Proceedings of ICFTBA 2025 Symposium: Financial Framework's Role in Economics and Management of Human-Centered Development
DOI: 10.54254/2754-1169/2025.GL30499

Enhancing the Fama-French Five-Factor Model with a Policy Factor: An Empirical Study on China's New Energy Vehicle Stock Returns

Ke Ding[1*†], Bowen Hua[2†]

[1]Macau University of Science and Technology, Macau, China
[2]Beijing New Oriental Foreign Language School at Yangzhou, Yangzhou, China
**Corresponding Author. Email:dingk684@gmail.com*
[†]These authors contributed equally to this work and should be considered as co-first author.

Abstract. This study addresses the Fama-French five-factor model's limitations in explaining stock returns of China's policy-sensitive new energy vehicle (NEV) industry by introducing a policy factor and applying machine learning techniques. Against global decarbonization goals and intensive domestic policy support, the traditional model fails to capture policy-driven return variations. Using 2019–2024 monthly data from the CSMAR database—including CSI NEV Industry Index returns, market five-factor data (covering major Chinese stock boards), and risk-free rates—we quantified policies across five dimensions (fiscal subsidies, taxation, production technology, infrastructure, end-use incentives) to build two indicators: Fiscal and Taxation Score (FTS) and Policy Type Score (PTS). These were synthesized into a 6:4 weighted Impact Intensity Index, with policy count retained as an auxiliary feature. Key findings: linear models show weak policy-return linearity (OLS R^2=0.64), non-linear models (especially Gradient Boosting, R^2=0.896) outperform linear ones, and the six-factor model (five factors + policy) exceeds the five-factor model in explanatory power. This study enriches asset pricing theory and offers insights for NEV investors and policymakers.

Keywords: Machine Learning, Five-Factor Model, New Energy Vehicle Industry

1. Introduction

Driven by global carbon neutrality commitments and domestic industrial policies, the new energy vehicle (NEV) industry has emerged as a strategic pillar of China's green economy and manufacturing upgrading. By 2024, China has maintained a leading global position in NEV production and sales, and the stock market performance of this industry is closely intertwined with policy changes—such as adjustments to fiscal subsidies, upgrades to emission regulations, and infrastructure construction plans. A case in point is the short-term volatility of NEV stocks triggered by the withdrawal of subsidies in 2022, while the 14th Five-Year Plan for NEV Industry Development has boosted long-term investor confidence.

DOI: 10.54254/2754-1169/2025.GL30499

However, it is challenging for traditional asset pricing models to account for these policy-driven return patterns. The Fama-French five-factor model [1], a cornerstone framework for cross-sectional return analysis in mature markets, focuses on market, size, value, profitability, and investment factors but overlooks the industry-specific impacts of policies. In China's NEV market, where policies directly shape production scales, consumer demand, and corporate profitability, this omission leads to incomplete explanations of stock returns, creating challenges for investors in risk pricing and for policymakers in market regulation.

2. Literature review

Regarding the application of the Fama-French model in China, Guo et al. [2] tested the five-factor model on China's A-share market using data from 1995 to 2015. They discovered that the size, value, and profitability factors worked effectively, in contrast to the redundant investment factor, with an improvement in R^2 of less than 5%. Huang [3] further verified the model's performance on Chinese individual stocks using data from 1994 to 2016, noting that the five-factor model outperformed the Capital Asset Pricing Model (CAPM) with an adjusted R^2 of 47.7%, but lacked industry-specific adjustments. Both studies highlighted the need to expand factor dimensions based on specific market contexts in China.

In research on NEV industry policies and stock returns, Yuan et al. [4] reviewed the development of China's NEV industry and emphasized its high dependence on policies—arguing that fiscal subsidies and infrastructure policies directly affect industry growth—but failed to link policy variables to asset pricing models. Subsequently, Su and Wang [5] took Chinese new energy vehicle concept stocks from 2011 to 2020 as the research sample and adopted a panel data model to explore the impact of new energy vehicle industry policy announcements on stock volatility. However, their research lacks exploration of policy factors in the context of stock price prediction and thus fails to provide direct references for the optimization of stock price prediction models.

In the field of machine learning applications in asset pricing, seminal studies [6] have established the superiority of nonlinear models, such as Random Forests and Gradient Boosting, for improving the accuracy of return prediction. However, most of these foundational studies did not target the NEV industry, nor did they combine policy features with traditional factors.

3. Research objectives and contributions

This study seeks to fill the existing research gaps, with specific objectives including constructing a policy quantification system tailored to China's NEV industry; integrating policy factors into asset pricing models and comparing the performance of linear and machine learning-based non-linear approaches; and providing empirical insights into the policy-return relationship for investors and policymakers.

