Stock Beta Prediction with Market Capitalization: Analyzing Cross-Sector Risk Patterns Through Firm-Level Financials

Koji Iwata

Supervisor: Dr. Nazanin Khazra

Department of Economics, University of Toronto, Toronto, ON, Canada
koji.iwata@mail.utoronto.ca
March 2025

This paper examines how market capitalization predicts stock Beta across industry sectors, using firm-level financial data from 3,700 publicly traded U.S. companies. Although firm size is typically viewed as an indicator of stability, this study finds a positive association between market capitalization and Beta, particularly in growth-oriented sectors, while the effect is notably weaker in defensive industries. Among the models evaluated, Random Forest Regression outperforms OLS and Decision Tree Regressors, underscoring the effectiveness of ensemble methods in capturing the non-linear dynamics of market risk.

1. INTRODUCTION

Stock market volatility is a fundamental indicator of financial risk, influencing investment decisions, valuation models, and portfolio strategies. This study examines how market capitalization predicts stock volatility, measured by Beta, across industry sectors. To conduct the analysis, we combine two datasets: S&P Capital IQ (2025), which provides firm-level financial performance metrics, and FactSet Workstation (2025), which offers market-oriented variables including market capitalization and Beta. Together, these datasets cover 3,700 publicly traded U.S. firms and allow for a detailed cross-sector comparison of firm size and market risk. This study hypothesizes that the relationship between firm size and Beta is not uniform, but shaped by sector-specific characteristics and underlying financial conditions.

Prior research examines the link between firm fundamentals and stock risk. Kumar and Patel (2020) find that larger firms are more sensitive to macroeconomic conditions, while Sridharan (2015) shows that metrics like dividend yield improve volatility forecasts. This study builds on these insights by using market capitalization as the main predictor of Beta, with other financial indicators included as controls to isolate the effect of firm size across sectors.

2. DATA

The S&P Capital IQ dataset provides detailed financial performance metrics such as Enterprise Value, EBITDA, Net Income, Operating Income, Dividend Yield, and the number of actively traded exchanges. In contrast, the FactSet Workstation dataset supplies key market-related information like Market Capitalization and 2-Year Beta, which are crucial for assessing market risk. Merging these datasets was necessary to combine operational and market metrics, allowing a comprehensive analysis of the sector-specific drivers of stock volatility. It is important to note that the data was collected at a single point in time during the research period, offering a cross-sectional snapshot rather than a time series analysis.

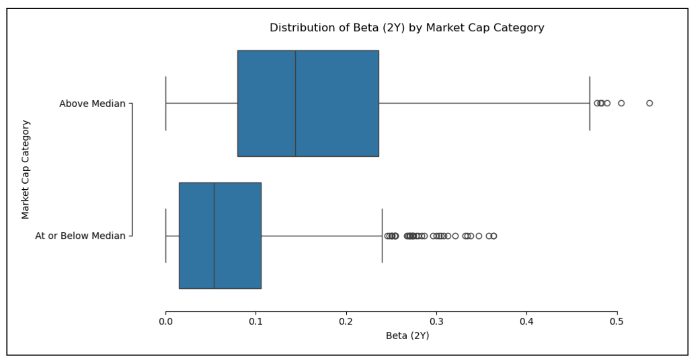

Fig. 1. 2-Year Beta distribution by market capitalization category. Firms are grouped into High and Low Market Cap using a binary classification.

3. SUMMARY STATISTICS

A. Market Capitalization

To investigate how firm size relates to market risk, firms were classified into High and Low Market Cap groups using a binary threshold. Figure 1 reveals distinct differences in the distribution of 2-Year Beta between the two groups. High Market Cap firms show a higher median Beta (0.144) and greater dispersion (mean = 0.1639, std = 0.1071), whereas Low Market Cap firms exhibit a lower central tendency (median = 0.054, mean = 0.0710) and tighter distribution (std = 0.0685).

These findings challenge the conventional view that larger firms are inherently more stable. One possible explanation is that large-cap firms are often concentrated in sectors characterized by high sensitivity to macroeconomic conditions, such as Information Technology, where volatility tends to be elevated. In contrast, smaller firms may operate in more insulated or niche markets, resulting in a more constrained and predictable Beta profile.

B. Beta Distribution by State

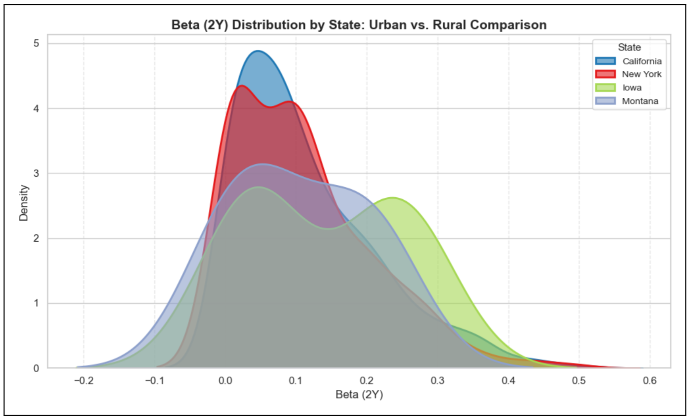

Fig. 2. Distribution of 2-Year Beta in selected urban (California, New York) and rural (Iowa, Montana) states.

