

Research on the Impact Mechanism of the Fusion Model on Stock Valuation Correction and Market Risk Forecasting in an Uncertain Environment

Yuxin Ge

*Department of Economics and Finance, Hang Seng University of Hong Kong, Hong Kong, China
s245225@hsu.edu.hk*

Abstract. Under geopolitical tensions and monetary policy shifts, global stock market volatility has intensified. Traditional valuation models and risk forecasting models, relying on historical static parameters and overlooking non-quantitative information such as management tone, suffer from systematic biases. This study uses the Chinese financial language model SilverSight and its ASSL framework as the core information-extraction tool and builds a fusion model. The integrated framework reconstructs asset pricing logic via a three-tier mechanism: extracting three types of complementary signals via ASSL, embedding them into DCF and GARCH-X as time-varying adjustments, and constructing a simultaneous equation system to reveal synergistic effects. Theoretical analysis indicates that the proposed framework transcends the conventional reliance on structured quantitative data and improves the accuracy of stock intrinsic value estimation and the forward-lookingness of market volatility forecasting through text-enhanced dynamic parameter adjustment. This study contributes a novel methodological pathway for the deep integration of large language models with asset pricing theory under uncertainty and a theoretical foundation for investors to leverage Chinese financial LLMs to enhance firm valuation and risk assessment.

Keywords: large language model, stock valuation, market risk prediction

1. Introduction

The international financial market has experienced unprecedented fluctuations and uncertainties. Geopolitical conflicts, sharp turns in monetary policy, supply chain restructuring, and other multiple shocks have formed a complex market environment that scholars have characterized as a polycrisis. Against this background, the volatility of major global stock markets has risen significantly, and the short-term violent fluctuations of asset prices coexist with the break of long-term trends. Investors are facing the severe dilemma of incomplete information and failure of forecasting.

In this volatile market environment, the emotional expression and information disclosure behavior of enterprise management have changed. Kwanjai and Nettayanun's empirical analysis of the disclosure documents filed with the U.S. Securities and Exchange Commission (SEC) by U.S.-listed companies shows that management significantly increases the expression of negative emotions

and uncertainty during recessions while reducing the use of binding language [1]. Pang further finds that Chinese A-share-listed companies tend to strategically exaggerate the optimistic tone of disclosures, forming an impression-management-oriented disclosure bias; this bias gradually corrects itself only when the risk subsides [2]. The study by Bardos, Mishra, and Somé, published in the *Journal of Corporate Finance*, finds a significant negative correlation between positive sentiment expressed during the financial report conference call and the implicit cost of equity capital [3]. Every time the positive sentiment increases by a standard deviation, the equity capital cost decreases by about 7 basis points. This means that the change in management tone will directly change the valuation results of stocks by influencing investors' expectations of discount rates. In terms of the risk dimension, Acheampong et al. studied the data of 114 European banks from 2005 to 2022 and found that negative narrative disclosure intonation was significantly positively correlated with credit risk, liquidity risk, operational risk, and market risk, and this effect in economic policy uncertainty was significantly amplified during the rising period [4]. These studies show that market uncertainty not only affects asset pricing through traditional macroeconomic variable channels but also acts on the information disclosure behavior of management and then is indirectly transmitted to the financial market through text emotions. Changes in management sentiment have a direct and significant impact on stock valuation and market risk.

At the same time, the rapid development of large language model (LLM) technology has opened a new path for quantitative analysis of financial text. The FinBERT model, developed by Huang, Wang, and Yang, significantly outperforms the traditional Loughran-McDonald dictionary method for emotion classification by pre-training on a financial corpus [5]. BloombergGPT, released by Wu et al., was trained on a hybrid corpus of more than 700 billion tokens, surpassing general large language models with a significant advantage in financial tasks [6]. It is particularly worth noting that the SilverSight model proposed by Zhou et al. is based on the adaptive semantic space learning (ASSL) framework [7]. Using semantic clustering and multi-task learning, SilverSight achieves performance comparable to full-data fine-tuning with only 10% of the training data, providing a more efficient approach to financial sentiment analysis.

Based on the above background, this article focuses on two core issues: first, under the increasing market uncertainty, why the traditional stock valuation model is prone to systemic deviations and why the traditional risk forecasting model is difficult to effectively identify nonlinear risks and extreme market changes. Second, how can the fusion model correct the systematic deviation in the traditional valuation model and improve the ability to predict market risks from the emotional level of analysts?

