric was designed to measure customer loyalty towards a brand, company, product, or service, based on the probability of recommending the product to peers. More recently, Van Velthoven (2014) linked NPS to net sentiment score (NSS), contrasting the already calculated NPS of Vistaprint (an online provider of printed and promotional material and marketing to small businesses and consumers) with comments mentioning the company on X (formerly Twitter). The study revealed that there is indeed a positive moderated correlation between NPS and NSS.

While firms and decision-makers generally agree on the direct impact of social media on sales (Donthu *et al*., 2021), the indirect effects on brand value and the workings of social media posts/trends-sales-brand popularity/value mechanism require further exploration. This study proposes a model that integrates economic, financial, linguistic, and psychological factors to capture the impact of social media comments on a firm's total value through stock prices. Additionally, we investigate how quickly this mechanism affects stock prices and company value.

In the era of digital communication, the influence of social media on public perception and behavior has become a significant area of interest. This study delves into the impact of social media interactions on the stock prices of publicly traded banks, a topic that has seen increasing scholarly attention in recent years. Notably, works such as Bollen *et al*. (2011), who investigated the relationship between the Twitter mood and the stock market, and Tetlock (2007), who examined the impact of media sentiment on securities markets, provide a foundational understanding of the nexus between public sentiment expressed online and market outcomes. Building upon this literature, our research aims to quantify the specific effects of social media sentiment—both positive and negative—on banking stocks, and thereby contribute to a more nuanced understanding of social media as a potent market mover.

Specifically, we seek to answer the following questions:

1) Is it possible to calculate the impact of social media interactions at a company level even when companies have a limited social media presence?

2) Is there an asymmetric effect between negative and positive interactions?

3) Is it possible to separate positive and negative sentiment for each company using robust linear model regression?

This paper is structured as follows: The "related work" section outlines similar work. The "problem definition" and "method" sections explain the proposed methodology. In "results and discussion" we show the experiment results and evaluation of the proposed method. Finally, in the "conclusions", some final assessments are made and direction for future research is provided.

2. Related Work

An interesting proposal focuses on measuring the impact of customer satisfaction on brand value. Colicev and O'Connor (2020) explore how social media data affects brand value by monitoring customer satisfaction, corporate reputation, word-of-mouth, and awareness using Partial Least Squares Path Modelling (PLS-PM). The study explores official content and user content to estimate latent variable scores, and concludes that social media marketing has a positive impact on customer satisfaction and that brand value is a precursor of sales and shareholder value.

Mendoza-Urdiales *et al*. (2021) present a methodology that categorizes and measures the impact of social media comments on the daily closing performance of 23 companies over a ten-year period. Their study obtained a success rate of over 80% in confirming the influence of social media on some of the world's largest publicly traded companies. Mendoza-Urdiales *et al*. (2022) propose a method to categorize positive and negative comments and their impact on daily performance by creating negative and positive variables. Using two methods, EGARCH and Transfer Entropy, the results indicated an asymmetric impact of negative comments compared to positive ones, persisting for 2-3 days after the comments had been posted. Núñez-Mora and Mendoza-Urdiales (2023) propose a big data approach in which they extracted all comments on social media that mentioned the 2557 most important publicly traded companies in the U.S. equity market. They categorized the comments as positive or negative, and were able to capture the asymmetric effect for the 508 largest publicly traded companies in the United States, with less than 30% missing data in the social media comments observations. Furthermore, the results show that the signal from social media impacts the stock market in under an hour.

Kirtac and Germano (2024) propose leveraging the ability of large language models (LLM) over traditional tools, like the Loughran-McDonald dictionary, to predict stock returns through sentiment analysis of financial news. Utilizing over 965,000 U.S. financial news articles from 2010 to 2023, they applied regression analysis and various performance metrics to evaluate the predictive accuracy of LLMs. The Open Pre-trained Transformer language model (OPT) showed exceptional performance with high accuracy and a notable Sharpe ratio in trading strategies. The study emphasizes the significant potential of LLMs in enhancing financial market prediction and investment strategy formulation, and advocates for their integration into financial analysis to improve market prediction and decision-making processes.

3. Problem Definition

The sample of banks included in our study were selected from the largest publicly traded companies in the United States. Specifically, we focused on companies that represent 99% of the total market capitalization, arranged in descending order of size, and totaling approximately 2557 entities. Within this universe, we targeted the 167 firms classified as banks under the industry RIC label. This selection criteria ensured that our analysis encompassed those banking institutions operating within the stock market that collectively represent 99% of the total market valuation on the U.S. stock exchange.

This methodical selection of banks allowed for the comprehensive examination of the impact of social media sentiment on stock prices across a significant portion of the banking sector. By focusing on such a substantial share of the market, the study aimed to develop insights that are both statistically robust and broadly applicable to major financial institutions within the United States. In Table 1, a sample of ten of the 167 banks is presented, with a brief description of the universe used in the study (see Table 1). Both global and local brands can be observed.

Table 1. Sample of International and Local U.S. Banks

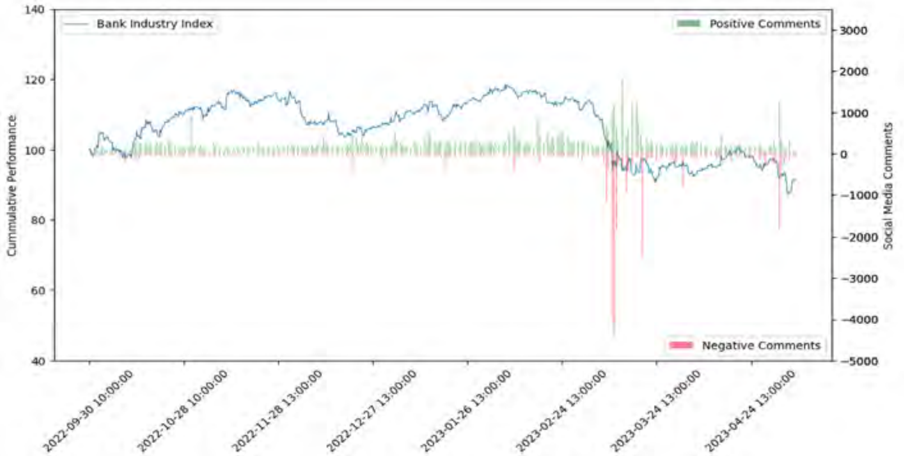
Bank Name	Reuters Instrument Code	Description
JPMorgan Chase & Co.	JPM.N	One of the largest and most influential global financial institutions, known for its extensive banking operations.
Bank of America Corp.	BAC.N	A major player in the global finance industry, serving millions of consumers and businesses worldwide.
Citigroup Inc.	C.N	A leading global bank with approximately 200 million customer accounts, operating in more than 160 countries.
Wells Fargo & Co.	WFC.N	One of the largest banks in the USA, known for its commercial and consumer banking services.
The PNC Financial Services Group, Inc.	PNC.N	Noted for its wealth management, asset management, and corporate banking services.
U.S. Bancorp	USB.N	The parent company of U.S. Bank, ranked as one of the largest commercial banks in the United States.

