# Application of Bayesian Regression Model in Financial Stock Market Forecasting

Xuejun Zhao[1,*]

[1] *College of Letter and Science, University of California, Santa Barbara, CA 93106, USA*
[*]*Corresponding author. Email: xuejun@ucsb.edu*

**ABSTRACT**

The Bayesian method is a statistics field targeting the Bayes theorem in interpreting probabilities. The Bayesian formula provides an insight into conditional probability based on present data and prior information. Due to the efficiency of the Bayesian model in predicting future outcomes, the model is integrated with regression analysis which is a set of statistical methods utilized for estimating relationships between dependent and independent variables. Bayesian regression analysis is a reliable model for investigating variables having a significant impact on the output of a particular process, such as financial stock market forecasting considered in this research. To fulfill the study's aim, the research adopts secondary research on published journals, case studies, and reports documented by scholars in the field. Due to the stochastic nature of stock market variables, inadequate data or poorly dispersed data can be addressed using Bayesian linear regression allowing investors to make better decisions and cut larger profit margins. The vector autoregression and the classical frequentist approach achieve a higher probability accuracy than non-Bayesian methods such as the Auto-Regressive and Moving Average Model time series models. The author found that by studying the vector Bayesian autoregressive prediction model, it is possible to analyze how investors use the Bayesian model to predict stock market volatility.

*Keywords: Bayes' theorem, ARMA model, Regression analysis, Frequentist approach, Financial stock market.*

## 1. INTRODUCTION

A country's economic and social structure is heavily reliant on the stock market. The huge demand and complexity of stock market forecasts are determined by factors such as the input of stock price time series, stocks, raw materials, non-parameters, and dynamics. A stock market is an open market where company stock is bought and sold at an agreed-upon price. A stock exchange is one of the major stock market components. A stock exchange is an organization that gives trading facilities for stock traders and brokers to trade equities. The author will analyze how investors predict stock market volatility by the Statistics method set.

Regression analysis is a statistical approach used to show how a variable of interest relates to a collection of already known variables. The objective is to construct a model that helps statisticians explain, control, and forecast the dependent variable based on the independent factors. Regression analysis might be superficial, multiple linear, nonlinear, or Bayesian[1]. If a regression modeling approach utilizes the Bayes' theorem, the process yields the Bayesian linear regression, which includes a prior likelihood distribution and a posterior distribution[2]. The Bayesian regression model calculates parameter values using the posterior distribution multiplied by the prior distribution. Under the standard distribution error assumption, a normally distributed error in a linear regression model employs the Ordinary Least Squares estimation.

The Bayes theorem was first announced nearly a year after the demise of its inventor Thomas Bayes in 1764[3]. Before 1764, Bayesian analysis had a tumultuous history due to the limited applicability of the method and the lack of computing devices. Due to limited computational ability, Bayes' models failed due to the technique's incapacity in dealing with prior probabilities

delaying its appreciation towards the later end of the 19th century[4]. As the computation devices continued to advance within the century, Bayesian analysis garnered more physics, statistics, and other fields. However, Bayesian achieved its full glory in the late 20th century when digital computers with decent processing specifications became widely available for scholars[2]. Many advances in the Bayesian method have followed from the development of global interest in Bayesian statistics. The application of probability to statistical issues is made easier using Bayesian statistics. It allows scholars to investigate the impact of several factors on the dependent variable. Since Bayes' theorem defines the conditional probability of an event, depending on facts and prior information, while regression analysis is a statistical method for estimating relationships between variables, the Bayesian regression analysis model can be an effective tool in forecasting the stock market. The model is essential for investors due to the stochasticity of stock market variables. Additionally, the best-known and most extensively utilized Bayesian economic forecasting model is the vector autoregression model (VAR). Vector autoregression models are built on the concept of replication which is a recurring theme in these models. This article will study the significance of the Bayesian regression model for a financial forecasting model in the stock market. The author will analyze how investors use the Bayesian model to predict stock market volatility by studying the vector Bayesian autoregressive prediction model.

## 2. LITERATURE REVIEW

Using a prior distribution is the most contentious feature in Bayesian regression analysis. Scholars utilize these prior distributions differently, either via the Conjugate or the Noninformative prior. The Conjugate prior is when the posterior distribution and the initial distribution have the same form, while the Noninformative is when they differ[5]. The Noninformative Prior distribution is utilized when scholars have limited previous knowledge or information. If an analyst can acquire adequate posterior distribution information or deduce accurate predictions, they use the Conjugate Priors, an example of conjugate predicates[6]. Hence, if the posterior is conjugate to the probability, the prior is referred to as a conjugate prior. When a conjugate prior already exists, the posterior distribution function can be found, and the parameters can be calculated using the posterior distribution. Loss functions in the exponential

family of distributions are the only ones that have a conjugate prior. A family with many joint distributions is known as the exponential family, especially if the probability distribution function has an exponential term, but Conjugate prior does not exist for Poisson's distribution[2]. Conjugate priors for Negative Binomial distributions exist if the parameter of a Poisson distribution follows a Gamma distribution[6]. The prior and the probability product is the exponential of the sum of the two parameters and is calculated after the posterior distribution parameter is found.

