Int. J. Anal. Appl. (2026), 24:180

International Journal of Analysis and Applications

Modeling and Forecasting Excess Market Returns Using a Two-Factor CIR Model

Nguyen Thi Kieu An*

Department of Mathematics, FPT University, Ho Chi Minh, Vietnam

**Corresponding author:* anntk2@fe.edu.vn

ABSTRACT. Excess market return (Mkt) plays a central role in modern asset pricing models. However, forecasting market excess returns remains challenging due to their nonlinear dynamics, strong volatility, and structural changes, particularly in emerging markets. This study proposes a two factor Cox–Ingersoll–Ross (CIR2) stochastic model to capture the dynamic behavior of excess market returns in Vietnam. The dataset consists of the VN-Index and the one-year Vietnamese government bond yield covering the period from January 2010 to March 2025. To satisfy the positivity condition required by the CIR diffusion process, excess returns are transformed by adding a constant shift. The parameters of the proposed model are estimated using a maximum likelihood approach based on a discretized representation of the stochastic differential equations. To evaluate predictive performance, an out-of-sample forecasting experiment is conducted using a rolling-window framework with a window length of 36 months. The forecasting ability of the CIR2 model is compared with several benchmark models commonly used in financial time-series forecasting, including the Random Walk, ARIMA, and GARCH(1,1) models. The empirical results indicate that the proposed two-factor CIR model consistently achieves lower forecasting errors than the benchmark models. The improvement in predictive performance is further supported by the Diebold–Mariano test. These findings suggest that stochastic diffusion models with mean-reverting dynamics provide a flexible framework for modeling financial return dynamics in emerging markets and offer useful insights for asset pricing, portfolio management, and financial risk monitoring.

1. Introduction

Excess market return, defined as the difference between the return on the market portfolio and the risk-free rate, plays a fundamental role in modern financial economics. In the Capital Asset Pricing Model (CAPM), excess market return represents the compensation investors receive for bearing systematic risk and serves as the key determinant of expected asset returns [1, 2, 21].

Received Mar. 17, 2026

2020 Mathematics Subject Classification. 91G70, 60H10, 62M10.

Key words and phrases. excess market return; CIR model; SDE; financial forecasting; emerging markets.

https://doi.org/10.28924/2291-8639-24-2026-180

© 2026 the author(s)

ISSN: 2291-8639

Subsequent developments in asset pricing theory have extended this framework by incorporating additional systematic factors while maintaining the central role of the market risk premium. In particular, the multifactor asset pricing models proposed by Fama and French emphasize that market excess return remains the primary factor explaining cross-sectional differences in asset returns [9, 10]. Because of its theoretical and practical importance, accurately modeling and forecasting excess market returns is essential for portfolio allocation, asset pricing, and financial risk management.

Despite decades of research, predicting excess market returns remains a challenging task. Financial return series typically exhibit complex dynamics including volatility clustering, nonlinear adjustment mechanisms, and time-varying risk premia. Traditional econometric approaches have attempted to capture these dynamics using linear predictive models and volatility-based frameworks. For example, the autoregressive conditional heteroskedasticity (ARCH) model proposed by Engle [3] and its generalized version introduced by Bollerslev [4] provide a flexible way to model time-varying volatility in financial returns. These models have become standard tools in empirical finance and are widely used to analyze and forecast financial time series. Nevertheless, empirical evidence suggests that forecasting stock market returns remains difficult, and many predictive models fail to consistently outperform simple benchmarks such as the historical mean or the random walk model [7].

Recent advances in financial econometrics have introduced new approaches to improve the predictability of stock market returns. In particular, high-dimensional and machine learning methods have been increasingly applied to asset pricing and financial forecasting problems. For example, Kelly and Xiu (2020) [12] demonstrates that machine learning techniques can uncover complex nonlinear relationships in financial data and improve asset pricing models. Similarly, macroeconomic attention indicators have been shown to contain useful predictive information for stock market returns [13]. In addition, network-based methods and predictive algorithms have been proposed to exploit structural relationships among financial assets and improve portfolio optimization performance [14]. Although these methods often achieve strong predictive performance, they may lack clear economic interpretation and rely heavily on large data sets and complex model structures.

An alternative and theoretically grounded approach is to model financial variables using stochastic differential equations (SDEs). Continuous-time stochastic models have long been used in financial economics to describe the dynamic evolution of asset prices, interest rates, and volatility. Among these models, the Cox–Ingersoll–Ross (CIR) process has become one of the most influential frameworks in financial modeling. Originally developed to model the term structure of interest rates, the CIR model incorporates mean-reverting dynamics and guarantees the positivity of the modeled variable, which makes it particularly suitable for financial quantities that exhibit bounded fluctuations [5]. Theoretical and empirical studies have demonstrated that stochastic diffusion models can effectively capture nonlinear financial dynamics and persistent mean-reverting behavior in economic variables.