Bank Name	Reuters Instrument Code	Description
Goldman Sachs Group Inc.	GSBC.OQ	Primarily known for investment banking, but also a significant player in other banking sectors.
Morgan Stanley	MSBI.OQ	Recognized for its investment banking, wealth management, and asset management services.
Citizens Financial Group, Inc.	CFG.N	One of the oldest and largest financial services firms in the United States.
KeyCorp	KEY.N	The parent company of KeyBank, a regional bank headquartered in Cleveland, Ohio.

Source: Prepared by the authors.

The market cap weighted cumulative performance of the industry was calculated and paired with the aggregated classified comments grouped in positive/negative, as shown in Figure 1 (see Figure 1). It shows a correlation in the short-term variation between type of comments and performance. A higher number of positive comments than negative comments can be observed, except for the month of March 2023, when the regional bank bubble occurred and Silicon Valley Bank liquidity issues arose. Additionally, during the same period, the performance dropped drastically.

Figure 1. Market Cap-Weighted Performance of 167 Banks Operating in the U.S. Stock Market, Paired with Real-time Comments Grouped According to Sentiment (September 30, 2022 - May 5, 2023)



Source: Quantum Analytics.

Figure 2 presents the total volume of comments per hour per company during the analyzed period (see Figure 2). On average, there are 500 comments per hour, with peaks reaching up to 2500 comments and above. This indicates continuous interactions between banks and third parties. Social media users were categorized into two groups.

3.1 Direct Relationship

This refers to users with direct interaction with the company, either as current or potential customers, or through a direct interest in the company's products or services.

1) *Customers:* They have purchased or are directly interested in purchasing the company's products or services.
2) *Prospective Customers:* Individuals interested in the company's offerings who may become customers in the future.
3) *Fans or Followers:* Users who actively engage with the company's content due to a genuine interest in the brand.
4) *Brand Advocates:* Satisfied customers who actively promote the company and its products or services.
5) *Employees or Team Members:* People working for the company who engage with the profile to support and promote the brand.

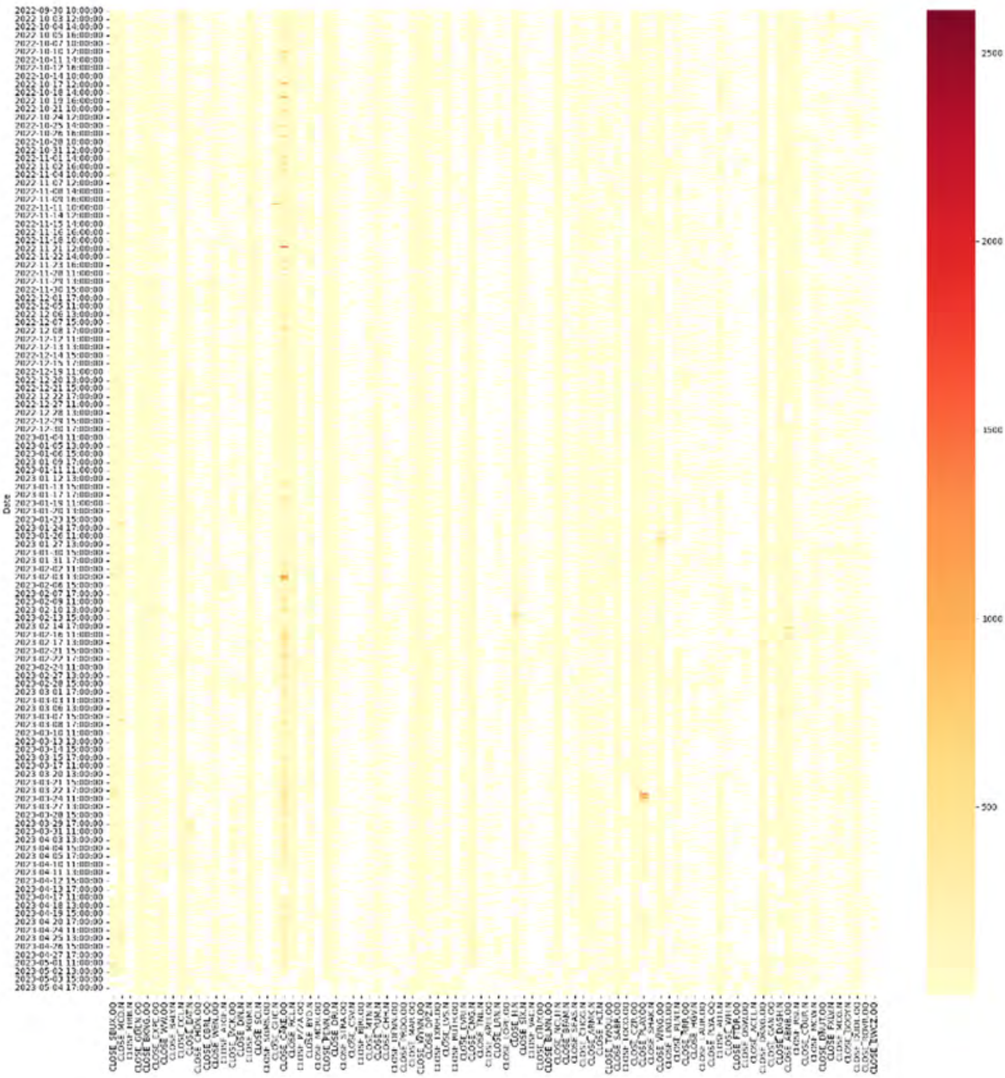
3.2 Indirect Relationship

These users may not have a direct commercial interest in the company but interact with the profile for various reasons.

1) *Trolls:* Users who engage with the profile to provoke or disrupt without a direct interest in the company's offerings.
2) *Competitors:* Other companies in the same industry that may monitor the profile but do not directly engage in business with the company.
3) *Influencers:* Individuals with a significant following who can impact the company's reputation, although their engagement might not be directly tied to a commercial interest in the company.

