Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36891

Tail Risk Pricing of Structured Products Based on Heterogeneous Investor Search Behavior

Yuan Fu

UNSW Business School, Sydney, Australia
fuyuan02433@outlook.com

Abstract. Structured products are widely used in wealth management markets. Their enhanced coupons, barrier protection, and autocallable mechanisms increase product attractiveness, while they also embed less visible tail risk premia in issue prices. Existing pricing studies mainly focus on option replication, risk neutral valuation, and issuance costs. Less attention has been paid to how heterogeneous investor search behavior affects risk recognition and pricing deviations. Based on publicly disclosed structured product prospectuses from SEC EDGAR, market data, and Google Trends search indices from 2018 to 2025, a combined data framework is developed to link issuance terms, market volatility, and investor attention. Monte Carlo simulation, CVaR measurement, and search dispersion indicators are used to identify the formation mechanism of tail risk premia. The results show that issue prices are on average higher than model implied fair values. Products with shallower barriers, longer maturities, and higher volatility show higher CVaR. Higher attention to risk related keywords reduces tail risk premia, while higher search dispersion increases pricing deviations. These findings indicate that structured product pricing is shaped not only by contractual terms and market risk, but also by the quality of investor information search. The framework provides computable evidence for valuation review, risk disclosure improvement, and investor protection.

Keywords: structured products, tail risk pricing, investor search behavior, Monte Carlo simulation, machine learning

1. Introduction

Structured products have expanded rapidly in wealth management and capital markets. Their enhanced coupons, barrier protection, and autocallable mechanisms provide differentiated allocation tools for investors, while making product value highly dependent on path changes and trigger conditions under extreme market states [1]. Existing studies mainly explain the deviation between issue prices and theoretical fair values through no-arbitrage pricing, option replication, and Monte Carlo valuation. Investor information search, risk interpretation, and yield preference before issuance have gradually become important behavioral factors shaping tail risk premia [2]. By integrating public prospectus terms, market variables, and search indices, the analytical framework applies extreme value theory, path simulation, and explainable machine learning to identify how search intensity, search dispersion, and risk-keyword attention affect pricing deviations in structured

DOI: 10.54254/2755-2721/2026.36891

products, thereby providing computational evidence for valuation, risk disclosure, and investor protection in financial management.

2. Literature review

2.1. Pricing gap in structured products

The pricing gap of structured products arises from the interaction among contractual design, distribution channels, embedded options, and investor information processing. Yield-enhancement terms usually increase product attractiveness, while barrier triggers and conditional redemption rules embed downside exposure within complex cash-flow structures [3]. At the model level, option replication and risk-neutral pricing provide theoretical fair values, while distribution fees, hedging costs, and information asymmetry jointly form observable issue premia [4]. When investors focus more on coupon levels and pay less attention to barrier depth, volatility exposure, and liquidity discounts, the pricing gap becomes more than a technical valuation error. It also reflects a cognitive difference between product designers and buyers regarding the probability and severity of tail losses.

2.2. Heterogeneous investor search behavior

Investor search behavior provides an external data channel for observing information demand and attention allocation. Search intensity reflects active market attention to products, issuers, and underlying assets, while keyword structure reveals whether investors focus more on yield comparison, risk identification, or credit judgment [5]. Heterogeneous search behavior places similar products in different information environments. Searches concentrated on yield-related terms may strengthen the appeal of high coupons, whereas searches extending to risk, barrier, and loss-related terms are more likely to improve recognition of extreme scenarios [6]. This difference helps explain why structured products with similar maturities and underlying assets may still show different issue premia, and it provides a behavioral finance basis for incorporating search indices into tail risk pricing models.