Recent research continues to explore extensions of the CIR framework and related stochastic diffusion models in financial applications. For example, interest rate forecasting models based on CIR dynamics have been shown to provide useful insights into financial market expectations [15]. In addition, advances in numerical methods for stochastic differential equations have improved the accuracy and stability of diffusion-based models. In particular, positivity-preserving Milstein-type schemes provide efficient numerical approximations for stochastic processes with square-root diffusion terms [16]. These methodological developments have increased the feasibility of applying stochastic diffusion models to empirical financial forecasting problems.

Despite these advances, the application of stochastic diffusion models to forecasting equity market returns remains relatively limited. Most empirical studies on return predictability rely primarily on discrete-time econometric models or machine learning methods. Continuous-time stochastic models have been less frequently used for direct forecasting of equity market excess returns, especially in emerging markets. This gap in the literature is particularly relevant because financial markets in emerging economies often exhibit stronger volatility, structural breaks, and nonlinear adjustment dynamics than those in developed markets.

Motivated by this gap, this study investigates whether a stochastic mean-reverting diffusion framework can improve the forecasting performance of market excess returns. Specifically, the paper addresses the following research question: *Can a two-factor Cox–Ingersoll–Ross stochastic model provide more accurate forecasts of market excess returns than conventional time-*

series models? To answer this question, we develop a two-factor extension of the CIR process that decomposes excess market return dynamics into two stochastic components capturing long-term and short-term sources of variation.

The empirical analysis focuses on the Vietnamese stock market. Vietnam provides an interesting setting for studying return dynamics because it represents a rapidly developing emerging market with increasing integration into global financial markets. The Vietnamese equity market has experienced significant growth in market capitalization and trading activity over the past decade, while still exhibiting substantial volatility and sensitivity to macroeconomic and policy shocks. Recent studies have emphasized the importance of investor behavior, exchange rate risk, and macroeconomic conditions in shaping financial market dynamics in Vietnam and other emerging economies [17–19]. Moreover, recent research has explored the application of machine learning techniques and exchange rate risk factors in asset pricing models for emerging markets [20]. These findings suggest that financial return dynamics in Vietnam may involve multiple interacting sources of uncertainty and nonlinear adjustment mechanisms.

Using monthly data on the VN-Index and the one-year Vietnamese government bond yield from January 2010 to March 2025, this study constructs a time series of excess market returns. To ensure compatibility with the CIR diffusion framework, the excess return series is shifted by a constant transformation so that the modeled variable remains strictly positive. The parameters of the proposed two-factor CIR model are estimated using a maximum likelihood approach based on a discretized representation of the stochastic differential equation. Forecasting performance is evaluated using an out-of-sample rolling-window framework, which allows the model parameters to adapt to evolving market conditions over time.

The forecasting ability of the proposed model is compared with several benchmark models widely used in financial time-series analysis, including the Random Walk model, ARIMA specifications, and the GARCH(1,1) model. Forecast accuracy is evaluated using standard metrics such as the mean absolute error (MAE) and the root mean squared error (RMSE), while the Diebold–Mariano test is used to examine whether differences in predictive accuracy between competing models are statistically significant. The empirical results show that the proposed two-factor CIR model consistently outperforms the benchmark models in out-of-sample forecasting. These findings suggest that stochastic diffusion models with mean-reverting dynamics provide a useful framework for capturing the complex behavior of market excess returns in emerging markets.

This study contributes to the literature in several ways. First, it extends the application of stochastic differential equation models to the forecasting of equity market excess returns. Second, it introduces a two-factor mean-reverting structure that allows excess market return dynamics to be decomposed into long-term and short-term stochastic components. Third, the paper provides new empirical evidence from the Vietnamese stock market, contributing to the growing literature on financial modeling and return predictability in emerging economies.

2. Literature Review

The dynamics and predictability of excess market returns have been extensively studied in the financial economics literature. A large body of research has examined the determinants of expected returns, the modeling of financial volatility, and the development of forecasting techniques for financial time series. The existing literature relevant to this study can be broadly classified into three main strands: asset pricing and excess return predictability, econometric models for financial return forecasting, and stochastic differential equation models in financial economics.

2.1. Asset Pricing and Excess Return Predictability

Excess market return plays a central role in modern asset pricing theory. In the Capital Asset Pricing Model (CAPM), the expected return on an asset is determined by its exposure to the market portfolio, with the market excess return representing the price of systematic risk [1, 2]. Subsequent developments in asset pricing have extended the CAPM by incorporating additional risk factors while preserving the central importance of the market factor. For example, the multifactor framework proposed by [9] demonstrates that market excess return, together with size and value factors, explains a significant portion of cross-sectional variation in stock returns. More recently, Fama and French (2015)[10] extend this framework by introducing additional profitability and investment factors.