Figure 2 compares the distribution of 2-Year Beta values between firms located in urban and rural states. The Beta distributions for urban states exhibit a pronounced central concentration, consistent with the presence of large, diversified firms typically found in metropolitan areas and financial hubs. These firms often operate in sectors such as technology and finance and benefit from greater liquidity, stable revenue streams, and institutional oversight, contributing to more predictable market risk profiles.

In contrast, firms in rural states display a wider dispersion in Beta values, indicating greater variability in systematic risk. This heterogeneity may stem from the concentration of firms in more cyclical industries, such as agriculture, manufacturing, and energy, which are inherently more sensitive to commodity prices, environmental factors, and global economic conditions. Additionally, limited analyst coverage and thinner trading volumes in these markets may introduce further volatility, resulting in the broader risk profiles observed.

C. Active Exchanges

Figure 3 demonstrates a clear relationship between the number of active exchanges a firm is listed on and its observed market risk. Firms listed on multiple exchanges tend to exhibit both higher median Beta values and wider dispersion, indicating elevated and more variable exposure to systematic risk. In contrast, firms listed on a single exchange show relatively concentrated and lower Beta values.

This pattern suggests that broader market participation may be associated with increased return volatility. Firms with wider listing exposure are likely subject to greater investor attention, cross-market arbitrage opportunities, and heightened sensitivity to global macroeconomic events. While broader exchange presence may improve liquidity and capital access, it may simultaneously introduce additional risk factors that contribute to the variability of Beta. These findings underscore the importance of considering market reach as a factor influencing firm-level volatility.

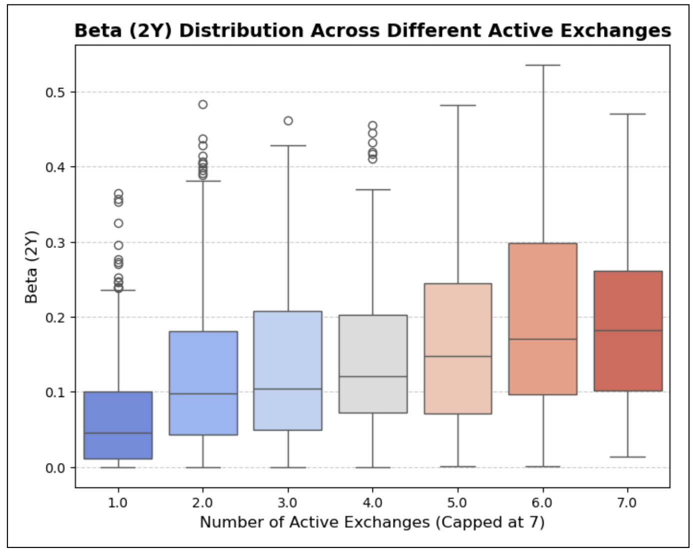

Fig. 3. 2-Year Beta distributions across firms grouped by the number of active exchanges.

D. Dividend Yield

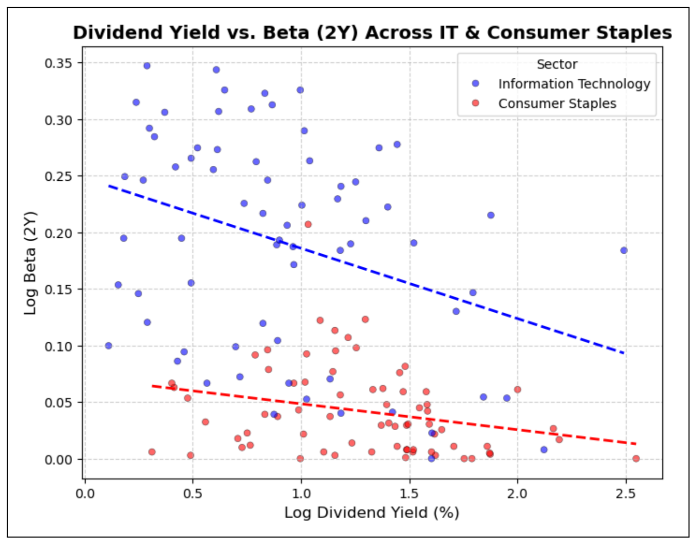

Fig. 4. Relationship between log-transformed Dividend Yield and 2-Year Beta for firms in the Information Technology and Consumer Staples sectors.

Figure 4 displays a clear inverse relationship between dividend yield and Beta within both the Information Technology and Consumer Staples sectors. However, the slope of the relationship is substantially steeper for Information Technology, indicating greater sensitivity of systematic risk to payout behavior in this sector.