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.10

Figure 2. Total Number of Comments for Each Analyzed Company (September 30, 2022 - May 5, 2023)



Source: Quantum Analytics.

4) *Critics:* Users who may voice concerns or criticism without being direct customers or having a direct commercial interest in the company.
5) *Neutral Observers:* Users who follow the profile out of curiosity or for informational purposes without actively engaging nor with a direct interest in the company's products or services.

Although several recent studies have focused on measuring the impact of social media comments on stock price performance (Li & Yang, 2024; Chung & Chang, 2024), it is still necessary to explore the impact of social media comments on individual company performance. Additionally, dividing the signal into negative and positive, aimed at capturing the asymmetric effect, is still a pending question to be solved in academia. The problem of missing data for social media has been addressed previously and is presented as a limitation (Núñez-Mora & Mendoza-Urdiales, 2023). In this study, we address this problem by running a robust linear regression model which avoids the manipulation of data and allows the use of raw data without imputation methods.

Figure 2 shows that certain companies maintain a consistent presence on social media, suggesting they are mainstream. In contrast, other banking institutions have minimal social media presence, posing a challenge from an individual analysis perspective (see Figure 2). This issue will be addressed through a combination of approaches in the following section.

4. Methodology

This study analyzes the industry level effect of social media comments on companies' performance. While the sentiment calculation and method to process and classify each comment as positive or negative is carried out individually for each company, the statistical modeling is a global approach in a single modeling using a robust linear regression.

When analyzing financial data, it is common to encounter anomalies and outliers, such as those observed during a market contagion or, as in this case, the regional banking crisis of 2023. These irregularities can frequently impact the variance of data, consequently affecting the accuracy of estimates and inferences derived from traditional panel data regression models. The implementation of robust statistical methods, especially robust linear regression (RLM), offers distinct advantages over conventional panel regression techniques in managing such data.

In both Huber (1973) and Huber (1981), the resilience of statistical estimators to deviations from typical model assumptions is highlighted, particularly in the presence of outliers and heteroscedasticity. These methods utilize alternative loss functions, such as the Huber loss, which are less sensitive to outliers than the squared error loss used in an ordinary least square (OLS) regression. This adaptability allows for more accurate estimations even when the data strays significantly from standard assumptions like normality and homoscedasticity.

Additionally, a robust regression is adept at managing the unique variations across different panel units by diminishing the impact of outliers, a frequent occurrence in economic and financial datasets. As highlighted by Croux and Rousseeuw (1992), these methodologies not only bolster the reliability of statistical inferences but also enhance computational efficiency. This dual benefit is essential for extensive panel data analyses common in contemporary research settings, providing a robust foundation for statistical modeling that effectively addresses the complexities and specificities of real-world data.

Therefore, robust linear models enhance traditional panel data analysis frameworks by offering robustness against data anomalies and flexibility in accommodating non-standard data distributions, thereby becoming vital tools in the statistical analysis of panel data across various fields.

The framework used includes extracting all the public interactions of the social media platform X (formerly Twitter) in real time with the 167 banks that operated in the U.S. stock market from September 30, 2022, to May 5, 2023. The comments were analyzed using natural language processing algorithms that individually assigned each comment a negative or positive grade [-1,1] in which -1 is a fully negative classification and 1 is a fully positive classification. The process included cleaning the text of spelling errors, removing stop words, and retaining words that gave meaning to each comment. In this way, the individual sentiment for each comment was calculated for each company. If a comment mentioned more than one monitored bank, it was used to construct the sentiment for both banks.

Subsequently, the aggregated sentiment and the number of comments classified as either positive or negative were added up in hourly frequencies for each bank during the observed period. This approach aimed to observe how the number of positive and negative comments influenced the NSS of the firms.

The explanatory variables constructed for the model are sentiment (net promoter score), negative interactions, and positive interactions. The sentiment variable represents the hourly aggregated individual sentiment for each bank, while the positive and negative variables represent the total number of positive and negative mentions for each bank during the same hour according to the previously mentioned classification method. This implies that for each observation and each bank, there should be a value for sentiment, positive interactions, and negative interactions. The dependent variable is the standardized hourly returns for the banks. Finally, these variables are aggregated into time series for each bank and used in the RLM regression. The model aims to explain the normalized return of each bank's stock prices ('ZRET') using the sentiment ('SENT') and presence of positive ('POS') and negative ('NEG') comments. The following equations were constructed:

1. The equation 1 defines the sentiment score for company *A* at time *t*, which is calculated as the sum of individual post sentiments on social media.
2. The equation 2 represents a time series of sentiment scores, capturing the evolution of social sentiment over time, which allows for dynamic analysis of how sentiment impacts stock prices.
3. The equation 3 calculates the return for company *A* at time *t* as the percentage change in price from the previous time period. It is a fundamental metric in financial analysis, used to assess investment performance. The returns are then standardized in the equation 4, and the time series was constructed with the standardized results in equation 5.
4. The equation 6 calculates the total positive sentiment for company *A* at time *t*, summing all positive post sentiments. It reflects the aggregate positive public perception at a given time. Conversely, the equation 7 calculates the total negative sentiment.
5. The positive variable (POS) for time *t* was created by aggregating all comments with positive sentiment (P >0) for time *t* (equation 8). Conversely, the negative variable (NEG) was created by aggregating all comments with negative sentiment (P<0) (equation 9).
6. Finally, the equation for the robust linear model regression is presented (equation 10).

$$SENT_{A,t} = \sum_{n=1}^{i} P_{n,t} \tag{1}$$

$$SENT_A = [SENT_{A,t}, SENT_{A,t+1}, SENT_{A,t+2}, \dots] \tag{2}$$

$$RET_{A,t} = \frac{PRICE_{A,t}}{PRICE_{A,t-1}} - 1 \tag{3}$$

$$ZRET_{A,t} = \frac{RET_{A,t} - \mu RET_A}{\sigma RET_A} \tag{4}$$

$$ZRET_A = [ZRET_{A,t}, ZRET_{A,t+1}, ZRET_{A,t+2}, \dots] \tag{5}$$

$$POS_{A,t} = \sum_{i=1}^{n} P_{i,t} \ \forall Pi > 0 \tag{6}$$

$$NEG_{A,t} = \sum_{i=1}^{n} P_{i,t} \ \forall Pi < 0 \tag{7}$$

$$POS_A = [POS_{A,t}, POS_{A,t+1}, POS_{A,t+2}, \dots] \tag{8}$$

$$NEG_A = [NEG_{A,t}, NEG_{A,t+1}, NEG_{A,t+2}, \dots] \tag{9}$$

$$ZRET_{i,t} = \alpha_{i,t} + \beta_1 SENT_{i,t} + \beta_2 POS_{i,t} + \beta_3 NEG_{i,t} + \epsilon_{i,t} \tag{10}$$

This methodology allowed for a comprehensive analysis of the impact of social media sentiment on stock returns across a large number of banks.