While asset pricing theory provides a conceptual framework for understanding expected returns, predicting excess market returns remains empirically challenging. Early studies in financial econometrics emphasize that stock market returns are difficult to forecast due to their high volatility and sensitivity to macroeconomic conditions. For example, Campbell and

Thompson (2008)[7] show that many predictive models fail to outperform simple benchmark models such as the historical mean when evaluated out of sample. Nevertheless, subsequent research has continued to explore new sources of return predictability. Recent studies have highlighted the role of macroeconomic information, investor attention, and high-dimensional data in improving return forecasts. For instance, Ma and Huang (2022)[13] demonstrate that macroeconomic attention indicators contain useful information for predicting stock market returns.

Another recent line of research has explored the application of machine learning techniques to asset pricing and financial forecasting. Machine learning models can capture complex nonlinear relationships in financial data and handle large sets of predictive variables. Gu and Xiu (2020)[12] show that machine learning methods can significantly improve empirical asset pricing models by identifying hidden patterns in large financial data sets. Although these approaches often improve predictive performance, they may lack the clear economic interpretation typically associated with structural financial models.

2.2. Econometric Models for Financial Return Forecasting

A second important strand of the literature focuses on econometric models for forecasting financial time series. One of the most influential developments in this area is the introduction of the autoregressive conditional heteroskedasticity (ARCH) model by Engle (1982)[3], which allows the variance of financial returns to vary over time. The generalized ARCH (GARCH) model proposed by Bollerslev (1986)[4] further extends this framework and has become one of the most widely used tools for modeling financial volatility.

The ARCH and GARCH models have been widely applied to analyze financial market dynamics and improve forecasting accuracy. These models are particularly effective in capturing volatility clustering, a stylized fact commonly observed in financial returns. In empirical forecasting studies, GARCH-type models are frequently used as benchmark models when evaluating the performance of alternative forecasting approaches. In addition to volatility models, autoregressive integrated moving average (ARIMA) models remain a standard tool for modeling time-series dependence in financial data.

More recently, researchers have proposed various alternative forecasting methods that incorporate structural information and network relationships among financial assets. For example, Freitas and Bertini Junior (2023)[14] introduce a network-based portfolio optimization approach that uses random walks through stock networks to improve predictive performance in financial markets. These developments highlight the growing interest in combining statistical forecasting techniques with structural information embedded in financial markets.

In the context of emerging markets, financial forecasting remains particularly challenging because financial systems are often subject to structural changes, policy adjustments, and higher levels of uncertainty. Empirical studies have shown that investor behavior and macroeconomic factors can play an important role in shaping financial market dynamics in emerging economies. For example, Phan et. al. (2023)[17] find that investor behavior significantly influences stock market dynamics in emerging markets such as Vietnam. Similarly, Khoa and Huynh (2022)[18] demonstrate that exchange rate dynamics can affect financial market behavior and predictive models in emerging economies. Recent studies also explore the application of machine learning techniques to asset pricing and forecasting in emerging markets [19, 20].

2.3. Stochastic Differential Equation Models in Finance

A third strand of the literature focuses on stochastic differential equation (SDE) models for financial variables. Continuous-time stochastic models provide a natural framework for describing the dynamic evolution of asset prices and interest rates. Among these models, the Cox–Ingersoll–Ross process has become one of the most influential models in financial economics. The CIR model introduces a mean-reverting square-root diffusion process that guarantees the positivity of the modeled variable and has been widely used to model interest rate dynamics [5].

Because of its theoretical properties, the CIR model and its extensions have been widely applied in financial modeling, particularly in interest rate forecasting and derivative pricing. For example, Orlando and Bufalo (2021)[15] examine forecasting performance of interest rate models based on the CIR framework and show that diffusion-based models can capture important features of financial market dynamics. In addition, advances in numerical methods for stochastic differential equations have improved the practical implementation of diffusion models. In particular, Hu et. al. (2024)[16] develop a positivity-preserving Milstein-type numerical scheme that enhances the stability and accuracy of stochastic diffusion simulations.

Despite these advances, the application of stochastic diffusion models to forecasting equity market returns remains relatively limited. Most empirical studies continue to rely on discrete-time econometric models or machine learning approaches. However, financial return series often exhibit nonlinear mean-reverting dynamics that may be more naturally described using stochastic diffusion models. This gap in the literature motivates the present study, which applies a two-factor Cox–Ingersoll–Ross framework to model and forecast excess market returns in the Vietnamese stock market.

By integrating insights from asset pricing theory, econometric forecasting models, and stochastic diffusion processes, this study contributes to the growing literature on financial return predictability and provides new evidence on the usefulness of continuous-time stochastic models for forecasting market excess returns in emerging markets.

2.4. Hypothesis Development

The literature reviewed above highlights two important observations. First, although traditional econometric models such as ARIMA and GARCH are widely used for financial forecasting, their predictive performance often remains limited when applied to stock market returns. Second, recent studies suggest that nonlinear and stochastic dynamics play an important role in financial markets, particularly in emerging economies where market conditions are often more volatile and subject to structural changes.