This result is economically significant. In Information Technology, a high-dividend firm is associated with markedly lower Beta, suggesting that dividend policy acts as a strong market signal of financial maturity and reduced uncertainty. This is particularly notable given the sector's high baseline volatility, driven by innovation cycles, intangible capital, and exposure to global macroeconomic conditions. Dividend issuance in this context appears to shift investor perception away from growth speculation toward stability, compressing the firm's risk premium.

In contrast, firms in Consumer Staples cluster tightly along both the Beta and dividend yield dimensions, with only a modest decline in Beta as dividend yield increases. The flatter slope likely reflects the homogeneity of operating environments in this sector, where firms already benefit from demand inelasticity and predictable revenue, limiting the informational value of dividends in altering perceived risk.

The divergence in slope magnitudes across the two sectors underscores a key finding of this study: while dividend yield is a consistent negative predictor of Beta, its marginal effect is conditional on sector-specific characteristics. In volatile, growth-oriented sectors, payout behavior plays a more informative role in shaping market expectations. In contrast, in defensively positioned sectors, Beta appears largely orthogonal to dividend yield, implying that other firm characteristics dominate risk perception.

This sectoral heterogeneity reinforces the importance of modeling financial predictors of volatility within industry context, rather than assuming homogeneous relationships across the cross-section.

4. MARKET CAPITALIZATION AND BETA ACROSS SECTORS

Figure 5 highlights the sector-specific elasticity of market risk with respect to firm size. While a consistently positive association between log market capitalization and log Beta (2Y) is evident across all eight sectors, the strength of this relationship, as indicated by the slope of the regression line, varies substantially by industry. This heterogeneity reveals that the role of firm size in predicting systematic risk is highly contextual and mediated by sector-specific fundamentals.

Among the growth-oriented sectors such as Information Technology, Industrials, and Consumer Discretionary, the slope is particularly steep. These sectors tend to be more exposed to macroeconomic uncertainty, technological disruption, and capital market sentiment. For example, in Information Technology, larger firms may derive scale from R&D investments or platform effects, but they are simultaneously more sensitive to innovation risk, valuation expectations, and global capital flows. Consequently, Beta rises more sharply with size, reflecting increased systematic exposure as firms scale.

By contrast, sectors such as Health Care, Energy, and Communication Services exhibit flatter slopes. The subdued gradient in Health Care may stem from inelastic demand and regulatory pricing mechanisms that insulate firms from economic shocks. In Energy, commodity pricing and geopolitical exposure affect all firms regardless of size, diluting the marginal effect of firm scale on Beta. Communication Services, although diverse, includes firms with entrenched infrastructure or subscription models that contribute to revenue stability, limiting Beta's sensitivity to firm size.

Financials and Materials occupy an intermediate position, with moderate slopes. In Financials, the positive slope may reflect the leverage-sensitive nature of larger institutions and their exposure to systemic banking risk. In Materials, cyclical demand patterns likely influence all firms, but larger firms may be more globally integrated, introducing greater covariance with broader market movements.

Economically, these results emphasize that market capitalization does not uniformly predict volatility across sectors. Instead, its effect is conditioned by the degree of operational leverage, exposure to external shocks, capital intensity, and macroeconomic cyclicality within each industry. This reinforces the central argument of this paper: sectoral context is essential when modeling how firm-specific fundamentals, such as size, translate into market risk. Neglecting such heterogeneity risks overstating the generalizability of firm-size effects on Beta across the corporate landscape.