5. Results

The robust linear model results with lag = 0 are shown in Table 2, where all exogenous variables (NSS, negative comments, and positive comments) yielded statistically significant coefficients (p-value < 0.05) (see Table 2). Additionally, an asymmetric effect between the coefficients of the negative and positive comments is evident, indicating that negative comments have a larger negative impact than the positive impact of positive comments. This is concordant with the asymmetric effect hypothesis.

Table 2. Results of Robust Linear Model Regression

Robust linear Model Regression Results

Dep. Variable:	Ret	No. Observations:	64661
Model:	RLM	Df Residuals:	64657
Method:	IRLS	Df Model:	3
Norm:	HuberT		
Scale Est.:	mad		
Cov Type:	H1		
Date:	Mon, 13 May 2024		
Time:	10:11:25		
No. Iterations:	36		

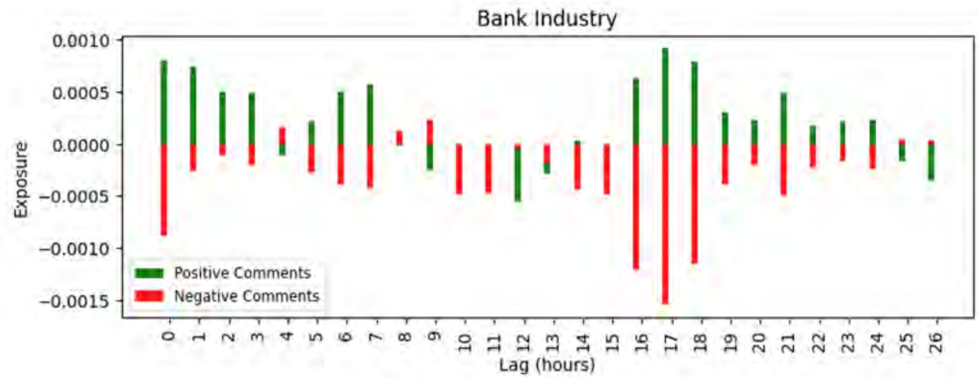
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0110	0.003	4.083	0.000	0.006	0.016
Sentiment	-0.0021	0.000	-4.705	0.000	-0.003	-0.001
Vader_Neg	-0.0009	0.000	-2.945	0.003	-0.001	-0.000
Vader_Pos	0.0008	0.000	2.941	0.003	0.000	0.001

Source: Prepared by the authors.

The statistically significant coefficients suggest that social media sentiment does indeed influence stock prices. The asymmetric effect observed, where negative comments have a larger impact than positive comments, supports the notion that investors and market participants may react more strongly to negative news or sentiment. This finding could have implications for firms in managing their social media presence and monitoring the sentiment of online discussions about their company. It highlights the importance of not only promoting positive sentiment but also actively managing and responding to negative comments to mitigate potential negative impacts on stock performance.

The analysis was repeated 24 times, each time incorporating a 1-hour lag to the exogenous variables to assess the evolution of the impact of positive and negative interactions on performance. The results, illustrated in Figure 3, show that performance exposure is maintained over time, that the signal intensity varies, and that the opposing effect between negative and positive comments is clearly sustained (see Figure 3). These results suggest that social media interactions not only have an almost immediate effect on stock performance, but that their impact can also be persistent over time with varying levels of intensity. Specifically, during the first eight hours after the comments are posted, their impact increases, and thus, the exposure of companies' performance to comment sentiment increases. Subsequently, the effect is dramatically reduced but continues to remain present for up to the following 24-26 hours. This finding suggests that monitoring and managing social media sentiment should be an ongoing effort rather than a one-time action, as the effects on stock performance are enduring and can vary in intensity over time.

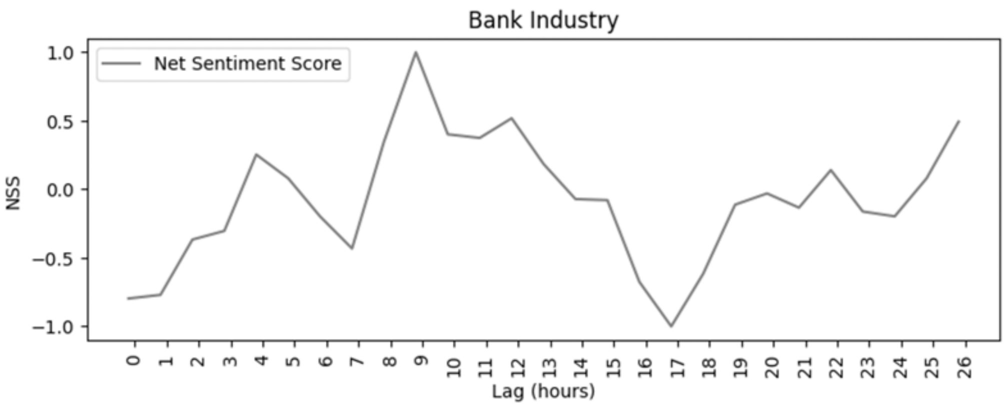
Figure 3. Influence of Positive and Negative Comments on Stock Performance Over Time



Source**:** Prepared by the authors.

The developed methodology enabled the calculation of the effect of the sentiment variable (SENT) in real time, which is widely known as the NSS. The results in Figure 4 intuitively demonstrate how NSS is aggregated using positive and negative comments, and how its lower scores are present when the coefficients of the negative comments increase negatively (see Figure 4). This highlights the importance of closely monitoring social media sentiment, particularly negative comments, as they can have a substantial impact on the overall sentiment score. This demonstration can aid in developing strategies to manage and mitigate negative sentiment to maintain a positive NSS.

Figure 4. Overall Net Sentiment Score Over Time



Source: Prepared by the authors.

6. Discussion

This study analyzed the impact of social media sentiment on stock performance, revealing that both positive and negative comments have a significant effect. Evans (1859) created a precedent regarding the influence of news on the stock market during the commercial crisis of 1857-1858, suggesting that information flow can impact market dynamics. However, the role of social media in shaping investor sentiment is a more recent phenomenon. Plourde (2023) discusses how consumers use social media as a platform for voicing their opinions about companies, which can influence brand perception and, indirectly, stock performance. Figure 1 captures the relationship of positive and negative comments in the performance of the banking industry. This is aligned with the findings of Reichheld (2011), who emphasized the importance of customer-driven strategies in enhancing company value.