The theoretical contribution is reflected in three levels: first, SilverSight's ASSL multitask learning framework is systematically introduced into asset pricing and risk forecasting research, which enriches the methodological boundaries of dynamic asset pricing theory in an uncertain environment. Second, the analysis chain of "SilverSight text signal extraction — valuation parameter dynamics — risk prediction prospective feedback mechanism" built in this paper provides a new theoretical path for exploring the deep integration of traditional financial theory and large language models. Third, the two-way causal linkage between valuation revision and risk forecasting revealed in this article studies the interplay between valuation and risk forecasting. At the practical level, the integrated model framework proposed in this paper provides a methodological reference for investors to use the Chinese financial language model to improve their ability to judge the value of stocks and identify risks.

2. Literature review

Stock valuation rests on discount rate estimation, a tradition that runs from single-factor to multi-factor specifications. Sharpe formulated the capital asset pricing model (CAPM), expressing expected return as a linear combination of the risk-free rate and the market risk premium, and in doing so laid the groundwork for modern asset pricing [8]. Fama and French added size and value factors to construct a three-factor model, showing that average returns across the cross-section are accounted for by market beta, firm size, and the book-to-market ratio [9]. Fama and French later incorporated profitability and investment factors into a five-factor version [10]. In parallel, the discounted cash flow (DCF) model values an asset as the discounted sum of expected future free cash flows and has become the standard framework for absolute valuation. Yet CAPM, its multi-factor extensions, and DCF all share a common constraint: parameter estimation draws on historical financial data, assumes market parameters hold steady over time, and admits only structured quantitative information while ignoring softer signals such as management tone. Loughran and McDonald made this point concretely, showing that roughly three-quarters of words classified as negative in standard dictionaries do not carry negative meaning in financial disclosures [11]. The semantic gap between everyday language and financial usage exposes a structural blind spot in purely quantitative models.

In financial risk prediction, Bollerslev's generalized autoregressive conditional heteroskedasticity (GARCH) model captures volatility clustering by modelling conditional variance as a function of lagged squared residuals and past conditional variances [12]. But in crisis periods, the statistical properties of historical returns offer little purchase on forward-looking risk signals that reside in text. This gap between what return series can reveal and what textual data contain is a persistent weakness of conventional risk forecasting.

Recent advances in financial large language models (FinLLMs) have begun to close that gap. Nie et al. reviewed the application-oriented literature and grouped it into five areas: linguistic tasks, sentiment analysis, financial time series modelling, financial reasoning, and agent-based simulation [13]. In sentiment analysis, FinBERT outperforms dictionary-based methods by drawing on contextual understanding to read management tone with greater accuracy. In information extraction, BloombergGPT's hybrid training strategy points to the value of domain-specific pre-training. Fabozzi and Nazemi add another layer: growth stocks are highly sensitive to discount rate shocks, value stocks to cash flow shocks, and news sentiment tracks the determinants of discount rates [14].

SilverSight, introduced by Zhou et al. in the Journal of the China Society for Scientific and Technical Information, is a Chinese financial multi-task large language model built around the Adaptive Semantic Space Learning (ASSL) framework [7]. ASSL has three distinguishing features. First, a sentence encoder embeds all instructions and input data into a semantic space, where K-Means clustering groups semantically similar tasks and separates conflicting ones, reducing the task interference that arises from multi-source heterogeneous training data. Second, each low-rank adaptation (LoRA) expert carries its cluster centre as a semantic anchor, so the system can route each query to the most semantically relevant expert without manual intervention. Third, a two-stage data rebalancing strategy operates as follows: an improved adaptive density-based spatial clustering of applications with noise (A-DBSCAN) algorithm performs nested clustering, downsampling high-density regions and upsampling low-density ones; then a maximal marginal relevance (MMR) utility function selects training data for both quality and coverage. SilverSight has assembled roughly 220,000 Chinese financial fine-tuning instances from 23 data sources. Results on CFLEB, a Chinese financial language understanding and generation evaluation benchmark, show that a model fine-

tuned on only 10% of the training data achieves performance comparable to that of the full-data model. Results on FinEval, a Chinese financial domain knowledge evaluation benchmark for large language models, further demonstrate the model's generalization ability; on the subject-determination subtask of the Financial Negative News and Subject Determination dataset (FinNSP-2), its F1 score, defined as the harmonic mean of precision and recall, is approximately twice that of the full-data model.

