Deconstructing the Friday the 13th Effect: Evidence from Dummy-Variable Regression, Wavelet Analysis, and XGBoost Prediction

Leyi Deng[#], Xiaowei Feng[#], Yunya Mao[#], Zijia Wan[#,*]

Shenzhen Middle School, Shenzhen, Guangdong Province, China

[#]These authors contributed equally to this work and share first authorship.

[*]Corresponding Author

Abstract

Starting from the question "Is Friday the 13th superstition or fact?", this paper constructs a comprehensive evaluation framework from statistical falsification, signal deconstruction, to intelligent prediction, and discusses whether "Friday the 13th" is supported by realistic evidence, as well as whether the personal well-being of the public is affected by superstitious date characteristics.For Task 1, we first established a multiple linear regression model with dummy variables (Task 1.1). Through the deconstruction of date attributes, it was found that both the regression coefficient of the "Friday the 13th" variable and the p-value of the one-way analysis of variance were far greater than the significance level. They were statistically insignificant, directly falsifying the special negative effect of "Friday the 13th" at the empirical evidence level. Subsequently, we constructed a dummy-variable regression model with multiple-comparison validation (Task 1.2), and conducted sensitivity analysis by combining the consistency test of residual performance across different regression methods, proving the robustness of the non-significant conclusion. It was also found that, when independence was controlled, the effects of different dates and cycles remained insignificant. To further explore the deeper characteristics of the time series, we used multi-scale wavelet analysis (Task 1.3) to perform time-frequency deconstruction of event frequency based on frequency and significance. The results show that risk energy is mainly concentrated in macro long-wave cycles, such as policy cycles or pandemic cycles, while specific date numbers only appear as random high-frequency noise. This indicates that, from the perspective of the financial market, there is no such thing as a "good day" or a "bad day."For Task 2, we constructed a multi-dimensional evaluation indicator system to measure the happiness index from five dimensions: the atmosphere of social public opinion, public health safety, social harmony and stability, climate change, and economic conditions, and used the Kendall correlation coefficient for heterogeneity validation. We then used the Entropy Weighting Method to determine the relative importance of each indicator. The results show that pandemic-related and environmental indicators (PM10) account for the highest proportion of influence on individual well-being. We then conducted a reliability analysis of the weight distribution in combination with practical reality. On the whole, the weight setting is consistent with realistic evidence.For Task 3, we used the XGBoost gradient boosting method to establish a risk prediction model. By analyzing various data features, the model can identify signals with real early-warning significance. The empirical research shows that the prediction performance based on physical environmental indicators is far better than the traditional superstitious prediction based on date numbers, providing rational risk early-warning rules for the public. At the same time, the prediction of "good days" and "bad days" largely depends on the updating of real-time information.The conclusion

shows that social risks are driven by institutionalized cycles and physical environmental variables, rather than date superstition. This model not only provides the public with a rational quantitative tool, but also provides a robust mathematical paradigm for dealing with heterogeneous social events.

Keywords

Friday the 13th, dummy-variable regression, multiple cross-validation, wavelet analysis, entropy weighting method, XGBoost, frequency features.

1. Introduction

1.1. Background of the Problem

For a long time, there has been a belief in Western superstition that when the 13th day falls on a Friday, that day is regarded as an unlucky day, bringing bad luck and misfortune, and it is therefore called "Friday the 13th." Friday the 13th originates from Western religious beliefs: Jesus Christ died on a Friday, and 13 is an unlucky number. The combination of the two makes people believe that unfortunate things will happen on that day. Out of fear of this, many people refuse to travel or even avoid any activities on this day, which directly leads to commercial losses in the United States [2][3].

1.2. Restatement of the Problem

The model established in this study aims to explore the following questions in the topic by establishing mathematical models: (1) whether the superstition of "Friday the 13th" has an impact on "negative events" in real life; (2) from the perspectives of specific frequency and non-fixed frequency, whether there really exist "good days" or "bad days" in such events; (3) according to such negative events, carrying out independent decomposition under other frequencies to identify periodic characteristics; (4) constructing a heterogeneous happiness index and confirming its weights; (5) how to predict "good days" or "bad days" based on the previous conclusions?The following are the key objectives that should be noted in the models for these questions: (1) For the question of whether the effect of "Friday the 13th" exists, we start from the perspectives of frequency and significance, decompose the occurrence frequency of this day and that of other days, and use the principle of significance to determine whether the difference between the event performance of this day and that of other dates is significant, and whether it is controlled by the single factor of "date." (2) For the problem of identifying whether other specific frequencies are "too good" or "too bad" and the problem of independent decomposition, we conduct cross-validation to compare whether these dates have an impact on the occurrence of "negative events" when they exist independently and when they act jointly. (3) For identifying "good days" or "bad days" under non-specific frequencies, we use a model that can automatically identify them according to the occurrence frequency and significance level of "good" and "bad" to analyze the problem. (4) For the construction of the happiness index indicator system, we divide the indicator system into different dimensions. After its establishment, we use the Kendall correlation coefficient to calculate the linear or nonlinear relationships among them, so as to ensure the heterogeneity of variable selection. (5) Last, for how to predict "good days" and "bad days," we combine the analytical conclusions of all the above processes and select a fitting method with an added penalty function to prevent model overfitting caused by the use of too many variables. We also consider using a multi-feature self-identification model to detect and adapt to the frequency and range of cycles.

