SBE
DOI 10.2478/sbe-2026-0005
Studies in Business and Economics: no. 21(1) 2026

SAMPLE SIZE MATTERS: A COMPARATIVE ANALYSIS OF THE I.I.D. ASSUMPTION AND RISK-RETURN ESTIMATIONS IN LATIN AMERICAN AND US STOCK MARKETS

María Inés BARBOSA CAMARGO
Universidad de La Salle, Bogota, Colombia

José Rodrigo VÉLEZ MOLANO
Universidad de La Salle, Bogota, Colombia

Jorge Mario SALCEDO MAYORGA
Universidad de La Salle, Bogota, Colombia

Abstract:
This study examines the impact of sample size on the independent and identically distributed (i.i.d.) assumption in financial time series and its subsequent effect on risk-return estimations. Focusing on Latin American and US stock market indices from August 2007 to December 2024, using 4,477 daily log-returns, our methodology employs moment estimations, i.i.d. tests (Ljung-Box and Augmented Dickey-Fuller), and a modified Capital Asset Pricing Model (CAPM) and Download CAPM across four rolling windows (60, 250, 500, and 1,000 days). Our findings show that financial returns consistently exhibit heavy tails and negative skewness, challenging the assumption of normality. Statistical properties and i.i.d. violations vary significantly with sample size, with tests becoming more stringent for larger samples. CAPM alpha and beta coefficients are also sample size dependent, revealing distinct risk-return profiles: the S&P 500 exhibits higher systematic risk (Beta > 1) with performance explained by the model (alpha ≈ 0), while Latin American markets are more defensive (Beta < 1) with more dispersed alphas, indicating greater influence from local factors. These results show the importance of considering sample size for robust financial modeling and informed investment decision-making, particularly in emerging markets, highlighting that no single model is universally applicable to all stock markets.

Key words: *Sample Size, Returns, High-order Moments, i.i.d. assumptions, Emerging markets.*

1. Introduction

The conceptualization and implementation of investment strategies present a complex challenge, further complicated by the need for accurate performance assessment. Practitioners and researchers have developed numerous models to estimate profitability and risk; however, these tools typically rely on statistical methods requiring strict assumptions regarding return behavior, particularly their distribution and sample moments.

For instance, Markowitz (1952) used mean and variance to approximate expected utility and optimize asset combinations. While influential, this model propagated "myths" regarding: i) proper return distributions; ii) mean-variance sufficiency; and iii) the stability of the mean-variance relationship. Markowitz (2019) acknowledged a "Great Confusion" arising from the belief that Gaussian distributions were a necessary condition for his model, rather than merely sufficient. Subsequent models, such as the Capital Asset Pricing Model (Sharpe, 1964) and Arbitrage Pricing Theory (Ross, 1976), also rely on the i.i.d. assumption, implying invariant distribution moments.

Despite extensive literature demonstrating that the i.i.d. assumption is frequently violated in financial markets, the direct impact of sample size on its validity and implications for financial models remains under-investigated. This gap is critical because assumption reliability varies with sample size, directly affecting risk and return accuracy. Consequently, this study determines how sample size influences i.i.d. validity and CAPM Alpha (α) and Beta (β) coefficients, particularly in emerging markets. This investigation aims to bridge the disconnect between identified "myths" and practical application by systematically evaluating the role of data length in validating foundational premises.

To address this, we compare stock indices from Latin America (Colombia's COLCAP, Mexico's IPC, Chile's IPSA, Peru's SPBVL, and Brazil's BOVESPA) with the United States (S&P 500), utilizing the MSCI World Index as the market portfolio. The methodology spans August 2007 to December 2024, yielding 4,477 observations and including events such as the 2008 financial crisis and the COVID-19 pandemic. We employ fixed-length moving windows of 60, 250, 500, and 1,000 days to assess how sample sizes affect data characteristics. Key techniques include moment estimations; i.i.d. tests, specifically Ljung-Box and Augmented Dickey-Fuller (ADF); and modified CAPM regressions to evaluate alpha and beta estimations. Additionally, D-CAPM is estimated to check robustness against model selection error.

This paper contributes to the literature in two ways. First, it details return distributions through numerous independent and identically distributed (i.i.d.) tests, offering novel insight into Latin American market complexities. Second, the findings assist researchers, investors, and policymakers in adjusting financial models by highlighting critical differences in market behaviors across sample sizes, particularly contrasting emerging markets with the United States.

The paper is structured as follows: Section 2 reviews the literature on financial return assumptions. Section 3 details the methodology, followed by the results in Section 4 and conclusions in Section 5.

2. Literature review

Quantitative finance bridges theoretical models and statistical methods to address practical applications like investment decisions and risk optimization. Early contributions, stemming from Markowitz's (1959) foundational work, laid the groundwork for pivotal models including the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966), Fama & French's (1993) three-factor model, and Ross's (1976) Arbitrage Pricing Theory (APT). Additionally, Autoregressive Integrated Moving Average (ARIMA) models have proven fundamental in capturing dynamic dependencies, guiding forecasting and investment strategies (Ahmed et al., 2023).

A core tenet of these models is the assumption that returns are independently and identically distributed (i.i.d.) and normal (Tsay, 2010). While facilitating data-driven modeling, empirical tests frequently reveal anomalies and deviations from these classical assumptions (Fu et al., 2024; Molero-Gonzalez et al., 2023). These deviations are particularly pronounced in emerging markets, characterized by high volatility and institutional instability (Azher & Iqbal, 2018).

The empirical reality of financial markets consistently diverges from idealized normal and i.i.d. distributions. Cont (2001) highlighted "stylized facts" common across assets, including heavy tails, volatility clustering, and leverage effects. This complexity suggests asset pricing models must vary across market regimes, underscoring the necessity for parsimonious models in emerging markets (Ali, 2022).

Empirical evidence consistently confirms that portfolio returns are not normally distributed, limiting investors' knowledge of future distributions (Al-Yahyaee et al., 2020; Chae & Lee, 2018; Pekar & Pcolar, 2022). Markets such as the Eurozone and the US exhibit heavy tails, skewness, and kurtosis (Anagnostidis et al., 2016; Andreu & Torra, 2009). The presence of heavy tails is critical; extreme events like financial crises can lead to severe risk underestimation under normal distribution assumptions (Eom et al., 2023; Jondeau et al., 2019). Consequently, relying on normality for Value-at-Risk (VaR) calculations can result in significantly higher actual losses, necessitating that investors account for non-normal distributions (Andreu & Torra, 2009; Chae & Lee, 2018).

Beyond distributional properties, the risk-return relationship manifests "long-term memory" (LTM), where past losses persist, highlighting the role of idiosyncratic risk (Eom et al., 2023; Dai & Harris, 2023). Since standard CAPM Beta relies on variance and fails to differentiate between losses and gains, alternative measures incorporating non-symmetric properties are being explored (Ding, 2023), including the time-frequency dynamics of beta (Mestre, 2023). Other studies compare factor model pricing ability (Fama & French, 2018; Ahmed et al., 2023) or identify factors using random matrix theory (Molero-Gonzalez et al., 2023). Evidence indicates that skewness and kurtosis significantly explain cross-sectional returns, though this relationship depends on the specific period (Jondeau et al., 2019; Dai & Harris, 2023).

In the context of market efficiency, both the Efficient Market Hypothesis (EMH) and the Box-Jenkins methodology require that returns exhibit no autocorrelation (Tsay, 2010; Fama, 1970). The Ljung-Box test is commonly employed to verify this zero-autocorrelation assumption (Muriel, 2025). However, real-world data often complicates this ideal, as inherent dependence challenges unpredictability assumptions. Furthermore, the Ljung-Box test relies on underlying assumptions—such as weak stationarity and asymptotic normality—that are often unmet in practice, impacting the robust assessment of market efficiency (Hassani & Yeyanegi, 2019; Muriel, 2025).

