

Article

Not peer-reviewed version

A Daily Funding Demand Forecasting Model for Fintech Investment Decisions and Its Impact on Investment Return Stability

[Xiaoyi Meng](#)* and Shaochun Liu

Posted Date: 11 March 2026

doi: 10.20944/preprints202603.0804.v1

Keywords: FinTech; funding demand forecasting; investment return optimization; risk forecasting; XGBoost; data-driven decision making



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Daily Funding Demand Forecasting Model for Fintech Investment Decisions and Its Impact on Investment Return Stability

Xiaoyi Meng ^{1,*} and Shaochun Liu ²

¹ University of California, Irvine, CA 92618, US
² Beijing JoinQuant Investment Management Co., Ltd, Beijing, 100020, China
* Correspondence: ariaaaameng@gmail.com

Abstract

The accuracy of financing demand prediction has a direct impact on the return on investment and risk exposure in fintech investment and asset allocation. Nevertheless, the real world financial transaction data often displays significant nonstationary features — for example, cyclical fluctuations, event shocks, and short-term anomalies — which make the traditional forecasting approach unstable in the real investment scenarios. This study builds a data set that includes 34 reproducible variables — including daily financing requirements, transaction peaks, capital occupation duration, and risk exposure levels — on the basis of 180 consecutive days of investment and operating data from a leading financial services firm. It systematically compares ARIMA, Prophet, Random Forest, and XGBoost models for financing demand forecasting. Empirical results show that XGBoost maintains a low forecast error (MAPE of 8.2%) in the case of market fluctuations and unusual events, which reduces the average error by about 22% in comparison with the baseline model. Based on these results, a model is built to analyze the effect of forecast errors on the stability of investment returns and the efficiency of capital turnover. Results show that keeping the forecast error under 10% significantly reduces the risk of capital misallocation in times of high volatility, while at the same time improving the stability of overall investment returns. This study provides a reusable model workflow and engineering reference for the establishment of the investment allocation and risk management system of the financial institutions.

Keywords: FinTech; funding demand forecasting; investment return optimization; risk forecasting; XGBoost; data-driven decision making

1. Introduction

In the fintech field, fund demand forecasting is critical, since precise capital allocation and efficient allocation of assets have a direct impact on the return on investment and risk reduction. In order to solve this problem, we present a framework for the integration of XGBoost and Random Forest models. Through the systematic analysis of the non-stationary character of financing demand, it can optimize the decision making. The introduction of Error Analysis and Funding Mismatch Risk Assessment has further improved the reliability of Prediction Results and System Response, and has provided a workable decision-making support tool for Financial Institutions.

2. Dataset Construction and Feature Engineering

2.1. Data Sources and Description

The indicator dataset for model training uses operational capital flow records from a major financial services institution as the core sample. Raw records span January 1, 2024 to June 30, 2024 (180 calendar days). To ensure temporal validity for financial modeling, weekends and 14 exchange-recognized holidays were removed, resulting in 126 effective trading-day observations. All derived indicators are

constructed using only $t-1$ and earlier historical data windows to avoid forward-looking leakage. The dataset therefore represents a daily indicator dataset, with intraday statistics sampled at 5-minute intervals and aggregated into daily features [1]. To quantify abnormal liquidity pressure on each trading day, the Daily Capital Stress Index (DCSI) is defined based on within-day sampled capital outflow distributions. For each effective trading day t , the raw index prior to scaling is:

$$DCSI_t = \frac{\max(x_{t,k}) - \bar{x}_t}{s(x_t)} \quad (1)$$

where $x_{t,k}$ denotes intraday capital outflow snapshots (million yuan) collected every 5 minutes, $\max(x_{t,k})$ is the maximum 5-minute outflow value on trading day t , $\bar{x}_t$ is the mean of all intraday snapshots on that trading day, and $s(x_t)$ is the sample standard deviation of intraday outflows on trading day t . This index measures the relative extremity of intraday funding stress. Before model input, all indicators including DCSI are Z-score standardized across the trading-day set. DCSI participates only in lagged form (DCSI $_{t-1}$) or as a feature derived from past windows. Subsequent experiments regress multi-horizon forecast errors on lagged DCSI to assess investment return stability under stress spikes [2].