The main contributions of this study are threefold. First, it proposes a novel policy quantification method that captures multi-dimensional policy impacts, filling the gap of insufficient policy measurement in NEV asset pricing research. Second, it provides empirical evidence for the superiority of non-linear models in policy-sensitive industries, confirming that machine learning better captures the complex policy-return relationship. Third, it develops a revised six-factor asset pricing framework that enhances the explanatory power for NEV stock returns, offering a more accurate tool for market participants and regulators.

The research methodology of this study proceeds as follows: first, policies are quantified across five dimensions to construct policy indicators; second, these indicators are integrated with

DOI: 10.54254/2754-1169/2025.GL30499

traditional factors, and both linear and non-linear models are used to analyze NEV stock returns; finally, model performances are compared to draw conclusions on the dynamic characteristics of policy-driven returns.

4. Data source and preliminary processing

Based on Fama-French's five-factor model, we added two policy factors: Policy Score and Number of Policy, thus forming a new six-factor formula. As shown in Table 1 is the six-factor formula and the meaning of each variable:

$$MonthlyStockExcessReturn - -\alpha + \beta 1(RiskPremium) + \beta 2SMBt + \beta 3HMLt + \beta 4RMW$$
$$Scoret + \beta 7NumberofPolicyt + eit$$

Table 1. Variables and definition

Variables	Definition
Monthly Stock Excess Return	The excess return of the CSI New Energy Vehicle Industry,calculated as the index return minus the risk-free rate of return).
Risk Premium	Market risk premium factor: market portfolio return minus the risk-free rate of return
SMBt	Size factor
HMLt	Book-to-market factor
RMWt	Profitability factor
CMAt	Investment factor
Policy Scoret	quantified score of policy impact
Number of Policyt	number of relevant policy documents released
εit	Random disturbance term
α	Intercept term (excess return that cannot be explained by the model factors)
β1-β7	Coefficients of each factor
t	A specified period

4.1. Data source

To study the overall situation of the new energy vehicle industry, we selected the CSI New Energy Vehicle Industry Index (930997) as the research object. This index includes Shanghai and Shenzhen A-share listed companies whose businesses are involved in the new energy vehicle industry as samples, with a wide investment coverage, including new energy vehicle manufacturers, charging piles, lithium battery equipment, motor controllers, battery materials, etc. Based on the six-factor formula, we select data in 2019-2024 from two websites (all the data are collected monthly):

CSMAR: A snapshot of the monthly return rate of the CSI New Energy Vehicle Industry Index, the risk-free rate, and the five-factor data (Risk Premium, SMB, HML, RMW, CMA).

AskCI: A snapshot of the crucial policy released, including the title, main content, and date.

DOI: 10.54254/2754-1169/2025.GL30499

4.2. Data cleaning

For the five-factor model data, screen the markets that cover the CSI New Energy Vehicle Industry Index (for short, the NEV Index): P9701, P9703, P9705, P9709, P9711, P9714

Screening Criteria: The constituent stocks of the NEV Index are derived from Shanghai A-shares and Shenzhen A-shares and include listed new energy vehicle companies on the ChiNext and Science and Technology Innovation Boards. Therefore, P9714 (Shanghai and Shenzhen A-shares, ChiNext, and Science and Technology Innovation Boards) is the core coverage sector of the index. The corresponding sectors are further subdivided as follows: the Shanghai A-share component covered by the index corresponds to P9701, and the Shenzhen A-share component corresponds to P9703; the ChiNext stocks included in the index correspond to P9705, and the Science and Technology Innovation Board stocks correspond to P9711. Meanwhile, the index also falls within the scope of P9709 (Shanghai and Shenzhen A-shares and ChiNext) and P9712 (Shanghai and Shenzhen A-shares and Science and Technology Innovation Board). However, the most comprehensive correspondence is the P9714 market.

Twelve rows of data containing missing values were removed. The cleaned data has a shape of 1182 rows and 11 columns.

4.3. Policy quantification

Policies collected from websites are compiled into a table and evaluated across five dimensions: fiscal support and subsidies, taxation and financial policies, production and technology policies, infrastructure construction policies, and end-use incentive policies. A value of 1 is assigned if a policy involves the dimension, and 0 if it does not.

Two indicators are established according to specific rules and weight ratios: Fiscal and Taxation Score (FTS), Policy Type Score (PTS).

These two scores are calculated in accordance with the following weight ratios:

Fiscal and Taxation Score = fiscal support and subsidies*2 + taxation and financial policies*1

Policy Type Score = production and technology policies + infrastructure construction policies + end-use incentive policies

Obtain the ultimate policy score:

Impact Intensity Index = Fiscal and Taxation Score*0.6+ Policy Type Score*0.4

Meanwhile, we added another factor: number of policies.