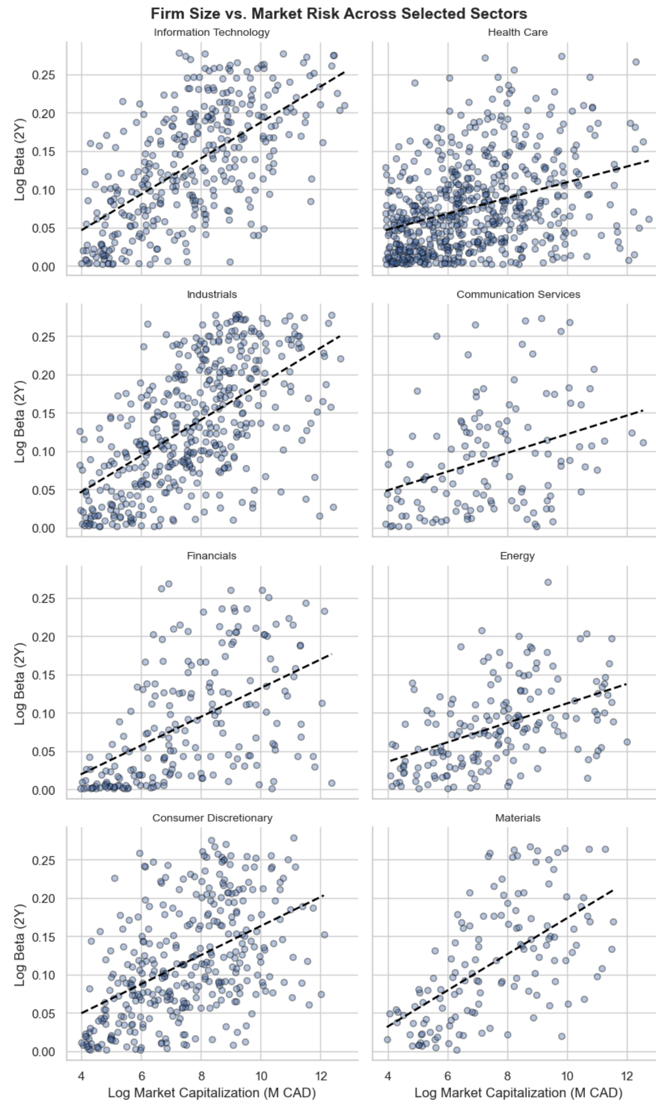

Fig. 5. Relationship between log-transformed Market Capitalization and log-transformed 2-Year Beta across eight major industry sectors. Each panel presents a fitted regression line capturing sector-specific sensitivity of market risk to firm size.

5. GEOSPATIAL VISUALIZATION

To examine regional differences in firm-level risk, we calculate the Beta Deviation from Sector Average (BDSA), which measures how a firm's beta compares to the average beta of its sector:

$$\mathrm{BDSA}_i = \beta_i - \overline{\beta}_s \quad (1)$$

where:

- β_i = 2-year beta of firm i
- $\overline{\beta}_s$ = average 2-year beta for sector s to which firm i belongs

This metric isolates the firm-specific component of volatility by removing sector-wide effects, allowing for a more meaningful comparison of relative risk across geographic regions. Figure 6 displays the average BDSA by U.S. state. States shaded in red represent firms with higher-than-average beta within their respective sectors, while those in blue indicate relatively lower intra-sector beta.

The map reveals a distinct regional divide. States in the eastern United States tend to exhibit positive deviations, suggesting that firms in these areas are more volatile relative to their sectoral peers. This may reflect the higher concentration of dynamic, growth-oriented, or macro-sensitive industries commonly found in the East. In contrast, states in the western half of the country more frequently display negative deviations, consistent with more conservative industry compositions and lower firm-specific exposure to systemic shocks. This geographic variation highlights how regional economic structures and firm characteristics can influence relative market risk within sectors.

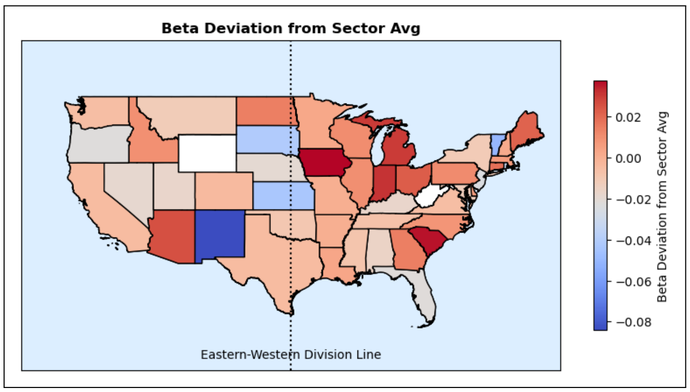

Fig. 6. Average Beta Deviation from Sector Mean by U.S. State. Positive values (red) indicate states where firms are riskier than their sector peers; negative values (blue) indicate states with relatively lower beta. The dashed vertical line marks the East–West division.

6. HTML-BASED WEBSCRAPING

A. Methods

To examine the relationship between consumer sentiment and sector-level market volatility, average customer review ratings were collected for a range of business sectors in California using a web scraping procedure applied to Google Maps. Each sector was associated with a targeted search query (e.g., "supermarkets in California" for the Consumer Staples sector), allowing for the construction of a proxy for real-world consumer experience and perceived service quality.

The scraping process was implemented using `Selenium` with a headless Chrome browser, automated via Python. The procedure is illustrated in Figure 7, which outlines the core stages of the `get_average_rating_from_url(url)` function:

- **Part 1:** Initializes the Chrome WebDriver and sets up headless browsing for performance efficiency.
- **Part 2:** Waits for the appearance of HTML elements containing review ratings, identifies the scrollable container on the Google Maps results page, and performs scroll-and-wait operations to dynamically load more listings.
- **Part 3:** Locates up to 100 rating elements, extracts their `aria-label` values, parses the numerical ratings, and calculates their average.

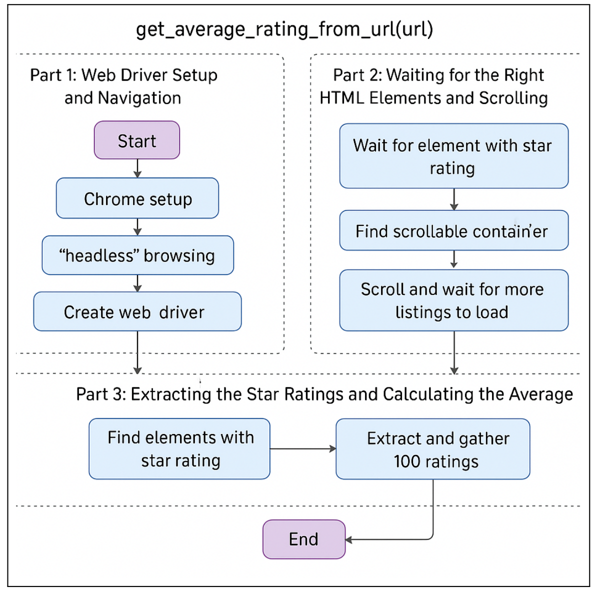

Fig. 7. Workflow for scraping and processing Google Maps ratings. The diagram shows the three-part structure of the `get_average_rating_from_url(url)` function, including web driver setup, dynamic content handling, and rating extraction.

