

# ***Research on Topic Evolution and Investors' Attitudes of Gold Price Investment Comments on Weibo — Based on Multilingual Text Analysis with BERTopic***

**Ziqi Ye**

*School of Management Science and Engineering, Baoding University of Technology, Baoding, China*

*yeziqu666@qq.com*

**Abstract.** Gold prices keep changing all the time. More and more people care about gold prices. Weibo is a main social platform in China. Many people share their ideas about gold price investment here. Investing in precious metals has become a hot topic on this open platform. This paper uses Python to find hot topics about gold prices. We also use the BERTopic multilingual topic model. We combine it with the belief framework of the Theory of Planned Behavior (TPB) to analyze topics. We sort out the features of the topics investors talk about, their inner thoughts and how their feelings change. Finally, we find that the change of gold prices makes many people care about it. Most investors are careful about gold investment. It shows that people think carefully when they invest in precious metals.

**Keywords:** Gold price, Investors' attitudes, BERTopic multilingual topic model, Topic evolution

## **1. Introduction**

The global economy goes up and down, and there are always regional conflicts. Gold's ability to avoid investment risks and its investment value become more and more important. Recently, the global gold price stays high and changes all the time. It makes ordinary investors talk about it a lot. Weibo has a huge number of users, and its information spreads very fast. It becomes a popular platform for investors to communicate. A lot of text materials about investors' thoughts, feelings and investment decisions are collected here. Many studies prove that there is a predictive relationship between the overall economic mood and the price bubble of precious metals [1]. But these statistical data are overall and not timely. It's hard to catch investors' real-time inner thoughts. Comments on social media show the real-time and personal feelings of investors, and also show how hot topics change over time. They can make up for the shortcomings of the overall research.

The Theory of Planned Behavior (TPB) provides a structured framework to analyze people's beliefs for understanding investors' decisions. The core of this theory is: a person's wish to do a certain investment behavior is mainly driven by three psychological factors, which are behavioral attitude, subjective norm and perceived behavioral control. Most researchers think that behavioral beliefs and control beliefs are the key factors that drive people's investment wishes, while the

influence of normative beliefs is not obvious [2]. This provides a direct reference for understanding the decision-making thoughts of ordinary investors. This paper applies the TPB framework to the analysis of Weibo comments. It can provide a theoretical basis for classifying hot topics, and avoid subjective and random topic classification. We use the Weibo topics identified by BERTopic to explore the personal inner psychological mechanisms that drive the discussions on Weibo. The significance of this study is: it adds text evidence of personal feelings, provides a lot of real-time and personal text evidence for overall financial research, and discusses the relationship between people's beliefs and hot topics. Applying the Theory of Planned Behavior to the text analysis of social media provides a tool for understanding the real-time psychological changes of investors. Combining the BERTopic model with traditional financial research methods (bubble test, factor analysis) provides a reference for the research on market mood.

## 2. Research design

### 2.1. Topic model

BERTopic is a modern topic modeling method that combines Transformer pre-trained model and unsupervised clustering algorithm. Successful cases where researchers use BERTopic with Arabic pre-trained model prove this method can be used for different languages [3]. Its core is to use BERT word embedding, Transformer model and c-TF-IDF technology to generate topic clusters. BERTopic is better than LDA in the depth and specificity of topic terms, semantic understanding and keeping the contextual semantic information. And it does not need complex word segmentation and dictionary preparation when dealing with multilingual texts.

### 2.2. Data source

This research is based on Weibo platform, and we collect users' comments under relevant topics on the platform. The key words for data collection include "gold price", "gold investment", "follow the trend", "long-term" and so on, covering the core dimensions such as users' investment methods, gold price trends and investment strategies. To comply with the ethical norms of social media data research, we first hide users' personal information such as nicknames, IDs and geographic locations, only keep the comment content, remove irrelevant content such as advertisements, marketing promotions and comments unrelated to the topic, and finally delete the same comments posted repeatedly. Finally, we get 14376 valid comments.

### 2.3. Data preprocessing

The core goal of data preprocessing is to turn the original comments into standardized texts that meet the model's input requirements. First, we clean the texts, remove irrelevant elements such as @users, topic tags and URL links, filter out special symbols and numbers, then change the texts into a unified expression mode. After that, we build a stop word list to filter redundant information in the word segmentation results and keep the core semantics. Finally, we connect the processed words with spaces to form text sequences that meet the input requirements of the BERTopic model, which are used for subsequent model modeling.

### 3. Results

#### 3.1. Topic correlation analysis

We use the BERTopic model and generate 27 topic clusters about gold price investment. We also make the data visible. The hierarchical clustering chart (as shown in Figure 1a) and the topic similarity matrix (as shown in Figure 1b) show similar information, so we can analyze and discuss them together.