Colicev and O'Connor (2020) argue that it is possible to differentiate the type of person making a comment depending on their relationship with the brand and calculate a direct influence. Nevertheless, our study results were conclusive without taking these distinctions into account. This suggests that while the source of a comment may influence its impact, the overall sentiment expressed in social media interactions remains a significant driver of brand value and stock performance.

Furthermore, Van Velthoven (2014) demonstrated a correlation between social media sentiment and net promoter score, indicating that positive online discussions can enhance brand loyalty and potentially affect stock prices. This is supported by the systematic review of Donthu *et al*. (2021), that mapped the electronic word-of-mouth research landscape and showed the growing significance of online sentiment in business outcomes. Figures 3 and 4 present the variation of the negative and positive comments impact through time and how this influences the NSC. Colicev and O'Connor (2020) found a positive impact of social media marketing on customer satisfaction, further emphasizing the link between online sentiment and company performance.

Mendoza-Urdiales *et al.* (2021) and Núñez-Mora and Mendoza-Urdiales (2023) provided empirical evidence of the direct relationship between social media comments and stock prices in individual firm analysis. Along with Kirtac and Germano (2024), this highlights the potential of leveraging large language models for sentiment trading. Furthermore, Li and Yang (2024) and Chung and Chang (2024) explored the broader implications of investor sentiment on market volatility and stock prices, respectively. Figure 2 presents the results of the combination of several approaches, in which the individual monitoring of several firms through a RLM captures the asymmetric effect by leveraging a robust natural language processing model.

7. Conclusion

The study reveals that firms are highly responsive to social media comments, with an immediate impact that persists over time and varies in intensity. It also shows that the NSS is more affected by negative comments than positive ones, suggesting a crucial insight for decision-makers monitoring this metric.

The study employed a bottom-up methodology to analyze the impact of social media on the stock performance of 167 publicly traded banks, revealing that both positive and negative comments have asymmetric effects on intraday data, with persistence in the signal post-commentary. The analysis also accounted for variability in the data, based on the popularity of each bank. Furthermore, the consistency of these findings across various types of studies, including different frequencies and modeling approaches (industry and firm-level), underscores the importance of natural language processing over the choice of statistical models in measuring the impact of social media interactions on stock performance.

This study provides a comprehensive understanding of social media sentiment's impact on publicly traded banks' stock performance. By employing a methodology that integrates natural language processing and RLM regression, the research demonstrates the significant and persistent influence of both positive and negative comments on stock prices. The findings highlight the importance of continuous monitoring and management of social media sentiment for firms, as it can substantially affect their stock performance and overall market perception.

Furthermore, the study reveals the asymmetric effect of positive and negative comments, with negative sentiment having a more pronounced impact on stock prices. This underscores the need for firms to develop strategies not only to promote positive sentiment but also to effectively address and mitigate negative comments to maintain a positive NSS.

The impact of social media comments and news on stock market performance has been extensively explored in recent decades. The 2017 Nobel Prize in Economics recipient, Richard H. Thaler, focused much of his research on how rationality influences decision-making, examining individual choices rather than their collective effect. These cognitive limitations can significantly impact financial markets. Thaler's work on behavioral economics (Thaler, 2017) highlights the role of psychological factors in economic decisions, which can be seen in how investors react to social media sentiments.

While this study observes how social media sentiment influences stock prices, it is important to acknowledge certain limitations that could affect the breadth and applicability of the findings. Firstly, the period under study may not fully capture long-term trends or the impact of extraordinary market events (seven months, from October 2022 to April 2024). Additionally, the reliance on social media data, which can be inherently biased, introduces potential variability in the sentiment. The information used in the study, extracted from public social media, may not represent all factors influencing stock market behavior, such as undisclosed financial information, which can also play a critical role. Acknowledging these constraints is crucial for a balanced interpretation of the results and for future research.

In conclusion, the intersection of behavioral economics and social media sentiment analysis presents a fertile ground for further research. Understanding the psychological underpinnings of investor reactions to online information can provide valuable insights into market behavior and inform strategies for managing the impact of social media on stock performance.

https://doi.org/10.36105/theanahuacjour.2024v24n1.10

Further work could focus on grouping comments according to user type to analyze whether a varying level of influence among those groups exists. It can also analyze the asymmetric effect of negative and positive comments over time across the broader financial market. Additionally, further investigation into the mechanisms behind the asymmetric effect of positive and negative sentiment could provide deeper insights into investors' psychology and market dynamics.

References

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Chung, M.H., & Chang, Y.K. (2024). Financial reporting complexity, investor sentiment, and stock prices. *Finance Research Letters*, 62(Part A), article 105026. https://doi.org/10.35692/07183992.13.1.8

Colicev, A., & O'Connor, P. (2020). How Social Media Impacts Brand Value: The Mediating Role of Customer Satisfaction. *Multidisciplinary Business Review*, 12(1), 15-26. https://doi.org/10.35692/07183992.13.1.8

Croux, C., & Rousseeuw, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. Computational Statistics, Physica-Verlag HD.

Donthu, N., Kumar, S., Pandey, N., Pandey, N., & Mishra, A. (2021). Mapping the Electronic Word-of-Mouth (eWOM) Research: A Systematic Review and Bibliometric Analysis. *Journal of Business Research*, 135, 758-773 https://doi.org/10.1016/j.jbusres.2021.07.015

Evans, D. M. (1859). *The history of the commercial crisis, 1857-58, and the stock exchange panic of 1859*. Groombridge and Sons.

Huber, P. J. (1973). The 1972 Wald Memorial Lectures: Robust regression: Asymptotics, conjectures, and Monte Carlo. *The Annals of Statistics*, 1(5), 799-821. https://www.jstor.org/stable/2958283

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons.