For each sector-specific URL, the function collects the average rating and appends it to a DataFrame. This dataset is then merged with 2-year Beta values (a measure of market risk) for each sector.

To enable comparative analysis, each sector was classified into one of two groups, High Rating or Low Rating, based on whether its average rating was above or below the sample median. This grouping created a dummy variable for group-wise comparison of Beta values using summary statistics.

B. Results

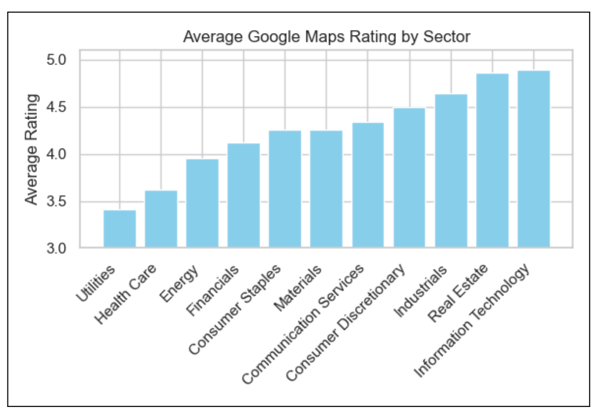

Fig. 8. Average consumer ratings for California-based sectors.

Table 1 presents the results of the web-scraped consumer ratings and their corresponding sector-level Beta values. Sectors such as Information Technology and Real Estate received higher average ratings, while sectors like Utilities and Health Care were rated less favorably.

Contrary to the intuitive assumption that higher consumer satisfaction correlates with lower perceived financial risk, the data reveal an unexpected relationship: sectors with higher average customer ratings also exhibit higher stock Beta values. Specifically, the High Rating group has an average Beta of 0.1179, compared to 0.0860 in the Low Rating group.

This result suggests that sectors perceived as delivering better consumer experiences may also be those with heightened exposure to systematic risk. One possible explanation is that sectors characterized by rapid innovation, competitive differentiation, and brand-sensitive markets (e.g., technology and real estate services) attract greater consumer approval but also exhibit more volatile market behavior due to investor speculation, shorter product cycles, and macroeconomic sensitivity. In contrast, sectors with more muted consumer sentiment such as Utilities and Health Care often operate in regulated, demand-stable environments that insulate them from market swings.

Economically, this finding highlights the distinction between operational or reputational quality (as perceived by consumers) and financial market volatility. It suggests that strong consumer sentiment may not be a reliable proxy for financial stability at the sector level. Instead, such sentiment may reflect characteristics like innovation, market momentum, or competitive churn that contribute to greater return volatility, particularly in growth-driven industries.

Table 1. Summary of Beta Statistics by Consumer Rating Group

Group	Avg. Rating	Mean	Med.	SD	Count
High	4.4200	0.1179	0.1250	0.0478	6
Low	3.7600	0.0860	0.0777	0.0250	5

Sectors are divided into *High* and *Low* groups based on whether their average Google Maps consumer rating is above or below the sample median. Beta values reflect the 2-Year Beta for each sector. Reported values are the mean, median, standard deviation, and count of Beta observations within each group.

7. OLS MULTIVARIATE REGRESSION ANALYSIS

Model 1 (Benchmark):	$\beta = \beta_0 + \beta_1 \cdot \mathrm{MC}$
Model 2:	$\beta = \beta_0 + \beta_1 \cdot \mathrm{MC} + \beta_2 \cdot \mathrm{DY}$
Model 3:	$\beta = \beta_0 + \beta_1 \cdot \mathrm{MC} + \beta_2 \cdot \mathrm{DY} + \beta_3 \cdot \mathrm{AE}$
Model 4:	$\beta = \beta_0 + \beta_1 \cdot \mathrm{MC} + \beta_2 \cdot \mathrm{DY} + \beta_3 \cdot \mathrm{AE} + \beta_4 \cdot \mathrm{OI}$

Table 2. Regression model specifications for estimating Beta. Abbreviations: MC = Market Cap (M CAD), DY = Dividend Yield (%), AE = Active Exchanges, OI = Operating Income (M CAD).

Table 2 outlines four nested OLS regression specifications designed to assess the relationship between firm-level financial indicators and stock Beta. Model 1 serves as the benchmark, incorporating only Market Capitalization, while Models 2 through 4 progressively add Dividend Yield, Active Exchange Count, and Operating Income.