Probability distributions are inclusive for linear regression in the Bayesian model. The response (dependent variable) is not expected to be chosen from a single probability distribution, but instead, the distribution itself is used to approximate the response[2]. The Bayesian Linear Regression model, in which the response variable is drawn from a normal distribution, is represented by,

$$y \sim N(\beta^\tau \chi, \sigma^2 l)$$

Y (response) is generated from a normal distribution, with a mean of zero and a variance equal to The inverse of the weight matrix multiplied by the predictor matrix transposes the matrix of inverse weights hence the linear regression[6]. Since the Bayesian regression model parameters are distributed, posterior probabilities for model parameters are computed conditionally from the training inputs and outputs.

$$P(y|\beta, X) = \frac{[P(\beta|y, X) * P(\beta|X)]}{[P(y|X)]}$$

The posterior probability distribution of model parameters is expressed by the above formula. The parameter (P(y|β, X)) is multiplied by the prior probability of the model (P(y|β, X)) and then divided by a normalization constant (P(y|β, X)).

$$Posterior = \frac{Liliihood * Prior}{Normalization}$$

In frequentist models, the domain knowledge is assumed, while in the Bayesian model's previous data is included. Before conducting quantitative estimation, Scholars apply noninformative priors, such as a normal distribution on the parameters. However, the posterior distribution of model parameters based on prior data is generated via Bayesian Linear Regression. The quantification of uncertainties in the Bayesian regression model facilitates accurate estimation of the model's distribution utilizing fewer data points.

## 3. METHODOLOGY

The data collection phase involves secondary research on published research journals from 2010 to 2021 on the Bayesian regression model in forecasting problems. The author searches and finds Bayesian method models, sorts out the results obtained, and leaves ten research papers that are suitable for financial forecasting using Bayesian regression models. The author also excluded duplicate papers, papers involving imprecise topics, papers not related to financial forecasts, and papers covering imprecise topics.

## 4. FINDINGS

Inadequate data or poorly dispersed data can be addressed using Bayesian linear regression. The statistical model allows a stock market analyst to attach prior information to the coefficients and noise in the absence of data, allowing the deduced priors to take over when the data is missing[7][8][9]. With a Bayesian linear regression, investors may use this tool to find variables with high confidence and others with low confidence[10]. The linear relation and the conviction on the relation are calculated utilizing Bayes' theorem[7][6]. The entire posterior distribution on the Bayesian regression model comprises the variables stock market analysts are uncertain about. The posterior distribution involves the current noise estimate and its probability distribution.

## 5. DISCUSSION

Sequential ordering and correlations frequently manifest in time series data analysis in stock market securities. The pattern in which sequential data evolves as the buying and selling processes progress involves time series analysis[3]. While classical statistics do not apply in many non-standard cases, Bayesian inference offers substantial advantages over such situations[7]. The ARMA time series models used in financial markets are highly unconventional even if analysts preserve normalcy assumptions. For instance, financial analysts rely on an array of asymptotic behaviors, consistency, asymptotic normalcy, and efficiency in their ARMA models to maintain normalcy[8]. Bayesian statistics are not significantly affected by small sample size because Bayesian regression models achieve a higher efficacy than standard time-series techniques.

The classical (frequentist) approach looks at the likelihood of an event given a long-term frequency. Stock market factors have random sampling attributes from the underlying variable population, with a real magnitude that is randomly distributed[9]. Inference rules incorporate universal constraints based on unbiased estimators, hypothesis testing, and confidence intervals, relying on large-sample approximations[5]. However, it is infrequent for financial data to adhere to these assumptions. For example, periods of economic crisis such as the Great Depression do not reoccur often enough to provide adequate samples of time-series data. In addition, volatility and correlation between various stock markets could either appreciate or depreciate in a crisis adding an element of stochasticity. A common assumption in these cases is that the underlying parameters are random variables[8]. It is crucial to bear in mind that many of these strong assumptions are completely irrelevant in a Bayesian context. In Bayesian statistics, the dataset is considered fixed, and the parameters are considered uncertain[5]. Hence, probability distributions can be altered when new information becomes available under the Bayesian framework.

The vector autoregression model use has expanded since it improves predictions and can be utilized to convey uncertainty[8]. The model is used for various purposes, including economic forecasting whereby Bayesian inference is commonly employed. Econometricians have discovered that non-Bayesian optimization is, in fact, more accessible to solve than the Bayesian integration issue[6]. There is an essential link between inference, computing, and models with practical applicability. By comparison, the significant advancements in posterior simulators in the past two decades have a notable impact on developing new models[5]. Since these models' applicability depends on available computational tools, powerful processors facilitate the making of influential previous assumptions, which are successful in compensating for a wasteful parameterization.