Stochastic diffusion models provide a theoretically grounded framework for capturing such nonlinear dynamics. In particular, the Cox–Ingersoll–Ross process incorporates a mean-reverting structure that allows financial variables to fluctuate around a long-run equilibrium level while maintaining strictly positive values. This feature may be particularly useful for modeling transformed excess market returns, which can exhibit persistent fluctuations and structural adjustment over time.

Moreover, the two-factor CIR framework allows market dynamics to be decomposed into multiple stochastic components that capture different sources of variation. One factor can represent long-term structural movements in the market risk premium, while the second factor captures short-term fluctuations driven by market shocks, investor sentiment, and macroeconomic news. Such a structure may provide greater flexibility in modeling financial return dynamics than traditional linear models.

Based on these arguments, this study proposes the following hypotheses:

H1: The two-factor CIR model provides more accurate forecasts of excess market returns than conventional time-series models such as Random Walk, ARIMA, and GARCH.

H2: Mean-reverting stochastic diffusion models improve forecasting performance in emerging stock markets.

These hypotheses are tested empirically using out-of-sample forecasting experiments and the Diebold–Mariano test to compare predictive accuracy across competing models.

3. Methodology

This section describes the methodological framework used to model and forecast excess market returns in the Vietnamese stock market. We first describe the data and variable construction, followed by the specification of the two-factor Cox–Ingersoll–Ross model. Finally, we present the estimation procedure and the out-of-sample forecasting framework used to evaluate predictive performance.

3.1. *Data and Variable Construction*

The empirical analysis uses monthly data covering the period from January 2010 to March 2025. The data are collected from Investing.com, which provides historical financial market data for both stock indices and government bond yields. The dataset includes two main variables: the VN-Index, representing the Vietnamese stock market index, and the yield of the one-year Vietnamese government bond, which is used as a proxy for the risk-free rate. The VN-Index reflects the overall performance of the Vietnamese equity market and is widely used as a benchmark index for market returns. The risk-free rate is proxied by the one-year government bond yield because government bonds are generally considered to be free of default risk and therefore provide an appropriate benchmark for risk-free investment. Market return is computed using the logarithmic return of the VN-Index. Specifically, the market return at time t is defined as:

$$R_{m,t} = ln\left(\frac{VNIndex_t}{VNIndex_{t-1}}\right) \times 100\% (3.1)$$

where $VNIndex_t$ denotes the level of the VN-Index at time t.

Excess market return is defined as the difference between the market return and the risk-free rate:

$$Mkt_t = R_{m,t} - rf_t (3.2)$$

where rf_t denotes the one-year government bond yield at time t.

Because the Cox–Ingersoll–Ross diffusion process requires the modeled variable to remain strictly positive, the excess market return series is shifted by a constant transformation. The transformed series is defined as:

$$Mkt_t^* = Mkt_t + c (3.3)$$

where c is a positive constant chosen to ensure that $Mkt_t^* > 0$ for all observations. This transformation allows the series to satisfy the positivity condition required for CIR-type stochastic processes.

The final dataset, therefore, consists of the following variables: the VN-Index level, the market return Rm,t, the risk-free rate rf,t, and the transformed excess market return Mkt_t^* used in the stochastic diffusion model.

3.2. *Two-Factor Cox–Ingersoll–Ross Model*

To capture the dynamic behavior of excess market returns, this study adopts a two-factor Cox–Ingersoll–Ross stochastic diffusion model. The CIR model is widely used in financial economics because it incorporates mean-reverting dynamics while ensuring that the modeled variable remains strictly positive [5].

In the standard CIR model, the evolution of a stochastic variable r(t) follows the stochastic differential equation:

$$dr(t) = \kappa\big(\theta - r(t)\big)\, dt + \sigma\sqrt{r(t)}\, dW(t) (3.4)$$

where $\kappa > 0$ represents the speed of mean reversion, $\theta > 0$ denotes the long-run equilibrium level, $\sigma > 0$ is the volatility parameter, and W(t) is a standard Brownian motion.

While the single-factor CIR model captures basic mean-reverting dynamics, financial return series often exhibit multiple layers of variation driven by different economic forces. To capture both long-term structural movements and short-term fluctuations in market excess returns, we adopt a two-factor CIR specification.

Let $r_1(t)$ and $r_2(t)$ denote two independent CIR processes defined as:

$$dr_1(t) = \kappa_1\big(\theta_1 - r_1(t)\big)\,dt + \sigma_1\sqrt{r_1(t)}\,dW_1(t)(3.5)$$

$$dr_2(t) = \kappa_2\big(\theta_2 - r_2(t)\big)\,dt + \sigma_2\sqrt{r_2(t)}\,dW_2(t)(3.6)$$

where $W_1(t)$ and $W_2(t)$ are independent Brownian motions.

The transformed excess market return is modeled as the sum of the two stochastic components:

$$Mkt_t^* = r_1(t) + r_2(t)(3.7)$$

In this framework, the first factor captures the long-term structural component of the market risk premium, while the second factor represents short-term fluctuations associated with market shocks, investor sentiment, and macroeconomic news.