4.4. Merge the ultimate dataset

Merge the above data by month. The dataset size: 1182 rows, 13 columns.

4.5. Statistical analysis

4.5.1. Descriptive analysis

We performed a descriptive analysis on all features from these perspectives: count, mean, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile, and maximum.

DOI: 10.54254/2754-1169/2025.GL30499

Table 2. Descriptive analysis

	Risk Premium	SMB	HML	RMW	CMA	Policy Score	Number of Policy
Count	1182	1182	1182	1182	1182	1182	1182
Mean	0.006387	0.003963	-0.000993	-0.001577	0.000933	0.305584	0.398477
Std.	0.060572	0.040770	0.026491	0.025025	0.021485	0.703170	0.691865
Min.	-0.231363	-0.140967	-0.084743	-0.144838	-0.116943	0.000000	0.000000
25%	-0.026756	-0.019200	-0.016630	-0.015596	-0.012907	0.000000	0.000000
50%	0.000781	0.004964	-0.002080	0.000170	0.001299	0.000000	0.000000
75%	0.028774	0.025539	0.014128	0.014195	0.013242	0.400000	1.000000
Max.	0.314596	0.121556	0.102598	0.110985	0.149542	5.000000	4.000000

4.5.2. Distribution analysis

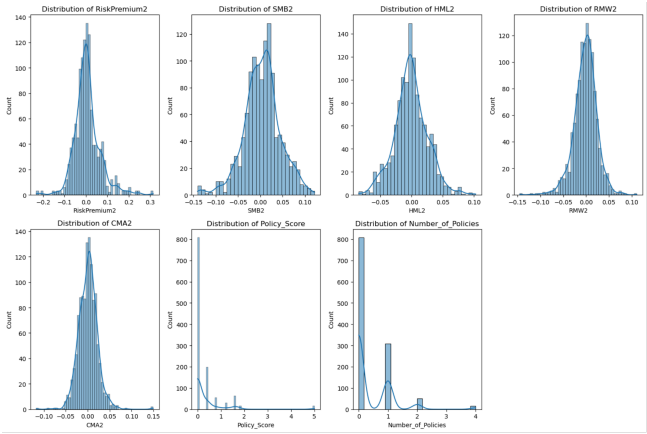

Figure 1. Distribution analysis

4.5.3. Feature correlation

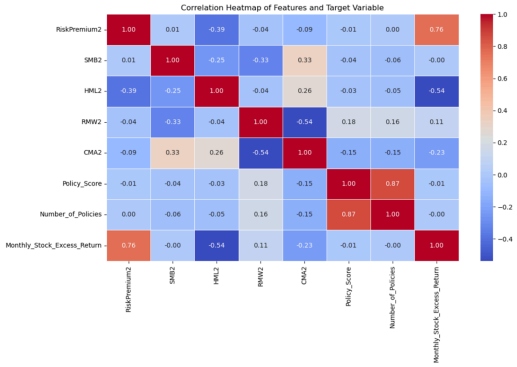

Figure 2. Feature correlation

DOI: 10.54254/2754-1169/2025.GL30499

5. Modeling and evaluation

5.1. Model introduction

Ordinary Least Squares (OLS) Regression: The fundamental principle of this regression model is to estimate the coefficient values corresponding to each independent variable in the linear regression equation. To achieve this, it minimizes the sum of squared errors, where the errors refer to the differences between the actual observed values of the dependent variable and the values predicted by the model. Through this process, a linear model is ultimately constructed that optimally aligns with the distribution trend of the target data.

Ridge Regression: The core principle of this model is to add a penalty term proportional to the sum of the squares of the independent variable coefficients to the Ordinary Least Squares (OLS) method. In this way, while retaining the model's ability to fit the data, it shrinks the absolute values of the coefficients, thereby reducing model complexity and the risk of overfitting, and ultimately obtaining more robust coefficient estimation results.

Lasso Regression: The core principle of this model is to add a penalty term proportional to the sum of the absolute values of the independent variable coefficients to the loss function of Ordinary Least Squares (OLS). This penalty mechanism can not only reduce model complexity and the risk of overfitting by shrinking the absolute values of coefficients (like Ridge Regression), but more importantly, when the penalty intensity is sufficient, it can shrink the coefficients of some independent variables that contribute less to the model to zero. In this way, it automatically realizes variable selection, eliminates redundant features, and ultimately obtains a more concise regression model with stronger generalization ability.