What emerges from this literature is a clear but unfilled opening. SilverSight, for all its demonstrated strength in sentiment analysis, has not been tested as a correction mechanism inside the DCF valuation framework or the GARCH risk model. Nor have conventional frameworks attended to the two-way feedback between valuation and risk, which operates through shifts in investor trading behaviour and market expectations. There is also little theoretical work on why a fusion model should improve DCF valuation accuracy or GARCH-type risk forecasts. This paper takes up those questions, using SilverSight's ASSL framework as the core extraction tool and constructing a joint framework that links traditional financial models with text-driven methods.

3. Theoretical framework

The discounted cash flow (DCF) model expresses the intrinsic value of the enterprise as the discounted sum of the expected future free cash flow:

$$V_0 = \sum_{t=1}^n \frac{FCF_t}{(1+r)^t} + \frac{TV_n}{(1+r)^n} \quad (1)$$

where V_0 is the current intrinsic value, FCF_t is the expected free cash flow of the t period, r is the discount rate, and TV_n is the final value. In traditional practice, both the cash flow growth rate and the discount rate are set as fixed values or static parameters based on historical average estimates. When market uncertainty rises sharply, the real growth rate of enterprises may deviate significantly from the historical trend, and the risk premium required by investors will also fluctuate significantly. The empirical research of Deng, Wang and Zhou shows that the predictability of expected returns of stocks is mainly realized through the discount rate channel and is consistent with the risk premium changes in the countercyclical period, which further proves the theoretical defect of setting the discount rate as an exogenous constant [15].

CAPM expresses the expected return of a single stock as:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] \quad (2)$$

The term $E(R_i)$ represents the expected return of the individual asset, while R_f denotes the return on a risk-free asset. The coefficient β_i captures the asset's market-related risk, and $E(R_m)$ is the expected return of the market portfolio; the difference $[E(R_m) - R_f]$ is the market risk premium. The Fama-French multi-factor model expands this framework by introducing factors of size, value, profitability, and investment factors. These models together face two fundamental limitations: first, they are essentially linear models and cannot capture nonlinear risk relationships. Second, the factor loadings are assumed to remain stable within the estimation window, but

empirical evidence shows that the beta coefficient fluctuates significantly during the uncertainty period. The latest study of Lim and Kim found that zero-shot volatility forecasting based on LLM can generate a more structural covariance matrix than the historical method, and the portfolio it builds is superior to the traditional method in the five major asset classes, which provides direct empirical support for this connection [16].

The standard form of the GARCH (p, q) model of Bollerslev [12] is:

$$\sigma^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (3)$$

where σ^2 is the conditional variance at time t , ω is the constant term, α_i are the autoregressive conditional heteroskedasticity (ARCH) coefficients capturing the impact of past squared residuals ε_{t-i}^2 , and β_j are the GARCH coefficients measuring the persistence of past conditional variances σ_{t-j}^2 . The parameters p and q denote the orders of the GARCH and ARCH components, respectively.

The core limitation of the model is that it relies entirely on historical income data and cannot incorporate risk signals from text information. Words such as "uncertainty", "risk" and "challenge" frequently used by management in financial reports cannot enter volatility forecasts through the standard GARCH framework. The value-at-risk (VaR) method is also subject to this underlying defect, and its calculation accuracy depends entirely on the estimation accuracy of the volatility σ .

The concept of emotional vulnerability proposed by Zhang et al. further expands the understanding of risk [17]. Not only does the direction of sentiment affect the market, but the stability of sentiment itself also constitutes an independent risk factor. They found that the emotional vulnerability factor can significantly explain the excess returns of stocks even after controlling the Fama-French five-factor model, which provides a theoretical basis for incorporating the multi-dimensional characteristics of text emotions into risk forecasting. This article uses SilverSight as the core text signal extraction tool of the fusion model. The ASSL framework of the model realizes the verticalization of financial text understanding through semantic space clustering. Among them, the LoRA expert system targets tasks such as sentiment analysis, topic recognition, and information extraction and adaptively selects the optimal processing path. In addition, SilverSight achieves near-full-data performance using only 10% of the training data to approach the full effect, which reduces the threshold of data demand in the actual deployment of the fusion model.