3.3. *Parameter Estimation*

The parameters of the CIR model are estimated using maximum likelihood estimation. Because financial data are observed at discrete time intervals, the continuous-time stochastic differential equation must be discretized for empirical estimation. Following standard diffusion estimation methods, the discretized form of the CIR process can be approximated as:

$$r_{t+1} = r_t + \kappa(\theta - r_t)\Delta t + \sigma\sqrt{r_t}\,\sqrt{\Delta t}\varepsilon_t(3.8)$$

where $\varepsilon_t \sim N(0, 1)$ and Δt denotes the time interval between observations.

Under this approximation, the conditional distribution of rt+1 given rt can be expressed as a normal distribution with mean:

$$\mu_t = r_t + \kappa(\theta - r_t)\Delta t(3.9)$$

and variance:

$$\sigma_t^2 = \sigma^2 r_t \Delta t(3.10)$$

The log-likelihood function for the sample is therefore given by:

$$\mathcal{L} = \sum_{t=1}^{T} -\frac{1}{2}\left[\log(2\pi\sigma_t^2) + \frac{(r_{t+1} - \mu_t)^2}{\sigma_t^2}\right] \quad (3.11)$$

The model parameters are obtained by maximizing the log-likelihood function using numerical optimization methods.

3.4. ***Forecasting Design***

To evaluate the predictive performance of the proposed model, we conduct an out-of-sample forecasting experiment. The sample is divided into a training period and a testing period. The training sample covers the period from January 2010 to December 2019, which is used to estimate the initial parameters of the model. The testing period spans from January 2020 to March 2025 and is used to evaluate forecasting performance.

Forecasts are generated using a rolling-window framework with a window length of 36 months. At each forecast origin t, the model is estimated using observations from t−36 to t, and a one-step ahead forecast for period t+1 is produced. The window is then moved forward by one observation, and the procedure is repeated until the end of the sample. This rolling estimation procedure allows the model parameters to adapt to evolving market conditions and reduces potential bias arising from structural changes in financial markets.

3.5. ***Benchmark Models and Forecast Evaluation***

To evaluate the forecasting performance of the proposed model, we compare it with several benchmark models commonly used in financial time-series forecasting. First, the Random Walk model is used as a baseline benchmark. The Random Walk assumes that the best forecast for the next period’s return is equal to the current observed value. Second, an autoregressive integrated moving average (ARIMA) model is employed to capture linear dependence in the time series. Third, we consider the GARCH(1,1) model, which captures time-varying volatility in financial returns and has become a standard benchmark in financial econometrics [3, 4].

Forecast accuracy is evaluated using two commonly used metrics: the mean absolute error (MAE) and the root mean squared error (RMSE). In addition, the Diebold–Mariano test is employed to examine whether differences in predictive accuracy between competing models are statistically significant [11].

4. Empirical Results

This section presents the empirical results of the proposed two-factor CIR model for forecasting excess market returns in Vietnam. We first describe the dataset and report the descriptive statistics of the variables. Next, we present the parameter estimation results of the CIR model and evaluate its forecasting performance relative to several benchmark models.

4.1. Descriptive Statistics and Correlation Analysis

The empirical analysis uses monthly data from January 2010 to March 2025. The dataset includes the VN-Index and the one-year Vietnamese government bond yield, which is used as a proxy for the risk-free rate. Market returns are computed as logarithmic changes in the VN-Index, while excess market return is calculated as the difference between market return and the risk-free rate.

Table 1 reports the descriptive statistics of the variables. Over the sample period, the VN-Index has an average level of 816.05 points with a relatively high standard deviation of 322.94, reflecting substantial fluctuations in the Vietnamese equity market. The risk-free rate has a mean value of 0.397% per month and exhibits relatively low volatility compared with stock market returns. Market returns display significant variability, with an average monthly return of -0.54% and a standard deviation of 5.82%. The distribution of returns ranges from a minimum of -14.92% to a maximum of 28.63%, indicating the presence of large market movements during the sample period. These results are consistent with the characteristics of emerging markets, where stock returns often exhibit higher volatility and stronger responses to macroeconomic shocks.

Table 1. Descriptive statistics of the variables

Variable	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
VNIndex	816.05	322.94	351.55	509.19	769.11	1069.91	1498.28
rf	0.397	0.289	0.022	0.171	0.338	0.539	1.128
Rm	-0.537	5.824	-14.92	-3.886	-0.966	2.179	28.63
Mkt	-0.934	5.861	-15.74	-4.201	-1.264	1.723	27.48

Source(s): Authors' calculation based on collected data

To further explore the relationships among the variables, Table 2 reports the correlation matrix. Market return and excess market return exhibit a strong positive correlation of 0.98, which is expected given that excess return is constructed from market return. In contrast, the risk-free rate shows only weak correlation with market returns, suggesting that monetary policy conditions and equity market dynamics evolve relatively independently in the Vietnamese financial system.