In financial risk management, forecasting procedures directly impact backtesting and estimation error (Barendse, Kole & Van Dijk, 2023). While VaR has limitations regarding asymmetric and heavy-tailed properties (Eom et al., 2023), robust models should incorporate higher-order moments (Lin et al., 2014). Incorporating conditional skewness and kurtosis improves performance over normal assumptions (Bali, Mo & Tang, 2008; Barendse, Kole & Van Dijk, 2023; Andreu & Torra, 2009). Analyzing these moments is key for understanding downside/upside risk connections and capturing fat-tail risk transmission during extreme events (Azimli, 2024; Min et al., 2023).

Despite extensive literature, research incorporating sample size impact on the i.i.d. assumption remains limited. This study addresses this gap. Statistical properties vary significantly between short-term (small samples) and long-term (large samples) data (Dutta & Kanti, 2023). Small samples often display leptokurtosis and negative skewness, whereas large samples tend toward Gaussian distributions. Recent studies employ rolling windows to capture these dynamic characteristics (Eom et al., 2023; Molero et al., 2023), revealing relationships between skewness and future returns (Jondeau et al., 2019; Dai & Harris, 2023; Ding, 2023).

These approaches reveal that while risk profiles can be time-robust, short-run betas may differ from long-run betas (Mestre, 2023). Furthermore, persistent negative skewness in developing markets reinforces the need for dynamic analysis (Pekar & Pcolar, 2022). Crucially, statistical tests such as Ljung-Box and Augmented Dickey-Fuller (ADF) become "more stringent" as observations increase (Muriel, 2025; Tsay, 2010).

This implies that conclusions regarding serial dependence and stationarity depend on sample size, directly impacting the validity of the i.i.d. assumption. By presenting numerous i.i.d. test iterations, this paper offers a new understanding of these dynamic properties and their implications for models like CAPM, particularly in emerging markets. This analysis bridges a significant gap by systematically evaluating the role of data length in validating foundational premises, serving as the basis for the following hypotheses:

H1: The assumption of independently and identically distributed (i.i.d.) returns is significantly influenced by sample size. Smaller samples exhibit greater deviations from classical assumptions compared to larger samples, particularly in emerging markets.

H2: Statistical test outcomes, such as autocorrelation (e.g., Ljung-Box) and stationarity (e.g., ADF), are sensitive to sample size in emerging markets. Different conclusions based on data length directly influence the perceived validity of the i.i.d. assumption and the effective application of financial models.

H3: Developing accurate risk models and informed investment decisions requires considering higher-order moments and sample size, as conclusions regarding market efficiency and model effectiveness are particularly sensitive in emerging markets.

3. Methodology

Data and subsamples

The data used are the daily closing prices of the following indexes: COLCAP (Colombia), IPC (Mexico), IPSA (Chile), SPBVL (Peru), BOVESPA (Brazil), the SP 500 (United States), and the MSCI World Index. The time series are downloaded from the stock exchange's official web pages and S&P/ MSCI web pages. The database contains 4,477 observations, ranging from August 16, 2007, to December 31, 2024. All index series exclude dividends. The period and the number of observations allow for the inclusion of several key events: the financial crisis, volatility clusters, the COVID-19 pandemic, and various market trends (bearish, bullish, and sideways).

As Tsay (2010) stated, when a significant amount of information (observations) is available, as in this case, 4,477, it is possible to conduct empirical and statistical tests with a subsample, which in turn would help assess the consistency of the results. Hence, the data analysis is divided into four fixed (in length) moving windows: 60, 250, 500, and 1,000 days. This division has two goals. The first objective is to enrich the description of the return distributions by showing the variability of the estimated moments across several samples. The second is to test at which lengths and consistency the classical assumptions of i.i.d. are held, and how this might impact the application of a CAPM-like model and the consistency of Beta and Jensen's alpha estimations.

The analysis is made using the daily log-returns (r_t), calculated with the closing prices (p_t, p_{t-1}) as:

$$r_t = [\ln(p_t) - \ln(p_{t-1})] \times 100 \quad for\ t = 1, \cdots, T \quad (1)$$

Moment Estimation

For each index and sample size, the moments are estimated using the widely used expressions (see Table 1).

Table 1. Moments' estimation formulas

MOMENT	FORMULA	NUMBER OF ESTIMATIONS FOR EACH MOMENT, INDEX AND SAMPLE SIZE
MEDIA	$\bar{r} = \frac{\sum r_t}{T_s}$ (2)	60: 4,416 250: 4,226 500: 3,976 1,000: 3,476
VARIANCE	$s^2 = \frac{\sum(r_t-\bar{r})^2}{T_s}$ (3)	
SKEWNESS	$s^3 = \frac{\sum(r_t-\bar{r})^3}{T_s}$ (4)	
KURTOSIS	$s^4 = \frac{\sum(r_t-\bar{r})^4}{T_s}$ (5)	

Source: Authors' calculations.

Where T_s is the sample size.

Independent and Identically Distributed Test

<u>Ljung-Box</u>

Given that independent random variables have no memory, the independence test is based on the autocovariance concept, explicitly trying to find no serial autocorrelation. To achieve this, the Ljung-Box test (LB) states as a Null Hypothesis that there are no significant autocorrelations in the whole series (Ljung & Box, 1978). Formally, the test statistic is:

$$Q = T_s(T_s + 2)\sum_{k=1}^{m} \frac{\hat{\gamma}_k^2}{T_s - k} \quad (6)$$

Where T_s is the sample size and $\hat{\gamma}_k$ is the estimated autocorrelation of grade k. Q is distributed as a Chi-squared with $m - p - q$ degrees of freedom, where m is the number of autocorrelations to be estimated, p the number of lags for autoregressive part and q the number of lags for the moving average part.

In the context of this paper, the test is performed on each index for every sample size for every date. As an example, for the S&P 500 and sample size 60, the test is performed on the first 60 observations (from 1 to 60), then on the next 60 (from 2 to 61), from 3 to 62, and so on; and for each of these rolling windows the test is run from lag 1 to 30, sequentially, to find out the lag from which the Null is rejected. This process illustrates how often the Null Hypothesis is rejected, in other words, how often the log-returns are independent and identically distributed (i.i.d.).

Augmented Dickey-Fuller

Said and Dickey (1984) redefined the Dickey-Fuller test to accommodate more general time-series structures, and it is now known as the Augmented Dickey-Fuller (ADF) test. This test has the advantage of greater power when testing stationarity in a time series, given the inclusion of higher autoregressive orders, as in ARMA models. The general model or equation is:

$$r_t - r_{t-1} = (\rho - 1)r_{t-1} + (\alpha + \beta)(Z_{t-1} - \beta Z_{t-2} + \beta^2 Z_{t-3} - \cdots) + e_t \quad (7)$$

Where $Z_t = r_t - r_{t-1}$ and e_t is an independent, identically distributed error. Under the Null Hypothesis, $\rho = 1$, the process has unit-roots; in other words, it is not stationary. Equation (7) can be modified to test unit roots, including a constant and/or a trend term.

Three types of tests are performed: with no constant, with no trend, with only a constant, and with both a constant and a trend. Additionally, the issue of specifying the max lag to use in equation (7) is calculated according to the rule of thumb defined by Schwert (1989) and Ng & Perron (2001):

$$\text{maxlag} = \left[12\left(\frac{T_s}{100}\right)^{0.25}\right] \quad (8)$$

Where $[x]$ represents the integer part of x. From here, a small loop is run to determine the optimal lag for the test. The condition is that if the last lag has a t-statistic in absolute value greater than 1.6, this lag is considered optimal. If no optimal lag is found, the routine defaults to one lag.