Decision Tree Regressor: This model is a regression model built based on a tree-like structure. Its core idea is to recursively divide the dataset into multiple sub-datasets according to a specific threshold of a certain feature. Each division corresponds to a branch of the tree, and this process continues until the sub-datasets meet the preset stopping conditions, eventually forming a tree-like model. Its advantages lie in that it does not require complex data preprocessing (such as feature standardization), has an intuitive and easy-to-understand model structure, and can naturally capture non-linear relationships between features. However, it also has characteristics such as being prone to overfitting and sensitive to minor changes in the training data.

Random Forest Regressor: As a regression model rooted in the ensemble learning framework, the Random Forest Regressor boosts its performance by building a set of decision trees and integrating the predictive outputs of these individual trees. This working mechanism not only preserves the inherent strengths of a single decision tree—such as its capability to capture nonlinear relationships between variables and its lack of need for complicated data preprocessing—but also significantly lowers the risk of overfitting. Additionally, it enhances the model’s stability, making it less sensitive to small fluctuations in the training dataset.

Gradient Boosting Regressor: The Gradient Boosting Regressor is an iterative regression model developed under the concept of ensemble learning. Its core logic lies in sequentially constructing a series of weak learners (most commonly shallow decision trees) to continuously correct the predictive errors made by the models built in previous iterations. This approach enables the model to effectively identify complex patterns and nonlinear correlations within the data. When compared to a standalone decision tree, the Gradient Boosting Regressor demonstrates stronger fitting capabilities and more robust generalization performance.

Model Evaluation Methodology: We selected R-squared and MSE to evaluate the model's performance by assessing its goodness of fit and error, respectively. Meanwhile, we used feature

DOI: 10.54254/2754-1169/2025.GL30499

importance to evaluate the degree of influence of each factor on the dependent variable.

5.2. Linear regression

We divided the training set and test set in an 8:2 ratio.

Here's the output:

Table 3. Linear regression

	OLS Regression	Ridge Regression	Lasso Regression
MSE	0.003855	0.002861	0.004171
R^2	0.641300	0.640737	0.611850

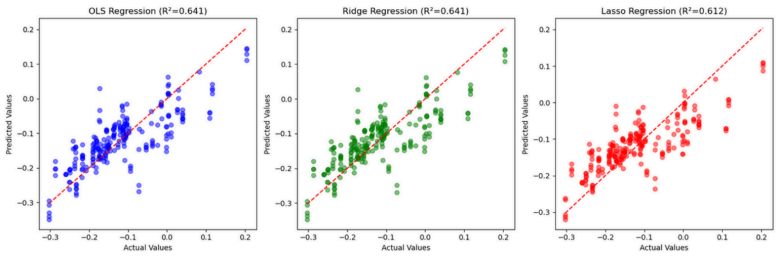

Figure 3. Linear regression

Feature Importance (sorted by OLS coefficients):

Table 4. LR-feature importance

Feature	OLS Regression	Ridge Regression	Lasso Regression
Risk Premium	0.065836	0.065147	0.058337
RMW	0.009019	0.008806	0.00482
Policy Score	0.000621	0.000423	-0.000000
SMB	-0.002907	-0.002884	-0.000000
Number of Policy	-0.004643	-0.004410	-0.000000
CMA	-0.005289	-0.005450	-0.003130
HML	-0.028355	-0.028270	-0.021075

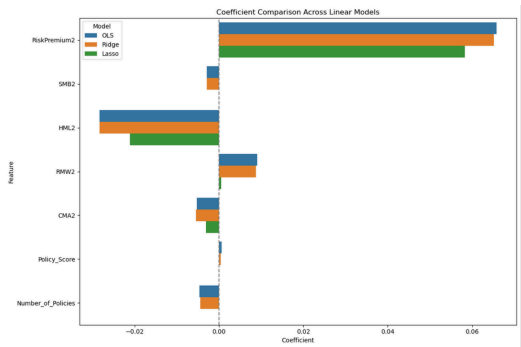

Figure 4. LR-feature importance

DOI: 10.54254/2754-1169/2025.GL30499

5.3. Non-linear regression

We still divided the training set and test set in an 8:2 ratio.

Here's the output:

Table 5. Non-linear regression

	Decision Tree	Random Forest	Gradient Boosting
MSE	0.002575	0.001258	0.001117
R^2	0.760409	0.882947	0.896058

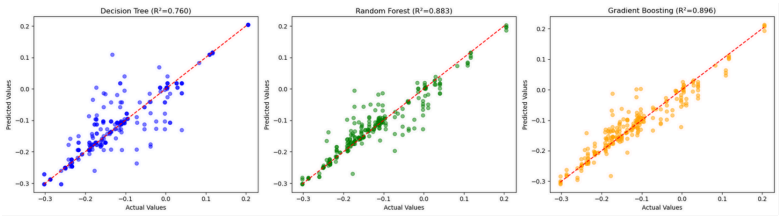

Figure 5. Non-linear regression

Feature Importance (sorted by Random Forest coefficients):