Kirtac, K., & Germano, G. (2024). Sentiment Trading with Large Language Models. *Finance Research Letters*, 62(Part B), article 105227. https://doi.org/10.1016/j.frl.2024.105227

Li, J., & Yang, J. (2024). Financial shocks, investor sentiment, and heterogeneous firms' output volatility: Evidence from credit asset securitization markets. *Finance Research Letters*, 60, article 104860. https://doi.org/10.1016/j.frl.2023.104860

Mendoza-Urdiales, R. A., García-Medina, A., & Núñez-Mora, J. A. (2021). Measuring information flux between social media and stock prices with Transfer Entropy. *PLOS ONE*, 16(9), article 0257686. https://doi.org/10.1371/journal.pone.0257686

Mendoza-Urdiales, R. A., Núñez-Mora, J. A., Santillán-Salgado, R. J., & Valencia-Herrera, H. (2022). Twitter Sentiment Analysis and Influence on Stock Performance Using Transfer Entropy and EGARCH Methods. *Entropy*, 24(7), 874. https://doi.org/10.3390/e24070874

Núñez-Mora, J. A., & Mendoza-Urdiales, R. A. (2023). Social sentiment and impact in US equity market: an automated approach. *Social Network Analysis and Mining*, 13, 111. https://doi.org/10.1007/s13278-023-01116-6

Plourde, A. (2023). Criticizing companies via social media and digital platforms: an alternative forum of justice for consumers? *Option consommateurs*. CID: 20.500.12592/m8xnp1. https://canadacommons.ca/artifacts/8246598/criticizing-companies-via-social-media-and-digital-platforms/9163537/

Reichheld, F. (2011). *The Ultimate Question 2.0 (revised and expanded edition): How Net Promoter Companies Thrive in a Customer-Driven World*. Harvard Business Review Press.

Tetlock, P. (2007). Giving content to Investor Sentiment: The role of Media in the Stock Market. *The Journal of Finance*, 62(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thaler, Richard H. (2017). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2017. Nobel Prize 2017.https://www.nobelprize.org/prizes/economic-sciences/2017/thaler/facts/

Van Velthoven, S.T.M. (2014). *Sentiment analysis on social media to predict Net Promoter Score* (master's thesis), Eindhoven University of Technology Research. https://pure.tue.nl/ws/portalfiles/portal/46988096/783003-1.pdf

About the authors

José Mendoza is a distinguished professional in the technology and financial sectors, with over three decades of experience in leadership roles and systems development. He demonstrates exceptional skill in integrating technological solutions into complex business operations. Throughout his career, José has led a variety of innovative projects, excelling in the implementation of biometric systems and customer identification platforms that have transformed banking and retail operations in various countries. His commitment to continuous technological advancement and his ability to adapt new technologies to business demands highlight his strategic vision and executive competence. José contributes not only his vast technical knowledge but also his practical experience in process optimization and the application of strategic technologies, which is essential for companies in the technology and financial sectors aspiring to lead in a highly competitive market.

mendoza@quantumanalytics.mx
https://orcid.org/0009-0006-7804-4849

Román Mendoza is a leader in the artificial intelligence industry within consulting, working on developing solutions, strengthening capabilities, and generating knowledge in the field. His work focuses on delivering cutting-edge projects that integrate big data analytics and natural language processing to drive data-driven decision-making. With a solid track record of over 12 years in consulting projects, Román has specialized in implementing advanced technologies for operations research.

roman@quantumanalytics.mx
https://orcid.org/0000-0003-2888-156X

Lineamientos para los autores

The Anáhuac Journal es una revista semestral de la Facultad de Economía y Negocios de la Universidad Anáhuac México.

The Anáhuac Journal recibe artículos inéditos que no se hayan publicado en ningún medio impreso o electrónico, ni que hayan sido postulados de forma simultánea para su publicación en otras revistas u órganos editoriales.

Los artículos deben ser producto de investigaciones y estudios con resultados originales en las siguientes líneas temáticas:

a) Derecho corporativo
b) Gobierno corporativo
c) Responsabilidad social empresarial
d) Ética empresarial
e) Liderazgo y dirección
f) Emprendimiento e innovación
g) Mercadotecnia
h) Economía y estrategia
i) Finanzas y contaduría
j) Toma de decisiones
k) Economía
l) Innovación

Sobre el proceso

1. *The Anáhuac Journal* utiliza el sistema de gestión editorial OJS (Open Journal Systems), de ahí que todos los manuscritos enviados por los autores deben ser ingresados en este sistema. Para ello, los autores necesitan registrarse y enviar sus artículos acompañados de lo siguiente:

   – Carta de originalidad. Todos los autores y coautores deben firmarla y otorgar sus datos.

   https://www.anahuac.mx/mexico/files/Carta-de-originalidad_Revistas-Univ-Anahuac-Mexico.docx

   – Autorización para comunicación pública de obra literaria (artículo)

   https://www.anahuac.mx/mexico/files/Autorizacion %20para %20comunicacion %20publica %20de %20obra %20literaria_Revistas %20Univ %20Anahuac %20Mexico.docx

2. Los artículos se enviarán a dos árbitros, quienes determinarán en forma anónima: a) publicarlo sin cambios, b) publicarlo cuando se hayan realizado correcciones menores, c) publicarlo una vez que se haya efectuado una revisión a fondo o d) rechazarlo. En caso de discrepancia entre ambos resultados, el texto se enviará a un tercer árbitro, cuya decisión definirá su publicación. Los resultados del proceso de dictamen académico serán inapelables en todos los casos. Los trabajos enviados por académicos serán siempre sometidos a consideración de árbitros externos a su institución.

3. Los trabajos autorizados por los árbitros se publicarán en la revista y se notificará a sus autores sobre esta decisión en un plazo máximo de cuatro meses a partir de la recepción del documento.

4. La revista se reserva el derecho de hacer los cambios editoriales que considere pertinentes.

Requisitos y aspectos formales para la presentación de los trabajos

Como parte del proceso de envío, los autores/as están obligados a comprobar que su envío satisfaga todos los elementos que se muestran a continuación. Se devolverá a los autores/as aquellos envíos que no cumplan estas directrices. Por políticas editoriales, todos los envíos deben hacerse en esta plataforma, por lo tanto, no se recibirán manuscritos enviados directamente a través de correo electrónico.