2.2. Design of Daily Fund Demand Indicator System

The construction of the daily fund demand indicator system aims to capture transaction intensity, volatility, and structural risk exposure. To quantify the daily funding gap, the Daily Funding Demand Index (FDI) is defined as a normalized expression of capital imbalance and time-weighted utilization pressure:

$$FDI_d = \frac{Q_d^{out} - Q_d^{in}}{\tau_d + \lambda_d} \quad (2)$$

Where: Q_d^{out} represents the total fund outflow across all accounts on day d (in millions of yuan), Q_d^{in} denotes the total fund inflow on the same day (in millions of yuan), τ_d indicates the average fund holding time within the day (in hours), and λ_d signifies the number of fund transfers (in instances). All terms are standardized before aggregation to ensure dimensional consistency. FDI serves as the main prediction target in model training [3]. To reflect non-linear transfer structures across accounts, the Capital Routing Complexity (CRC) is introduced. It is computed as:

$$CRC_d = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\frac{L_i^2}{C_i + 1} \right) \quad (3)$$

Where N_d denotes the number of valid fund paths occurring on day d , L_i represents the number of jumps in path i , and C_i indicates the number of fund reversals between accounts involved in that path. Paths are extracted from transaction graphs via depth-limited traversal of the fund flow network. A higher CRC implies lower routing efficiency and increased distribution uncertainty [4]. In subsequent analysis, FDI (as target) and CRC (as input feature) jointly support error sensitivity modeling, particularly under stress and network congestion condition.

2.3. Anomaly Handling and Time Series Preprocessing Methods

Unexpected events, nonoperational transfers, or data-logging errors can produce local extremal values in the day-to-day financing requirements, thus undermining the underlying assumptions of the model. In order to improve the adaptive ability of the model, an adaptive Z-score adjusting algorithm is proposed, which is defined as follows:

$$Z_k^{adj} = \frac{F_k - M_w}{S_w + \delta} \quad (4)$$

where F_k denotes the funding demand value on day k (in millions of yuan), M_w and S_w represent the moving average and standard deviation over a window length of $w=7$ days, respectively, and δ is a minimal constant introduced to prevent a zero denominator (set to 10^{-6} in this paper). When $|Z_k^{adj}| > 3$, the value is flagged as an outlier and corrected using local linear interpolation. This ensures the data sequence retains its actual volatility trend while mitigating noise interference that could compromise model stability [5].

To address non-continuous trading days in financial markets and complete the continuous time index sequence required for forecasting, a funding demand filling algorithm based on holiday compensation weights is employed. For missing dates k , the fill value H_k^{fill} is defined as:

$$H_k^{fill} = \sum_{j=1}^{N_p} \frac{C_j}{L_j} \times W_h \tag{5}$$

where N_p denotes the number of valid funding paths during the corresponding historical period, C_j represents the net funding flow of path j , L_j indicates the funding occupancy duration for this path, and W_h is the holiday weight coefficient based on temporal proximity. The resulting imputation values are treated equivalently to original observations in model inputs.

To eliminate the impact of long-term trend drift on model parameter convergence, the capital demand sequence undergoes logarithmic transformation followed by differencing [6]. The resulting differenced sequence is tested for stationarity via the ADF test, whose statistic is defined as follows:

$$A_{ADF} = \frac{\sum_{i=1}^{T-1} (Y_{i+1} - Y_i) \times Y_i}{\sum_{i=1}^{T-1} Y_i^2} \tag{6}$$

where Y_i denotes the log-transformed fund demand value on day i , and T represents the sequence length. If A_{ADF} is less than the significance threshold, the null hypothesis of sequence stationarity is accepted. The preprocessed data is directly employed for rolling model training and backtesting experiments. It forms a rigorous correlation with prediction errors, fund mismatch rates, and investment return volatility metrics, ensuring the stability and economic interpretability of model inputs during engineering deployment.

2.4. Feature Engineering and Variable Importance Assessment

To enhance the generalization capability and interpretability of the capital demand prediction model, a derived feature set encompassing dimensions such as transaction intensity, volatility, capital structure, and risk exposure is constructed based on the original variables. Within the transaction intensity dimension, the time-weighted flow rate indicator R_d^{adj} is introduced, defined as follows:

$$R_d^{adj} = \frac{1}{T_d} \sum_{t=1}^{T_d} \frac{Q_{d,t} \times \omega_t}{\tau_t + \eta} \tag{7}$$

Among these, $Q_{d,t}$ represents the net capital flow during the t time period on the d th day, ω_t denotes the trading period weight, τ_t indicates the average trading interval time for that period, and η serves as a positive constant to prevent the denominator from becoming zero. With respect to the volatility characteristics, the HFT is designed to describe unusual fluctuations in trading patterns [7]. In addition, a network metric called the "Capital Node Dispersion Index" is proposed in order to capture the effectiveness of capital flows in an account structure. Higher values suggest a lower concentration of capital allocation, indicating a higher possible risk of mismatch risk. In the construction of the XGBoost model, all 34 variables were ranked using information gain-based feature importance assessment. The first five variables contributed to more than 10% of the final MAPE, as illustrated in Table 1. This ranking will be used for feature pruning and engineering simplification during subsequent model compression and deployment.