Table 6. NLR- feature importance

Feature	Decision Tree	Random Forest	Gradient Boosting
Risk Premium	0.684103	0.660565	0.668633
HML	0.118210	0.115645	0.108821
RMW	0.071468	0.085879	0.074368
SMB	0.045452	0.054661	0.061289
CMA	0.035613	0.047784	0.045153
Policy Score	0.32927	0.027097	0.036261
Number of Policies	0.012227	0.008369	0.005475

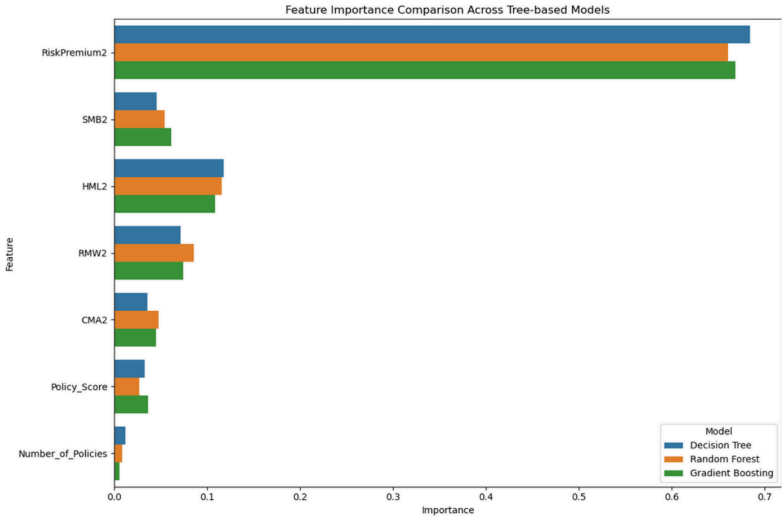

Figure 6. NLR- feature importance

DOI: 10.54254/2754-1169/2025.GL30499

In the above nonlinear regression tests, the feature importance values of the policy factors were relatively low across all three models. Previously, our quantification of policies was rather crude, so we subsequently chose the two best-performing models (Random Forest Regressor and Gradient Boosting Regressor) to conduct parameter optimization and incorporated the lag effect of policy impacts into consideration.

5.3.1. Optimization 1

Assume a policy has a lag effect, and its impact will persist for two months. Propagate the policy scores for the current month backward for two months, with scores cumulatively added while the number of policies remains constant each month.

Here's the output:

Table 7. Optimization 1

	Random Tree	Gradient Boosting
MSE	0.001116	0.000923
R^2	0.896162	0.914091

Feature Importance (sorted by Random Forest coefficients):

Table 8. OP1- feature importance

	Random Tree	Gradient Boosting
Risk Premium	0.650327	0.651039
HML	0.108208	0.101685
RMW	0.072367	0.069616
Policy Score	0.066351	0.088118
SMB	0.053468	0.051665
CMA	0.041072	0.032116
Number of Policy	0.008207	0.005760

5.3.2. Optimization 2

To make the result more realistic, we assume a policy has a decaying lag effect. Set the decay coefficients: 66% in month t+1 and 33% in month t+2, with effects cumulatively added to the original monthly scores while the policy count remains constant each month.

Here's the output:

Table 9. Optimization 2

	Random Tree	Gradient Boosting
MSE	0.001110	0.001134
R^2	0.896714	0.894517

DOI: 10.54254/2754-1169/2025.GL30499

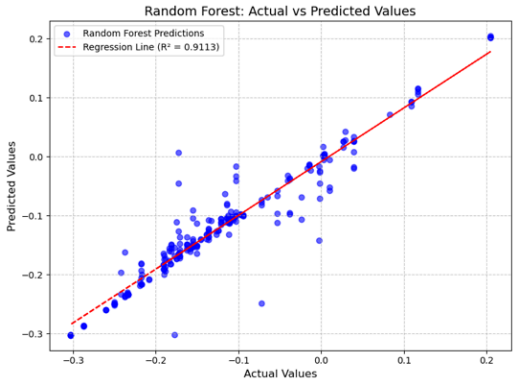

Figure 7. Optimization 2-random forest

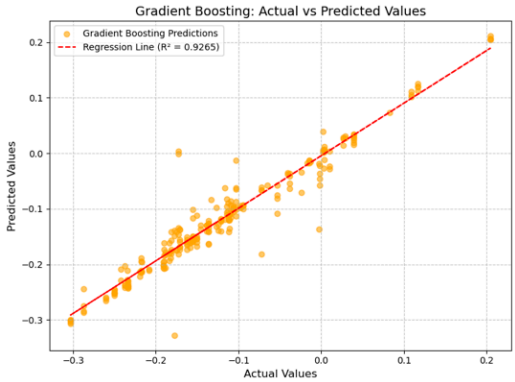

Figure 8. Optimization 2-gradientboosting

Feature Importance (sorted by Random Forest coefficients):