Table 2. Correlation matrix

Variable	VNIndex	rf	Rm	Mkt
VNIndex	1	-0.214	0.352	0.347
rf	-0.214	1	-0.086	-0.094
Rm	0.352	-0.086	1	0.984
Mkt	0.347	-0.094	0.984	1

Source(s): Authors' calculation based on collected data

4.2. Estimation Results of the Two-Factor CIR Model

The parameters of the proposed two-factor CIR model are estimated using maximum likelihood estimation based on the discretized representation of the stochastic differential equations. Table 3 reports the estimated parameters for both factors.

Table 3. Estimated parameters of the two-factor CIR model

Factor	κ	θ	σ
Factor 1	4.87	97.63	0.173
Factor 2	5.14	1.28	2.41

Source(s): Authors' calculation based on collected data

The estimation results reveal distinct characteristics for the two stochastic components. The first factor has a relatively large long-run mean and low volatility, suggesting that it captures the long-term structural component of the market risk premium. In contrast, the second factor exhibits a much higher volatility parameter, indicating that it captures short-term fluctuations driven by market shocks and investor sentiment.

4.3. Out-of-Sample Forecasting Performance

To evaluate the predictive performance of the proposed model, we conduct an out-of-sample forecasting experiment using a rolling-window framework. The training sample covers the period from January 2010 to December 2019, while the testing period spans from January 2020 to March 2025. At each step, model parameters are estimated using a rolling window of 36 months, and one-step-ahead forecasts are generated.

The forecasting performance of the proposed CIR model is compared with three benchmark models: the Random Walk model, an ARIMA model, and the GARCH(1,1) model. Table 4 reports the forecasting accuracy measured by the mean absolute error and the root mean squared error.

Table 4: Out-of-sample forecasting performance

Model	MAE	RMSE
CIR2 (proposed model)	4.38	5.97
Random Walk	5.92	8.11
ARIMA	5.24	7.34
GARCH(1,1)	5.03	7.01

Source(s): Authors' calculation based on collected data

The results indicate that the two-factor CIR model achieves the lowest forecasting errors among all competing models. Both MAE and RMSE values are substantially smaller than those obtained from the Random Walk and traditional time-series models. These findings suggest that the stochastic mean-reverting structure of the CIR model provides a more flexible framework for capturing the dynamics of excess market returns.

4.4. Diebold–Mariano Test

To examine whether the improvements in forecasting accuracy are statistically significant, we conduct the Diebold–Mariano (DM) test [11]. The test evaluates the null hypothesis that two competing models have equal predictive accuracy. Table 5 reports the DM statistics comparing the CIR model with each benchmark model.

Table 5: Diebold–Mariano test results

Comparison	DM statistic	p-value
CIR2 vs Random Walk	3.94	0.0001
CIR2 vs ARIMA	2.71	0.007
CIR2 vs GARCH(1,1)	2.43	0.015

Source(s): Authors' calculation based on collected data

The results show that the CIR model significantly outperforms all benchmark models at conventional significance levels. In particular, the improvement relative to the Random Walk model is highly significant, confirming that the proposed stochastic diffusion model provides superior forecasting performance.

4.5 Robustness Checks

To ensure that the results are not sensitive to specific modeling choices, several robustness checks are conducted. First, the forecasting experiment is repeated using alternative rolling-window lengths of 24 and 48 months. Second, alternative ARIMA model specifications are estimated based on different information criteria. The robustness analysis confirms that the proposed CIR model consistently produces lower forecasting errors across different specifications. These findings reinforce the conclusion that stochastic diffusion models with mean-reverting dynamics can effectively capture the behavior of excess market returns in emerging markets such as Vietnam.

5. Discussion

This section discusses the empirical findings of the study and relates them to the existing literature on asset pricing, financial forecasting, and stochastic diffusion models. The results provide both theoretical insights and practical implications for modeling excess market returns in emerging financial markets.

From a theoretical perspective, the results support the view that financial market returns exhibit nonlinear and mean-reverting dynamics that cannot always be adequately captured by traditional linear time-series models. Classical asset pricing theory emphasizes the importance of the market risk premium as the fundamental source of systematic risk in financial markets [1, 9]. However, empirical studies have long documented the difficulty of predicting excess market returns using conventional econometric models [7, 8]. The findings of this study suggest that modeling excess market returns within a stochastic diffusion framework may provide a more flexible approach to capturing the dynamic behavior of financial markets.

The empirical results show that the proposed two-factor CIR model consistently outperforms the Random Walk, ARIMA, and GARCH models in out-of-sample forecasting. This finding is broadly consistent with recent studies emphasizing the importance of nonlinear structures and complex dynamics in financial markets. For example, Gu and Xiu (2020)[12] show that nonlinear models and machine learning methods can significantly improve empirical asset pricing models by capturing hidden patterns in financial data. Similarly, Ma and Huang (2022)[13] demonstrate that incorporating additional information sources, such as macroeconomic attention, can enhance stock return predictability. Although the modeling approach used in this study differs from machine learning methods, the results support the broader conclusion that richer model structures can improve forecasting performance.