In Model 1, Market Cap displays a positive and statistically significant coefficient (0.022, $p < 0.01$), suggesting that larger firms are associated with higher systematic risk. This result is consistent with the hypothesis that scale may increase exposure to macroeconomic fluctuations and global financial cycles.

Model 2 introduces Dividend Yield as an explanatory variable. Its coefficient is negative and significant (-0.033, $p < 0.01$), indicating that firms with higher shareholder payouts tend to have lower Beta, likely reflecting greater earnings stability and reduced market sensitivity. Notably, the Market Cap coefficient decreases to 0.016, suggesting that dividend policy partially offsets the risk associated with firm size.

In Model 3, the number of active exchanges a firm is listed on is added. The coefficient on Active Exchanges is negative (-0.031, $p < 0.01$), suggesting that firms traded on more exchanges experience reduced Beta, possibly due to enhanced liquidity, broader investor bases, or improved price discovery mechanisms. Market Cap's coefficient rebounds to 0.022, reflecting continued relevance in explaining volatility.

Model 4 adds Operating Income, which enters with a marginally negative coefficient (-0.007, $p < 0.1$). This result suggests that strong core earnings performance further dampens volatility. In this final model, Market Cap increases to 0.027 and remains strongly significant, while Dividend Yield and Active Exchanges retain their negative associations (-0.035 and -0.031, respectively), underscoring their role in moderating systematic risk.

Taken together, the models demonstrate that Market Capitalization consistently exerts an upward influence on Beta, even as other firm-level characteristics are controlled for. At the same time, dividend payments, liquidity through broader exchange listings, and operational performance appear to mitigate this effect. The empirical findings reinforce that while firm size is a key driver of market risk, its impact is conditioned by broader financial policies and firm characteristics.

Table 3. OLS Regression Results for Sector Beta (2Y) Using Firm-Level Financial Variables

	Dependent Variable: Beta (2Y)			
	(1)	(2)	(3)	(4)
Constant	-0.053***	0.033**	0.024*	0.026
	(0.005)	(0.014)	(0.014)	(0.018)
Market Cap (M CAD)	0.022***	0.016***	0.022***	0.027***
	(0.001)	(0.001)	(0.002)	(0.004)
Dividend Yield (%)		-0.033***	-0.033***	-0.035***
		(0.004)	(0.004)	(0.005)
Active Exchanges			-0.031***	-0.031***
			(0.008)	(0.009)
Operating Income (M CAD)				-0.007*
				(0.004)
Observations	2882	1060	1059	1003
R^2	0.266	0.168	0.179	0.177
Adj. R^2	0.265	0.166	0.177	0.174
Residual Std. Error	0.075	0.086	0.085	0.086
F Statistic	1042.1***	106.5***	76.7***	53.6***

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

8. MACHINE LEARNING APPROACHES

A. Decision Trees

The regression tree model offers a nonparametric alternative to linear regression, capable of capturing complex interactions and nonlinearities in the relationship between firm-level financial indicators and stock Beta. Unlike ordinary least squares models, which impose linearity and constant marginal effects, regression trees partition the predictor space into discrete regions based on variable thresholds that minimize within-node variance of the outcome. The algorithm recursively divides the data according to optimal splitting rules that reduce the total squared prediction error across terminal nodes.

Formally, the regression tree solves the following objective function:

$$\min_{s,X} \sum_{j=1}^{T} \sum_{i \in R_j} (\beta_i - \bar{\beta}_{R_j})^2 \tag{2}$$

where:

- β_i denotes the observed 2-Year Beta for firm i,
- $\bar{\beta}_{R_j}$ represents the average Beta in region (leaf node) R_j,
- R_j is the set of observations falling into terminal node j,
- T is the number of terminal nodes (leaves),
- s refers to the split thresholds identified by the algorithm,
- $X \in \{\mathrm{MC}, \mathrm{DY}, \mathrm{NI}, \mathrm{OI}, \mathrm{EV}\}$ is the set of predictors, where MC = Market Cap, DY = Dividend Yield, NI = Net Income, OI = Operating Income, and EV = Enterprise Value.

This objective function guides the algorithm to select partitioning rules that minimize the within-node variance of Beta, thereby constructing a decision structure that reflects conditional associations among the input variables.

Figure 9 presents the regression tree trained on the complete set of financial indicators. The root node confirms that Market Capitalization is the most influential predictor of Beta, with an initial split at approximately 2.85 billion CAD. Firms with market capitalization below this threshold are assigned a notably lower average Beta (mean = 0.146), which aligns with the theoretical proposition that smaller firms are less sensitive to broad market fluctuations.

Subsequent splits on the left-hand branch, corresponding to firms with lower market capitalization, are based on further thresholds involving Market Capitalization and Enterprise Value. One terminal node contains firms with both low market capitalization and low enterprise value, recording the lowest Beta in the entire tree (0.019). Economically, this suggests that firms with limited scale and low total valuation, often indicative of low financial leverage and reduced market exposure, are associated with minimal systematic risk.