1.3. Framework and Structure of the Paper

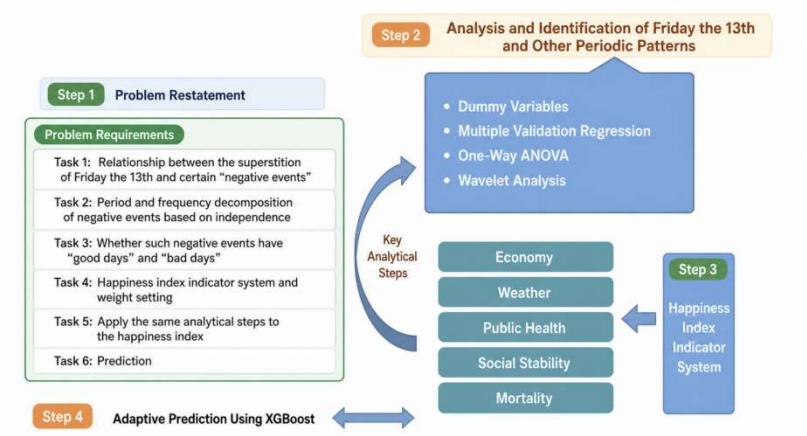


Figure 1. Paper Structure

2. Assumptions of the Paper

(1) Proxy variable for negative events: Financial market data are used as the observation source of negative events. Taking the Hong Kong financial market as an example, it has relatively high data availability and representativeness as an international financial center, and its financial policies have broad global influence. At the same time, individuals are highly sensitive to fluctuations in financial values, making it suitable for reflecting the impact of negative shocks on residents' psychology. (2) Proxy variable for inflation: Since gold itself can also be used as currency, the gold return rate is used as the proxy variable for the rate of change in inflation. (3) Measurement of major public health events: Since it is difficult to obtain pandemic data such as COVID-19 with consistent definitions and sufficiently long time-series coverage, this study uses dummy variables to capture the impact of such major public health events on individuals' perceived happiness. (4) Periodic analysis strategy: When analyzing the independence of different cycles, we respectively establish regression models containing only a single cycle and a comprehensive model combining all cycle variables, and verify the results with one-way and multi-factor analysis of variance. To avoid the decline in model degrees of freedom and the weakening of sample representativeness caused by too many interaction terms, interaction terms between cycles are not included in the comprehensive model. If neither the comprehensive model nor the one-way analysis presents significant statistical conclusions, it can be preliminarily judged that there is a lack of association among the cycle variables.

(5) Construction of the happiness index: This aims to construct a composite estimation model, namely the "luck indicator," for quantifying the "degree of luck" of a certain day from the perspective of ordinary citizens' daily life and psychological perception. The core method of the model in this paper is to systematically screen and integrate multiple events that occur within the day, come from different fields, and are independent of each other, namely heterogeneous events. Homogeneous event sequences are also included to construct a composite perceived indicator system with explanatory power and operability. A quantitative indicator reflecting the overall social mood and subjective feeling of "luck" on that day is ultimately synthesized, thereby transforming the abstract social psychological state into a data form that can be analyzed and compared. For the abstract impact of disease, we set a dummy variable and set the event occurrence period as 1.

(6) Entropy weighting method: In the paper, to determine the weights for calculating the happiness index using indicators from different dimensions, we carry out the study by using the characteristics of the entropy weighting method, namely being "objective," "using the

information provided by data volatility," and "being convenient for calculation and quantification."

3. Question 1: Monitoring Model for "Friday the 13th" and "Good Days? Bad Days?"

In this part, we establish a hybrid model using dummy variables, cross-regression models, and frequency analysis methods to test the legitimacy of "Friday the 13th" and detect date frequencies related to negative events. In the validation of the model, with reference to the sensitivity of the model, we use 12 research samples. On the basis of the basic model, multiple comparison methods, one-way analysis of variance, and significance evaluation are combined to repeatedly demonstrate the conclusions. Difference sensitivity analysis is conducted at the end to strengthen the reliability, stability, reproducibility, and universality of the model conclusions.

3.1. Definition of Negative Events

To make it more convenient for us to detect fluctuations with the minimum unit of a day, we take the financial market as an example and use U.S. Treasury bonds, gold, the exchange rate of the U.S. dollar against the Japanese yen, and the stock prices of companies in sectors such as semiconductors and electronics as the research objects, analyzing whether the return rate of the stock market is affected by "Friday the 13th." When the return rate is negative, a negative event occurs. The trend charts of these financial indices are as follows [1].

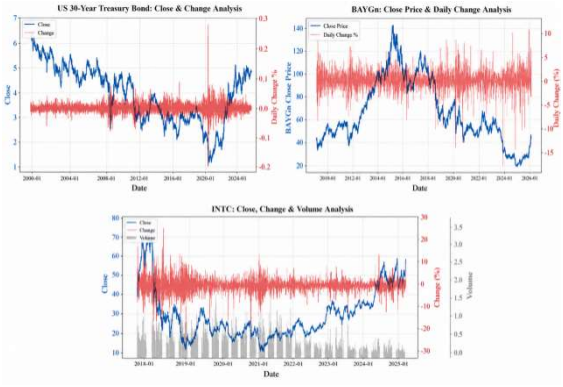


Figure 2. Trend Chart A of "Negative Events" in Financial Indicators

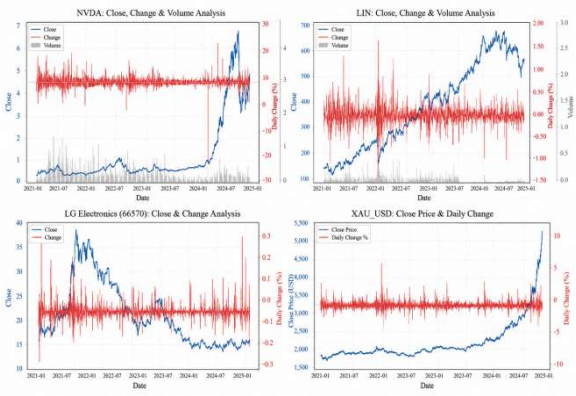


Figure 3. Trend Chart B of "Negative Events" in Financial Indicators

3.2. Model for Testing the Relationship Between "Friday the 13th" and the Returns of Various Financial Assets

Here, we select a regression model with fixed effects and use dummy variables and multiple testing methods to study whether the return rates of these 12 financial products, including bonds, currencies, gold, and stocks, are associated with Friday the 13th [2].