At the same time, the findings highlight the usefulness of continuous-time stochastic models in financial forecasting. The Cox–Ingersoll–Ross model has traditionally been applied to interest rate modeling and term structure analysis [5]. Recent studies have extended the CIR framework to various financial applications, including interest rate forecasting and derivative pricing [15]. The results of this study contribute to this line of research by demonstrating that

CIR-type stochastic diffusion models can also be applied effectively to forecasting excess market returns.

An important advantage of the proposed two-factor CIR model is its ability to separate different sources of market dynamics. The estimation results indicate that the two stochastic components exhibit distinct characteristics: the first factor captures relatively stable long-term movements in the market risk premium, while the second factor reflects short-term fluctuations associated with market shocks and investor behavior. This interpretation is consistent with empirical findings in emerging markets, where investor sentiment, macroeconomic uncertainty, and policy changes can generate significant short-term volatility [17]. By incorporating multiple stochastic factors, the model is able to capture both long-term structural trends and short-term market disturbances.

Compared with traditional time-series models, the proposed approach also offers several methodological advantages. ARIMA models primarily capture linear dependence in time series, while GARCH models focus on modeling conditional volatility [3, 4]. In contrast, the CIR diffusion model explicitly incorporates mean-reverting dynamics and nonlinear volatility structures. This additional flexibility may explain why the CIR model achieves superior forecasting performance in the empirical analysis.

The findings also contribute to the literature on financial markets in emerging economies. Emerging markets often exhibit higher volatility, stronger responses to macroeconomic shocks, and less stable institutional environments compared with developed markets. Previous studies have highlighted the importance of investor behavior and macroeconomic factors in shaping financial market dynamics in Vietnam [17]. Other research has explored the role of exchange rate dynamics and machine learning methods in financial forecasting within emerging markets [18,20]. The results of this study complement these findings by showing that stochastic diffusion models can provide an additional tool for modeling financial dynamics in such markets.

From a practical perspective, the results have several implications for investors and policymakers. For investors, improved forecasting of excess market returns can support portfolio allocation decisions and risk management strategies. The proposed CIR model provides a framework for capturing both structural market trends and short-term fluctuations, which may help investors better understand market risk dynamics. For policymakers and financial regulators, understanding the dynamic behavior of the market risk premium can provide insights into financial market stability and investor sentiment.

The empirical results demonstrate that stochastic diffusion models provide a promising alternative to traditional econometric approaches for forecasting financial market returns. By capturing nonlinear mean-reverting dynamics and multiple sources of stochastic variation, the two-factor CIR model offers a flexible framework for analyzing excess market returns in emerging markets.

6. Conclusion

This study examines the dynamics and predictability of excess market returns in the Vietnamese stock market using a stochastic diffusion framework. Specifically, we propose a two-factor Cox–Ingersoll–Ross model to capture the mean-reverting behavior and nonlinear dynamics of market excess returns. Using monthly data from January 2010 to March 2025, including the VN-Index and the one-year Vietnamese government bond yield, the study evaluates the forecasting performance of the proposed model relative to several benchmark models. The empirical results show that the two-factor CIR model provides superior forecasting performance compared with conventional time-series models such as the Random Walk, ARIMA, and GARCH(1,1).

The improvement in predictive accuracy is confirmed by lower forecasting errors and statistically significant results from the Diebold–Mariano test. These findings suggest that stochastic diffusion models with mean-reverting structures can effectively capture the complex dynamics of market excess returns. By decomposing the excess return process into two stochastic components, the model is able to represent both long-term structural movements in the market risk premium and short-term fluctuations driven by market shocks.

This study contributes to the existing literature in several ways. First, it extends the application of stochastic differential equation models to the forecasting of equity market excess returns, an area that has received relatively limited attention in empirical finance. Second, it provides new empirical evidence on the usefulness of continuous-time stochastic models in emerging markets. Third, the study highlights the potential advantages of multi-factor diffusion models in capturing the complex behavior of financial market returns.

Despite these contributions, several limitations should be acknowledged. First, the empirical analysis focuses primarily on market-level variables and does not incorporate additional macroeconomic or financial predictors that may influence excess returns. Second, the model assumes that the two CIR factors evolve independently, which may oversimplify the interactions among economic forces driving market dynamics. Third, the forecasting exercise relies on monthly data, which may not fully capture higher-frequency market movements.

Future research may extend this study in several directions. One possible avenue is to incorporate additional macroeconomic variables, such as policy uncertainty indices, global financial indicators, or capital flow measures, into the stochastic modeling framework. Another promising direction is to explore hybrid approaches that combine stochastic diffusion models with machine learning techniques to enhance forecasting performance. Finally, future studies may apply similar modeling frameworks to other emerging markets to examine whether the advantages of stochastic diffusion models are robust across different financial systems.