Profiling markets

As an application or example for the implications of sample size beyond moment estimation, a modified version of the Capital Asset Pricing Model (CAPM) [Treynor (1962), Sharpe (1964), Lintner (1965), and Mossin (1966)] is estimated for profiling markets (Mestre, 2023). The profiling process starts with the CAPM main regression model:

$$r_t - r_f = \alpha + \beta\left(r_{m,t} - r_f\right) + e_t \quad (9)$$

Where r_f is a risk-free rate, $r_{m,t}$ is the return of market portfolio, e_t is an error term. The estimation process gives $\hat{\alpha}$ (Jensen's Alpha) that represents the excess of expected return, when positive it usually means that the *manager* has a better performance than predicted by the market and when it is negative otherwise. $\hat{\beta}$, also known simply as beta, is a measured of risk or correlation between the asset and the market, when it is greater than one means the asset is riskier than the market and when it is less than one otherwise.

To analyze the effects of sample size on estimating these metrics across the selected indexes, we used the MSCI World Index as the market portfolio and the indexes as the assets. For a detailed discussion of the "world market portfolio," see Vassalou (2000) and Clark & Kassimatis (2011). Regarding the risk-free rate, we choose to set it equal to zero, given that any other alternative would lead to a discussion about an "international-all-compassing" interest rate, which is not the goal of this paper. Hence, the regression model used is:

$$r_t = \alpha + \beta r_{m,t} + e_t \quad (10)$$

Again, equation (10) is estimated for each index and for each sample size in a rolling window. Then, the second step for the profiling process is to test the following hypothesis for each $\hat{\alpha}$ and $\hat{\beta}$:

$$H_0: \hat{\alpha} = 0, H_a: \hat{\alpha} < 0;\ H_0: \hat{\alpha} = 0, H_a: \hat{\alpha} > 0 \quad (11)$$
$$H_0: \hat{\beta} = 1, H_a: \hat{\beta} < 1;\ H_0: \hat{\beta} = 1, H_a: \hat{\beta} > 1 \quad (12)$$

Using a standard one-tail t-statistic where the degrees of freedom are equal to the sample size (60, 250, 500, and 1000) minus 2 (parameters in 10), and the significance level were set to 5%. The final step, according to Mestre (2023), is to count the times of rejection and no rejection for each parameter. Results are reported in Table 4.

D-CAPM

Although the CAPM remains a popular model for evaluating the risk-return relationship it has its shortcomings. Specifically, for the purpose of this study, the notorious one is that the CAPM assumes a symmetric significance when calculating risk to returns above the mean and returns below the mean. These could lead to a false interpretation that a bull market is as risky as a bear market.

To check the robusteness of the results, the Downside CAPM (D-CAPM) proposed by Estrada (2002) is implemented. A key advantange of this model is that it adjusts systematic risked based on the asset's behavior during negative market periods. This adaptation captures a more relevant component ok risk for emerging markets, providing a more realistic view of index sensitivity under adverse conditions without departing from the interpretative structure of the CAPM.

The main parameter of the D-CAPM model is the downside beta β_D, defined as the covariance between the asset and the market only on the days when the market records returns below zero (or below the risk-free rate), divided by the market's semivariance over those same periods. In this way, β_D measures the asset's sensitivity specifically under adverse conditions, providing a more realistic approach to systematic risk when the returns distribution exhibits asymmetries, leptokutosis, or extreme events.

The functional form of the D-CAPM preserves the linear structure of the CAPM but replaces the standard beta with its downside version:

$$\beta_D = \frac{E\{Min[(r_i - \mu_i),0]\}E\{Min[(r_M - \mu_M),0]\}}{E\{Min[(r_M - \mu_M),0]^2\}} \quad (13)$$

This approach is particularly relevant for the Latin American indices analyzed, whose returns series exhibit negative skweness and fat tails in shorter windows, features that challenge the validity of the traditional CAPM. Results are reported in Table 5.

Finally, the process described is implemented in R open-source software (R Core Team, 2023) and following libraries were used: *moments* to estimated moments (Komsta, 2022), *ggplot2* for graphics (Wickham, 2016), *fUnitRoots* for the ADF test (Wuertz et al, 2022), and *stats* for the Ljung-Box test and the regression of the equation (10) and tests in expressions (11)-(12) (R Core Team, 2023).

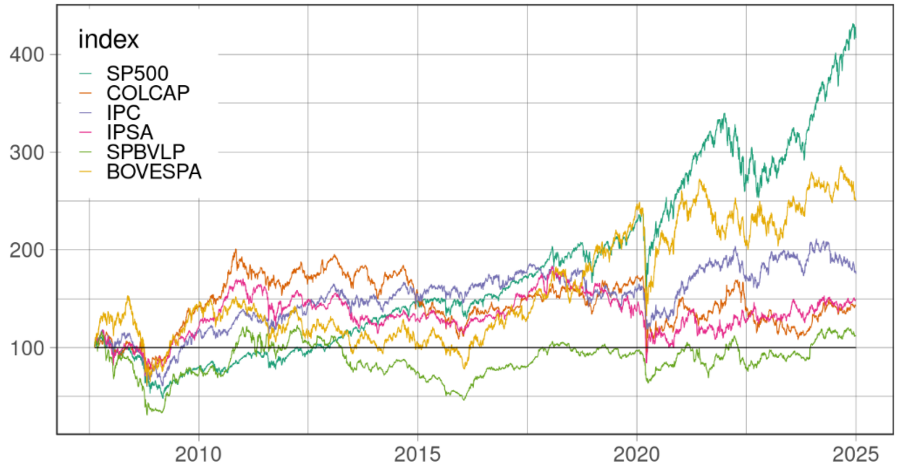
4. Results and discussion

This section presents empirical findings regarding the statistical properties of daily log-returns for the selected stock market indices, the outcomes of the independent and identically distributed (i.i.d.) tests, and the estimations of a modified Capital Asset Pricing Model (CAPM) across varying sample sizes. The analysis aims to illuminate how sample size influences return distributions, the validity of classical financial assumptions, and the resultant risk-return metrics.

Descriptive Analysis

Figure 1 illustrates the historical evolution of the leading American stock market indices included in this study from August 2007 to December 2024. For comparative purposes, all indices were normalized to a base value of 100 on August 16, 2007. The figure illustrates the typical behavior of financial time series, exhibiting varying performance across different markets. Notably, the S&P 500 consistently exhibits the strongest performance by returns, followed by BOVESPA, IPC, and IPSA. The superior performance of the S&P 500 may be attributed to its size and diversified composition, which encompasses 500 companies that offer broader sector diversification compared to indices like COLCAP, which represents only 20 companies. Furthermore, the significant outperformance of US indices could be linked to recent market optimism surrounding artificial intelligence, with major US-based technology companies playing a central role.

Figure 1. Main America's indexes over the period 2007 - 2024



Note: For comparison purposes all indexes were set to 100 on August 16th/2007. **Source:** Authors' calculations**.**

Table 2. Selected markets indicators

Concept	US**	Brazil	Mexico	Colombia	Chile	Peru
Market Classification*	Developed	Emerging	Emerging	Emerging	Emerging	Emerging
Number of listed companies	5,421	339	132	61	291	192
Number of listed companies (Domestic)	4,009	331	128	60	230	176
Number of listed companies (Foreing)	1,412	8	4	1	61	16
Stock market turnover ratio (%)***	109.97	162.65	23.26	9.5	14.44	1.96
Total Equity Risk Premium (%) ****	4.5	6.1	5.2	6.0	5.8	22.0

*According to Morgan Stanley Capital International (2026).