4. The impact mechanism of the fusion model on stock valuation correction and market risk forecasting

The fusion model has three layers. The first is signal extraction, where SilverSight's ASSL multi-task framework pulls three kinds of signals from unstructured financial text—annual reports, quarterly reports, earnings call transcripts, and the management's discussion and analysis (MD&A) sections within them:

Table 1. Textual signal taxonomy and extraction outcomes

Signal Type	Extraction Method	Output Form	Information Dimension
Management Tone	Contextual Sentiment Analysis	$Sentiment_t \in [-1,1]$	Sentiment Orientation
Risk Topic Frequency	Topic Identification and Classification	$RiskTopic_t \in [0,1]$	Risk Attention
Forward-Looking Statements	Information Extraction and Quantification	g_t^{LLM}	Forward-Looking Guidance

Management tone reflects the sentiment orientation of corporate disclosures; shifts in tone feed into investors' risk perception and discount rate expectations, which in turn produce time-varying adjustments in the cost of equity capital. Risk topic frequency picks up management's latent references to firm risk exposure. These references function as a forward-looking signal of future fundamental volatility and tend to lead realized return volatility. Forward-looking statements are the most direct of the three: management's explicit guidance on future business growth, including internal assessments of industry cycles, demand shifts, and cost pressures. Because such statements reach beyond historical financial data, they form the core basis for revising cash flow expectations.

The ASSL framework's clustering mechanism matters here for a particular reason. Sentiment analysis, topic recognition, and information extraction are semantically proximate tasks, and ASSL routes them to adjacent LoRA experts automatically. Conventional multi-task learning often suffers performance degradation from task interference; the clustering design reduces that risk.

The second layer takes these text-derived signals as dynamic adjustment terms and embeds them into the parameter structures of the DCF valuation model and the GARCH risk prediction model. The third layer links valuation and risk into a two-way feedback system built on the revised estimates.

Conventional models work only with structured quantitative data and miss management's subjective assessment of business prospects and internal risk. SilverSight addresses this by mapping signals from heterogeneous financial texts into quantifiable vectors in a unified semantic space—soft information converted into hard parameters. In practice, the sentence encoder first encodes each financial text into a high-dimensional semantic vector; K-Means clustering then identifies the semantic category of each text segment, separating risk-related paragraphs from forward-looking ones without any manual predefinition of task types. This avoids the subjective bias that manual annotation introduces. The roughly 220,000 Chinese financial fine-tuning instances SilverSight has drawn from 23 sources provide the breadth needed to build such a semantic space.

4.1. Dynamic correction of DCF model

Traditional DCF models fix the growth rate at a historical average. SilverSight instead extracts growth rate predictors from forward-looking text, replacing that backward-looking constant:

$$\widehat{FCF}_t = FCF_t^{historical} \times (1 + g_t^{LLM}) \quad (4)$$

The shift matters because it changes what the growth rate g actually represents. A historical average reflects past operating performance and nothing more—it cannot absorb contemporaneous industry shocks, policy changes, or strategic adjustments management has already begun to make. Growth expectations g_t^{LLM} pulled from forward-looking statements carry management's real-time assessments of future revenues, costs, and capital expenditures. Private information, in other words, becomes a quantifiable growth parameter. Chen et al. offer empirical backing: combining LLM-extracted corporate relationship knowledge graphs with sentiment analysis raises the average R^2 of continuous price forecasting to roughly 0.97, suggesting that forward-looking information extracted by LLMs improves prediction accuracy considerably [18].

The dynamic discount rate equation is specified as:

$$\hat{r}_t = r_0 - \alpha \bullet Sentiment_t, \alpha > 0 \quad (5)$$

Here r_0 is the baseline discount rate and $Sentiment_t$, is the management tone score extracted by SilverSight, and α captures the sensitivity of the discount rate to tone shifts. When management tone turns negative ($Sentiment_t < 0$), investors read this as rising operational uncertainty and demand higher risk compensation. The cost of equity capital goes up, and the intrinsic valuation of the stock falls.

The modified DCF valuation formula is thus expressed as:

$$\widehat{V}_0 = \sum_{t=1}^n \frac{\widehat{FCF}_t}{(1 + \hat{r}_t)^t} + \frac{\widehat{TV}_n}{(1 + \hat{r}_n)^n} \quad (6)$$

where $\widehat{V}_0$ is the estimated current intrinsic value of the stock, $\widehat{FCF}_t$ is the estimated free cash flow generated at the end of period t , $\widehat{TV}_n$ is the terminal value at the end of the forecast horizon n , and $\hat{r}_n$ is the discount rate applied to the terminal value, with n representing the number of years in the explicit forecast period.