The market structure shifts over time, and occasionally, structural change happens suddenly. To successfully manage and reconcile the concept of probability, it is imperative to use dynamic Bayesian modeling processes[5]. These methods control the process of updating expectations while incorporating new information with careful respect to fundamental laws of probability[1]. Bayesian forecasting requires cautious attention to

parameters, variables, and modeling structure. The Fundamental forecasting model incorporates three market-related elements, including the shape of the yield curve, the performance of stocks against bonds, and the central bank's current monetary policies[8]. By utilizing forecasting models, excess return estimates, forecasting error volatility, and the correlation structure, the variables are fed into the modified mean-variance optimization system to generate Fundamental Currency and Fundamental Global Macro portfolios[6]. The technical forecasting model employs three market-related parameters and two technical indicators derived from two different daily moving averages to provide trading indications from historical price data. Excess return forecasting errors, forecasting error volatility, and the correlation structure of the errors are utilized as inputs to the modified mean-variance optimization system to create stock market securities portfolios[2]. These portfolios act as a guidebook for investors trading in various investment instruments such as equities, government bonds, commodities, and currencies to realize higher returns.

## 6. CONCLUSION

Chaotic Researching stock market prediction is an essential task for financial researchers since investing in the stock market has a higher risk. While this is feasible because of recent advances in artificial intelligence, the majority of the danger has been reduced. Forecasts are created recursively by way of a VAR. Forecasts are generated for each variable used in the system. While there are various situations in which VARs can be used, forecasting related variables where a straightforward interpretation is not required is just one of them. Additionally, impulse response analysis is used to determine whether one variable can be used to forecast another, and forecast error variance decomposition is used to uncover the variance of a forecast's variables due to the variability of the other variables. Fluctuation in the financial markets affects asset pricing, risk management, portfolio management, and the overall economy. It is consequently critical to better understand the driving forces behind financial market volatility. Studying the vector Bayesian autoregressive prediction model, it is possible to analyze how investors use the Bayesian model to predict stock market volatility. The vector autoregression models are built on the concept of replication which is a recurring theme in these models, and its use has

expanded since it improves predictions and can be utilized to convey uncertainty.

## AUTHORS' CONTRIBUTIONS

This paper is independently completed by Xuejun Zhao.

## REFERENCES

[1] Permai, S. D., & Tanty, H. (2018). Linear regression model using the Bayesian approach for energy performance of the residential building. Procedia Computer Science, 135(1), pp. 671-677. Retrieved from https://doi.org/10.1016/j.procs.2018.08.219

[2] DelSole, T. (2007). A Bayesian Framework for Multimodel Regression. Journal of Climate, 2810-2826. Retrieved from https://doi.org/10.1175/JCLI4179.1

[3] Beck, K., Niendorf, B., & Peterson, P. (2012). The use of Bayesian methods in economic research. Investment Management and Financial Innovations, 68-75. Retrieved from https://www.businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/4750/imfi_en_2012_03_Beck.pdf

[4] Andernach, I. E., Hunewald, O. E., & Muller, C. P. (2013). Bayesian Inference of the Evolution of HBV/E. PloS ONE, 8(11), 1-12. Retrieved from https://doi.org/10.1371/journal.pone.0081690

[5] Chan, J. C. (2019). Asymmetric Conjugate Priors for LargeBayesian VARs. Purdue University. Retrieved from https://economics.indiana.edu/documents/asymmetric-conjugate-priors-for-large-bayesian-vars.pdf

[6] Dey, S., & Dey, T. (2012). Bayesian estimation and prediction intervals for a Rayleigh distribution under a conjugate prior. Journal of Statistical Computation and Simulation, 82(11), 1651-1660. Retrieved

from
https://doi.org/10.1080/00949655.2011.59080
8

[7] Huang, W., & Chang, M.-S. (2021, February 03). Gold and Government Bonds as Safe-Haven Assets Against Stock Market Turbulence in China. Macroeconomics and Monetary Economics, 11(1), pp. 1-12. Retrieved from https://doi.org/10.1177/2158244021990655

[8] Ulf, H., & Raimond, M. (2006). Portfolio Choice and Estimation Risk. A Comparison of Bayesian to Heuristic Approaches. ASTIN Bulletin: The Journal of the IAA, 36(1), 135-160. Retrieved from https://doi.org/10.1017/S0515036100014434

[9] Nonejad, N. (2017). Modeling and forecasting aggregate stock market volatility in unstable environments using mixture innovation regressions. Journal of Forecasting, 36(6), 718-740. Retrieved from https://doi.org/10.1002/for.2466

[10] Ouysse, R., & Kohn, R. (2010, December 01). Bayesian variable selection and model averaging in the arbitrage pricing theory model. Computational Statistics and Data Analysis, 54(12), pp. 3249-3268. Retrieved from https://doi.org/10.1016/j.csda.2009.09.034