** Includes NSYE and Nasdaq.

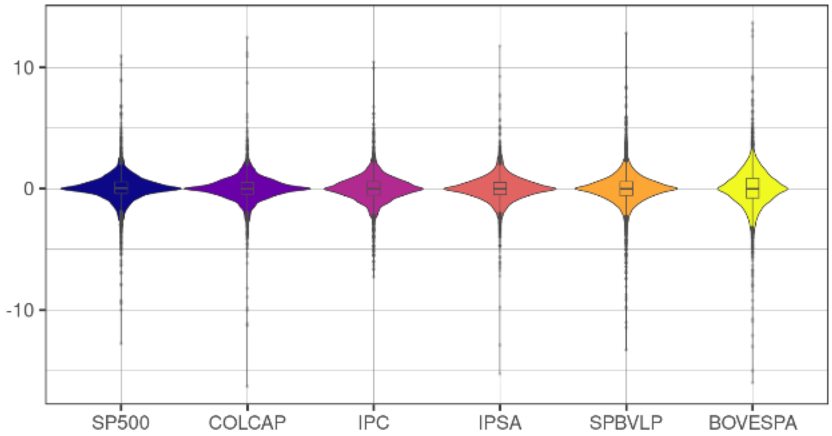
*** As reported by The World Bank Database, for the most recent years avaiable (2022).

**** As reported by Damodaran (2025).

Table 2 contextualizes the market studied by contrasting the 'Developed' US market against five 'Emerging' Latin American markets. Despite their shared classification, the regional markets exhibit significant heterogeneity that impacts the analysis. Market depth varies considerably, with listed companies ranging from 61 in Colombia to over 5,000 in the US. Structural divergences are also evident in liquidity and risk; while Brazil displays high turnover (162.65%), Peru faces distinct illiquidity (1.96%) and a significantly higher Equity Risk Premium (22.0%) compared to the US benchmark (4.5%). These features highlight the unique size, risk, and liquidity constraints inherent to these markets.

Table 3 complements this by presenting the descriptive statistics for the daily log-returns across the entire sample period. A critical finding from this table, as further visualized in Figure 2, is that all daily log-return series exhibit heavy tails, evidenced by high kurtosis values ranging from 13.4 to 28.5. Additionally, all indices, except one, display negative skewness, indicating that losses from outliers are generally larger than profits from outliers. These characteristics of heavy tails and negative skewness directly challenge the traditional assumption of normally distributed returns in financial modeling, underscoring the empirical reality that financial markets frequently deviate from idealized distributions.

Figure 2. Distribution and descriptive statistics for the daily log returns



Sources: Official web pages for each stock exchange. Author's Calculations.

Table 3. Descriptive Statistics for the daily log returns

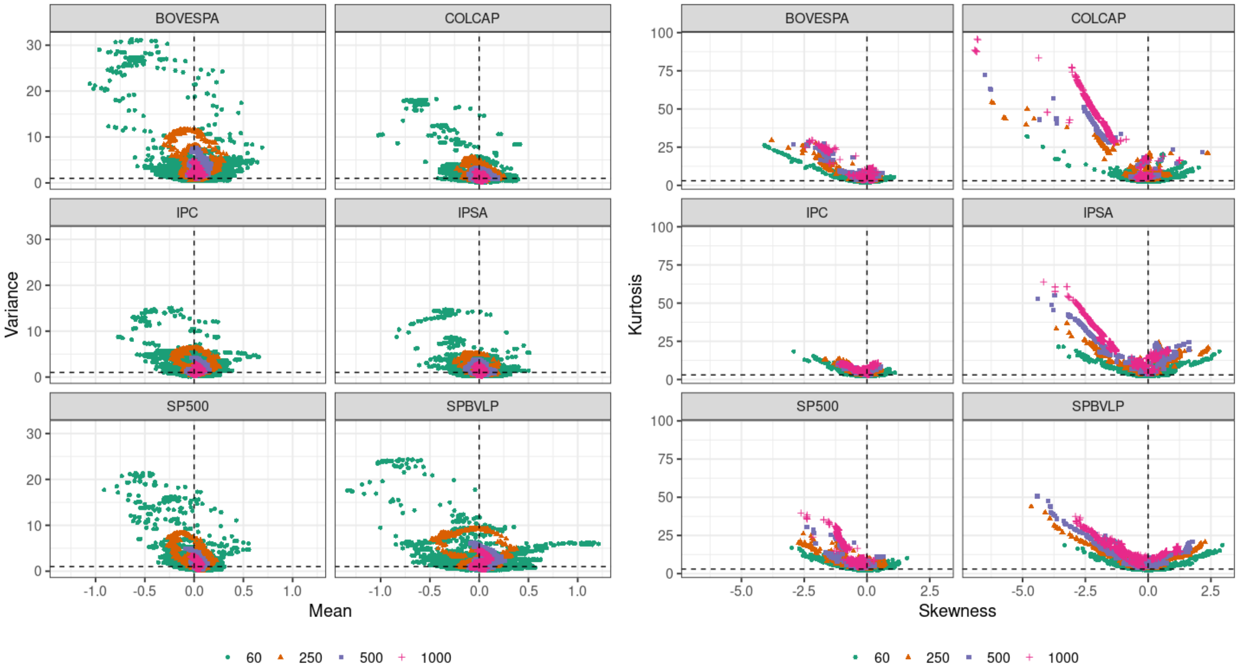
	SP500	COLCAP	IPC	IPSA	SPBVLP	BOVESPA
Median	0.042	0.000	0.000	0.000	0.000	0.000
Mean	0.032	0.008	0.013	0.009	0.002	0.021
Std. Dev.	1.255	1.152	1.142	1.143	1.372	1.654
Skewness	-0.520	-0.861	-0.014	-0.731	-0.546	-0.430
Kurtosis	15.517	28.537	10.298	22.482	15.636	13.858
Minimum	-12.765	-16.290	-7.266	-15.262	-13.291	-15.993
Maximum	10.957	12.470	10.441	11.785	12.816	13.677
Range	23.722	28.760	17.707	27.047	26.106	29.670

Sources: Official web pages for each stock exchange. Author's Calculations.

Sample Moments

Figure 3 visually demonstrates the variability of estimated moments—mean, variance, skewness, and kurtosis—for each index across the four distinct sample sizes (60, 250, 500, and 1,000 days). The results clearly illustrate that the statistical properties of return data vary significantly with sample size. Specifically, for smaller samples (e.g., 60 days), the data tends to exhibit higher leptokurtosis (heavy tails) and more pronounced negative skewness. As the sample size increases, there is a tendency for the distributions to approach a more Gaussian-like shape, although they still retain some characteristics of non-normality. This observed dynamic highlights how the choice of sample length can significantly alter the perceived distributional characteristics of financial time series, consistent with the findings of Dutta & Kanti (2023) and Pekar & Pcolar (2022).

Figure 3. Media, Variance, Skewness and Kurtosis moments for each sample size and index



Sources. Official web pages for each stock exchange. Author's Calculations.

Independent and Identically Distributed (I.I.D.) Tests

To thoroughly assess the i.i.d. assumption, both the Ljung-Box (LB) and Augmented Dickey-Fuller (ADF) tests were applied. Figure 4 presents the percentage of times the Ljung-Box test, designed to detect the absence of serial autocorrelation, did not reject its null hypothesis across different lags and sample sizes. A key insight from these results is that for small sample sizes (n = 60), the Ljung-Box test frequently fails to reject the null hypothesis, suggesting that the Index Return often behaves as white noise. However, as the sample size increases, an apparent emergence of autocorrelated series, particularly of order one and sometimes even higher orders, is observed. This implies that conclusions about serial independence are susceptible to the chosen sample length, directly impacting on the validity of the i.i.d. assumption. The paper's comprehensive presentation of these test iterations provides a new understanding of these dynamic properties.

Table 4 reports the rejection and no-rejection rates for the Augmented Dickey-Fuller (ADF) test, which assesses the stationarity of the time series by testing for the presence of unit roots. The results show a clear trend: as the sample size increases, the percentage of rejections of the stationarity test also increases for all indices. This indicates that larger samples are more likely to lead to the conclusion of stationarity (i.e., rejection of the unit root hypothesis). Notably, this trend is less pronounced for the smallest sample size (n = 60), where the possibility of a non-stationary series is higher. These findings reinforce the crucial importance of the selected sample size for stationarity analysis across all markets, as statistical tests like Ljung-Box and ADF become more stringent and discriminating with a greater number of observations, as previously highlighted by Muriel (2025) and Tsay (2010).