On the right-hand branch, which captures larger firms, Dividend Yield plays a prominent role in segmenting risk profiles. Firms with low dividend yields are further partitioned based on Net Income, whereas those with higher payouts are differentiated by Operating Income. This structure indicates that among large firms, systematic risk is increasingly influenced by profitability and payout policies. Notably, the highest Beta observed in the model (0.22) corresponds to large firms characterized by low or negative income and weak dividend performance, highlighting their heightened vulnerability to market dynamics.

In summary, the regression tree uncovers systematic heterogeneity in the financial determinants of Beta. It provides evidence of interaction effects, such as the compounding influence of low income and low dividend payout among large-cap firms, which are not easily captured in linear regression frameworks. While the model's mean squared error of 0.00881 reflects only a modest improvement over simpler specifications, its interpretive clarity adds considerable value. The resulting tree offers a transparent taxonomy of firm-level risk grounded in observable financial characteristics.

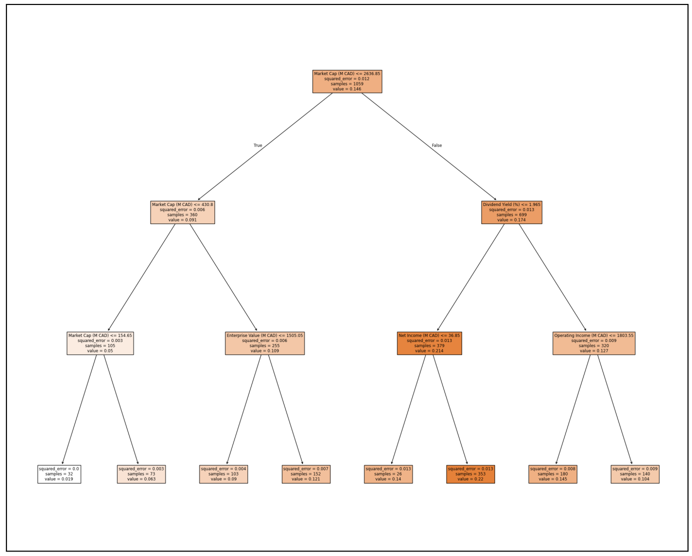

Fig. 9. Regression tree predicting 2-Year Beta using firm-level financial indicators.

B. Random Forest Model

The Random Forest model demonstrates superior predictive performance relative to both Ordinary Least Squares and single-tree regression approaches. By aggregating the outputs of numerous decision trees, the ensemble method captures complex, nonlinear interactions among firm-level financial variables. The model achieves a mean squared error of 0.00135, indicating a substantial improvement in out-of-sample accuracy compared to prior models.

Figure 10 presents the feature importance scores derived from the fitted Random Forest model. Market Capitalization and Dividend Yield emerge as the most influential predictors, each contributing over 20 percent to the model's explanatory power. This finding directly reinforces the core hypothesis of this study: that firm size and payout behavior play central roles in determining a firm's exposure to systematic risk.

The strong importance of Market Capitalization confirms that larger firms tend to exhibit more pronounced Beta values, albeit with context-specific variation captured in the tree-based architecture. Economically, this may be explained by the fact that larger firms are more deeply integrated into global markets, more heavily tracked by institutional investors, and more sensitive to macroeconomic cycles. The elevated role of Dividend Yield implies that firms signaling maturity and stability through shareholder distributions are systematically less volatile, aligning with classical financial theory on dividend signaling and risk aversion.

Enterprise Value ranks third in importance, suggesting that market risk is not driven by equity valuation alone, but also reflects the debt structure of firms. This supports the interpretation that firms with higher enterprise values are typically more leveraged and therefore more sensitive to external shocks and interest rate movements.

In contrast, traditional income-based measures such as Net Income, EBITDA, and Operating Income are less informative in the presence of size and payout controls. Their relatively lower importance suggests that short-term profitability has a diminished role in predicting market Beta when structural and policy variables are already accounted for. Active Exchanges ranks lowest, indicating that the breadth of market listing adds minimal marginal information to the prediction task, once other financial characteristics are included.

These results contribute to the broader literature by demonstrating that volatility is primarily a function of firm structure and investor-facing signals, rather than transient earnings metrics. This hierarchy of predictors also illustrates how nonlinear models like Random Forests can reveal more nuanced relationships than standard linear regression, which assumes homogeneity in marginal effects.

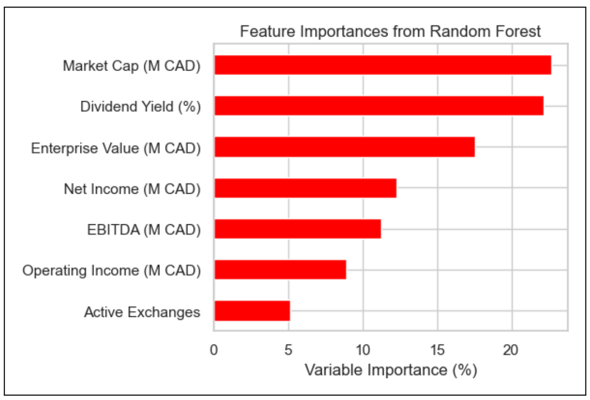

Fig. 10. Feature importance scores from Random Forest model predicting 2-Year Beta.