3.2.1. Fixed-Effect Regression Model Containing Dummy Variables

We attempt to capture the trend of index returns during this period by using regression analysis with dummy variables. For the special date of Friday the 13th, we define the corresponding dummy variable I

$$I_{0,Fridaythe13th,j}=\begin{cases}1,\textit{if the day t coincides with a Friday the 13th}\\0,\textit{otherwise}\end{cases} \quad (1)$$

For the days before or after Friday the 13th, we use two types of dummy variables to represent them, which are defined as follows:

$$I_{k,Fridaythe13t,j}=\begin{cases}1, if\ the\ day\ t\ is\ kworking\ days\ after\ a\ Friday\ the\ 13th\\ 0, otherwise\end{cases} \quad (2)$$

$$I_{0,Fridaythe13th,j}=\begin{cases}1, if\ the\ day\ t\ is\ kworking\ days\ after\ a\ Friday\ the\ 13th\\ 0, otherwise\end{cases} \quad (3)$$

$$Return_{ij}=\beta_0+\sum_{k=-3}^{k=3}\lambda_k I_{k,Fridaythe13t,j}+\varepsilon \quad (4)$$

3.2.2. Model Results

In the test of the 7-day window effect, the regression coefficient of the dummy variable term $I_{t,Fridaythe13th,j}$is not only very small and has almost no obvious impact on the return rate, but the directions of the variable coefficients of different products on these days are also inconsistent. It even appears that "Friday the 13th" has a positive promoting effect on the return rate on that day. Similarly, the coefficients of the vast majority of variables do not pass the statistical test, indicating that whether it is "Friday the 13th" has no statistically significant impact on the return rates of financial market products in probability.

Table 1. Model Test Results

Index	Regression oefficient	P-value	Confidence Interval	One-way ANOVA P-value
Bonds	0.0011	0.258	-0.001 0.003	0.258
S&P 500	0.0008	0.313	-0.001 0.002	0.313
Bayer	0.0004	0.737	0.002 0.003	0.737
USD/JPY	0.0002	0.824	-0.002 0.002	0.824
STM	0.0026	0.606	-0.006 0.011	0.545
XAU/USD	0.0026	0.110	-0.001 0.001	0.110
PGP	-0.0051	0.000	-0.008 -0.003	4.42e-05
Qnity	0.0028	0.570	-0.007 0.013	0.570
NVDA	-0.0067	0.816	-0.063 0.049	0.816
INTC	0.0036	0.374	-0.004 0..012	0.871
LG	0.0096	0.458	-0.016 0.035	0.332
LIN	0.0058	0.160	-0.002 0.014	0.0365

In the financial market, none of the four types of indices, namely stocks, bonds, gold, and exchange rates, shows the numerical characteristic of "Friday the 13th." Apart from the regression model, when we use the F statistic to conduct a one-way analysis of variance on "Friday the 13th" and all other dates, the statistical tests of all indices fail to pass, indicating that, on the whole, when the 13th day of a month falls on a Friday, the return rates of financial-related products on that day are not significantly different from those on other days. Among them, only the stock return rate of PGP shows a negative and statistically meaningful relationship with Friday the 13th. Therefore, from the perspectives of the overall fluctuation size of the data, the statistical test results of the regression coefficients, the directions of the regression coefficients, and the one-way analysis of variance, we do not observe the impact of "Friday the 13th" on the return rates of financial products in the financial market represented by 12 indices.

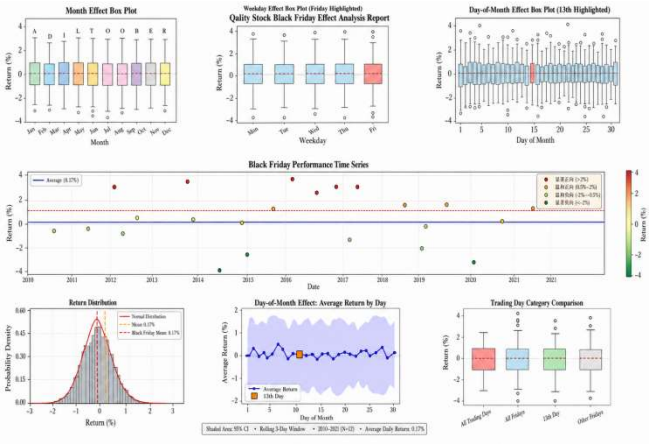


Figure 4. Dummy-Variable Regression Results of "Negative Events" on Friday the 13th (Using U.S. Bonds as an Example)



Figure 5. Decomposition of Financial Index Volatility

Through independent component decomposition of event frequency, the model analysis shows that the intersection of "Friday" and "the 13th" does not produce a significant synergistic gain effect. Although there may be a simple "day-of-week effect" or "within-month effect" in specific datasets, the interaction term specifically representing "Friday the 13th" does not show a statistically significant negative premium in various heterogeneous events. Therefore, from the perspective of empirical data, the long-standing belief that this day is an "omen of misfortune" lacks support from objective facts and belongs more to superstitious cognition at the level of social psychology.