Figure 4. No Rejection percentage for the LB Test

Note. The percentage (over the row total) corresponds to the ratio between the number of times of No Rejection over the number of tests carried out. For samples of 60 observations 4,416 tests were conducted; for 250 observations, 4,226; for 500 observations, 3,976 and for 1,000 observations, 3,476. Source: Authors' calculations.

Table 4. Rejection (R) and No Rejection (NR) percentage rate for ADF test

	S&P		COLCAP		IPC		IPSA		SPBVL		BOVESPA	
	R	NR	R	NR	R	NR	R	NR	R	NR	R	NR
No Intercept and No Trend												
n = 60	83.5%	16.5 %	84.8 %	15.2 %	85.7%	14.3 %	84.4%	15.6 %	81.9%	18.1 %	86.4%	13.6 %
250	99.6%	0.4%	99.7 %	0.3%	99.9%	0.1%	99.8%	0.2%	99.7%	0.3%	99.7%	0.3%
500	99.9%	0.1%	99.8 %	0.2%	100.0 %	0.0%	99.9%	0.1%	100.0 %	0.0%	99.8%	0.2%
1,000	100.0 %	0.0%	99.9 %	0.1%	100.0 %	0.0%	100.0 %	0.0%	100.0 %	0.0%	100.0 %	0.0%
Intercept and No Trend												
n = 60	77.1%	22.9 %	73.7 %	26.3 %	77.6%	22.4 %	74.0%	26.0 %	73.9%	26.1 %	76.4%	23.6 %

250	99.1%	0.9%	98.7%	1.3%	98.5%	1.5%	98.8%	1.2%	94.3%	5.7%	98.9%	1.1%
500	99.8%	0.2%	99.7%	0.3%	99.9%	0.1%	99.8%	0.2%	99.8%	0.2%	99.8%	0.2%
1,000	100.0%	0.0%	99.8%	0.2%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%
Intercept and Trend												
n = 60	70.0%	30.0%	69.2%	30.8%	71.9%	28.1%	71.4%	28.6%	67.9%	32.1%	71.3%	28.7%
250	94.6%	5.4%	95.1%	4.9%	95.3%	4.7%	94.8%	5.2%	91.6%	8.4%	96.3%	3.7%
500	99.8%	0.2%	99.6%	0.4%	99.8%	0.2%	99.8%	0.2%	99.5%	0.5%	99.7%	0.3%
1,000	100.0%	0.0%	99.8%	0.2%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%

Note. The percentage (over the row total) corresponds to the ratio between the number of times of R or NR over the number of tests carried out. The number of tests performed are the same as in figure 4. Source: Authors' calculations.

A key practical implication in investment modeling relates to the dependence of model efficacy on the size of the data sample. In short samples, return series exhibit limited autocorrelation, often approximating a random walk or white noise. Consequently, complex factor models (e.g., CAPM, Fama and French, APT) are likely to fail. Conversely, large samples allow the underlying time series structure to fully materialize, providing the necessary complexity and "memory" for these sophisticated explanatory models to perform successfully. Empirical evidence confirms this dependency, Pham & Phuoc (2020), for instance, found that estimating the CAPM using a 12-year daily return sample (approximately 3,000 observations) significantly reduced the Beta standard deviation and yielded a tighter 95% confidence interval compared to using a short, 3-year sample (approximately 756 observations).

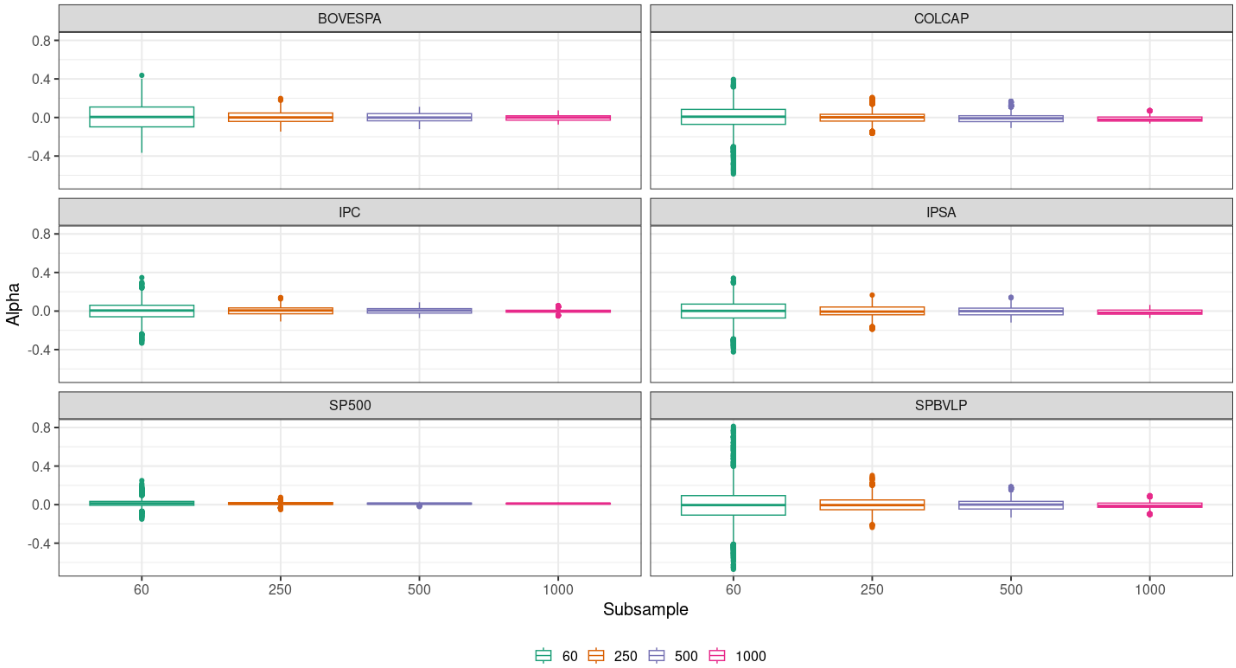
CAPM Estimations and Market Profiling

To demonstrate the implications of sample size beyond mere moment estimation and i.i.d. testing, a modified Capital Asset Pricing Model (CAPM) was employed to profile the markets.

Figure 5 illustrates the estimated alpha coefficients for each stock index across the four different time windows (60, 250, 500, and 1,000 days). The S&P 500 exhibits a high concentration of alpha values close to zero, particularly as the sample size increases. This suggests that the performance of the S&P 500 is primarily explained by the risk factors embedded within the CAPM model, with minimal generation of abnormal returns relative to the global index. In contrast, Latin American indices, such as those from Colombia, Peru, and Chile, show greater dispersion in their alpha values, especially for smaller sample sizes (60 and 250 days).

This broader dispersion signifies a greater influence of structural or local factors not captured by the CAPM, leading to a higher incidence of both positive and negative abnormal returns. While this dispersion attenuates with larger samples for some Latin American markets (Mexico, Brazil), it persists for others (Colombia, Chile, Peru), indicating that even larger samples can reveal the continued influence of idiosyncratic factors. Table 5 further quantifies this by showing the distribution of alpha values, confirming the high proportion of zero alphas for S&P 500 and IPC, compared to the greater dispersion observed in COLCAP, IPSA, and SPBVLP.

Figure 5. Estimated Alpha for each sample size and index



Sources. Official web pages for each stock exchange. Author's Calculations.

Table 5. Distribution of $\hat{\alpha}$ and $\hat{\beta}$ for key values.