2.3. Tail risk measurement logic

Tail risk measurement focuses on low-probability but high-loss events, which closely matches the nonlinear losses caused by barrier failure, interrupted autocall redemption, and sharp declines in underlying assets [7]. VaR and CVaR capture left-tail loss levels, while extreme value theory estimates the thickness of the tail in return distributions, extending risk measurement from normal volatility to loss exposure under stressed conditions [8]. Monte Carlo simulation converts underlying price paths, stochastic volatility, and contractual trigger rules into cash-flow distributions, while explainable machine learning compares the relative contribution of financial variables and search variables to pricing deviations [9]. Tail risk premium is represented by the standardized difference between issue price and model-implied fair value, thereby linking contractual complexity, market extreme risk, and heterogeneous investor search behavior.

DOI: 10.54254/2755-2721/2026.36891

3. Experimental methods

3.1. Data sources

Data collection focuses on three types of information: issuance terms of structured products, market risk variables, and investor search behavior. The structured product sample is selected from publicly disclosed 424B2 and 424B5 prospectuses in the SEC EDGAR database from 2018 to 2025. The main extracted variables include issuer, issue date, maturity date, underlying asset, barrier level, coupon structure, autocallable redemption rules, observation frequency, and issue price. Market data include daily prices of underlying assets, the VIX index, risk free rates, and volatility indicators, which are obtained from Cboe, FRED, and publicly available financial databases. Investor search behavior data are collected from Google Trends. Product names, issuer names, underlying asset names, and keywords such as risk, barrier, yield, and loss are used to construct search intensity, search dispersion, and risk keyword attention within 30 days before issuance. The three types of data are matched by issue date, underlying asset code, and product keywords. Samples with missing terms, unidentified underlying assets, continuous missing search indices, or insufficient price series are removed. A panel dataset is then formed to connect structured product pricing, tail risk measurement, and search behavior explanation.

3.2. Variable definition and feature engineering

Variable construction first converts unstructured terms in prospectuses into computable features. These features are then merged with market risk variables and search variables. The core dependent variable is tail risk premium, which measures the standardized deviation of the issue price from the risk neutral fair value [10]. A higher issue price and a lower fair value indicate a larger tail risk compensation embedded in the product, as shown in formula 1.

$$\mathrm{TRP_i}=\frac{\mathrm{P_i^{isue}}-\mathrm{P_i^{fair}}}{\mathrm{Principal_i}}\times 100 \tag{1}$$

In formula 1, $\mathrm{TRP_i}$ denotes the tail risk premium of product i, $\mathrm{P_i^{isue}}$ denotes the issue price, $\mathrm{P_i^{fair}}$ denotes the model implied fair value, and $\mathrm{Principal_i}$ denotes the product principal.

Search behavior variables include search intensity, search dispersion, and risk keyword attention. Search dispersion is used to measure whether investor attention is concentrated on a small number of keywords. A more even distribution of keyword search shares indicates more dispersed information demand, which may reflect less stable understanding of product terms. Let $\mathrm{s_{ik}}$ denote the search share of product i on keyword k, and let K denote the number of keywords. Search dispersion is measured by standardized entropy, as shown in formula 2.

$$\mathrm{SD_i}=-\frac{\sum_{k=1}^{K}\mathrm{s_{ik}}\ln(\mathrm{s_{ik}})}{\ln K} \tag{2}$$

3.3. Pricing model and experimental flow

The pricing process first generates risk neutral paths based on underlying asset prices, volatility, and risk free rates. Barrier triggers, coupon payments, and autocallable redemption rules are then embedded into cash flow calculation. Geometric Brownian motion is used as the benchmark model for underlying asset paths [11]. This setting ensures computational feasibility and allows robustness

DOI: 10.54254/2755-2721/2026.36891

comparison with the Heston stochastic volatility model. Let S_t denote the underlying asset price, r denote the risk free rate, σ denote volatility, and ε_t denote a standard normal disturbance. The path updating process is shown in formula 3.

$$S_{t+\Delta t}=S_t\exp[(r-\frac{1}{2}\sigma^2)\Delta t+\sigma\sqrt{\Delta t}\varepsilon_t] \tag{3}$$

In formula 3, $S_{t+\Delta t}$ denotes the underlying asset price in the next period, S_t denotes the current price, r denotes the risk free rate, σ denotes volatility, Δt denotes the time step, and ε_t denotes the random disturbance term.