Table 1. Feature Importance Evaluation Results (Based on XGBoost Information Gain Metric).

Feature Name	Feature Type	Relative Importance (Gain)
FDI (Daily Funding Demand Index)	Target Variable Derived	0.241
R^{adj} (Intraday Flow Rate)	Derived Variable	0.196
DCSI (Funding Peak Pressure)	Defined Indicator	0.162
CRC (Funding Path Complexity)	Network Structure Variable	0.113
FNDI (Fund Node Dispersion Index)	Derived Network Metrics	0.098

3. Construction of the Fund Demand Forecasting Model

3.1. Model Selection Logic and Modeling Process

In order to improve the accuracy and robustness of financing demand prediction, a hybrid model is proposed, which takes into account both linear trends and nonlinear disturbances, as

illustrated in Figure 1. The core modeling objective is to minimize multi-step rolling forecast error, defined by the following objective function:

$$L_{\text{total}} = \sum_{h=1}^H \lambda_h \times |\hat{F}_{t+h} - F_{t+h}|^\gamma \quad (8)$$

where $\hat{F}_{t+h}$ denotes the forecast value for day $t+h$, F_{t+h} represents the corresponding actual value, λ_h is the step-length weighting coefficient (weighting nearer future time points), γ is the deviation penalty order (set to 1.2 in this paper to enhance anomaly suppression), and H is the forecast window length. Different model structures are set as parallel branches within the model ensemble, with weighted averaging of outputs forming the final prediction [8]. This objective function is uniformly applied throughout all model training stages to ensure consistent error constraint measurement logic in capital return simulations.

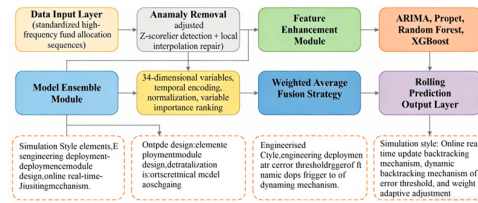


Figure 1. Capital Demand Forecasting Modeling Flowchart.

3.2. Benchmark Model Design

To establish a baseline performance reference system, this study selects the classic time series forecasting model ARIMA and the complex trend-oriented Prophet model as comparison models. Their output error intervals serve as performance boundaries for the ensemble model. The ARIMA model employs a combined structure of differencing, autoregression, and moving average. Its forecasting process can be formalized as:

$$\Phi(B)(Y_t - \mu) = \Theta(B)\varepsilon_t \quad (9)$$

where Y_t represents the smoothed funding demand sequence on day t , μ denotes the sequence mean, $\Phi(B)$ and $\Theta(B)$ denote the autoregressive and moving average polynomial operators for lagged terms, respectively, and ε_t is the white noise term. Parameter order is determined using the AIC minimum criterion, and the residual sequence must satisfy the Ljung-Box test conditions. In contrast, the Prophet model enhances the model's ability to express irregular periodic disturbances by decoupling the trend term $g(t)$, seasonal terms $s(t)$, and holiday term $h(t)$ [9]. Its form is:

$$\hat{Y}_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (10)$$

where $\hat{Y}_t$ represents the fitted forecast, $g(t)$ denotes the piecewise linear trend function, $s(t)$ is the periodic Fourier series expansion term, $h(t)$ indicates the holiday shock term, and ε_t represents the error disturbance. As shown in Figure 2, this model demonstrates strong fitting performance during the end-of-month period with a significant holiday effect. In the rolling prediction experiment, Prophet outperformed ARIMA during the stable period, but exhibited significant error propagation during abnormal fluctuations.

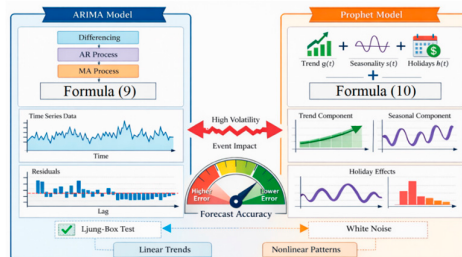


Figure 2. Comparison of fitting errors between ARIMA and Prophet models within the sample interval.