	$\hat{\alpha}$			$\hat{\beta}$		
Index - Sample Size	**< 0**	**= 0**	**> 0**	**< 1**	**= 1**	**> 1**
S&P 500						
60	0.4%	97.5%	2.1%	12.4%	44.9%	42.7%
250	0.0%	99.6%	0.4%	14.6%	25.3%	60.1%
500	0.0%	99.6%	0.4%	18.9%	18.8%	62.4%
1000	0.0%	98.4%	1.6%	13.7%	23.8%	62.5%
COLCAP						
60	5.1%	88.6%	6.3%	91.4%	8.6%	0.0%
250	6.2%	86.3%	7.5%	99.7%	0.3%	0.0%
500	6.6%	85.5%	7.9%	100.0%	0.0%	0.0%
1,000	5.1%	86.3%	8.6%	100.0%	0.0%	0.0%
IPC						

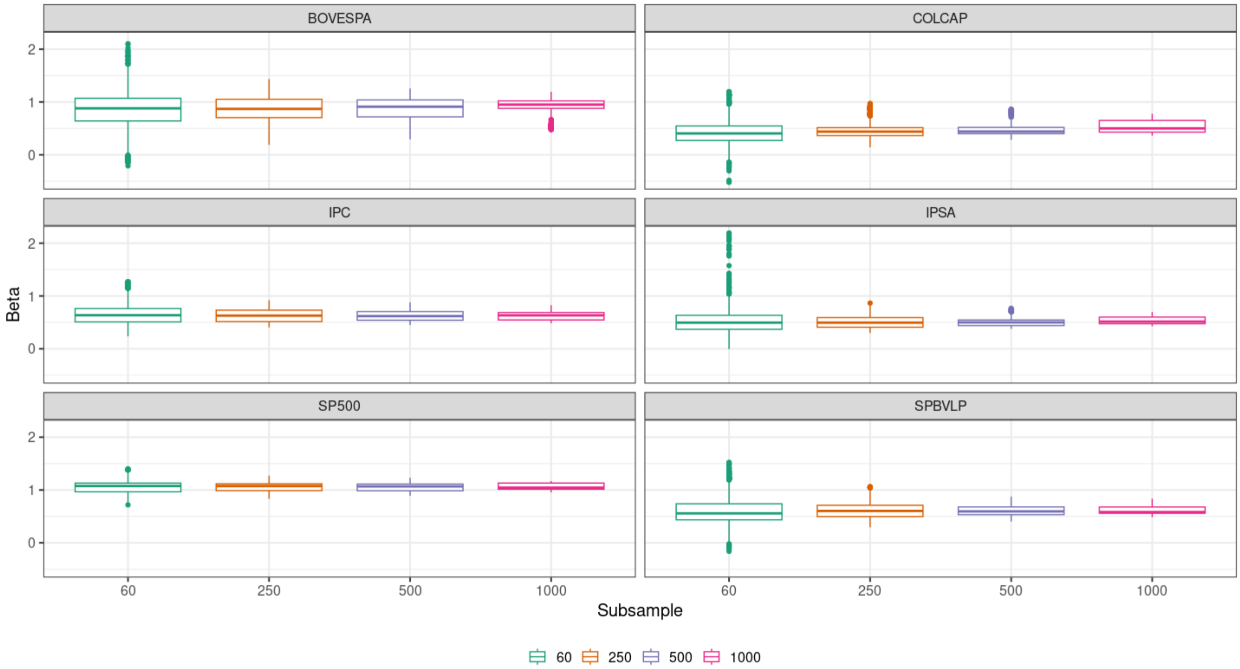
	60	2.3%	95.1%	2.6%	74.8%	25.0%	0.2%
	250	0.3%	99.4%	0.3%	98.5%	1.5%	0.0%
	500	0.0%	99.2%	0.8%	100.0%	0.0%	0.0%
	1,000	0.4%	96.6%	3.0%	100.0%	0.0%	0.0%
IPSA							
	60	4.7%	88.0%	7.3%	86.0%	13.8%	0.3%
	250	3.8%	88.6%	7.5%	99.3%	0.7%	0.0%
	500	5.2%	84.3%	10.5%	100.0%	0.0%	0.0%
	1,000	1.9%	94.8%	3.2%	100.0%	0.0%	0.0%
SPBVLP							
	60	11.6%	76.6%	11.8%	74.1%	25.1%	0.8%
	250	13.3%	76.1%	10.6%	96.8%	3.2%	0.0%
	500	17.0%	73.6%	9.4%	100.0%	0.0%	0.0%
	1,000	11.7%	82.1%	6.2%	100.0%	0.0%	0.0%
BOVESPA							
	60	2.6%	92.5%	4.8%	29.9%	61.9%	8.2%
	250	3.4%	96.2%	0.4%	47.2%	39.2%	13.7%
	500	2.7%	97.3%	0.1%	46.8%	33.2%	20.0%
	1,000	5.3%	93.4%	1.3%	45.4%	37.1%	17.6%

Sources: Official web pages for each stock exchange. Author's Calculations.

Figure 6 depicts how the beta coefficients, a measure of systematic risk and correlation with the market, change with different sample sizes. For the S&P 500, there is a clear tendency for beta values to be greater than one as the sample size increases, rising from 42.7% to 62.5% for the 1,000-observation sample. This indicates that the S&P 500 is more sensitive and reacts with greater intensity to global market movements, reflecting its high level of financial integration and dominant role. This sensitivity is consistent with greater exposure to systemic or global events. Conversely, most Latin American markets exhibit betas below one, implying lower sensitivity to the global index. This lower sensitivity generally persists and becomes more pronounced with increasing sample sizes, suggesting a defensive nature or weaker financial integration at the international level.

For instance, Colombia, Chile, Mexico, and Peru show a strong concentration of beta coefficients below one, with an extremely low frequency of betas greater than one. This behavior aligns with the concept of defensive assets, where markets are less affected by global economic cycles due to structural or institutional factors that limit their exposure to these cycles. Brazil presents an intermediate pattern, with an observed increase in both betas below one and above one as sample size grows, potentially indicating alternating periods of lower and higher exposure to global risk factors. These findings are consistent with Mestre (2023) and Božović (2022) regarding the sensitivity of estimated coefficients to sample size. The observed patterns reinforce the robustness of these market characteristics over more extended periods, emphasizing the importance of sample size in understanding systematic risk exposure.

Figure 6. Estimated Beta for each sample size and index



Sources: Official web pages for each stock exchange. Author's Calculations.

Downside CAPM

Having established the results under the tradicional CAPM framework, the analysis now turns to the downside specification. As explained in the methodology, the D-CAPM relies exclusively on observations in wich the global market exhibits negative returns, which naturally reduces the number of data points avaible within each rolling window. In our case, the average number of downside observations per window is approximately 26.8 for 60-day windows, 111.4 for 250-day windows, 222.2 for 500-day windows and 442.9 for 1.000-day windows. These differences imply that the downside beta is estimated under a markedly more restrictive information set, emphasizing market behavior during adverse conditions rather than across the full return distribution. The results obtained under this specification are summarized in Table 6.

Table 6. Distribution of $\hat{\beta}$ for key values D-CAPM model

Index - Sample Size		**< 1**	**1**	**> 1**
S&P 500				
	60	0.00%	100.00%	0.00%
	250	0.00%	100.00%	0.00%
	500	0.00%	100.00%	0.00%
	1000	0.00%	100.00%	0.00%
COLCAP				

	60	25.95%	74.05%	0.00%
	250	6.81%	93.19%	0.00%
	500	3.32%	96.68%	0.00%
	1,000	0.00%	100.00%	0.00%
IPC				
	60	11.86%	88.14%	0.00%
	250	3.69%	96.31%	0.00%
	500	1.63%	98.37%	0.00%
	1,000	0.00%	100.00%	0.00%
IPSA				
	60	23.70%	76.30%	0.00%
	250	6.65%	93.35%	0.00%
	500	5.46%	94.54%	0.00%
	1,000	0.00%	100.00%	0.00%
SPBVLP				
	60	21.98%	77.90%	0.11%
	250	9.82%	90.18%	0.00%
	500	0.00%	100.00%	0.00%
	1,000	0.00%	100.00%	0.00%
BOVESPA				
	60	10.14%	87.39%	2.47%
	250	0.66%	99.34%	0.00%
	500	0.00%	100.00%	0.00%
	1,000	0.00%	100.00%	0.00%

Sources: Official web pages for each stock exchange. Author's Calculations.