4.2. Dynamic correction of GARCH risk model

The standard GARCH (1,1) model estimates volatility from historical returns alone. The text-enhanced GARCH model with exogenous variables (GARCH-X) adds two exogenous variables to the conditional variance equation—the sentiment and risk-topic factors extracted by SilverSight:

$$\sigma_t^2 = \omega + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2 - \gamma \bullet Sentiment_{t-1} + \delta \bullet RiskTopic_{t-1} \quad (7)$$

Here $\gamma > 0$ and $\delta > 0$. If the management sentiment in the previous period is negative ($Sentiment_{t-1} < 0$) or risk keywords frequency is high ($RiskTopic_{t-1}$ is large), the conditional variance forecast adjusts upward. The limitation of the standard GARCH framework is that it extrapolates past price fluctuations statistically, which is a backward-looking characterization of risk. During sudden market shocks, volatility clustering takes multiple periods to fully materialize, and the model systematically underestimates contemporaneous risk. Negative tone and elevated risk topic frequency are different in kind: they lead the risk. When management's risk discourse intensifies in financial reports, it signals heightened future operational uncertainty, and this information reaches the market before price fluctuations catch up.

The modified Value at Risk (VaR) follows:

$$VaR_t^\alpha = \mu_t + z_\alpha \bullet \sigma_t \quad (8)$$

The multi-expert architecture of SilverSight's ASSL framework fits this specification. The system handles sentiment analysis and topic recognition simultaneously, and its semantic clustering keeps outputs consistent at the embedding level. Sentiment analysis and risk topic identification tend to fall in adjacent semantic clusters, so their LoRA expert embeddings sit close together—fewer semantic conflicts arise when the signals are fused.

4.3. Two-way linkage mechanism between valuation and risk

The fusion model departs from single-model approaches in one central respect: it treats valuation correction and risk prediction as a two-way causal system, formalized as a simultaneous equation framework. The transmission from valuation to risk runs through the transaction differentiation effect under heterogeneous beliefs.

Let $\widehat{V}_t$ denote the fundamental value derived from the text-adjusted DCF model, and P_t the current market price. The valuation deviation is defined as $\Delta V_t = \widehat{V}_t - P_t$.

When there is a significant deviation between the internal valuation of the stock and the market price, the difference in value judgement of market investors will widen: fundamental investors will reverse arbitrage based on the valuation deviation, buy undervaluation, sell overvaluation, and trend traders will follow the current price trend to chase up and fall. The divergence of the trading direction of the two investors groups amplifies market turnover and price volatility, ultimately elevating the realized volatility of the stock. The larger the valuation deviation, the more severe the dispersion in market expectations, the more heterogeneous the resulting trading behavior, and the higher the corresponding market risk.

Risk feeds back into valuation through the risk premium. When lagged risk rises, the discount rate adjusts endogenously:

$$\hat{r}_t^{risk\ adjustment} = r_0 - \alpha \bullet Sentiment_t + \lambda \bullet \hat{\sigma}_t^{prediction} \quad (9)$$

$\lambda > 0$ is the risk-transmission coefficient, measuring the marginal effect of lagged risk on the current discount rate.

The simultaneous equation system is specified as follows:

$$\begin{cases} \widehat{V}_t = \sum_{t=1}^n \frac{\widehat{FCF}_t}{(1+r_0-\alpha \bullet \text{Sentiment}_t + \lambda \bullet \sigma_t)^t} \\ \sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 - \gamma \bullet \text{Sentiment}_{t-1} + \delta \bullet \text{RiskTopic}_{t-1} \end{cases} \quad (10)$$

What this system captures, in theoretical terms, is how the fusion model reconfigures asset pricing logic under uncertainty. SilverSight's text signals enter the valuation equation and the risk equation at the same time, and valuation and risk interact through the parameter λ and the trading-behavior channel. The model thus adjusts asset pricing dynamically as new information arrives, improving both the accuracy of intrinsic value estimation and the forward-looking capacity of volatility forecasting.

5. Conclusion

Financial markets today face persistent uncertainty, and conventional stock valuation models and risk prediction frameworks carry structural limitations that become harder to ignore in such conditions. This paper has used SilverSight and its adaptive semantic space learning (ASSL) multi-task framework as the core information extraction tool to build a fusion model. The main findings fall into three areas.