Figure 1a shows the similarity between text topics. It combines similar topics layer by layer with a tree structure. This makes the meaning more direct and easy to analyze. The core behavioral belief topics, such as "attention to price fluctuations" and "cautious investment mood", have the highest similarity. It shows a strong semantic link: price fluctuations → risk perception → cautious attitude. It means when investors talk about gold price changes on social platforms, they will have the thought of investing carefully. We can also see from Figure 1a that the two clusters of "gold investment" and "gold jewelry consumption" are very far apart. This proves the "investment-consumption dual cognitive structure" mentioned before.

In Figure 1b, the similarity between price fluctuation topics and cautious investment mood topics is higher than 0.42. So it proves the existence of the strong semantic link again. The similarity between "consumption pressure and price complaints" topics and all kinds of investment decision topics is generally low. It confirms the separation of the two cognitive structures of "investment-consumption" in discussions from a quantitative perspective. We can see the conclusions of hierarchical clustering and similarity matrix match perfectly.

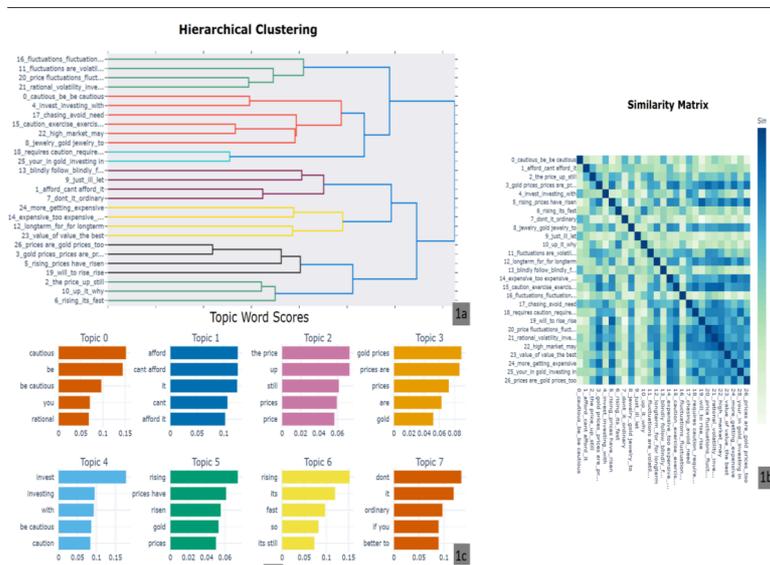


Figure 1. Topic modeling results of Weibo gold price comment content

#### 3.2. Topic evolution analysis

We use the BERTopic model to do topic modeling on the Weibo comments about gold price investment. We get 27 detailed topics in total. To analyze the key points of the discussions in Weibo comments deeply, as shown in Figure 1c, we mainly show the key words and their score distribution of the top 10 important topics. It shows the semantic core of each topic in a direct way. To understand the overall thoughts of investors' discussions deeply, we manually sort and sum up the 27

model topics into 6 primary topics and 15 secondary topics according to the main semantic correlation and the belief framework of the Theory of Planned Behavior. The specific results are shown in Table 1, which systematically reveals the macro narrative map of the discussions about gold price investment on Weibo.

Table 1. Hierarchical classification of themes

Level 1 Theme	Level 2 Theme	Model Theme ID
Price Dynamics	Price Fluctuations	20_price fluctuations_fluctuation
	Upward Price Trends	5_rising prices have_risen6_rising_its_fast
	Investment Behavior	4_invest_investing_with
Investment Strategy	Cautious Investment	0_cautious_be_cautious15_caution_exercise_exerc is...
	Long-Term Investment Perspective	12_longterm_for_for longterm
	Market Volatility	21_rational_volatility_investment
Market Analysis	Value of Precious Metals	23_value_of_value_the_best
	High-Price Market	22_high_market_may
Jewelry and Precious Metals	Application of Precious Metals in Jewelry	8_jewelry_gold jewelry_to
	Herding Behavior	13_blindly_follow_blindly_following
Market Psychology	Market Sentiment and Decision-Making	11_fluctuations_are_volatile
Future Expectations	Expectations of Continued Price Increases	19_will_to_rise_rise
	Increased Precious Metals Investment	25_your_in_gold_investing_in

Figure 2 reveals the dynamic evolution rule of the discussions about gold price investment on Weibo, and provides empirical support for studying the development trend of the topics. Each line stands for the changing track of the popularity of different topics over time. Although the popularity curves of all topics show the feature of interweaving and fluctuating, this truly reflects the immediate and diverse responses of social media discussions to market events.