3.3. Ensemble Model Design

In order to improve the capability of identification of nonlinear disturbance and interaction, a framework of Random Forest and XGBoost is established. This approach is able to overcome the limitations of traditional models in dealing with high dimensional and temporal dependencies. Random Forest uses all input features to construct deep decision trees, and uses Bagging to reduce the risk of overfitting. It can be used to capture nonlinear cross-effects between multidimensional variables.XGBoost is able to use gradient weighting and regularization to improve the stability of the model. Both models incorporate variable importance evaluation mechanisms, which can be used to improve the performance of the model. The key inputs are the capital demand index, the path complexity, and the flow rate [10]. Parameter tuning takes place through multiple rounds of cross validation and rolling training across time windows. As illustrated in Figure 3, a full closed loop from feature input to forecast output is achieved. It also supports the integration of the investment return simulation module to measure the marginal effect of the forecast error on the capital mismatch.

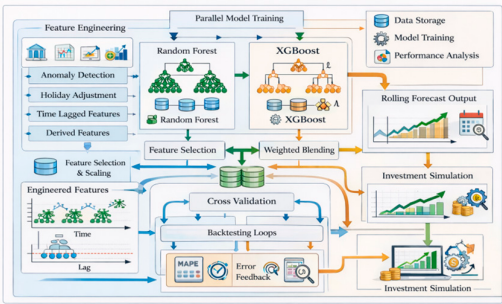


Figure 3. Integrated Model Construction and Rolling Forecast Process.

4. Experimental Design and Results Analysis

4.1. Investment Allocation Simulation System Development

To evaluate how funding-demand forecasts influence the stability of investment returns, a daily-granularity investment allocation simulation framework was constructed. Model inputs include funding-demand forecasts for day t and lagged intraday indicators aggregated from only $t-1$ and earlier data. Intraday capital flow snapshots are sampled at 5-minute intervals and aggregated to daily features using time-weighted mean based on interval transfer volume, ensuring no forward-looking leakage. The simulation adopts a rolling backtest design with parameters fixed for reproducibility: training window = 60 trading days, validation window = 20 trading days, forecast horizon = 5 days, and roll step size = 10 trading days. Capital allocation decisions are simulated for each horizon forecast, and return stability is measured via 30-day rolling standard deviation of portfolio returns. Funding mismatch risk is formalized by Mismatch Rate (MR) defined in Eq.(12) (computed using only past allocation deviations). Allocation strategy switching is triggered by forecast-error tiers: 0-5% (Tier-1), 5-10% (Tier-2), 10-15% (Tier-3), each mapping to a quantitatively bounded allocation regime that modulates the lagged high-frequency transfer ratio (HFR_ $t-1$) and risk-capital weight (RCW_ $t-1$). All simulated inputs/outputs are logged in structured tables for verification.

4.2. Impact of Forecast Error on Capital Allocation Stability

Forecast deviations exhibit non-linear effects on capital allocation stability. Based on the 5-day rolling average MAPE, prediction errors are stratified into four tiers: Tier 1, Tier 2, Tier 3, and Tier 4. For each tier, we compute the Mismatch Rate (MR) and Turnover Rate Shift (TRS) to quantify capital allocation disruptions. The MR measures the absolute deviation between predicted and simulated allocation amounts, while the TRS captures the variance of daily turnover rate over a rolling window. As shown in Figure 4, when forecast error remains below 10%, the MR stays under 7%, and TRS is stable within a $\pm$

5% band. During high-volatility days in this range, the Redundancy Ratio (RR)—defined as the proportion of idle capital—declines steadily, indicating efficient utilization. However, in Tier 4, MR increases to over 14.6%, and TRS variance exceeds 11.3%, signaling unstable redistribution. These patterns reveal a non-linear degradation of system stability beyond the 10% error threshold. The results confirm that error-aware tiered strategies embedded in the simulation module effectively reduce mismatch risk in low-error intervals and suppress instability propagation under moderate volatility. This justifies the integration of forecast deviation thresholds into dynamic allocation switching logic.

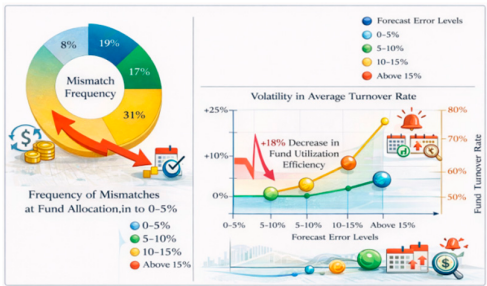


Figure 4. Impact of Prediction Error on Allocation Stability.