The first concerns information integration. Conventional models overlook soft information; the fusion model does not. SilverSight's ASSL framework, through semantic clustering and multi-expert adaptive selection, extracts three complementary signals—management tone, risk topic frequency, and forward-looking statements—from heterogeneous texts such as annual reports and earnings call transcripts. The second concerns parameter dynamics. Conventional models treat parameters as static; the fusion model embeds text-derived signals as dynamic adjustment terms in the DCF valuation equation and adds sentiment and risk-topic factors to the conditional variance equation of the GARCH-X model. Parameters become time-varying rather than fixed. The third concerns feedback. The simultaneous equation system constructed in this paper shows that text-augmented valuation and text-augmented risk prediction are not independent exercises. They interact, and the interaction itself carries information.

This study is, however, primarily theoretical. The proposed framework has not yet been tested against real stock market data, and key parameters require calibration with market-specific estimates. The model also assumes a linear relationship between management tone and the discount rate, while actual markets may respond nonlinearly. Future work could take four directions: empirical validation using real-world data; nonlinear specifications for the tone–discount rate relationship; comparative studies that benchmark SilverSight against other financial LLMs within the same integrated framework; and development of a real-time dynamic asset pricing system based on the proposed architecture.

References

- [1] Kwanjai, M., & Nettayanun, S. (2025). Sentiment and tone analysis of corporate communication in different market conditions. *SAGE Open*, 15(3), 1–12. <https://doi.org/10.1177/21582440251363860>
- [2] Pang, L. (2026). Geopolitical risk and corporate disclosure tone: Evidence from China. *Applied Economics Letters*, advance online publication. <https://doi.org/10.1080/13504851.2026.2706083>
- [3] Bardos, K. S., Mishra, D. R., & Somé, H. Y. (2025). Firm-level climate sentiments, climate politics and implied cost of equity capital. *Journal of Corporate Finance*, 94, 102846. <https://doi.org/10.1016/j.jcorpfin.2025.102846>

- [4] Acheampong, A., Ibeji, N., Owusu, F., & Nguyen, T. (2025). Narrative disclosure tone and bank risk: The role of economic policy uncertainty. *Managerial and Decision Economics*, 46(6), 3428–3447. <https://doi.org/10.1002/mde.4539>
- [5] Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2), 806–841. <https://doi.org/10.1111/1911-3846.12832>
- [6] Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. *arXiv*, 2303.17564. <https://doi.org/10.48550/arXiv.2303.17564>
- [7] Zhou, Y., Li, Z., Ye, G., Tian, S., Ni, Y., Liu, X., & Chai, H. (2025). SilverSight: A multi-task Chinese financial large language model based on adaptive semantic space learning. *Journal of Chinese Information Processing*, 39(8), 42–52.
- [8] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- [9] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- [10] 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>
- [11] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- [12] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [13] Nie, Y., Kong, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., & Zohren, S. (2024). A survey of large language models for financial applications: Progress, prospects and challenges. *arXiv*, 2406.11903. <https://doi.org/10.48550/arXiv.2406.11903>
- [14] Fabozzi, F. A., & Nazemi, A. (2023). News-based sentiment and the value premium. *Journal of International Money and Finance*, 136, 102864. <https://doi.org/10.1016/j.jimonfin.2023.102864>
- [15] Deng, Y., Wang, Y., & Zhou, T. (2025). Macroeconomic expectations and expected returns. *Journal of Financial and Quantitative Analysis*, 60(4), 1760–1796. <https://doi.org/10.1017/S0022109024000279>
- [16] Lim, T., & Kim, J. H. (2026). Measuring volatility with LLMs: Implications for covariance structure and portfolio construction. *Finance Research Letters*, 106, 110293. <https://doi.org/10.1016/j.frl.2026.110293>
- [17] Chen, J., Chen, Z., Zhang, N., & Zheng, Y. (2026). Construction of the emotional vulnerability factor based on network and large language models. *Risk Sciences*, 2, 100049. <https://doi.org/10.1016/j.risk.2025.100049>
- [18] Chen, W., Zheng, J., Li, Y., Zhang, X., Wang, H., & Al-Qudah, I. (2026). Beyond sentiment: Knowledge graph-driven financial market forecasting with large language models-extracted enterprise relations and adaptive residual networks. *International Journal of Data Science and Analytics*, 22(1), 84. <https://doi.org/10.1007/s41060-026-01064-2>