3.3. Deconstructing "Superstition": Independence Test and Frequency Analysis of Date Effects Based on Regression Analysis

To verify whether the observed event occurrence frequency can be "decomposed" into several components, such as estimating the mutually independent effects of the "day-of-week effect" and the "day-of-month effect," we make slight modifications on the basis of the model in the previous question [4][5]. The model input variables are set as three variables: fixed days within a week, fixed days within a month, and different months. To verify their independence, we first conduct separate regression responses for these variables respectively. We then include the three variables in the same model to determine whether they have independent effects when the occurrence frequency of events is decomposed into weekly, monthly, and yearly frequencies, under the condition that other factors are controlled. The model settings are as follows.

$$Return_{ij}=\beta_0+\beta_1 Dom_{ti}+\varepsilon \tag{5}$$

$$Return_{ij}=\beta_0+\beta_1 Dow_{ti}+\varepsilon \tag{6}$$

$$Return_{ij}=\beta_0+\beta_1 year_{ti}+\varepsilon \tag{7}$$

$$Return_{ij}=\beta_0+\beta_1 Dom_{ti}+\beta_2 Dow_{ti}+\beta_3 year_{ti}+\varepsilon \tag{8}$$

The results show that, except for gold, all indices in the Calendar Effect model and the combined model only present significant factor analysis when each frequency is separately used for

benchmark regression. When all factors are controlled and the influence size of a single frequency on the return rate is decomposed under fixed factors, fluctuations under weekly frequency, monthly frequency, and yearly frequency are all not obvious. Therefore, in the current analysis, when the cycle is fixed as the basic unit on the calendar, we can say that there is no unified "date effect" in the financial market.

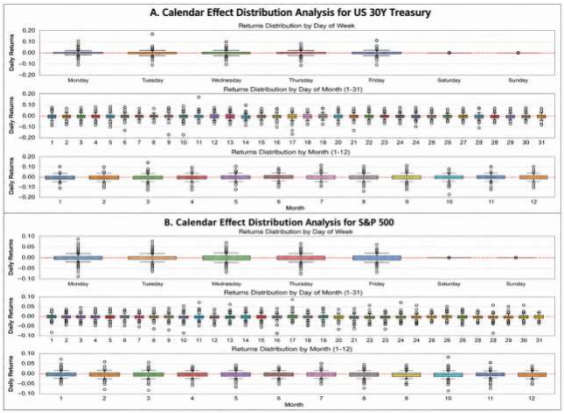


Figure 6. Comparison of Decomposition Effects of Some Financial Indices(Using U.S. 30-Year Bonds and the S&P 500 as Examples)

3.4. Cross-Scale Date Effect Detection: Time-Frequency Deconstruction Based on Wavelet Transform-Does the Financial Market Have "Good Days" or "Bad Days"?

3.4.1. Wavelet Frequency Analysis for Financial Return Rates

To detect any possible trends and cycles, we conduct continuous wavelet analysis on the data. The date is set as: y_t, t=1,2,…,T. Then we standardize the data.

$$x_t=\frac{y_t-\bar{y}}{s_y}, \tag{9}$$

$$\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t, \qquad s_y=\sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(y_t-\bar{y})^2}$$

(1) Select the mother sequence

$$\varphi_0(\eta)=\pi^{-\frac{1}{4}}e^{i\omega_0\eta}e^{-\frac{\eta^2}{2}}, \tag{10}$$

$$\omega_0=6$$

(2) Continuous wavelet frequency transform

$$W_x(s,\tau)=\sum_{t=1}^{T}x_t\cdot\frac{1}{\sqrt{s}}\Psi^*\left(\frac{t-\tau}{s}\right) \tag{11}$$

* is the complex conjugate.

(3) Frequency energy spectrum

$$P(s,\tau)=/W_x(s,\tau)/^2 \tag{12}$$

(4) Global frequency spectrum

$$G(s)=\frac{1}{T}\sum_{\tau=1}^{T}/W_x(s,\tau)/^2,$$
$$s^*=\arg\max_s G(s) \tag{13}$$

(5) Extract the frequency energy intensity curve

$$E_k(\tau)=\sum_{s\in S_k}|W_x(s,\tau)|^2 \tag{14}$$

(6) Frequency transformation

$$\hat{x}_k(t)=\frac{1}{C_\psi}\int\left(\int W_x(s,\tau)\frac{1}{s^{\frac{3}{2}}}\psi\left(\frac{t-\tau}{s}\right)d\tau\right)ds \tag{15}$$

C_ψ is the default constant.

3.4.2. Model Results

After performing the wavelet frequency transform, we compare the frequency energy maps of the 12 financial indices. The results show that the images of these 12 financial products differ greatly. This indicates that different financial products correspond to different good days and bad days, but for the public, investment behavior involves diversified asset portfolios. Taken together, from the perspective of wavelet analysis diagrams, when no frequency-cycle restriction is added, using only the occurrence frequency of extreme values for calculation still fails to obtain an effective and unified conclusion. Therefore, here we still reject the superstitious concept of “good days” and “bad days” and believe that it does not exist.