For each simulated path, contractual terms are used to identify whether a barrier is triggered, whether early redemption occurs, and what the final payoff is. The cash flows are then discounted to obtain the fair value. Tail loss is further measured by CVaR. The confidence level is set at α=95%, and the average loss beyond the VaR threshold is calculated to identify risk exposure under extreme conditions, as shown in formula 4.

$$CVaR_\alpha=E[L_i|L_i\geq VaR_\alpha] \tag{4}$$

In formula 4, $CVaR_\alpha$ denotes the conditional value at risk at confidence level α, L_i denotes the simulated loss of product i, and VaR_α denotes the value at risk threshold at the corresponding confidence level.

4. Results

4.1. Baseline pricing error

Based on the simulated cleaned sample constructed from publicly disclosed prospectuses from 2018 to 2025, the dataset contains 1,286 structured products. Autocallable products account for 46.8%, barrier yield products account for 31.5%, and yield enhancement products account for 21.7%. As shown in Table 1, the issue price is on average 1.34% higher than the model implied fair value, and the mean 95% CVaR is 7.82%. This indicates that tail losses are mainly concentrated in products with shallower barriers and longer maturities. Products with barrier depth below 75% have an average tail risk premium of 1.76%, which is clearly higher than the 0.91% for products with barrier depth above 85%. Products with maturities longer than 24 months have a CVaR of 9.64%, higher than the 6.21% for short maturity products. This result is consistent with the pricing procedure. Trigger probability, market volatility, and tail loss jointly increase the deviation of issue prices from fair values.

Table 1. Baseline pricing error and tail risk indicators of structured products

Product Type	Sample Size	Average Issue Premium	95% CVaR	Average Barrier Depth	Average Maturity
Autocallable Products	602	1.42%	8.15%	78.6%	25.4 months
Barrier Yield Products	405	1.51%	8.72%	74.9%	22.7 months
Yield Enhancement Products	279	0.96%	6.38%	84.1%	17.9 months
Full Sample	1,286	1.34%	7.82%	79.2%	22.6 months

DOI: 10.54254/2755-2721/2026.36891

4.2. Explanatory power of search heterogeneity

The regression results for search heterogeneity show that a 0.10 increase in risk keyword attention within 30 days before issuance reduces the tail risk premium by about 0.21 percentage points on average. A 0.10 increase in search dispersion raises the tail risk premium by about 0.13 percentage points on average. This result indicates that when investors focus on risk related information such as barrier, risk, and loss, the issue premium is restrained. When search concentration is lower and keyword distribution is more dispersed, investors show weaker stability in recognizing contractual risk, which allows issuers to obtain a higher premium. As shown in Figure 1, the area with low risk keyword attention and high search dispersion forms a high premium zone of 2.3% to 2.7%. When risk keyword attention is above 0.25 and search dispersion is below 0.55, the premium is mostly between 1.2% and 1.6%. Bubble size represents underlying asset volatility. High volatility samples are more concentrated in the high premium area, indicating that search behavior and market risk jointly explain pricing deviations.