Table 10. OP2- feature importance

	Random Tree	Gradient Boosting
Risk Premium	0.639513	0.640720
Policy Score	0.116362	0.139165
HML	0.100245	0.094545
RMW	0.063478	0.050367
SMB	0.045867	0.04594
CMA	0.026224	0.022760
Number of Policies	0.008311	0.002849

5.4. Comparison analysis

To investigate whether our six-factor model exhibits superior predictive performance in the new energy vehicle sector, we conducted a comparative analysis. Specifically, we applied the identical dataset and modeling approach to assess the performance of the five-factor model for comparison. Here's the result:

DOI: 10.54254/2754-1169/2025.GL30499

5.4.1. Linear regression

Table 11. Comparison (linear)

	Five-Factor Model			Six-Factor Model		
	OLS	Ridge	Lasso	OLS	Ridge	Lasso
R^2	0.639811	0.639263	0.61185	0.6413	0.640737	0.61185
MSE	0.003871	0.003876	0.004171	0.003855	0.003861	0.004171

5.4.2. Non-linear regression

In the parameter optimization of the six-factor model, we only selected Random Forest and Gradient Boosting; therefore, we also only used these two models for testing the five-factor model:

Table 12. Comparison (non-linear)

	Five-Factor Model		Six-Factor Model	
	RF	GB	RF	GB
R^2	0.873234	0.88146	0.910883	0.925705
MSE	0.001362	0.88146	0.000958	0.000878

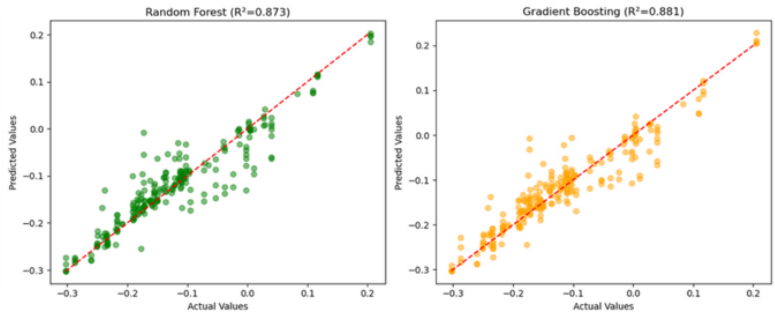

Figure 9. Non-linear (five factor model)

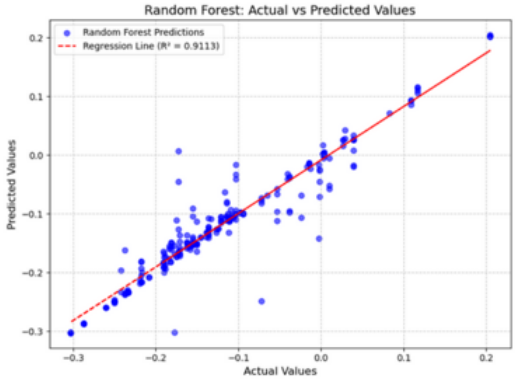

Figure 10. Random forest (six-factor model)

DOI: 10.54254/2754-1169/2025.GL30499

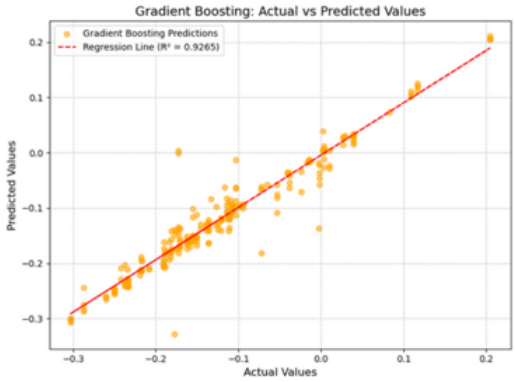

Figure 11. Gradient boosting (six-factor model)

6. Result and discussion

6.1. Hypothesis 1

Policy factors have a linear relationship with the stock returns of the new energy vehicle industry.

In the test of the linear regression model, the R-squared values of all three models (OLS: 0.641300, Ridge: 0.640737, Lasso: 0.611850) are less than 0.7, which indicates that there is no significant linear relationship between the policy factors and the stock returns in the NEV market.

6.2. Hypothesis 2

Policy factors have a significant impact on the stock returns of the new energy vehicle industry.