Estructura mínima del trabajo:

- Formato WORD (formato.doc o .docx), tipografía Arial de 12 puntos, interlineado de 1.5 cm.
- La extensión máxima es de 25 hojas tamaño carta (215.9 x 279.4 mm). incluyendo tablas, figuras, referencias y apéndices.
- Los artículos pueden ser escritos en español o inglés y deben incluir el título y un resumen en ambos idiomas. En todos los casos, incluyendo los artículos en español, la calidad del estilo de la versión final es total responsabilidad del autor o autores.
- El resumen será de hasta 180 palabras; deberá incluir máximo 5 palabras clave y la clasificación JEL (Journal of Economic Literature), que puede descargarse de la siguiente liga: https://www.aeaweb.org/econlit/jelCodes.php?view=jel
- El título, resumen y palabras clave deben ir al comienzo del artículo, en ambos idiomas.
- Deberá incluir una introducción que refleje con claridad los antecedentes del trabajo, su desarrollo y conclusiones.
- Las notas de pie de página deberán estar en la hoja correspondiente y deberán usarse para aclarar conceptos o cuestiones editoriales, pero no para las referencias bibliográficas.
- Las referencias bibliográficas se harán según las normas de la APA 7.ª edición. No deben extenderse de modo innecesario y deberán aparecer completas en páginas separadas, ordenadas alfabéticamente y, para cada autor, en orden cronológico, del más antiguo al más reciente. Deben tener la información completa sobre la fuente respectiva, incluyendo el DOI (Digital Object Indentifier) cuando esté disponible, y deben insertarse al final del artículo, antes de cualquier apéndice. Los autores deben asegurarse de que haya una correspondencia estricta entre los nombres y años reconocidos en el texto y aquellos listados en la bibliografía, es decir, todos los trabajos citados deberán aparecer en las referencias bibliográficas.
- Las referencias bibliográficas se harán según las normas de la APA 7.ª edición que establecen, entre otras, lo siguiente:

  a) *Libros*: Autor (apellido e inicial del nombre). Año de publicación (entre paréntesis). Título del libro (en cursivas), número de edición (entre paréntesis y solo si se considera importante). Editorial.

  Ejemplo: Castel, R. (1997). *Las metamorfosis de la cuestión social. Una crónica del asalariado* (1.ª ed.). Paidós.

  Libro con más de un autor: Autor(es) (apellido e inicial del nombre). Año de publicación (entre paréntesis). Título del libro (en cursivas). Edición (entre paréntesis), Lugar de edición (:), Editorial: De Mattos, C. y Ducci, M.E. (2005). Santiago en la globalización: ¿una nueva ciudad? (2.ª ed.). Lom.

  Ejemplo: De Mattos, C. y Ducci, M.E. (2005). *Santiago en la globalización: ¿una nueva ciudad?* (2.ª ed.). Lom.

Note que si el libro está escrito por 3 más autores, deberá indicarse con un *et al.* después del primer autor (del latín *et alia,* y otros).

Ejemplo: Dellanegra, G. *et al*. (1983). *Los países del Atlántico Sur: geopolítica de la Cuenca de la Plata*. Pleamar.

b) Artículo de revista impresa: Apellido, inicial del nombre. Año de publicación (entre paréntesis). Título del artículo en redondas y entre comillas. Nombre de la revista (en cursivas), Volumen de la revista, número de edición (entre paréntesis), intervalo de páginas en el que se encuentra el artículo.

Ejemplo: Oszlak, O. (2009). «El Estado transversal». *Encrucijadas UBA*, 8 (26), 2-4.

c) Artículo de revista en internet: Apellido, inicial del nombre. Año de publicación (entre paréntesis). Título del artículo (en redondas y entre comillas). Nombre de la revista (en cursivas). Volumen de la revista, número de edición (entre paréntesis). Página de internet. Ejemplo: Gadner, H. (1983). «La teoría de las inteligencias múltiples». *Revista Española de Investigación en Educación,* 9 (2). http://urlinventada.es

Cualquier otra fuente, como base de datos o enciclopedia en línea debe integrar datos detallados del texto citado: autor si lo hay, y si no, directamente el título del «texto entre comillas», año si lo hay y si no lo hay, colocar (s/f) (sin fecha). Y la dirección de internet después de punto.

En el caso de que el texto tenga DOI, incluirlo. Esta es la página para búsqueda de DOI: http://search.crossref.org/?q=

Ejemplo: Morey, C. C. *et al.* (2015). «The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory». *Archives of Scientific Psychology*, 3, 18-29. https://doi.org/10.1037/arc0000014

- Las ilustraciones, fotografías, mapas, diagramas, dibujos, deben tener sus leyendas correspondientes, títulos, una numeración consecutiva, la fuente de donde se obtuvo la información (en caso de ser elaboración propia, hay que señalarlo de esa manera) y deben estar libres de derechos. En caso de imágenes, ilustraciones o dibujos deben estar en alta resolución (300 dpi); en caso de tablas y gráficas deberá enviarse aparte el archivo original en el que fueron creadas (excel, power point, etcétera).

- Los títulos de cuadros y gráficas deben realizarse en tablas de word o excel e ir secuenciados en números arábigos. El cuadro o gráfica debe llevar encabezado en letra negrita tipo Times New Roman y al centro. En la parte inferior debe anotarse la fuente con letra de 10 puntos. En los cuadros debe citarse la fuente (autor y año); con el siguiente formato: bordes (líneas) internos y sencillos, sin bordes izquierdo y derecho en los títulos de cada columna. Las gráficas con borde suave y las barras y líneas en tonos grises. OJO: las fuentes de tablas o gráficos que no sean elaboración propia deben también ir incluidos en las Referencias bibliográficas.

- Las pruebas matemáticas largas y tablas muy detalladas y extensas deberán estar en un apéndice o, en su caso, omitirse. Los autores deberán hacer un esfuerzo por explicar los resultados del significado de las pruebas matemáticas.

- Las ecuaciones deberán presentarse en líneas separadas y centradas. Deberán estar numeradas consecutivamente, en el margen derecho, usando números arábigos entre paréntesis.

- Información del autor(es): en la última página deberán incluirse los datos generales del autor (es): nombre completo, centro o departamento al que se encuentra(n) adscrito(s) laboralmente, dirección postal institucional, dirección de correo electrónico, orcid y un breve resumen de su experiencia académica (no mayor a 250 palabras).

- Los manuscritos deben ir acompañados de la Carta de originalidad y la Autorización para comunicación pública de obra literaria (artículo).

Acerca de este sistema de publicación

Esta revista utiliza Open Journal Systems (https://openjournalsystems.com), que es un gestor de revistas de acceso abierto y un *software* desarrollado, financiado y distribuido de forma gratuita por el proyecto Public Knowledge Project sujeto a la Licencia General Pública de GNU.

Guidelines for authors

The Anáhuac Journal is a semiannual publication
from the School of Business and Economics
at Universidad Anáhuac Mexico.

The Anáhuac Journal receives unpublished articles that have not appeared in any print or electronic media, nor have been simultaneously proposed for publication in other journals or editorial entities.