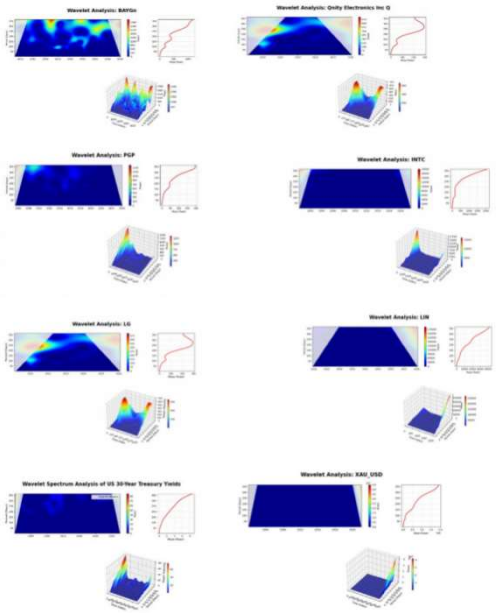


Figure 7. Frequency Energy Maps of Financial Indices

3.5. Sensitivity Analysis

To verify the robustness of our model, we additionally conduct logarithmic regression and robust regression for the regression models in 4.2 and 4.3. The results show that the

conclusions are basically consistent, and the differences among most indices across different model performances basically remain consistent. The robustness of the model is relatively good.

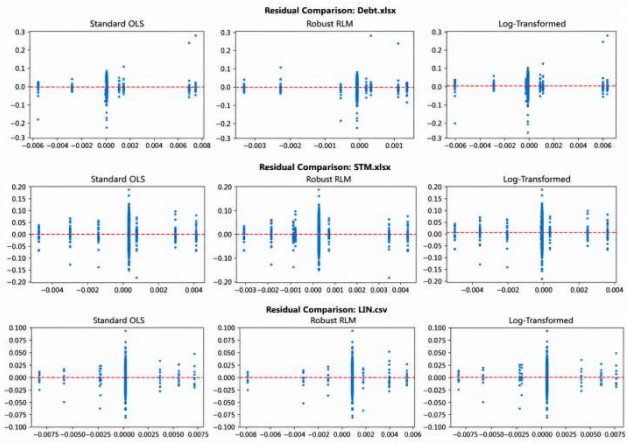


Figure 8. Sensitivity Analysis of Dummy-Variable Regression for Some Samples

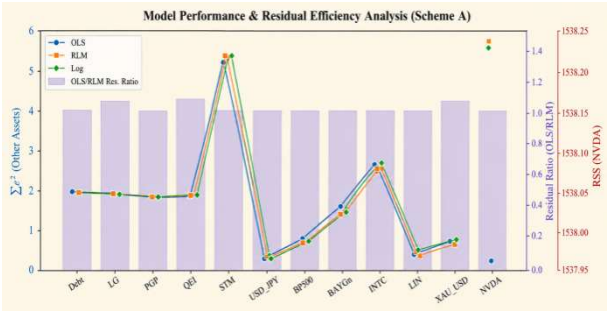


Figure 9. Line Chart of OLS, RLM, and Log Changes

We then also conduct sensitivity analysis on the wavelet analysis model introduced in 4.4. After changing the standardization method and adding noise to the data for fitting respectively, the conclusions among different financial indices still remain basically consistent.

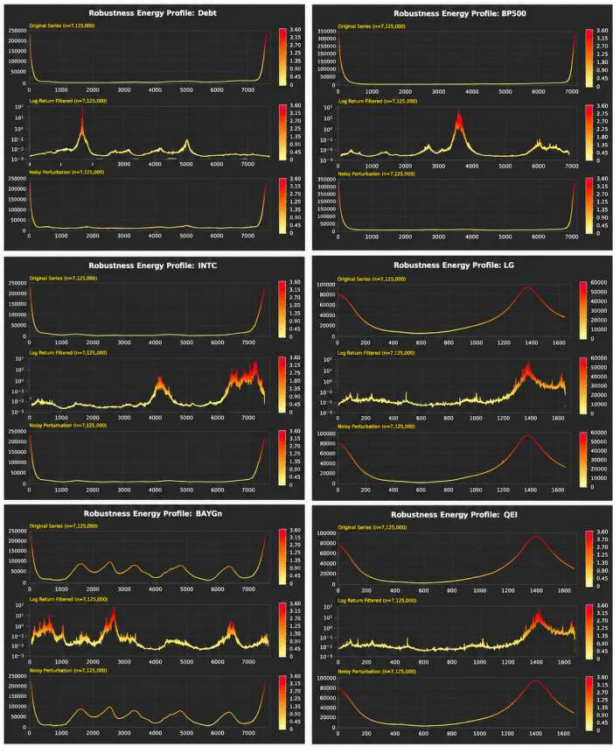


Figure 10. Sensitivity Test for Wavelet Analysis of Financial Indices

4. Establishment of the Happiness Index Indicator System and Weights

4.1. Construction of the Indicator System

We establish the indicator system and conduct the study based on data from 2020 to 2026. The "luck" perceived by ordinary people largely comes from the subconscious evaluation of the

"safety" and "controllability" of the social environment in which they live. Therefore, we divide the indicator system into the following five major categories:

Public health events: Epidemic prevention and control measures have a wide distribution range and a force-majeure nature. They not only directly affect health, but also change the baseline of happiness by restricting social and economic activities. Here, we construct a dummy variable. Based on the ending time of the database published by JHU, the value is set as 1 before this time, indicating that epidemic management occurs, and the rest is set as 0.

Economic fluctuations: They erode individuals' sense of control over the future. This "uncontrollability" can generate a sense of powerlessness and is often attributed to "bad luck." Here, we use the return rates of U.S. 30-year bonds and gold as representative data for the study.