From October 2025 to the beginning of 2026, the frequency of topics about gold price investment began to rise rapidly. It reflects the abnormal changes in the prices of the gold market, which then leads to the concentrated and explosive discussions of the public. We can see from Figure 2 that the discussions during the explosive period can be sorted into three main dimensions, including gold price fluctuation, investment risk alert and rational investment strategy. Among them, the topics of "price fluctuation perception" are the basis of the discussions, and directly lead to the high-frequency appearance of the topics of "investment risk alert". The topics of "rational investment strategy" are the responsive discussions of investors to market fluctuations. The three form a complete logical relationship, which shows that Weibo users do not act impulsively but think rationally when the gold market changes abnormally.

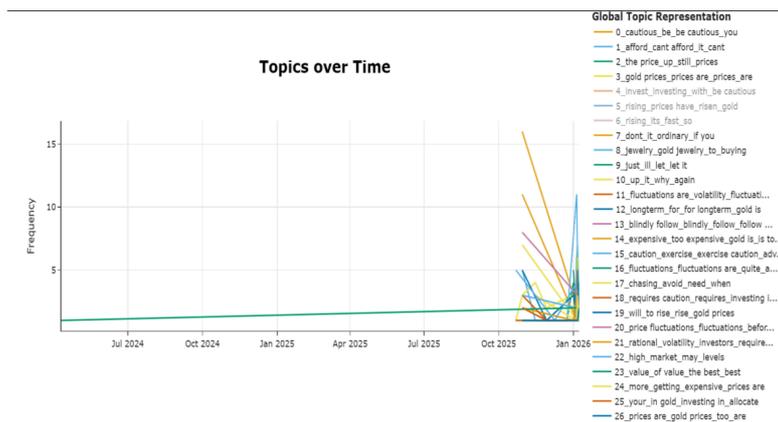


Figure 2. Topic evolution trend

## 4. Conclusions and suggestions

### 4.1. Research conclusions

This research takes the gold price investment comments on Weibo as the research object. It uses the BERTopic multilingual topic model to analyze the topic features, the correlation rules between topics and investors' emotional tendencies. The core topics can be summed up into 5 types: "cautious investment mood", "attention to price fluctuations", "investment behaviors and tools", "price perception and purchasing power" and "warning of follow-the-trend risks". Among them, "cautious investment mood" and "negative complaint mood" account for a high proportion. This reflects that the overall market mood is conservative, which is highly consistent with the market environment of high gold prices. There is a strong logical relationship among the topics: gold price fluctuations, investors' cautious decision-making and vigilance against following the trend. The similarity between price fluctuation topics and cautious investment topics reaches 0.42. The "price perception and purchasing power" topic has the lowest correlation with other investment topics, which reflects the separation of cognitive structures between investment and consumption. The cautious investment topic is mainly about risk aversion mood, the price perception topic is mainly about negative complaint mood, and the investment tool topic is mainly about rational mood. This corresponding relationship reveals the emotional distribution rule of folk discussions on gold prices.

### 4.2. Research suggestions

Investors should establish a rational investment awareness, make clear the boundary between the risk-averse attribute and consumption attribute of gold investment, and avoid blindly buying gold jewelry following the trend because of "rising gold prices". If there is a demand for capital preservation and appreciation, investors can choose gold ETFs, investment gold bars and other tools first, weaken the interference of short-term price fluctuations, reduce excessive attention to the daily rise and fall of gold prices, and have a long-term perspective. For banks, funds or large investment groups, they should be more cautious when carrying out large-scale precious metal asset allocation. Existing studies show that there are significant differences in the "safe haven" attributes of different precious metals [4]. For example, platinum shows stronger anti-volatility and "safe haven" attributes because of its stable demand and relatively safe producing areas. Palladium, however, has an asymmetric response to volatility shocks and a weak risk-averse function. Therefore, for large investment groups, when formulating bulk precious metal investment strategies, they must first

carefully evaluate and distinguish the risk-averse characteristics of different metals, and cannot treat them all the same. They should also focus on the impact of government policies, geopolitical risks and changes in the global supply and demand structure. These factors all have a huge impact on the prices of different metals [5].

## References

- [1] Maghyereh, A., Abdoh, H. Can news-based economic sentiment predict bubbles in precious metal markets?. *Financ Innov* 8, 35 (2022).
- [2] Verghese, J., Chin, P.N. Factors affecting investors' intention to purchase gold and silver bullion: evidence from Malaysia. *J Financ Serv Mark* 27, 41–51 (2022).
- [3] Jiang, S., Li, H., & Gan, D. (2025). Technology acceptance model for online education: identifying interdisciplinary topics and their evolution based on BERTopic model. *Social Sciences & Humanities Open*, 12, 101831.
- [4] Diaz, John Francis T., Do Scarce Precious Metals Equate to Safe Harbor Investments? The Case of Platinum and Palladium, *Economics Research International*, 2016, 2361954, 7 pages, 2016.
- [5] Kangalli Uyar, S.G., Uyar, U. & Balkan, E. Fundamental predictors of price bubbles in precious metals: a machine learning analysis. *Miner Econ* 37, 65–87 (2024).