Comparing the CAPM with the downside CAPM (D-CAPM) shows that isolating negative global returns substantially reshapes estimated betas. By focusing only on adverse market conditions, the D-CAPM reveals sensitivities that the traditional model obscures.

For the S&P 500, the CAPM indicates increasing frequencies of betas above one as the sample grows, reflecting strong global integration. Under the D-CAPM this becomes even clearer: all downside betas equal one across every window, indicating perfectly proportional responses to global downturns. This underscores the U.S. market's consistently high exposure and strongly pro-cyclical behavior during negative global episodes.

Latin American markets, however, present a markedly different configuration. Under the traditional CAPM, indices such as COLCAP, IPC, IPSA, and SPBVLP already show predominantly sub-unitary betas, suggesting weaker sensitivity to the global cycle. When applying the D-CAPM, this defensive pattern strengthens: in small windows a relevant fraction of betas remains below one, but in larger samples, downside betas converge almost entirely to one, and observations above one disappear. This indicates that these markets not only exhibit limited exposure to global risk under normal conditions but also absorb negative global shocks more weakly, acting as defensive markets during international stress.

Brazil represents an intermediate case. Under the CAPM, it periodically shows betas above one, indicating shifting degrees of integration. Under the D-CAPM these episodes diminish: betas converge to one in large windows, showing a more moderate and stable sensitivity to global downturns.

5. Conclusions

This paper conducted a comprehensive comparative analysis between the S&P 500 and five selected Latin American emerging market indices (COLCAP, IPC, IPSA, SPBVL, BOVESPA), utilizing an extensive dataset of 4,477 daily observations from August 16, 2007, to December 31, 2024. Our primary objectives were threefold: to provide a richer description of return distributions by examining the variability of estimated moments across different sample sizes, to evaluate the consistency with which classical independently and identically distributed (i.i.d.) assumptions hold across varying window lengths, and to assess the potential impact of deviations from these assumptions on the application of the CAPM and the stability of its α and β estimations.

The findings conclusively demonstrate that sample size is not merely a technical detail, but a fundamental factor that influences the analysis of financial series, particularly in emerging markets. We observed that statistical properties such as independence and the distribution of returns vary substantially according to the number of observations under different window sizes, directly affecting the interpretation of metrics like CAPM α and β. As shown by Pham & Phuoc (2020) and the references in it, one direct consequence is that practical investment modeling should change, with longer time windows recommended to improve the fitness of these parametric models.

Consistent with existing literature (Dutta & Kanti, 2023; Pekar & Pcolar, 2022), returns in smaller samples exhibited higher levels of leptokurtosis (heavy tails) and negative skewness, fundamentally invalidating the assumption of normality in short periods. Furthermore, the Ljung-Box test for these smaller samples often failed to reject the null hypothesis, suggesting white noise behavior, which contrasts sharply with the apparent emergence of autocorrelation in larger samples. This evidence supports the assertion that smaller samples exhibit greater deviations from classical assumptions regarding independence, verifying our H1. This aligns with findings that distribution uncertainty can distort investors' expectations of returns and risk (Chae & Lee, 2018).

Moreover, the independence (Ljung-Box) and stationarity (Augmented Dickey-Fuller) tests revealed that their outcomes are highly sensitive to the sample size chosen. Specifically, both tests become more stringent and discriminating as the number of observations increases For instance, the ADF test showed that larger samples are more likely to lead to the conclusion of stationarity, while the Ljung-Box test showed that the convergence to serial correlation emerges only with larger samples. This phenomenon, previously highlighted by Muriel (2025) and Tsay (2010), demonstrates how different conclusions based on data length directly influence the perceived validity of the i.i.d. assumption. Thus, the evidence confirms our H2.

The analysis of asset pricing models further highlights this dependence. In applying the CAPM, our results demonstrate that the estimated α and β coefficients are highly sensitive to sample size, corroborating the work of Mestre (2023) and Božović (2022). For the S&P 500, β frequently exceeded one, with increasing frequencies as the sample grows, indicating a strong sensitivity integration. Conversely, Latin American indices generally exhibited β values below one, suggesting lower sensitivity or structural defensiveness, a characteristic that often persisted and even became more pronounced with larger sample sizes (Azher & Iqbal, 2018).

A crucial insight emerges from comparing the CAPM with the Downside CAPM (D-CAPM). By isolating negative global returns, the D-CAPM provides a more robust measure of market integration during stress periods. For the S&P 500, the D-CAPM shows downside betas of unity across all sample windows, indicating a perfectly proportional and consistently high exposure to global downturns. Conversely, Latin American emerging markets confirm their structural defensiveness: the predominantly sub-unitary β values found under the CAPM are reinforced by the D-CAPM, downside beta values converge almost entirely to one in larger samples, with observations above one disappearing. This suggests these markets absorb negative global shocks less intensely, effectively acting as defensive assets during international stress. clarifies this behavior by exhibiting downside betas of unity across all sample windows, indicating a perfectly proportional and consistently high exposure to global downturns.

The observed variability of α and β estimates, highly dependent on sample size and the distinction between up/down-side risks, demonstrates that drawing stable conclusions about market efficiency requires incorporating the nuances of higher-order moments (previously identified heavy tails and skewness) and the chosen data length. The D-CAPM, by focusing on downside risk, offers a more robust and less ambiguous measure of integration during market crises. The sensitivity of the models' outcomes to the number of observations and the specific market conditions (normal vs. downturns) provides strong empirical evidence that conclusions are highly context-dependent in emerging markets. Consequently, the results confirm our H3.

The overarching implication of our research is that commonly used metrics for evaluating financial assets can vary substantially depending on the chosen sample size and the specific market analyzed. Directly applying traditional models, such as the CAPM, without explicitly considering these variations, can lead to inaccurate assessments of asset returns or risks. This concern is particularly acute in emerging markets characterized by unique structural and institutional factors. Therefore, reevaluating fundamental assumptions such as independence or normality is not merely a technical exercise, but a necessary condition for correctly interpreting financial data and ensuring model validity.

This paper makes a significant contribution by presenting detailed descriptions of return distributions and comprehensive results from numerous independent and identically distributed (i.i.d.) test iterations across varying sample sizes, offering a novel understanding of the nuances of Latin American stock markets. Our findings can aid researchers, investors, and policymakers in better adjusting models and metrics by providing a robust comparison with a mature market like the United States.

While our study offers significant insights, it is subject to certain limitations. The selection of specific stock market indices and the period, though extensive, may not capture the full diversity of all global markets. The decision to set the risk-free rate to zero simplifies the application of the CAPM but may abstract from the real-world complexities of international interest rates. Future research could explore the impact of other methodological choices, such as alternative rolling window lengths or the inclusion of higher-order moments in portfolio optimization, as suggested by the literature (Athayde & Flôres, 2003; Min et al., 2022). Adapting methodologies to better incorporate unique market characteristics, especially in emerging economies, could further refine financial modeling for more robust investment decisions, avoiding errors derived from a homogeneous approach to non-homogeneous markets.

7. References

Ahmed, S., Bu, Z., Symeonidis, L., & Tsvetanov, D. (2023). Which factor model? A systematic return covariation perspective. *Journal of International Money and Finance*, 136, 102865. https://doi.org/10.1016/j.jimonfin.2023.102865

Andreu, J. & Torra, S. (2009). Optimal market indices using value at-risk: a first empirical approach for three stock markets, *Applied Financial Economics*, 19(14), 1163-1170, https://doi.org/10.1080/09603100802360024

Azher, S. & Iqbal, J. (2018). Testing Conditional Asset Pricing in Pakistan: The Role of Value-at-Risk and Illiquidity Factors. *Journal of Emerging Market Finance*, 17(2_suppl), S259-S281.