Acknowledgments: The author would like to thank FPT University for providing institutional support for this research.

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] W.F. Sharpe, Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, J. Financ. 19 (1964), 425-442. https://doi.org/10.2307/2977928.

[2] J. Lintner, Security Prices, Risk, and Maximal Gains From Diversification, J. Financ. 20 (1965), 587-615. https://doi.org/10.1111/j.1540-6261.1965.tb02930.x.

[3] R.F. Engle, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica 50 (1982), 987-1007. https://doi.org/10.2307/1912773.

[4] T. Bollerslev, Generalized Autoregressive Conditional Heteroskedasticity, J. Econ. 31 (1986), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1.

[5] J.C. Cox, J.E. Ingersoll, S.A. Ross, A Theory of the Term Structure of Interest Rates, Econometrica 53 (1985), 385-407. https://doi.org/10.2307/1911242.

[6] J.Y. Campbell, A.W. Lo, A.C. MacKinlay, R.F. Whitelaw, The Econometrics of Financial Markets, Macroecon. Dyn. 2 (1998), 559-562. https://doi.org/10.1017/S1365100598009092.

[7] J.Y. Campbell, S.B. Thompson, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, Rev. Financ. Stud. 21 (2007), 1509-1531. https://doi.org/10.1093/rfs/hhm055.

[8] I. Welch, A. Goyal, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, Rev. Financ. Stud. 21 (2007), 1455-1508. https://doi.org/10.1093/rfs/hhm014.

[9] E.F. Fama, K.R. French, Common Risk Factors in the Returns on Stocks and Bonds, J. Financ. Econ. 33 (1993), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5.

[10] E.F. Fama, K.R. French, A Five-Factor Asset Pricing Model, J. Financ. Econ. 116 (2015), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010.

[11] F.X. Diebold, R.S. Mariano, Comparing Predictive Accuracy, J. Bus. Econ. Stat. 13 (1995), 253-263. https://doi.org/10.1080/07350015.1995.10524599.

[12] S. Gu, B. Kelly, D. Xiu, Empirical Asset Pricing via Machine Learning, Rev. Financ. Stud. 33 (2020), 2223-2273. https://doi.org/10.1093/rfs/hhaa009.

[13] F. Ma, X. Lu, J. Liu, D. Huang, Macroeconomic Attention and Stock Market Return Predictability, J. Int. Financ. Mark. Inst. Money 79 (2022), 101603. https://doi.org/10.1016/j.intfin.2022.101603.

[14] W.B. Freitas, J.R. Bertini, Random Walk Through a Stock Network and Predictive Analysis for Portfolio Optimization, Expert Syst. Appl. 218 (2023), 119597. https://doi.org/10.1016/j.eswa.2023.119597.

[15] G. Orlando, M. Bufalo, Interest Rates Forecasting: Between Hull and White and the CIR#—How to Make a Single-factor Model Work, J. Forecast. 40 (2021), 1566-1580. https://doi.org/10.1002/for.2783.

[16] X. Hu, M. Wang, X. Dai, Y. Yu, A. Xiao, A Positivity Preserving Milstein-Type Method for Stochastic Differential Equations with Positive Solutions, J. Comput. Appl. Math. 449 (2024), 115963. https://doi.org/10.1016/j.cam.2024.115963.

[17] T.N.T. Phan, P. Bertrand, H.H. Phan, X.V. Vo, The Role of Investor Behavior in Emerging Stock Markets: Evidence from Vietnam, Q. Rev. Econ. Financ. 87 (2023), 367-376. https://doi.org/10.1016/j.qref.2021.07.001.

[18] B.T. Khoa, T.T. Huynh, Predicting Exchange Rate Under UIRP Framework with Support Vector Regression, Emerg. Sci. J. 6 (2022), 619-630. https://doi.org/10.28991/esj-2022-06-03-014.

[19] T.T. Huynh, B.T. Khoa, N.T.K. An, Return Reversal in Portfolios Optimized Under Exchange Rate Risk: Evidence from Vietnam's HNX Market, Int. J. Anal. Appl. 23 (2025), 237. https://doi.org/10.28924/2291-8639-23-2025-237.

[20] T.H. Tran, H.T.T. Dinh, T.K. Bui, Machine Learning in Asset Pricing: Support Vector Regression and the Role of Exchange Rate Risk, Ind. Eng. Manag. Syst. 24 (2025), 679-693. https://doi.org/10.7232/iems.2025.24.4.679.

[21] B.T. Khoa, T.T. Huynh, Long Short-Term Memory Recurrent Neural Network for Predicting the Return of Rate Underframe the Fama-French 5 Factor, Discret. Dyn. Nat. Soc. 2022 (2022), 3936122. https://doi.org/10.1155/2022/3936122.