In the first test of nonlinear regression models, the fitting performance of the Decision Tree was significantly worse than that of the other two models (Decision Tree: 0.760409, Random Forest: 0.882947, Gradient Boosting: 0.896058). The possible reasons are as follows: 1. The Decision Tree has a high risk of overfitting and tends to overlearn the noise in the data (such as short-term fluctuations in stock returns); 2. The Decision Tree is relatively sensitive to minor changes in the training dataset, resulting in poor stability 3. The Decision Tree has difficulty capturing complex multi-feature interactions.

Regarding feature importance, the values of Policy Score and Number of Policy were the lowest across all three models. In particular, the value for Number of Policy was as low as 0.005475. It can be seen from this that the number of police has almost no impact on the returns of NEV stocks.

Since our previous quantification of the Policy Score was relatively crude, there was a significant error in the feature importance of the Policy Score. Therefore, we chose Random Forest and Gradient Boosting and conducted two rounds of parameter optimization to facilitate analysis.

6.2.1. Optimization 1

Table 13. Comparison result in optimization 1

R^2	Random Forest	Gradient Boosting
Before	0.882947	0.896058
After	0.896162	0.914091

DOI: 10.54254/2754-1169/2025.GL30499

After parameter optimization, the fitting degree of both models increased.

Table 14. Comparison feature importance in optimization 1

Feature Importance of Policy Score	Random Forest	Gradient Boosting
Before	0.027097	0.036261
After	0.066351	0.088118

After parameter optimization, the feature importance of the Policy Score increased significantly. To make the model more in line with real-world scenarios, we conducted a second round of parameter optimization.

6.2.2. Optimization 2

Table 15. Comparison feature importance in optimization 2

Feature Importance of Policy Score	Random Forest	Gradient Boosting
Before	0.066351	0.088118
After	0.116362	0.139165

After considering the policy lag effect, the feature importance value of the Policy Score became even higher. In both models, the relative influence of the Policy Score ranked second among all factors. This indicates that policies have a significant impact on NEV stock returns.

6.3. Hypothesis 3

The Six-Factor Model has a better explanatory power than the Five-Factor Model.

Linear: The six-factor model performs slightly better than the five-factor model in terms of goodness of fit and error. In Hypothesis 1, it has been concluded that there is no linear relationship between policy factors and NEV stock returns, so this result is reasonable.

Non-Linear: In the tests of the Random Forest and Gradient Boosting models, the six-factor model performed significantly better than the five-factor model in terms of goodness of fit (R-squared) and error (MSE). This indicates that the six-factor model has better explanatory power than the five-factor model.

7. Conclusion

7.1. Summary of methodology and key findings

This study addresses the limitations of the Fama-French five-factor model in explaining stock returns in China’s policy-sensitive NEV industry. It incorporates a policy factor into the traditional model and applies a multi-method machine learning framework, following a systematic research design.

First, a comprehensive policy quantification system was constructed. Using monthly data from 2019 to 2024 from the CSMAR database—including returns of the CSI NEV Industry Index, market five-factor data, and risk-free rates—and policy texts from official sources, policies were evaluated across five dimensions: fiscal subsidies, taxation, production technology, infrastructure, and end-use incentives. This process yielded two core indicators: the Fiscal and Taxation Score (FTS) and the

DOI: 10.54254/2754-1169/2025.GL30499

Policy Type Score (PTS). These two indicators were synthesized into a weighted Impact Intensity Index with a 6:4 ratio, while policy count was retained as an auxiliary feature.

Second, iterative modeling and evaluation were conducted. Three linear models—Ordinary Least Squares (OLS), Ridge Regression, and Lasso Regression—and three non-linear models—Decision Tree, Random Forest, and Gradient Boosting—were tested. Data was divided into training and test sets with an 8:2 ratio, and model parameters were optimized through cross-validation. Additionally, the top-performing non-linear models were further optimized by incorporating policy lag effects (simple cumulative lag and decaying lag).

Three key findings emerged from the study. First, Policy factors and NEV stock returns do not exhibit a significant linear relationship. Linear models exhibited weak explanatory power, with an R^2 of 0.64 for OLS and 0.61 for Lasso, and policy factors had minimal linear influence (the coefficient of Policy Score in OLS was only 0.000621), confirming that the impact of policies on returns is non-linear. Second, the non-linear impact of policy factors becomes significant after optimization. In initial non-linear models (Random Forest and Gradient Boosting), the feature importance of policy factors was low (Policy Score importance < 4%). However, after integrating lag effects, the importance of policy factors increased significantly. Under the decaying lag setting (66% weight for t+1 and 33% weight for t+2), the feature importance of Policy Score ranked second, at 0.116 in Random Forest and 0.139 in Gradient Boosting—indicating that when the time-dependent nature of policies is considered, their driving effect on NEV returns becomes evident. Third, the six-factor model outperforms the five-factor model. In linear models, the improvement in fit by the six-factor model was marginal (the R^2 of OLS increased from 0.640 for the five-factor model to 0.641 for the six-factor model). In contrast, the gap widened in non-linear models: the optimized Gradient Boosting model for the six-factor framework achieved an R^2 of 0.914 and a Mean Squared Error (MSE) of 0.000923, significantly outperforming the five-factor model (R^2 = 0.881, MSE = 0.88146). This proves that adding a policy factor effectively enhances the explanatory power for returns in policy-sensitive industries.