Azimli, A. (2024). Time-varying spillovers in high-order moments among cryptocurrencies. *Financial Innovation*, 10(96). https://doi.org/10.1186/s40854-024-00612-8

Bali, T.G., Mo, H., & Tang, Yi. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *Journal of Banking & Finance*, 32, 269-282.

Bao, Y., & Ullah, A. (2006). Moments of the estimated Sharpe ratio when the observations are not IID. *Finance Research Letters*, 3(1), 49-56. https://doi.org/10.1016/j.frl.2005.11.001

Barendse, S., Kole, E., & van Dijk, D. (2023). Backtesting Value-at-Risk and Expected Shortfall in the Presence of Estimation Error. *Journal of Financial Econometrics*, 21(2), 528-568. https://doi.org/10.1093/jjfinec/nbab008

Božović, M. (2022). A common pattern across asset pricing anomalies. *Finance Research Letters*, 48, 103004. https://doi.org/10.1016/j.frl.2022.103004

Chae, J. & Lee, E.J. (2018). Distribution uncertainty and expected stock returns. *Finance Research Letters*, 25, 55-61.

Clark, E., & Kassimatis, K. (2011). An alternative measure of the “world market portfolio”: Determinants, efficiency, and information content. *Journal of International Money and Finance*, 30(5), 724-748.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223-236.

Dai, Y., & Harris, R. D. (2023). Average tail risk and aggregate stock returns. *Journal of International Financial Markets, Institutions and Money*, 82, 101699. https://doi.org/10.1016/j.intfin.2022.101699

Damodaran, A. (2025)., Country Risk: Determinants, Measures and Implications -The 2025 Edition. Available at SSRN: https://ssrn.com/abstract=5354459 or http://dx.doi.org/10.2139/ssrn.5354459

De Athayde, G. M., & Flôres, R. G. (2003). Incorporating skewness and kurtosis in portfolio optimization: A multidimensional efficient set. In *Advances in portfolio construction and implementation* (pp. 243-257). Butterworth-Heinemann.

Ding, R. (2023). f-Betas and portfolio optimization with f-divergence induced risk measures, *Quantitative Finance*, 23(10), 1483-1496. https://doi.org/10.1080/14697688.2023.2230629

Dutta, S., & Kanti, T. (2023). Modeling Long Term Return Distribution and Nonparametric Market Risk Estimation. *Sankhya B*, 1-33.

Eom, C., Eom, Y., & Park, J. W. (2023). Left-tail momentum and tail properties of return distributions: A case of Korea. *International Review of Financial Analysis*, 87, 102570. https://doi.org/10.1016/j.irfa.2023.102570

Estrada, J. (2002). Systematic risk in emerging markets: The D-CAPM. *Emerging Markets Review*, 3(4), 365-379. https://doi.org/10.1016/S1566-0141(02)00042-0

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427-465. https://doi.org/10.2307/2329112

Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fu, J., Zhang, X., Zhou, W., & Lyu, Y. (2024). A continuous heterogeneous agent model for multi-asset pricing and portfolio construction under market matching friction. *International Review of Economics & Finance*, 89, 267-283. https://doi.org/10.1016/j.iref.2023.07.049

Harvey, C., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5-68.

Hassani, H., & Yeganegi, M. R. (2019). Sum of squared ACF and the Ljung–Box statistics. *Physica A: Statistical Mechanics and its Applications*, 520, 81-86. https://doi.org/10.1016/j.physa.2018.12.028

Jondeau, E., Zhang, Q., & Zhu, X. (2019). Average skewness matters. *Journal of Financial Economics*, 134(1), 29-47. https://doi.org/10.1016/j.jfineco.2019.03.003

Komsta, L., & Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and related tests. *R package version*, *14*(1). https://CRAN.R-project.org/package=moments

Lin, C.H., Changchien, C.C., Kao, T.C., & Kao, W.S. (2014). High-order moments and extreme value approach for value-at-risk. *Journal of Empirical Finance*, 29, 421-434. http://dx.doi.org/10.1016/j.jempfin.2014.10.001

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4), 587-615.

Lintner, J. (1965). The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 13–37.

Ljung, G. M. & Box, G. E. P. (1978), On a measure of lack of fit in time series models. *Biometrika*, 65, 297--303. 10.2307/2335207

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91.

Markowitz, H. (1959). Portfolio Selection: Efficient Diversification of Investments. Yale University Press.

Markowitz, H. (2019). From Portfolio Theory to Practice: An AQR interview with Harry Markowitz. AQR Capital Management.

Mestre, R. (2023). Stock profiling using time–frequency-varying systematic risk measure. *Financial Innovation*, 9(1), 9-52. https://doi.org/10.1186/s40854-023-00457-7

Min, L., Liu, D., Huang, X., & Dong, J. (2022). Worst-case Mean-VaR Portfolio Optimization with Higher-Order Moments. *Engineering Letters*, 30(1).

Molero-González, L., Trinidad-Segovia, J. E., Sánchez-Granero, M. A., & García-Medina, A. (2023). Market Beta is not dead: An approach from Random Matrix Theory. *Finance Research Letters*, 103816. https://doi.org/10.1016/j.frl.2023.103816

Morgan Stanley Capital International website (2026). *MSCI Market Classification*, available online at https://www.msci.com/indexes/index-resources/market-classification

Mossin, J. (1966). Equilibrium in a capital asset market. Econometrica. *Journal of the econometric society*, 768-783.

Mossin, J. (1966). Theory of Financial Markets. Englewood Cliffs: Prentice Hall.

Muriel, N. (2025). Weighted portmanteau statistics for testing for zero autocorrelation in dependent data. *Journal of Applied Statistics*, 1-19. https://doi.org/10.1080/02664763.2024.2449413

Ng, S., & Perron, P. (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica*, 69(6), 1519–1554. http://www.jstor.org/stable/2692266

Pekar, J., & Pcolar, M. (2022). Empirical distribution of daily stock returns of selected developing and emerging markets with application to financial risk management. *Central European Journal of Operations Research*, 30, 699-731. https://doi.org/10.1007/s10100-021-00771-4

Pham, C., & Phuoc, L. (2020). Is estimating the capital asset pricing model using monthly and short-horizon data a good choice? *Heliyon*, 6(7), 1-14. https://doi.org/10.1016/j.heliyon.2020.e04339

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13, 341–360.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Said, S. & Dickey, D. (1984). Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order. *Biometrika*, 71(3), pp. 599-607. https://www.jstor.org/stable/2336570

Schwert, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," Journal of Business and Economic Statistics, 7, 147-160.

Schuster, M., & Auer, B. R. (2012). A note on empirical Sharpe ratio dynamics. *Economics letters*, 116(1), 124-128. https://doi.org/10.1016/j.econlet.2012.02.005

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425-442.

Treynor, J.L. (1962). Towards a Theory of Market Value of Risky Assets. Unpublished manuscript, A final version was published in 1999 in Asset Pricing and Portfolio Performance, ed. Robert A. Korajczyk, 99: 15–22. London: Risk Books

Tsay, R. S. (2010). Analysis of financial time series. Wiley Series in Probability and Statistics, Hoboken, New Jersey, 3rd edition.

Vassalou, M. (2000). Exchange rate and foreign inflation risk premiums in global equity returns. *Journal of International Money and Finance,* 19, 433–470

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wuertz, D. Setz, T. & Chalabi, Y. (2022). fUnitRoots: Rmetrics - Modelling Trends and Unit Roots_. R package version 4021.80, https://CRAN.R-project.org/package=fUnitRoots