C. Model Comparison and Discussion

The comparative results across models are summarized in Table 4. The OLS model provides a global linear approximation, yielding interpretable coefficients but limited in its ability to accommodate variable interactions or threshold effects. Its low R^2 value of 0.17694 highlights its inadequacy in capturing the complexity of the data-generating process.

The regression tree offers improved interpretability through rule-based segmentation and captures key interaction terms missed by OLS. For example, it shows that Dividend Yield only moderates Beta in specific market cap ranges. However, its predictive accuracy remains modest due to the high variance inherent in single-tree models.

The Random Forest outperforms both methods by averaging over multiple trees, thereby reducing variance while retaining flexibility. With an R^2 of 0.88952 and the lowest recorded MSE of 0.00135, it demonstrates a strong capacity to model firm-level Beta with precision. While this comes at the cost of interpretability, the gain in explanatory power makes it particularly useful for high-dimensional financial applications.

In economic terms, the Random Forest model validates the notion that systematic risk is governed by a set of interdependent structural factors, especially firm size, valuation, and payout policy. The results emphasize the importance of considering nonlinear effects and cross-variable interactions when modeling market volatility at the firm level.

Table 4. Model Performance Comparison

Metric	OLS	Decision Tree	Random Forest
MSE	0.00729	0.00881	0.00135
MAE	0.07065	0.07485	0.02894
RMSE	0.08537	0.09385	0.03672
MAPE (%)	∞	221.73	∞
R^2	0.17694	0.27848	0.88952
Adjusted R^2	0.17364	0.27368	0.88879

9. CONCLUSION

This study examined the relationship between market capitalization and stock Beta across U.S. industry sectors, using firm-level financial and market data. The results show that while larger firms generally exhibit higher Beta, the strength and direction of this relationship vary significantly by sector. Growth-oriented industries, such as Information Technology, display a strong positive size-risk association, whereas defensive sectors like Health Care show weaker or negligible effects.

Regression and machine learning models yielded consistent findings. OLS confirmed the positive impact of firm size and the mitigating effects of dividend yield, exchange listings, and profitability. The regression tree highlighted non-linear interactions, and the Random Forest model achieved the highest predictive accuracy, identifying Market Capitalization and Dividend Yield as the most influential predictors.

Additional analyses reveal that sectors with higher consumer ratings, such as Technology and Real Estate, exhibit higher Beta, challenging the assumption that favorable sentiment signals lower risk. Geospatial analysis results show firms in the eastern U.S. are riskier relative to sector peers, suggesting that location-based economic dynamics also shape volatility. Together, these findings confirm that the effect of firm size on Beta is conditional on sectoral, behavioral, and regional context, and best captured through models that allow for interaction and nonlinearity.

Future research should consider dynamic modeling over time and the incorporation of additional firm-level and macroeconomic variables to further refine risk prediction frameworks.

REFERENCES

1. Ahmed, A., & Smith, B. (2018). Corporate finance and stock volatility: A review. *Journal of Financial Studies*, 45(2), 123–140.
2. Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1), 259–299.
3. Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1), 1–43.
4. Dennis, P., Mayhew, S., & Stivers, C. (2006). Stock returns and implied volatility. *Journal of Financial and Quantitative Analysis*, 41(2), 381–406.
5. FactSet. (2025). *FactSet Workstation.*

6. Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427–465.
7. Grullon, G., Lyandres, E., & Zhdanov, A. (2012). Real options and stock returns. *The Journal of Finance*, 67(4), 1499–1537.
8. Johnson, M., Rodriguez, P., & Singh, A. (2021). Financial indicators and risk management. *Journal of Risk and Financial Management*, 14(3), 120–137.
9. Kim, J., & Lee, S. (2017). Leverage and market risk: Evidence from US firms. *Financial Analysis Journal*, 32(4), 210–229.
10. Kumar, R., & Patel, D. (2020). Financial structure and stock volatility. *Journal of Corporate Finance*, 54, 101–115.
11. Li, H., & Chen, L. (2019). Corporate performance and stock risk. *Journal of Financial Markets*, 20, 50–68.
12. Peterson, D. R., & Smedema, A. R. (2013). Volatility covariance and expected returns. *Financial Management*, 42(3), 517–536.
13. S&P Capital IQ. (2025). *S&P Capital IQ Pro Database*. Retrieved from https://www.spglobal.com/marketintelligence/en/solutions/sp-capital-iq-pro
14. Sridharan, S. A. (2015). Forecasting volatility using financial statements. *The Accounting Review*, 90(5), 2079–2106.
15. Zhao, Y., & Wang, X. (2016). Debt policy and volatility: A multi-industry perspective. *Journal of Banking and Finance*, 68, 85–102.