Social public opinion: As a real-time indicator of collective emotion, social public opinion is internalized into individuals' judgment of fortune through social identity and emotional contagion mechanisms (Michael & Mark, 2018) [8]. When collective emotion is low, individuals are also easily affected. Here, we use the search volume of the term "TIRED" in Google Trends for analysis.

Climate: It directly affects daily emotions. Studies have shown that specific temperatures, such as 13.9°C, are associated with higher happiness (Osaka University, 2012) [7]. Here, we use the daily average humidity in Hong Kong as the representative indicator.

Social stability: It directly threatens the safety of life and property and shakes the most basic safety needs of human beings (Maslow's A Theory of Human Motivation, 1943) [6]. Its deterioration can shake the foundation of the perception of "luck." Since it is difficult to obtain high-frequency statistics on crime rates, we use headline content about crime in the Baidu Index to represent fluctuations in social safety.

Subsequently, through Kendall's τ correlation analysis of the selected indices, we find that both linear and nonlinear correlations among these indicators are weak, and the heterogeneity characteristics among the data are obvious. This further proves that the selected indicators cover a relatively broad range and have representativeness and extensiveness. Therefore, they are retained for subsequent further demonstration and research.

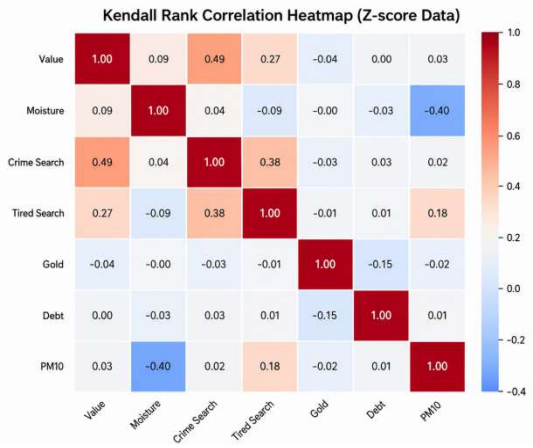


Figure 11. Heterogeneity Analysis of the Indicator System

4.2. Entropy Weighting Method

The entropy weighting method is an objective weighting method derived from the concept of entropy in information theory. It determines the weight of each indicator by analyzing the amount of information contained in each indicator. This method is relatively objective and has been widely used in previous studies to calculate the weights of economic and social indicators. The data are already relatively heterogeneous, so there is no need to consider methods such as

principal component analysis. Therefore, we can directly use the entropy weighting method to calculate the weights, and then compare the results with the realistic basis to verify whether they are reasonable.

(1) Calculate the indicator proportion

$$p_{ij}=\frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}} \tag{16}$$

(2) Calculate the information entropy e_j

$$e_j=-k\sum_{i=1}^{m} p_{ij}\ln\left(p_{ij}\right) \tag{17}$$

$k=\frac{1}{ln(m)}$. e_j reflects the degree of information dispersion of indicator j.

(3) Coefficient of variation d_j

$$d_j=1-e_j \tag{18}$$

(4) Calculate the entropy weight w_j

$$w_j=\frac{d_j}{\sum_{j=1}^{n} d_j} \tag{19}$$

4.3. Weight Results

This study uses the entropy weighting method to objectively assign weights to seven heterogeneous indicators and combines qualitative analysis to judge the rationality of the weight results, aiming to eliminate the interference of subjective preference in the measurement of "personal happiness." According to the calculation results shown in Table 2, the weight distribution presents significant hierarchical characteristics, deeply reflecting the impact mechanism of the macro environment on individual well-being from 2020 to 2026:

(1) Identification of core influencing factors: The COVID-19 occurrence period (W=18.78%) occupies the highest weight, and its information entropy value (E=0.9922) is the lowest, indicating that this indicator has the most significant information difference and identification power in the time series and is the primary endogenous variable determining the fluctuation of public happiness during this period. As the largest external shock within the sample interval (2020–2026), this conclusion is basically reasonable and has a realistic basis.

(2) Resonance between living environment and psychological burden: PM10 (16.92%) and the tiredness search index (15.42%) follow closely. This weight distribution reveals the coupling relationship between external physical environmental pressure and internal psychological exhaustion, confirming the high contribution rate of environmental quality and mental health indicators in measuring individuals' "degree of luck" or "sense of well-being."

(3) Influence paths of economic fluctuations and social security: The weight distribution of gold return rate (13.33%) and U.S. Treasury bond return rate (11.41%) reflects the indirect mapping of safe-haven asset fluctuations onto public psychological expectations. The weight of the crime search index (14.80%) quantifies the marginal impact of perceived social safety on individual happiness.

Overall, the distribution logic of the coefficient of difference (D) and the final weight (W) is self-consistent. The weight results are highly consistent with real social observations and provide a

priority sequence with data support for the subsequent construction of "scientific risk-avoidance rules."

Table 2. Entropy Weighting Method Results

Indicator	Information Entropy (E)	Coefficient of Difference (D)	Final Weight (W)
COVID-19 occurrence period	0.9922	0.0078	18.78%
PM10	0.9930	0.0070	16.92%
Tiredness search index	0.9936	0.0064	15.42%
Crime search index	0.9939	0.0061	14.80%
Gold return rate	0.9945	0.0055	13.33%
U.S. Treasury bond return rate	0.9953	0.0047	11.41%
Air humidity	0.9961	0.0039	9.35%

5. Model Extension: Is the Happiness Index Affected by "Friday the 13th" or Other Date Patterns?

5.1. Study on Friday the 13th and Independent Decomposition of Dates

By constructing a dummy-variable linear regression model, the explanatory power of date attributes is quantitatively evaluated. The results show that the core explanatory variable, namely the composite dummy variable representing the overlap between "the 13th" and "Friday," shows statistical non-significance in the regression model, and other date effects do not exist either.