7.2. Theoretical and practical contributions

7.2.1. Theoretical contributions

First, this study extends the Fama-French five-factor model to policy-sensitive industries. By introducing a policy factor tailored to the context of China's NEV industry, it addresses the model's neglect of industry-specific policy impacts in emerging markets like China. It also provides a replicable framework for integrating policy variables into asset pricing models, enriching the theory of cross-sectional return analysis. Second, the study validates the value of non-linear models in policy-return analysis. By comparing the performance of Gradient Boosting, Random Forest, and linear models, it demonstrates that non-linear models better capture policy effects, highlighting the need to move beyond linear assumptions when studying policy-driven financial markets and offering new methodological insights for asset pricing research.

7.2.2. Practical contributions

For investors, the findings provide guidance for policy-aware investment strategies. For instance, investors can prioritize NEV stocks during periods of high-impact policies (e.g., upgrades to fiscal subsidies) and account for policy lag effects of approximately two months to avoid timing errors. Additionally, the six-factor model offers a more accurate tool for risk pricing and return forecasting.

DOI: 10.54254/2754-1169/2025.GL30499

For policymakers, the study emphasizes the need for stable and predictable policy design. Since policy lag effects shape market expectations, abrupt policy changes (such as sudden subsidy withdrawals) may increase market volatility. Therefore, policymakers should adopt phased adjustments and clarify long-term policy directions to maintain market stability.

7.3. Limitations and future research directions

This study has three primary limitations. First, the granularity of policy quantification is insufficient. The current system uses a binary scoring method (1 = covering the dimension, 0 = not covering the dimension) and fixed weights, which fail to capture differences in policy intensity (e.g., variations in subsidy amounts) and dynamic weight adjustments over policy cycles. Second, the data scope is limited. The sample period (2019–2024) excludes the early stages of the NEV industry (e.g., the peak subsidy period of 2015–2018), and the data only covers constituents of the CSI NEV Index, omitting small and medium-cap NEV stocks on niche boards. Third, the generalizability of the model remains untested. The six-factor model has only been tested in the NEV industry, and its performance in other policy-sensitive sectors (e.g., photovoltaics, semiconductors) is yet to be verified.

Future research can address these gaps in three ways. First, enhance policy quantification by introducing continuous variables (e.g., subsidy amounts, tax reduction ratios) and utilizing machine learning methods such as Latent Dirichlet Allocation (LDA) topic modeling to dynamically assign weights based on the sentiment and intensity of policy texts. Second, expand the data scope and methodological approaches by extending the sample period to 2015–2024, including small and medium-cap stocks, and testing advanced algorithms such as XGBoost-LSTM hybrid models to capture both the non-linear and time-series dynamic characteristics of returns. Third, test cross-industry generalizability by applying the six-factor model to other policy-sensitive sectors to verify its universality, or develop industry-specific policy factors to further improve model fit.

References

[1] Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1-22. https: //doi.org/https: //doi.org/10.1016/j.jfineco.2014.10.010

[2] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. The Review of Financial Studies, 33(5), 2223-2273. https: //doi.org/10.1093/rfs/hhaa009

[3] Guo, B., Zhang, W., Zhang, Y., & Zhang, H. (2017). The five-factor asset pricing model tests for the Chinese stock market. Pacific-Basin Finance Journal, 43, 84-106.https: //doi.org/https: //doi.org/10.1016/j.pacfin.2017.02.001

[4] Huang, T.-L. (2019). Is the Fama and French five-factor model robust in the Chinese stock market? Asia Pacific Management Review, 24(3), 278-289. https: //doi.org/https: //doi.org/10.1016/j.apmrv.2018.10.002

[5] Su, M., & Wang, C. (2022). Policy Announcement, Investor Attention, and Stock Volatility: Evidence From the New Energy Vehicle Industry. Front Psychol, 13, 838588. https: //doi.org/10.3389/fpsyg.2022.838588

[6] Yuan, X., Liu, X., & Zuo, J. (2015). The development of new energy vehicles for a sustainable future: A review. Renewable and Sustainable Energy Reviews, 42, 298-305. https: //doi.org/https: //doi.org/10.1016/j.rser.2014.10.016