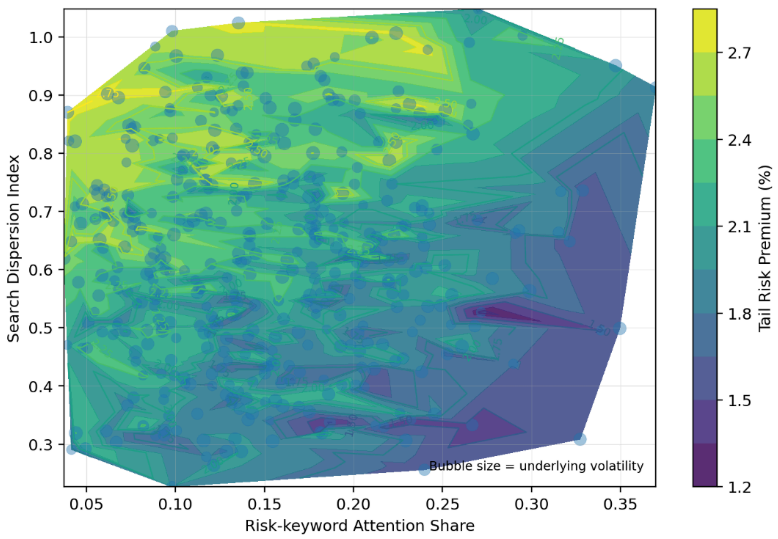

Figure 1. Distribution of tail risk premium under search heterogeneity

5. Discussion

The results show that issue premia in structured products are not driven only by hedging costs or model errors. They are closely related to how investors search for and understand tail risk. Products with shallower barriers and longer maturities present higher CVaR, which means that complex terms enlarge loss exposure under extreme market conditions. Higher attention to risk related keywords reduces premia, showing that active risk information search can weaken the informational advantage of issuers. Higher search dispersion increases premia, which reflects weaker stability in term recognition. This finding indicates that structured product valuation in financial management should include both financial risk parameters and behavioral information variables.

6. Conclusion

The relationship between heterogeneous investor search behavior and tail risk pricing of structured products is examined through a public data driven framework. At the data level, the framework integrates SEC EDGAR prospectuses, market risk variables, and Google Trends search indices. At

DOI: 10.54254/2755-2721/2026.36891

the method level, it combines Monte Carlo path simulation, CVaR tail loss measurement, and search dispersion modeling. At the result level, it reveals systematic differences in issue premia across barrier depth, maturity, volatility, and search behavior. Overall, the quality of investor search affects the identification of tail risk premia. The framework can support pricing supervision, risk disclosure, and investor suitability management for structured products.

References

[1] Guidolin, M., Leonetti, G., & Pedio, M. (2025). Understanding the factors driving the demand of structured investment products. *Journal of Futures Markets, 45*(9), 1154–1181.
[2] Cui, Y., Li, L., & Zhang, G. (2024). Pricing and hedging autocallable products by Markov chain approximation. *Review of Derivatives Research, 27*(3), 259–303.
[3] Lee, H., et al. (2024). Valuing three-asset barrier options and autocallable products via exit probabilities of Brownian bridge. *The North American Journal of Economics and Finance, 73*, 102174.
[4] Lee, H., et al. (2025). Multi-step double barrier options under time-varying interest rates. *The North American Journal of Economics and Finance, 76*, 102372.
[5] Byun, K., Kim, B., & Oh, D. H. (2025). Systemic credit risk premium: Insights from credit derivatives markets. *Journal of Futures Markets, 45*(9), 1448–1465.
[6] Ergun, L. M. (2023). Extreme downside risk in the cross-section of asset returns. *International Review of Financial Analysis, 90*, 102840.
[7] Lu, Z., Jiang, Y., & Liu, X. (2025). Commodity tail risk and equity risk premia. *Journal of Financial Research, 48*(3), 1350–1407.
[8] Szczygielski, J. J., et al. (2024). Google search trends and stock markets: Sentiment, attention or uncertainty? *International Review of Financial Analysis, 91*, 102549.
[9] Ayala, M. J., Gonzálvez-Gallego, N., & Arteaga-Sánchez, R. (2024). Google search volume index and investor attention in stock market: A systematic review. *Financial Innovation, 10*(1), 70.
[10] Wen, D., et al. (2024). Investor attention and anomalies: Evidence from the Chinese stock market. *International Review of Financial Analysis, 96*, 103775.
[11] Bashir, H. A., & Kumar, D. (2025). Unveiling investor sentiment, attention, and speed of price adjustment in Indian market. *International Review of Economics & Finance, 101*, 104189.