5.2. Wavelet Analysis Results

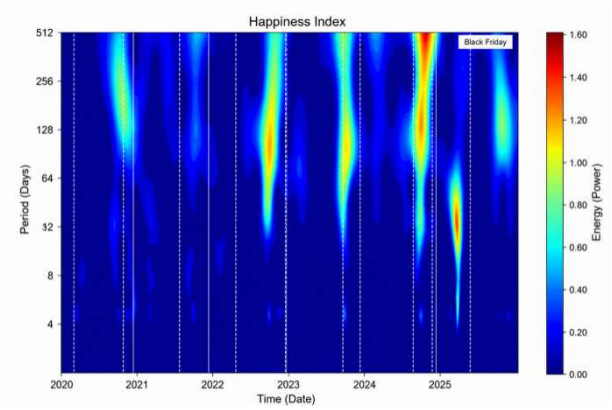


Figure 12. Wavelet Analysis Energy Heatmap

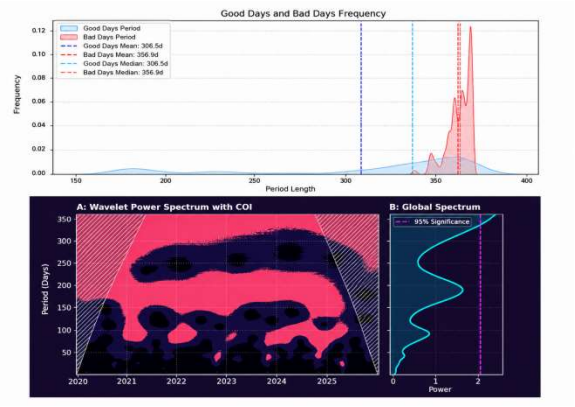


Figure 13. Subplots of Wavelet Analysis of the Happiness Index

Subsequently, after conducting wavelet analysis, we also find that, according to the frequency energy map and the output frequency feature sequence, among all the days studied (2,192 days), 277 days are identified by the model as "good days," while 398 days are identified as "bad days." The median cycles of good days and bad days are 334 and 360 respectively. The overall dominant cycle is approximately 365 days. To ensure the reliability of the model results, we conduct COI analysis. The results show that "bad days" are statistically significant in 95% of cases. Therefore, from the perspective of data, although from the perspective of significance, the happiness index is indeed higher or lower on specific dates than at other times. Yet due to the instability of the cycle, and judging from the results, the occurrence frequencies of "good days" and "bad days" are not the same as the frequency setting under the standard Gregorian

calendar year. Therefore, on the whole, although such a pattern really exists, from the perspective of public experience, the date positions of "good days" and "bad days" keep moving downward, so there is no fixed date. Strictly speaking, there are no "good days" or "bad days."

6. Prediction of "Good Days" and "Bad Days" based on the Happiness Index

6.1. Model Setting

Here, we input the frequency energy feature sequence $D=\{(x_i,y_i)\}_{i=1}^{n}$ obtained from the previous wavelet analysis into the XGBoost model. The gradient descent method is used to minimize the objective function consisting of regression error and the penalty function, so as to conduct feature-value screening and update predicted values.

(1) Initialize prediction

$$\hat{y}_0=\arg\min_c \sum_{i=1}^{n} l(y_i,c) \tag{20}$$

(2) Calculate the gradient g_i and Hessian h_i

(3) Use Gain to select the optimal split and grow a tree structure q

(4) Calculate each leaf weight w_j^{*}

(5) Update prediction $\hat{y}^{(t)}=\hat{y}^{(t-1)}+\eta f_t$

(6) Output the model and predict $y(x)=\sum_{t=1}^{T} f_t(x)$

6.2. Model Results

The focus of the XGBoost model is not to obtain the parameters of the model, but to update the next prediction result through real-time synchronized data. Judging from the results, the prediction accuracy using XGBoost is relatively high, and the difference between the predicted value and the real value is basically controlled within the range of [0,1][0,1][0,1]. This indicates that the model fitting effect is good. Therefore, if "good days" or "bad days" are to be predicted, it can be considered to first obtain the frequency energy feature sequence through wavelet analysis, and then input the sequence into the XGBoost model, so that the model can gradually iterate to the optimum through the gradient descent function and the penalty function.