4.3. Analysis of Capital Misallocation Risk and Return Volatility

To further assess the impact of forecast-driven capital misallocation, this section analyzes the relationship between allocation mismatch rates and return volatility. Misallocation events are defined as periods where the mismatch rate exceeds 20% for at least two consecutive trading days. These events are grouped based on frequency and duration. Concurrently, the daily return series is used to compute the Sharpe Ratio (SR), representing risk-adjusted return, and the End-of-Day Deviation (EOD), reflecting deviations in final-period liquidity levels. As illustrated in Figure 5, the investment return volatility under high mismatch conditions shows an increase of over 27% in standard deviation compared to the benchmark strategy. These spikes are not randomly distributed; they exhibit clustering aligned with forecast deviation peaks. Once the mismatch rate surpasses the 20% threshold, the system demonstrates a decline in hedging effectiveness and a delayed return to equilibrium. Figure 5 further reveals overlapping clusters between mismatch rate bands and elevated return volatility zones, suggesting a direct coupling effect. This pattern supports the interpretation that forecast inaccuracy induces systemic misallocation risk, which in turn amplifies return instability.

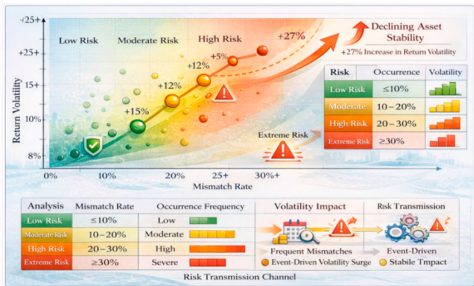


Figure 5. Correlation Analysis of Fund Misallocation Rate and Return Volatility Based on Predictive Deviation Stratification.

5. Conclusions

The proposed prediction framework based on daily funding demand indicators and intraday-derived features effectively captures the non-linear dynamics of capital requirements in financial operations. Experimental results show that the integrated model achieves a forecast error (MAPE) as low as 8.2%, and maintains allocation stability even under volatility, with mismatch rates controlled below

10% in most cases. The tiered error-aware allocation mechanism significantly reduces capital misallocation risk and smooths return volatility, particularly when forecast deviations are within the 10% threshold. These findings confirm the importance of model robustness and feature interpretability in real-world allocation tasks. Furthermore, the framework can be extended to heterogeneous asset pools by incorporating asset-specific liquidity metrics and multi-horizon planning. A modular multi-level coordination mechanism is also feasible for scaling allocation logic under complex asset constraints.

References

1. Zheng X., Dwyer, V. M., Barrett, L. A., Derakhshani, M., & Ha, S. (2023). Rapid vital sign extraction for real-time opto-physiological monitoring at varying physical activity intensity levels. *IEEE Journal of Biomedical and Health Informatics*, 27(7), 3107-3118.
2. Meher B K, Singh M, Birau R, et al. Forecasting stock prices of fintech companies of India using random forest with high-frequency data[J]. *Journal of Open Innovation: Technology, Market, and Complexity*, 2024, 10(1): 100180.
3. Afshan S, Leong K Y, Najmi A, et al. Fintech advancements for financial resilience: Analysing exchange rates and digital currencies during oil and financial risk[J]. *Resources Policy*, 2024, 88: 104432.
4. Zheng Z, He J, Yang Y, et al. Does financial leverage volatility induce systemic financial risk? Empirical insight based on the Chinese fintech sector[J]. *Managerial and Decision Economics*, 2023, 44(2): 1142-1161.
5. Özlem Ş, Tan O F. Predicting cash holdings using supervised machine learning algorithms[J]. *Financial Innovation*, 2022, 8(1): 44.
6. Engin E, Fakhouri D İ. Comparison of machine learning algorithms for predicting financial risk in cash flow statements[J]. *Turkish Journal of Forecasting*, 2024, 8(1): 1-12.
7. Antar M, Tayachi T. Partial dependence analysis of financial ratios in predicting company defaults: random forest vs XGBoost models[J]. *Digital Finance*, 2025, (7)997-1012.
8. Kostecka Z, Ślepaczuk R. Improving Realized LGD approximation: A Novel Framework with XGBoost for handling missing cash-flow data[J]. *arXiv preprint arXiv:2406.17308*, 2024.
9. Ben Jabeur S, Stef N, Carmona P. Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering[J]. *Computational Economics*, 2023, 61(2): 715-741.
10. Yang J, Li Y, Harper D, et al. Macro Financial Prediction of Cross Border Real Estate Returns Using XGBoost LSTM Models[J]. *Journal of Artificial Intelligence and Information*, 2025, 2: 113-118.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.