Articles should come from research and studies, and offer original results in the following subject areas:

a) Corporate law
b) Corporate government
c) Corporate social responsibility
d) Corporate ethics
e) Leadership and Management
f) Entrepreneurship and Innovation
g) Marketing
h) Economics and Strategy
i) Finance and Accounting
j) Decision making
k) Economics
l) Innovation

About the process

1. *The Anáhuac Journal* uses the OPJ (Open Journal Systems) editorial management system, and all manuscripts submitted shall be uploaded to this system. Authors are required to register and submit their articles along with:

   – Letter of Originality

   All authors and co-authors sign it and provide its data.

   https://www.anahuac.mx/mexico/files/Statement %20of %20Originality.docx

   – Authorization for public communication of literary works (article)

   https://www.anahuac.mx/mexico/files/Authorization %20for %20public %20communication %20of %20literary %20work_Revistas %20Univ %20Anahuac %20Mexico.docx

2. Works are submitted to two judges, who will anonymously determine whether: a) to publish the work without changes, b) to publish the work with minor corrections, c) to publish the work after a complete review, or d) to reject the work. In the event the judges are not in agreement in their decision, the text will be sent to a third judge, whose decision will be final. All works submitted will be considered by judges not affiliated with the author's institution.

3. Accepted articles will be published in the journal and authors will be notified of this decision within four months from the submission of the original document.

4. The journal reserves the right to make the editorial changes it deems necessary.

Requirements and formal aspects for submitting works

As part of the submission process, authors are required to confirm their article meets all the following elements. Submissions that do not meet these guidelines will be returned to the author. In accordance with editorial policies, all submissions shall be received through this platform; manuscripts that are sent directly by e-mail will not be received.

Minimum structure of the work:

- WORD format (.doc or .docx), Arial Font, 12 points and 1.5 cm spacing.
- The maximum length is 25 pages letter size (215.9 x 279.4 mm), including tables, charts, references and appendices.
- Articles may be written in Spanish or English and should include title of the article and an abstract in both languages. In all cases, including articles in Spanish, the stylistic quality of the final version is the responsibility of the author or authors.
- Abstract of up to 180 words, which should include 5 keywords maximum and the JEL (Journal of Economic Literature) classification, which can be downloaded from: https://www.aeaweb.org/econlit/jelCodes.php?view=jel
- The title, abstract and keywords in Spanish and English should be placed at the beginning of the article.
- An introduction should be included that clearly reflects the work's background, development and conclusions.
- Footnotes should appear on the corresponding page and should not be used for bibliographical references.
- Bibliographical references should follow APA 7th ed. rules. They should not be extended unnecessarily, and should be complete and on separate pages, in alphabetical order, and each author should appear chronologically from the earliest to the most recent. They should have all the information for the corresponding source, including the DOI (Digital Object Identifier), when available, and they should be inserted at the end of the article, before any appendix. Authors should ensure that names and years that appear in the text and those listed in the bibliography match exactly that is, all the works cited in the text should appear in the bibliographical references.
- The bibliographical references should follow APA 7th ed. rules (briefly mentioned following, for further information, visit www.apastyle.org):

    a) *Books*. Author (last name and initials). Year of publication (in brackets). Title of the book (in italics), edition (in brackets and only if important). Publisher. If there is no publisher, write [s.n.], from the Latin *sine nomine*, which means «without name».

    Example: Castel, R. (1997). *Las metamorfosis de la cuestión social. Una crónica del asalariado* (1st ed.). Paidós.

    Books that have more than one author: Author(s) (last name and initials). Year of publication (in brackets). Title of the book (in italics). number of edition (in brackets), Publisher.

    Example: De Mattos, C. and Ducci, M.E. (2005). *Santiago en la globalización: ¿una nueva ciudad?* (2nd ed.). Lom.

Note that if the article or book is written by 3 authors or more, after the first author write *et al*., from the Latin *et alia* (and others):

Example: Dellanegra, G. *et al*. (1983). *Los países del Atlántico Sur: geopolítica de la Cuenca de la Plata*. Pleamar.

b) Article in a print publication: Last name, initials. Year of publication (in brackets). «Title of the article» (in quotation marks). Name of the publication (in italics), volume, issue (in brackets), pages on which the article appears.

   Example: Oszlak, O. (2009). «El Estado transversal». *Encrucijadas UBA,* 8 (26), 2-4.

c) Article published online: Last name, initials. Year of publication (in brackets). Title of the article (in quotation marks). Name of the publication (in italics). Volume, issue (in brackets). website.

   Example: Gadner, H. (1983). «La teoría de las inteligencias múltiples.» *Revista Española de Investigación en Educación*, 9 (2). http://urlinventada.es

   Any other source, such as a data base or online encyclopedia should include detailed data about the cited text: author (if any), if there is no author, write the title of the text, year (if any, in brackets; if there is no year, n/y), and the website address after the dot.

   If the text has a DOI, please be sure to include it at the end of the reference. The DOI can be searched at: www.crossref.org

   Example: Morey, C. C. *et al.* (2015). «The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory.» *Archives of Scientific Psychology, 3*, 18-29. https://doi.org/10.1037/arc0000014

- Illustrations, photographs, maps, diagrams and drawings must have their corresponding legends, titles, sequential numbering, and the source of information (when created by the author, this should be indicated) and all images should be free of rights. Images, illustrations or drawings should be in high resolution (300 dpi); tables and charts should be sent separately in the original file in which they were created (Excel, Power Point, etc.).

- Titles of tables and charts should be made in Word or Excel tables with a numeric sequence in Arabic numbers. The heading for the table or chart should be centered and in Times New Roman. Sources should be in font size 10, at the bottom. Tables must cite their source (author and year). The format should be the following: internal and simple borders (lines), without left and right borders on the titles of each column. Charts should have a thin border, and bars and lines should have grey tones. NOTE: the sources for tables or charts that are not created by the author should also be included in the bibliographical references.

- Long mathematical proofs and highly detailed and long tables should be placed in an appendix, or left out. Authors should make an effort to explain the result of mathematical proofs.

- Equations should be presented on separate and centered lines. They should be numbered and sequenced, and the number placed in the right margin in Arabic numbers, in brackets.

- Author information: general information about the author(s) should be included on the last page: full name, center or department and/or university, zip code of the institution, e-mail address, orcid, and a summary of their academic experience (no more than 250 words).

- Manuscripts should be accompanied by the Letter of Originality and Authorization for public communication of literary works (article).

About this publication system

This journal uses Open Journal Systems (https://openjournalsystems.com), an open access journal manager and a software developed, funded and distributed freely by the Public Knowledge Project subject to the GNU Public General License.