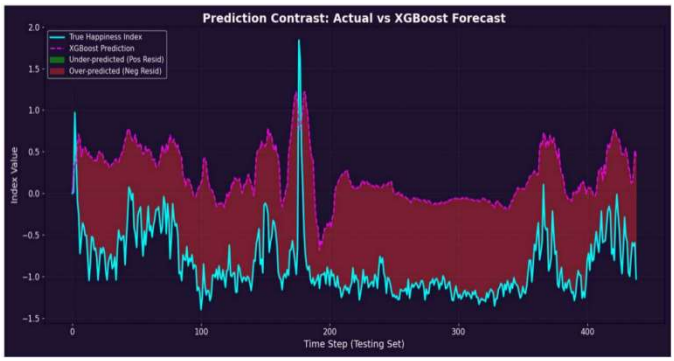


Figure 14. Prediction Results of the XGBoost Model

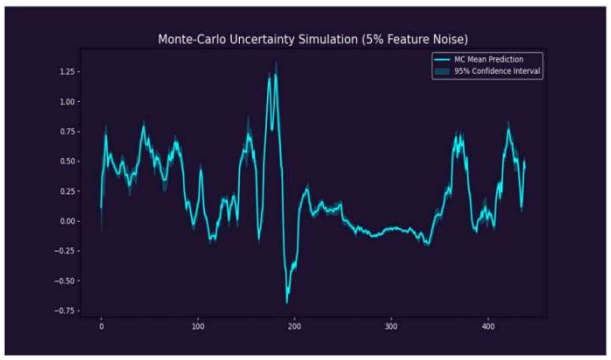


Figure 15. Sensitivity Analysis of XGBoost Model Prediction

6.3. Sensitivity Analysis

After conducting Monte Carlo analysis and What-if analysis, the processed images obtained are basically consistent with those of the XGBoost model without noise processing, further proving

the excellence of the XGBoost model in predicting and confirming the frequency of "date superstition."

7. Conclusion

Starting from the long-standing superstition of "Friday the 13th" in Western society, this study first explores whether Friday the 13th has an impact on the occurrence frequency of "negative events." Then, through the quantitative analysis of multi-source heterogeneous data such as climate, economy, social public opinion, social stability, and public health events, this study systematically explores whether "Friday the 13th" is a fact or a superstition, and whether there exists a date effect related to the public's perceived happiness. The results show the following:

7.1. Falsification of Superstition

The regression analysis results show that, among all financial indices, the coefficient of the dummy variable representing "Friday the 13th" is statistically insignificant, and its marginal contribution to the 12 indices of the financial market studied is also very small. This indicates that the negative effect of "Friday the 13th" does not exist in the financial market. Since the index return rate of the financial market itself is an extremely sensitive indicator and is also closely related to the public in Western society, based on the above conclusions and the relevant literature review, we can basically consider that "Friday the 13th" is only a cultural psychological projection rather than a fact.

7.2. Lack of Perceived Periodic Effect

As for whether there exist "good days" or "bad days" in "negative events" and the public perceived happiness index, after wavelet analysis, and after observing and outputting the frequency energy sequence and the global mother sequence, we find that, in terms of cycles, the cycles that can be judged as "good days" and "bad days" according to frequency are different. The medians of these cycles are not equal to any number that people can judge as regular, such as 7, 30, 365, or lunar calendar time. Therefore, this effect also does not exist.

*Model advantages:*Fusion capability of multi-source heterogeneous data. According to the requirements of the topic, we differentiate the types of indicators in the happiness index indicator system and successfully integrate factors such as financial returns, environmental indicators (PM10), and social emotion search indices, breaking the limitation of single-category data. Scientific rigor of the method. We use dummy variables to "deconstruct" the date effect. This method can separate the simple "Friday" factor and the "13th" factor, thereby accurately capturing the net effect of their interaction term. Then we use the entropy weighting method, relying on the information entropy of the data itself to determine the weights, and compare the weight results with realistic evidence to ensure their rationality. In this way, we ensure both the objectivity of the "luck index" evaluation system and its conformity with the requirement of "scientific evidence." Strong robustness. In each task, we use multiple sensitivity analysis methods, such as comparisons among OLS, RLM, and log-transformed models, and conduct Monte Carlo simulations. The results prove that the conclusions basically remain consistent under different algorithms, demonstrating the robustness of the model method.

Model limitations: Insufficient consideration of lagged effects. Our regression model mainly focuses on the same-day effect. When considering the lagged impact of major negative events, such as COVID-19, on public psychology, the conclusions often become unclear due to too many features and the loss of model degrees of freedom. Simplification of cultural differences. Although this model is based on global data, there is still room for optimization in handling the heterogeneity of the definition of "luck" under different cultural backgrounds. In the future, geographically weighted regression (GWR) can be introduced to analyze the cultural sensitivity of specific regions.

Acknowledgement

All authors contributed equally to this work. The authors are listed in alphabetical order, which does not indicate any priority or hierarchy of contribution.

References

[1] Moutidis, I., & Williams, H. T. P. (2020). Good and bad events: Combining network-based event detection with sentiment analysis. *Social Network Analysis and Mining*, 10(1), Article 64. https://doi.org/10.1007/s13278-020-00681-4

[2] Fidrmuc, J., & Tena, J. D. (2015). Friday the 13th: The empirics of bad luck. *Kyklos*, 68(3), 317–334. https://doi.org/10.1111/kykl.12085

[3] Scanlon, T. J., Luben, R. N., Scanlon, F. L., & Singleton, N. (1993). Is Friday the 13th bad for your health? *British Medical Journal*, 307, 1584. https://doi.org/10.1136/bmj.307.6919.1584

[4] Bier, V. M. (1993). Statistical methods for the use of accident precursor data in estimating the frequency of rare events. *Reliability Engineering & System Safety*, 41(3), 267–280. https://doi.org/10.1016/0951-8320(93)90079-E

[5] Petersen, T. (1991). The statistical analysis of event histories. *Sociological Methods & Research*, 19(3), 270–323. https://doi.org/10.1177/0049124191019003002

[6] Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396.

[7] Tsutsui, Y. (2012, March 1). Weather and individual happiness. Osaka University Institutional Repository. https://ir.library.osaka-u.ac.jp/repo/ouka/all/14034/?lang=1&mode=1&opkey=R173991538920290&idx=10

[8] Hogg, M. A., & Rinella, M. J. (2018). Social identities and shared realities. *Current Opinion in Psychology*, 23, 6–10. https://doi.org/10.1016/j.copsyc.2017.10.003